PART ONE -INTRODUCTION TO CLINICAL MEDICINE

1. THE PRACTICE OF MEDICINE - *The Editors*

WHAT IS EXPECTED OF THE PHYSICIAN

The practice of medicine combines both science and art. The role of *science in medicine* is clear. Science-based technology and deductive reasoning form the foundation for the solution to many clinical problems; the spectacular advances in genetics, biochemistry, and imaging techniques allow access to the innermost parts of the cell and the most remote recesses of the body. Highly advanced therapeutic maneuvers are increasingly a major part of medical practice. Yet skill in the most sophisticated application of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician. One must be able to identify the crucial elements in a complex history and physical examination and extract the key laboratory results from the crowded computer printouts of laboratory data in order to determine in a difficult case whether to "treat" or to "watch." Deciding when a clinical clue is worth pursuing, or when it should be dismissed as a "red herring," and estimating in any given patient whether a proposed treatment entails a greater risk than the disease are essential to the decision-making process that the skilled clinician must exercise many times each day. This combination of medical knowledge, intuition, and judgment defines the *art of medicine*, which is as necessary to the practice of medicine as is a sound scientific base.

The editors of the first edition of this book articulated what is expected of the physician in words that, although they reflect the gender bias of that era, still ring true as a universal principle:

No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering he needs technical skill, scientific knowledge, and human understanding. He who uses these with courage, with humility, and with wisdom will provide a unique service for his fellow man, and will build an enduring edifice of character within himself. The physician should ask of his destiny no more than this; he should be content with no less.

Tact, sympathy and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. He is human, fearful, and hopeful, seeking relief, help and reassurance.

THE PATIENT-PHYSICIAN RELATIONSHIP

It may seem trite to emphasize that physicians need to approach patients not as "cases" or "diseases" but as individuals whose problems all too often transcend their physical complaints. Most patients are anxious and frightened. Physicians should instill confidence and reassurance, overtly and in their demeanor, but without an air of arrogance. A professional attitude, coupled with warmth and openness, can do much to alleviate the patients' anxiety and to encourage them to share parts of their history that may be embarrassing. Some patients "use" illness to gain attention or to serve as a

crutch to extricate themselves from a stressful situation; some even feign physical illness; others may be openly hostile. Whatever the patient's attitude, the physician needs to consider the setting in which an illness occurs -- in terms not only of the patients themselves but also of their families and social and cultural backgrounds. The ideal patient-physician relationship is based on thorough knowledge of the patient, on mutual trust, and on the ability to communicate with one another.

The direct, one-to-one patient-physician relationship, which has traditionally characterized the practice of medicine, is increasingly in jeopardy because of the increasing complexity of medicine and change in health care delivery systems. Often the management of the individual patient is a team effort involving a number of several different physicians and professional personnel. The patient can benefit greatly from such collaboration, but *it is the duty of the patient's principal physician to guide them through an illness*. To carry out this difficult task, this physician must be familiar with the techniques, skills, and objectives of specialist physicians and of colleagues in the fields allied to medicine. In giving the patient an opportunity to benefit from scientific advances, the primary physician must, in the last analysis, retain responsibility for the major decisions concerning diagnosis and treatment.

Patients are increasingly cared for by groups of physicians in clinics, hospitals, integrated health care delivery systems, and health maintenance organizations (HMOs). Whatever the potential advantages of such organized medical groups, there are also drawbacks, chiefly the loss of the clear identification of the physician who is primarily and continuously responsible for the patient. Even under these circumstances, it is essential for each patient to have a physician who has an overview of the problems and who is familiar with the patient's reaction to the illness, to the drugs given, and to the challenges that the patient faces.

The practice of medicine in a "managed care" setting puts additional stress on the classic paradigm of the patient-physician relationship. Many physicians must deal with a patient within a restricted time frame, with limited access to specialists, and under organizational guidelines that may compromise their ability to exercise their individual clinical judgment. As difficult as these restrictions may be, it is the ultimate responsibility of the physician to determine what is best for the patient. This responsibility cannot be relinquished in the name of compliance with organizational guidelines.

The physician must also bear in mind that the modern hospital constitutes an intimidating environment for most patients. Lying in a bed surrounded by air jets, buttons, and lights; invaded by tubes and wires; beset by the numerous members of the health care team -- nurses, nurses' aides, physicians' assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others; sharing rooms with other patients who have their own problems, visitors, and physicians; transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel -- it is little wonder that patients may lose their sense of reality. In fact, the physician is often the only tenuous link between the patient and the real world, and a strong personal relationship with the physician helps to sustain the patient in such a stressful situation.

Many trends in contemporary society tend to make medical care impersonal. Some of these have been mentioned already and include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed care programs, which are intended to reduce costs but in which the patient may have little choice in selecting a physician; (3) increasing reliance on technologic advances and computerization for many aspects of diagnosis and treatment; (4) increased geographic mobility of both patients and physicians; (5) the need for numerous physicians to be involved in the care of most patients who are seriously ill; and (6) an increasing tendency on the part of patients to express their frustrations with the health care system by legal means (i.e., by malpractice litigation). Given these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine has defined humanistic qualities as encompassing integrity, respect, and compassion. Availability, the expression of sincere concern, the willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of the humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative (or strongly positive) emotional responses. Physicians should be alert to their own reactions to such patients and situations and should consciously monitor and control their behavior so that the patients' best interests remain the principal motivation for their actions at all times.

An important aspect of patient care involves an appreciation of the "quality of life," a subjective assessment of what each patient values most. Such an assessment requires detailed, sometimes intimate knowledge of the patient, which can usually be obtained only through deliberate, unhurried, and often repeated conversations. It is in these situations that the time constraints of a managed care setting may prove problematic.

The famous statement of Dr. Francis Peabody is even more relevant today than when delivered more than three-quarters of a century ago:

The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity, ***for the secret of the care of the patient is in caring for the patient.***

CLINICAL SKILLS

History Taking The written history of an illness should embody all the facts of medical significance in the life of the patient. Recent events should be given the most attention. The patient should, at some point, have the opportunity to tell his or her own story of the illness without frequent interruption and, when appropriate, receive expressions of interest, encouragement, and empathy from the physician. The physician must be alert to the possibility that any event related by the patient, however trivial or apparently remote, may be the key to the solution of the medical problem.

An informative history is more than an orderly listing of symptoms; something is always gained by listening to patients and noting the way in which they describe their

symptoms. Inflections of voice, facial expression, gestures, and attitude may reveal important clues to the meaning of the symptoms to the patient. Taking history often involves much data gathering. Patients vary in their medical sophistication and ability to recall facts. Medical history should therefore be corroborated whenever possible. The family and social history can also provide important insights into the types of diseases that should be considered. In listening to the history, the physician discovers not only something about the disease but also something about the patient. The process of history taking provides an opportunity to observe the patient's behavior and to watch for features to be pursued more thoroughly during the physical examination.

The very act of eliciting the history provides the physician with the opportunity to establish or enhance the unique bond that is the basis for the ideal patient-physician relationship. It is helpful to develop an appreciation of the patient's perception of the illness, the patient's expectations of the physician and the medical care system, and the financial and social implications of the illness to the patient. The confidentiality of the patient-physician relationship should be emphasized, and the patient should be given the opportunity to identify any aspects of the history that should not be disclosed.

Physical Examination Physical signs are objective indications of disease whose significance is enhanced when they confirm a functional or structural change already suggested by the patient's history. At times, however, the physical signs may be the only evidence of disease.

The physical examination should be performed methodically and thoroughly, with consideration for the patient's comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities. Unless the physical examination is systematic, important segments may be omitted. The results of the examination, like the details of the history, should be recorded at the time they are elicited, not hours later when they are subject to the distortions of memory. Skill in physical diagnosis is acquired with experience, but it is not merely technique that determines success in eliciting signs. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not a question of keener eyes and ears or more sensitive fingers but of a mind alert to these findings. Since physical findings are subject to changes, the physical examination should be repeated as frequently as the clinical situation warrants.

Laboratory Tests The availability of a wide array of laboratory tests has increased our reliance on these studies for the solution of clinical problems. The accumulation of laboratory data does not relieve the physician from the responsibility of careful observation, examination, and study of the patient. It is also essential to bear in mind the limitations of such tests. By virtue of their impersonal quality, complexity, and apparent precision, they often gain an aura of authority regardless of the fallibility of the tests themselves, the instruments used in the tests, and the individuals performing or interpreting them. Physicians must weigh the expense involved in the laboratory procedures they order relative to the value of the information they are likely to provide.

Single laboratory tests are rarely ordered. Rather, they are generally obtained as "batteries" of multiple tests, which are often useful. For example, abnormalities of

hepatic function may provide the clue to such nonspecific symptoms as generalized weakness and increased fatigability, suggesting the diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to particular diseases, such as hyperparathyroidism or underlying malignancy.

The thoughtful use of screening tests should not be confused with indiscriminate laboratory testing. The use of screening tests is based on the fact that a group of laboratory determinations can be carried out conveniently on a single specimen of blood at relatively low cost. Screening tests are most useful when they are directed towards common diseases or disorders in which the result directs other useful tests or interventions that would otherwise be costly to perform. Biochemical measurements, together with simple laboratory examinations such as blood count, urinalysis, and sedimentation rate, often provide the major clue to the presence of a pathologic process. At the same time, the physician must learn to evaluate occasional abnormalities among the screening tests that may not necessarily connote significant disease. An in-depth workup following a report of an isolated laboratory abnormality in a person who is otherwise well is almost invariably wasteful and unproductive. Among the more than 40 tests that are routinely performed on patients, one or two are often slightly abnormal. If there is no suspicion of an underlying illness, these tests are ordinarily repeated to ensure that the abnormality does not represent a laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient's condition and other test results.

Imaging Techniques The availability of ultrasonography, a variety of scans that employ isotopes to visualize organs heretofore inaccessible, computed tomography, and magnetic resonance imaging has opened new diagnostic vistas and has benefited patients because these new techniques have largely supplanted more invasive ones. While the enthusiasm for noninvasive technology is understandable, the expense entailed in performing these tests is often substantial and should be considered when assessing the potential benefits of the information provided.

PRINCIPLES OF PATIENT CARE

Medical Decision-Making Both during and in particular after the physician has taken the history, performed the physical examination, and reviewed the laboratory and imaging data, the challenging process of the differential diagnosis and medical decision-making begins. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases and to understand the significance of missing diagnoses that may be less likely. Arriving at a diagnosis requires the application of the scientific method. Hypotheses are formed, data are collected, and objective conclusions are reached concerning whether to accept or reject a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately. Medical decision-making occurs throughout the diagnostic and treatment process. It involves the ordering of additional tests, requests for consults, and decisions regarding prognosis and treatment. This process requires an in-depth understanding of the natural history and pathophysiology of disease, explaining why these features are strongly emphasized in this textbook. As described below, medical decision-making

should be evidence-based, thereby ensuring that patients derive the full benefit of the scientific knowledge available to physicians.

Evidence-Based Medicine Sackett has defined evidence-based medicine as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." Rigorously obtained evidence is contrasted with anecdotal experience, which is often biased. Even the most experienced physicians can be influenced by recent experiences with selected patients, unless they are attuned to the importance of using larger, more objective studies for making decisions. The prospectively designed, double-blind, randomized clinical trial represents the "gold standard" for providing evidence regarding therapeutic decisions, but it is not the only source. Valuable evidence about the natural history of disease and prognosis can come from prospective cohort studies and analytic surveys. Persuasive evidence on the accuracy of diagnostic tests can be derived from cross-sectional studies of patients in whom a specific disorder is suspected. Evidence is strengthened immensely when it has been confirmed by multiple investigations, which can be compared with one another and presented in a meta-analysis or systemic overview.

In failing to apply the best and most current evidence, the physician places the patient at unnecessary risk. However, a knowledge of or rapid access to the best available evidence is not sufficient for optimal care. The physician must know whether the evidence is relevant to the patient in question and, when it is, the consequences of applying it in any particular situation. The skills and judgment required to apply sound evidence represent an increasing challenge. Indeed, one might redefine a "good doctor" as one who uses the ever-growing body of rigorously obtained evidence (the science of medicine) in a sensible, compassionate manner (the art of medicine).

While an understanding of biologic and physiologic mechanisms forms the basis of contemporary medicine, when a therapeutic modality is selected, the highest priority must often be placed on improving *clinical outcome* rather than interrupting what is believed to be the underlying process. For example, for decades patients who had suffered myocardial infarction were treated intuitively with drugs that suppress frequent ventricular extrasystoles, since these were believed to be harbingers of ventricular fibrillation and sudden death. Clinical trials, however, have provided firm evidence that the antiarrhythmic agents actually increase the risk of death in such patients. This finding suggests that the extrasystoles are *markers* of high risk rather than the *cause* of fatal events.

Practice Guidelines Physicians are faced with a large, increasing, and often bewildering body of evidence pointing to potentially useful diagnostic techniques and therapeutic choices. The intelligent and cost-effective practice of medicine consists of making selections most appropriate to a particular patient and clinical situation. Professional organizations and government agencies are developing formal clinical practice guidelines in an effort to aid physicians and other caregivers in this endeavor. When guidelines are current and properly applied, they can provide a useful framework for managing patients with particular diagnoses or symptoms. They can protect patients -- particularly those with inadequate health care benefits -- from receiving substandard care. Guidelines can also protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of

medical resources. On the other hand, clinical guidelines tend to oversimplify the complexities of medicine. Different groups with differing perspectives may develop divergent recommendations regarding issues as basic as the need for periodic sigmoidoscopy in middle-aged persons. Furthermore, guidelines do not -- and cannot be expected to -- take into account the uniqueness of each individual and of his or her illness. The challenge for the physician is to integrate into clinical practice the useful recommendations offered by the experts who prepare clinical practice guidelines without accepting them blindly or being inappropriately constrained by them.

Assessing the Outcome of Treatment Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. For example, findings on physical or laboratory examination -- such as the level of blood pressure, the patency of a coronary artery on an angiogram, or the size of a mass on a radiologic examination -- can provide information of critical importance. However, patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, to preserve or regain function, and to enjoy life. The components of a patient's health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important areas can be assessed by means of structured interviews or specially designed questionnaires. Such assessments also provide useful parameters by which the physician can judge the patient's subjective view of his or her disability and the response to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Care of the Elderly Over the next several decades, the practice of medicine will be greatly influenced by the health care needs of the growing elderly population. In the United States the population over age 65 will almost triple over the next 30 years. It is essential that we understand and appreciate the physiologic processes associated with aging; the different responses of the elderly to common diseases; and disorders that occur commonly with aging, such as depression, dementia, frailty, urinary incontinence, and fractures. The elderly have more adverse reactions to drugs, in large part due to altered pharmacokinetics and pharmacodynamics. Commonly used medications such as digoxin and aminoglycosides have prolonged half-lives in the elderly, and tissues such as the central nervous system are more sensitive to certain drugs, such as the benzodiazepines and narcotics. The large number of drugs used by the elderly increases the risk of unwanted interactions, especially when care is provided by several physicians in an uncoordinated manner.

Diseases in Women versus Men In the past, many epidemiologic studies and clinical trials focused on men. It is now appreciated that there are significant gender differences in diseases that afflict both men and women. Mortality rates are substantially higher in women than in men under the age of 50 suffering acute myocardial infarction. Hypertension is more prevalent in African-American women than in their male counterparts (and in African-American than in white males); osteoporosis is more common in women, reflecting the menopausal loss of estrogen; diseases involving the immune system, such as lupus erythematosus, multiple sclerosis, and primary biliary cirrhosis, occur more frequently in women; and the average life expectancy of women is greater than that of men. Recently, considerable attention has been paid to women's

health issues, a subject that regrettably did not receive sufficient attention in the past. Ongoing study should enhance our understanding of the mechanisms of gender differences in the course and outcome of certain diseases.

Iatrogenic Disorders In an *iatrogenic disorder*, the deleterious effects of a therapeutic or diagnostic maneuver cause pathology independent of the condition for which the intervention was performed. Adverse drug reactions occur in at least 5% of hospitalized patients, and the incidence increases with use of a large number of drugs. No matter what the clinical situation, it is the responsibility of the physician to use powerful therapeutic measures wisely, with due regard for their beneficial action, potential dangers, and cost. Every medical procedure, whether diagnostic or therapeutic, has the potential for harm, but it would be impossible to provide the benefits of modern scientific medicine if reasonable steps in diagnosis and therapy were withheld because of possible risks. *Reasonable* implies that the physician has weighed the pros and cons of a procedure and has concluded, on the basis of objective evidence whenever possible, that it is necessary for establishing a diagnosis, for the relief of discomfort, or for the cure of disease. However, the harm that a physician can do is not limited to the imprudent use of medication or procedures. Equally important are ill-considered or unjustified remarks. Many a patient has developed a cardiac neurosis because the physician ventured a grave prognosis on the basis of a misinterpreted finding of a heart murmur. Not only the diagnostic procedure or the treatment but the physician's words and behavior are capable of causing injury.

Informed Consent Patients often require diagnostic and therapeutic procedures that are painful and that pose some risk. For many such procedures, patients are required to sign a consent form. The patient must understand clearly the risks entailed in these procedures; this is the definition of *informed consent*. It is incumbent on the physician to explain the procedures in a clear and understandable manner and to ascertain that the patient comprehends both the nature of the procedure and the attendant risks. The dread of the unknown that is inherent in hospitalization can be mitigated by such explanations.

Incurability and Death No problem is more distressing than that presented by the patient with an incurable disease, particularly when premature death is inevitable. What should the patient and family be told, what measures should be taken to maintain life, what can be done to maintain the quality of life, and how is death to be defined?

The concept of incurable illness and terminal care often evokes examples of cancer. However, patients with many other end-stage diseases including chronic obstructive pulmonary disease, congestive heart failure, renal or hepatic failure, and overwhelming infection face similar issues. The same principles of terminal care should be applied in each of these cases. Doing seemingly small things, focused on the needs of the patient, can do much to restore comfort or dignity during a person's final weeks or days. In the same way that pain should be attentively managed with analgesia, every effort should be made to alleviate shortness of breath and to provide good skin care.

Although some would argue otherwise, there is no ironclad rule that the patient must immediately be told "everything," even if the patient is an adult with substantial family responsibilities. How much is told should depend on the individual's ability to deal with

the possibility of imminent death; often this capacity grows with time, and whenever possible, gradual rather than abrupt disclosure is the best strategy. A wise and insightful physician is often guided by an understanding of what a patient wants to know and when he or she wants to know it. The patient's religious beliefs may also be taken into consideration. The patient must be given an opportunity to talk with the physician and ask questions. Patients may find it easier to share their feelings about death with their physician, who is likely to be more objective and less emotional, than with family members. As William Osler wrote:

One thing is certain; it is not for you to don the black cap and, assuming the judicial function, take hope away from any patient...hope that comes to us all.

Even when the patient directly inquires, "Am I dying?" the physician must attempt to determine whether this is a request for information or a demand for reassurance. Only open communication between the patient and the physician can resolve this question and guide the physician in what to say and how to say it.

The physician should provide or arrange for emotional, physical, and spiritual support and must be compassionate, unhurried, and open. There is much to be gained by the laying on of hands. Pain should be adequately controlled, human dignity maintained, and isolation from the family avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining apparatus can so easily detract from attention to the whole person and encourage concentration instead on the life-threatening disease, against which the battle will ultimately be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care*, in the broadest sense of the term. In offering care to the dying patient, the physician must be prepared to provide information to family members and to deal with their guilt and grief. It is important for the doctor to assure the family that everything possible has been done.

"Do Not Resuscitate" Orders and Cessation of Therapy When carried out in a timely and expert manner, cardiopulmonary resuscitation is often useful in the prevention of sudden, unexpected death. However, unless there are reasons to the contrary, this procedure should not be used merely to prolong the life of a patient with terminal, incurable disease. The decision whether or not to resuscitate or even to treat an incurably and terminally ill patient must be reviewed frequently and must take into consideration any unexpected changes in the patient's condition. In this context, the administration of fluids or food is considered therapy that may be withdrawn or withheld. These decisions must also take into account both the underlying medical condition, especially its reversibility, and the wishes of the patient, especially if these have been expressed in a living will or advance directive. If the patient's wishes cannot be ascertained directly, a close relative or another surrogate who can be relied on to transmit the patient's wishes and to be guided by the patient's best interests should be consulted. The patient's autonomy -- whether the choice is to continue or discontinue treatment or to be resuscitated or not in the event of a cardiopulmonary arrest -- must be paramount. The courts have ruled that competent patients may refuse therapy and that an incompetent patient's previously stated wishes regarding life support should therefore be respected. The issues involving death and dying are among the most difficult in medicine. In approaching them rationally and consistently, the physician must

combine both the science and the art of medicine.

THE EXPANDING ROLE OF THE PHYSICIAN

Genetics and Medicine The genomic era is likely to lead to a revolution in the practice of medicine. Obtaining the DNA sequence of the entire human genome may help to elucidate the genetic components of common chronic diseases -- hypertension, diabetes, atherosclerosis, many cancers, dementias, and behavioral and autoimmune disorders. This information should make it possible to determine individual susceptibility to these conditions early in life and to implement individualized prevention programs. Subclassification of many diseases on a genetic basis may allow the selection of appropriate therapy for each patient. As the response to drugs becomes more predictable, pharmacotherapy should become more rational. In short, the completion of the Human Genome Project is likely to lead to a substantial increase in physicians' ability to influence their patient's health and well-being.

Patients will be best served if physicians play an active role in applying this powerful, sensitive new information rather than being passive bystanders who are intimidated by the new technology. This is a rapidly evolving field, and physicians and other health care professionals must remain updated to apply this new knowledge. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals.

Medicine on the Internet The explosion in use of the Internet through personal computers is having an important impact on many practicing physicians. The Internet makes a wide range of information available to physicians almost instantaneously at any time of the day or night and from anywhere in the world. This medium holds enormous potential for delivering up-to-date information, practice guidelines, state-of-the-art conferences, journal contents, textbooks (including this text), and direct communications with other physicians and specialists, thereby expanding the depth and breadth of information available to the physician about the diagnosis and care of patients. Most medical journals are now accessible on-line, providing rapid and comprehensive sources of information. Patients, too, are turning to the Internet in increasing numbers to derive information about their illnesses and therapies and to join Internet-based support groups. Physicians are increasingly challenged by dealing with patients who are becoming more sophisticated in their understanding of illness. At this time, there is one critically important caveat. Virtually anything can be published on the Internet, thus circumventing the peer-review process that is an essential feature of quality publications. Physicians or patients who search the Internet for medical information must be aware of this danger. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and is a positive force in the practice of medicine.

Delivering Cost-Effective Medical Care As the cost of medical care has risen, it has become necessary to establish priorities in the expenditure of resources. In some instances, preventive measures offer the greatest return for the expenditure; outstanding examples include vaccination, improved sanitation, reduction in accidents and occupational hazards, and biochemical- and DNA-based screening of newborns. For example, the detection of phenylketonuria by newborn screening may result in a net

saving of many thousands of dollars.

As resources become increasingly constrained, the physician must weigh the possible benefits of performing costly procedures that provide only a limited life expectancy against the pressing need for more primary care for those persons who do not have adequate access to medical services. For the individual patient, it is important to reduce costly hospital admissions as much as possible if total health care is to be provided at a cost that most can afford. This policy, of course, implies and depends on close cooperation among patients, their physicians, employers, payers, and government. It is equally important for physicians to know the cost of the diagnostic procedures they order and the drugs and other therapies they prescribe and to monitor both costs and effectiveness. The medical profession should provide leadership and guidance to the public in matters of cost control, and physicians must take this responsibility seriously without being or seeming to be self-serving. However, the economic aspects of health care delivery must not interfere with the welfare of patients. The patient must be able to rely on the individual physician as his or her principal advocate in matters of health care.

Accountability Medicine is a satisfying but demanding profession. Physicians must understand the characteristics of the populations they serve, and they must appreciate their patients' social and cultural attitudes to health, disease, and death. As the public has become more educated and more sophisticated regarding health matters, their expectations of the health system in general and of their physicians in particular have risen. Physicians are expected to maintain mastery of their rapidly advancing fields (the *science* of medicine) while considering their patient's unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care that they provide but also for their patient's satisfaction with the delivery and costs of care.

In the United States, there are increasing demands for physicians to account for the way in which they practice medicine by meeting certain standards prescribed by federal and state governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus the physician must defend the cause for and duration of a patient's hospitalization if it falls outside certain "average" standards. Authorization for reimbursement is increasingly based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. The purpose of these regulations is both to improve standards of health care and to contain spiraling health care costs. This type of review is being extended to all phases of medical practice and is profoundly altering the practice of medicine. Physicians are also expected to give evidence of their continuing competence through mandatory continuing education, patient-record audits, recertification by examination, or relicensing.

Continued Learning The conscientious physician must be a perpetual student because the body of medical knowledge is constantly expanding and being refined. The profession of medicine should be inherently linked to a career-long thirst for new knowledge that can be used for the good of the patient. It is the responsibility of a physician to pursue continually the acquisition of new knowledge by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, such a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

Research and Teaching The title *doctor* is derived from the Latin *docere*, "to teach," and physicians should share information and medical knowledge with colleagues, with students of medicine and related professions, and with their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information, i.e., on research, which often involves patients; improved medical care requires the transmission of this information. As part of broader societal responsibilities, the physician should encourage patients to participate in ethical and properly approved clinical investigations if they do not impose undue hazard, discomfort, or inconvenience. To quote Osler once more:

To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease -- these are our ambitions.

(Bibliography omitted in Palm version)

Back to Table of Contents

2. ETHICAL ISSUES IN CLINICAL MEDICINE - *Bernard Lo*

Physicians frequently confront ethical issues in clinical practice that are perplexing, time-consuming, and emotionally draining. Experience, common sense, and simply being a good person do not guarantee that physicians can identify or resolve ethical dilemmas. Knowledge about common ethical dilemmas is also essential.

FUNDAMENTAL ETHICAL GUIDELINES

Physicians should follow two fundamental but frequently conflicting ethical guidelines: respecting patient autonomy and acting in the patient's best interests.

RESPECTING PATIENT AUTONOMY

Competent, informed patients may refuse recommended interventions and choose among reasonable alternatives.

Informed Consent Informed consent requires physicians to discuss with patients the nature of the proposed care, the alternatives, the risks and benefits of each, the likely consequences, and to obtain the patient's agreement to care. Informed consent involves more than obtaining signatures on consent forms. Physicians need to educate patients, answer questions, make recommendations, and help them deliberate. Patients can be overwhelmed with medical jargon, needlessly complicated explanations, or too much information at once.

Nondisclosure of Information Physicians may consider withholding a serious diagnosis, misrepresenting it, or limiting discussions of prognosis or risks because they fear that a patient will develop severe anxiety or depression or refuse needed care. Patients should not be forced to receive information against their will. Most people, however, want to know their diagnosis and prognosis, even if they are terminally ill. Generally, physicians should provide relevant information, offer empathy and hope, and help patients cope with bad news.

Emergency Care Informed consent is not required when patients cannot give consent and when delay of treatment would place their life or health in peril. People are presumed to want such emergency care, unless they have previously indicated otherwise.

Futile Interventions Autonomy does not entitle patients to insist on whatever care they want. Physicians are not obligated to provide futile interventions that have no physiologic rationale or have already failed. For example, cardiopulmonary resuscitation would be futile in a patient with progressive hypotension despite maximal therapy. But physicians should be wary of using the term "futile" in looser senses to justify unilateral decisions to forego interventions when they believe that the probability of success is too low, no worthwhile goals can be achieved, the patient's quality of life is unacceptable, or the costs are too high. Such looser usages of the term are problematic because they may be inconsistent and mask value judgments.

ACTING IN THE BEST INTERESTS OF PATIENTS

The guideline of *beneficence* requires physicians to act for the patient's benefit. Laypeople do not possess medical expertise and may be vulnerable because of their illness. They justifiably rely on physicians to provide sound advice and to promote their well-being. Physicians encourage such trust. Hence, physicians have a fiduciary duty to act in the best interests of their patients. The interests of the patient should prevail over physicians' self-interest or the interests of third parties, such as hospitals or insurers. These fiduciary obligations of physicians contrast sharply with business relationships, which are characterized by "let the buyer beware," not by trust and reliance. The guideline of "*do no harm*" forbids physicians from providing ineffective interventions or acting without due care. This precept, while often cited, provides only limited guidance, because many beneficial interventions also have serious risks.

CONFLICTS BETWEEN BENEFICENCE AND AUTONOMY

Patients' refusals of care may thwart their own goals or cause them serious harm. For example, a young man with asthma may refuse mechanical ventilation for reversible respiratory failure. Simply to accept such refusals, in the name of respecting autonomy, seems morally constricted. Physicians can elicit patients' expectations and concerns, correct misunderstandings, and try to persuade them to accept beneficial therapies. If disagreements persist after discussions, the patient's informed choices and view of his or her best interests should prevail. While refusing recommended care does not render a patient incompetent, it may lead the physician to probe further to ensure that the patient is able to make informed decisions.

PATIENTS WHO LACK DECISION-MAKING CAPACITY

Patients may not be able to make informed decisions because of unconsciousness, dementia, delirium, or other conditions. Physicians should ask two questions regarding such patients: Who is the appropriate surrogate? What would the patient want done?

ASSESSING CAPACITY TO MAKE MEDICAL DECISIONS

All adults are considered legally competent unless declared incompetent by a court. In practice, physicians usually determine that patients lack the capacity to make health care decisions and arrange for surrogates to make them, without involving the courts. By definition, competent patients can express a choice and appreciate the medical situation, the nature of the proposed care, the alternatives, and the risks, benefits, and consequences of each. Their choices should be consistent with their values and should not result from delusions or hallucinations. Psychiatrists may help in difficult cases because they are skilled at interviewing mentally impaired patients and can identify treatable depression or psychosis. When impairments are fluctuating or reversible, decisions should be postponed if possible until the patient recovers decision-making capacity.

CHOICE OF SURROGATE

If a patient lacks decision-making capacity, physicians routinely ask family members to serve as surrogates. Most patients want their family members to be surrogates, and

family members generally know the patient's preferences and have the patient's best interests at heart. Patients may designate a particular individual to serve as proxy; such choices should be respected. Some states have established a prioritized list of which relative may serve as surrogate if the patient has not designated a proxy.

STANDARDS FOR SURROGATE DECISION MAKING

Advance Directives These are statements by competent patients to direct care if they lose decision-making capacity. They may indicate (1) what interventions they would refuse or accept or (2) who should serve as surrogate. Following the patient's advance directives, surrogate respects patients' autonomy.

Oral conversations are the most frequent form of advance directives. While such conversations are customarily followed in clinical practice, casual or vague comments may not be trustworthy.

Living wills direct physicians to forego or provide life-sustaining interventions if the patient develops a terminal condition or persistent vegetative state. Generally patients may refuse only interventions that "merely prolong the process of dying."

A health care proxy is someone appointed by the patient to make health care decisions if he or she loses decision-making capacity. It is more flexible and comprehensive than the living will, applying whenever the patient is unable to make decisions.

Physicians can encourage patients to provide advance directives, to indicate both what they would want and who should be surrogate, and to discuss their preferences with surrogates. In discussions with patients, physicians can ensure that advance directives are informed, up-to-date, and address likely clinical scenarios. Such discussions are best carried out in the ambulatory setting. The federal Patient Self-Determination Act requires hospitals and health maintenance organizations to inform patients of their right to make health care decisions and to provide advance directives.

Substituted Judgment In the absence of clear advance directives, surrogates and physicians should try to decide as the patient would under the circumstances, using all information that they know about the patient. While such substituted judgments try to respect the patient's values, they may be speculative or inaccurate. A surrogate may be mistaken about the patient's preferences, particularly when they have not been discussed explicitly.

Best Interests When the patient's preferences are unclear or unknown, decisions should be based on the patient's best interests. Patients generally take into account the quality of life as well as the duration of life when making decisions for themselves. It is understandable that surrogates would also consider quality of life of patients who lack decision-making capacity. Judgments about quality of life are appropriate if they reflect the patient's own values. Bias or discrimination may occur, however, if others project their values onto the patient or weigh the perceived social worth of the patient. Most patients with chronic illness rate their quality of life higher than their family members and physicians do.

Legal Issues Physicians need to know pertinent state laws regarding patients who lack decision-making capacity. A few state courts allow doctors to forego life-sustaining interventions only if patients have provided written advance directives or very specific oral ones.

Disagreements Disagreements may occur among potential surrogates or between the physician and surrogate. Physicians can remind everyone to base decisions on what the patient would want, not what they would want for themselves. Consultation with the hospital ethics committee or with another physician often helps resolve disputes. Such consultation is also helpful when patients have no surrogate and no advance directives. The courts should be used only as a last resort when disagreements cannot be resolved in the clinical setting.

DECISIONS ABOUT LIFE-SUSTAINING INTERVENTIONS

Although medical technology can save lives, it can also prolong the process of dying. Competent, informed patients may refuse life-sustaining interventions. Such interventions may also be withheld from patients who lack decision-making capacity on the basis of advance directives or decisions by appropriate surrogates. Courts have ruled that foregoing life-sustaining interventions is neither suicide nor murder.

MISLEADING DISTINCTIONS

People commonly draw distinctions that are intuitively plausible but prove untenable on closer analysis.

Extraordinary and Ordinary Care Some physicians are willing to forego "extraordinary" or "heroic" interventions, such as surgery, mechanical ventilation, or renal dialysis, but insist on providing "ordinary" ones, such as antibiotics, intravenous fluids, or feeding tubes. However, this distinction is not logical because all medical interventions have both risks and benefits. Any intervention may be withheld, if the burdens for the individual patient outweigh the benefits.

Withdrawing and Withholding Interventions Many health care providers find it more difficult to discontinue interventions than to withhold them in the first place. Although such emotions need to be acknowledged, there is no logical distinction between the two acts. Justifications for withholding interventions, such as refusal by patients or surrogates, are also justifications for withdrawing them. In addition, an intervention may prove unsuccessful or new information about the patient's preferences or condition may become available after the intervention is started. If interventions could not be discontinued, patients and surrogates might not even attempt treatments that might prove beneficial.

DO NOT RESUSCITATE (DNR) ORDERS

When a patient suffers a cardiopulmonary arrest, cardiopulmonary resuscitation (CPR) is initiated unless a DNR order has been made. Although CPR can restore people to vigorous health, it can also disrupt a peaceful death. After CPR is attempted on a general hospital service, only 14% of patients survive to discharge, and even fewer in

certain subgroups. DNR orders are appropriate if the patient or surrogate requests them or if CPR would be futile. To prevent misunderstandings, physicians should write DNR orders and the reasons for them in the medical record. "Slow" or "show" codes that merely appear to provide CPR are deceptive and therefore unacceptable. Although a DNR order signifies only that CPR will be withheld, the reasons that justify DNR orders may lead to a reconsideration of other plans for care.

ASSISTED SUICIDE AND ACTIVE EUTHANASIA

Proponents of these controversial acts believe that competent, terminally ill patients should have control over the end of life and that physicians should relieve refractory suffering. Opponents assert that such actions violate the sanctity of life, that suffering can generally be relieved, that abuses are inevitable, and that such actions are outside the physician's proper role. These actions are illegal throughout the United States, except that physician-assisted suicide is legal in Oregon under certain circumstances. Whatever their personal views, physicians should respond to patients' inquiries with compassion and concern. Physicians should elicit and address any underlying problems, such as physical symptoms, loss of control, or depression. Often, additional efforts to relieve distress are successful, and after this is done patients generally withdraw their requests for these acts.

CARE OF DYING PATIENTS

Patients often suffer unrelieved pain and other symptoms during their final days of life. Physicians may hesitate to order high doses of narcotics and sedatives, fearing they will hasten death. Relieving pain in terminal illness and alleviating dyspnea when patients forego mechanical ventilation enhances patient comfort and dignity. If lower doses of narcotics and sedatives have failed to relieve suffering, increasing the dose to levels that may suppress respiratory drive is ethically appropriate because the physician's intention is to relieve suffering, not hasten death. Physicians can also relieve suffering by spending time with dying patients, listening to them, and attending to their psychological distress.

CONFLICTS OF INTEREST

Acting in the patient's best interests may conflict with the physician's self-interest or the interests of third parties such as insurers or hospitals. The ethical ideal is to keep the patient's interests paramount. Even the appearance of a conflict of interest may undermine trust in the profession.

FINANCIAL INCENTIVES

In managed care systems, physicians may serve as gatekeepers or bear financial risk for expenditures. Although such incentives are intended to reduce inefficiency and waste, there is concern that physicians may withhold beneficial care in order to control costs. In contrast, physicians have incentives to provide more care than indicated when they receive fee-for-service reimbursement or when they refer patients to medical facilities in which they have invested. Regardless of financial incentives, physicians should recommend available care that is in the patient's best interests -- no more and

no less.

DENIALS OF COVERAGE

Utilization review programs designed to reduce unnecessary services may also deny coverage for care that the physician believes will benefit the patient. Physicians should inform patients when a plan is not covering standard care and act as patient advocates by appealing such denials of coverage. Patients may ask physicians to misrepresent their condition to help them obtain insurance coverage or disability. While physicians understandably want to help patients, such misrepresentation undermines physicians' credibility and violates their integrity.

GIFTS FROM PHARMACEUTICAL COMPANIES

Physicians may be offered gifts ranging from pens and notepads to lavish entertainment. Critics worry that any gift from drug companies may impair objectivity, increase the cost of health care, and give the appearance of conflict of interest. A helpful rule of thumb is to consider whether patients would approve if they knew physicians had accepted such gifts.

OCCUPATIONAL RISKS

Some health care workers, fearing fatal occupational infections, refuse to care for persons with HIV infection or multidrug-resistant tuberculosis. Such fears about personal safety need to be acknowledged, and institutions should reduce occupational risk by providing proper training, equipment, and supervision. Physicians should provide appropriate care within their clinical expertise, despite personal risk.

MISTAKES

Mistakes are inevitable in clinical medicine. They may cause serious harm to patients or result in substantial changes in management. Physicians and students may fear that disclosing such mistakes could damage their careers. Without disclosure, however, patients cannot understand their clinical situation or make informed choices about subsequent care. Similarly, unless attending physicians are informed of trainees' mistakes, they cannot provide optimal care and help trainees learn from mistakes.

LEARNING CLINICAL SKILLS

Learning clinical medicine, particularly learning to perform invasive procedures, may present inconvenience or risk to patients. To ensure patient cooperation, students may be introduced as physicians, or patients may not be told that trainees will be performing procedures. Such misrepresentation undermines trust, may lead to more elaborate deception, and makes it difficult for patients to make informed choices about their care. Patients should be told who is providing care, what benefits and burdens can be attributed to trainees, and how trainees are supervised. Most patients, when informed, allow trainees to play an active role in their care.

IMPAIRED PHYSICIANS

Physicians may hesitate to intervene when colleagues impaired by alcohol abuse, drug abuse, or psychiatric or medical illness place patients at risk. However, society relies on physicians to regulate themselves. If colleagues of an impaired physician do not take steps to protect patients, no one else may be in a position to do so.

CONFLICTS FOR TRAINEES

Medical students and residents may fear that they will receive poor grades or evaluations if they act on the patient's behalf by disclosing mistakes, avoiding misrepresentation of their role, and reporting impaired colleagues. Discussing such dilemmas with more senior physicians can help trainees check their interpretation of the situation and obtain advice and assistance.

ADDITIONAL ETHICAL ISSUES

MAINTAINING CONFIDENTIALITY

Maintaining the confidentiality of medical information respects patients' autonomy and privacy, encourages them to seek treatment and to discuss their problems candidly, and prevents discrimination. Physicians need to guard against inadvertent breaches of confidentiality, as when talking about patients in elevators. Maintaining confidentiality is not an absolute rule. The law may require physicians to override confidentiality in order to protect third parties, for example, reporting to government officials persons with specified infectious conditions, such as tuberculosis and syphilis; persons with gunshot wounds; and victims of elder abuse and domestic violence. Computerized medical records raise additional concerns because breaches of confidentiality may affect many patients.

ALLOCATING RESOURCES JUSTLY

Allocation of limited health care resources is problematic. Ideally, allocation decisions should be made as public policy, with physician input. At the bedside, physicians generally should act as patient advocates within constraints set by society, reasonable insurance coverage, and sound practice. *Ad hoc* rationing by the individual physician at the bedside may be inconsistent, discriminatory, and ineffective. In some cases, however, two patients may compete for the same limited resources, such as physician time or a bed in intensive care. When this occurs, physicians should ration their time and resources according to patients' medical needs and the probability of benefit.

ASSISTANCE WITH ETHICAL ISSUES

Discussing perplexing ethical issues with other members of the health care team, colleagues, or the hospital ethics committee often clarifies issues and suggests ways to improve communication and to deal with strong emotions. When struggling with difficult ethical issues, physicians may need to reevaluate their basic convictions, tolerate uncertainty, and maintain their integrity while respecting the opinions of others.

(Bibliography omitted in Palm version)

Back to Table of Contents

3. DECISION-MAKING IN CLINICAL MEDICINE - *Daniel B. Mark*

To the medical student who requires 2 h to collect a patient's history and perform a physical examination, and several additional hours to organize them into a coherent presentation, the experienced clinician's ability to reach a diagnosis and decide on a management plan in a fraction of the time seems extraordinary. While medical knowledge and experience play a significant role in the senior clinician's ability to arrive at a differential diagnosis and plan quickly, much of the process involves skill in clinical decision-making. The first goal of this chapter is to provide an introduction to the study of clinical reasoning.

Equally bewildering to the student are the proper use of diagnostic tests and the integration of the results into the clinical assessment. The novice medical practitioner typically uses a "shotgun" approach to testing, hoping to a hit a target without knowing exactly what that target is. The expert, on the other hand, usually has a specific target in mind and efficiently adjusts the testing strategy to it. The second goal of this chapter is to review briefly some of the crucial basic statistical concepts that govern the proper interpretation and use of diagnostic tests; quantitative tools available to assist in clinical decision-making will also be discussed.

CLINICAL DECISION-MAKING

CLINICAL REASONING

The most important clinical actions are not procedures or prescriptions but the judgments from which all other aspects of clinical medicine flow. In the modern era of large randomized trials, it is easy to overlook the importance of this elusive mental activity and focus instead on the algorithmic practice guidelines constructed to improve care. One reason for this apparent neglect is that much more research has been done on how doctors *should* make decisions (e.g., using a Bayesian model discussed below) than on how they actually *do*. Thus, much of what we know about clinical reasoning comes from empirical studies of nonmedical problem-solving behavior.

Despite the great technological advances of the twentieth century, uncertainty still plays a pivotal role in all aspects of medical decision-making. We may know that a patient does not have long to live, but we cannot be certain how long. We may prescribe a potent new receptor blocker to reverse the course of a patient's illness, but we cannot be certain that the therapy will do so without side effects. Uncertainty in medical outcomes creates the need for probabilities and other mathematical/statistical tools to help guide decision-making. (These tools are reviewed later in the chapter.)

Uncertainty is compounded by the information overload that characterizes modern medicine. Today's experienced clinician needs close to 2 million pieces of information to practice medicine. Doctors subscribe to an average of 7 journals, representing over 2500 new articles each year. Computers offer the obvious solution both for management of information and for better quantitation and management of the daily uncertainties of medical care. While the technology to computerize medical practice is available, many practical problems remain to be solved before patient information can be standardized and integrated with medical evidence on a single electronic platform.

The following three examples introduce the subject of clinical reasoning:

· A 46-year-old man presents to his internist with a chief complaint of hemoptysis. The physician knows that the differential diagnosis of hemoptysis includes over 100 different conditions, including cancer and tuberculosis (Chap. 33). The examination begins with some general background questions, and the patient is asked to describe his symptoms and their chronology. By the time the examination is completed, and even before any tests are run, the physician has formulated a working diagnostic hypothesis and planned a series of steps to test it. In an otherwise healthy and nonsmoking patient recovering from a viral bronchitis, the doctor's hypothesis would be that the acute bronchitis is responsible for the small amount of blood-streaked sputum the patient observed. In this case, a chest x-ray and purified protein derivative (PPD) skin test may be sufficient.

· A second 46-year-old patient with the same chief complaint who has a 100-pack-year smoking history, a productive morning cough, and episodes of blood-streaked sputum may generate the principal diagnostic hypothesis of carcinoma of the lung. Consequently, along with the chest x-ray andPPD skin test, the physician refers this patient for bronchoscopy.

· A third 46-year-old patient with hemoptysis who is from a developing country is evaluated with an echocardiogram as well, because the physician thinks she hears a soft diastolic rumble at the apex on cardiac auscultation, suggesting rheumatic mitral stenosis.

These three vignettes illustrate two aspects of expert clinical reasoning: (1) the use of cognitive shortcuts, or *heuristics*, as a way to organize the complex unstructured material that is collected in the clinical evaluation; and (2) the use of diagnostic hypotheses to consolidate the information and indicate appropriate management steps.

THE USE OF COGNITIVE SHORTCUTS

Heuristics reduce the complexity of a problem to a manageable level. Psychologists have found that people rely on three basic types of heuristics. For example, when assessing a patient, clinicians often weigh the probability that this patient's clinical features match those of the class of patients with the leading diagnostic hypotheses being considered. In other words, the clinician is searching for the diagnosis for which the patient appears to be a representative example; this cognitive shortcut is called the *representativeness heuristic*. It may take only a few characteristics from the history for an expert clinician using the representativeness heuristic to arrive at a sound diagnostic hypothesis. For example, an elderly patient with new-onset fever, cough productive of copious sputum, unilateral pleuritic chest pain, and dyspnea is readily identified as fitting the pattern for acute pneumonia, probably of bacterial origin. Evidence of focal pulmonary consolidation on the physical examination will increase the clinician's confidence in the diagnosis because it fits the expected pattern of acute bacterial pneumonia. Knowing this allows the experienced clinician to conduct an efficient, directed, and therapeutically productive patient evaluation although there may be little else in the history or physical examination of direct relevance. The inexperienced medical student or resident, who has not yet learned the patterns most prevalent in

clinical medicine, must work much harder to achieve the same result and is often at risk of missing the important clinical problem in a sea of compulsively collected but unhelpful data.

However, physicians using the representativeness heuristic can reach erroneous conclusions if they fail to consider the underlying prevalence of two competing diagnoses. Consider a patient with pleuritic chest pain, dyspnea, and a low-grade fever. A clinician might consider acute pneumonia and acute pulmonary embolism to be the two leading diagnostic alternatives. Clinicians using the representativeness heuristic might judge both diagnostic candidates to be equally likely, although to do so would be wrong if pneumonia was much more prevalent in the underlying population. Mistakes may also result from a failure to consider that a pattern based on a small number of prior observations will likely be less reliable than one based on larger samples.

A second commonly used cognitive shortcut, the *availability heuristic*, involves judgments made on the basis of how easily prior similar cases or outcomes can be brought to mind. For example, the experienced clinician may recall 20 elderly patients seen over the past few years who presented with painless dyspnea of acute onset and were found to have acute myocardial infarction. The novice clinician may spend valuable time seeking a pulmonary cause for the symptoms before considering and discovering the cardiac diagnosis. In this situation, the patient's clinical pattern does not fit the expected pattern of acute myocardial infarction, but experience with this atypical presentation, and the ability to recall it, can help direct the physician to the diagnosis.

Errors with the availability heuristic can come from several sources of recall bias. For example, rare catastrophes are likely to be remembered with a clarity and force out of proportion to their value, and recent experience is, of course, easier to recall and therefore more influential on clinical judgments.

The third commonly used cognitive shortcut, the *anchoring heuristic*, involves estimating a probability by starting from a familiar point (the anchor) and adjusting to the new case from there. For example, a clinician may judge the probability of colorectal cancer to be extremely high after an elevated screening carcinoembryonic antigen (CEA) result because the prediction of colorectal cancer is anchored to the test result. Yet, as discussed below, this prediction would be inaccurate if the clinical picture of the patient being tested indicates a low probability of disease (for example, a 30-year-old woman with no risk factors). Anchoring can be a powerful tool for diagnosis but is often used incorrectly (see "Measures of Disease Probability and Bayes' Theorem," below).

DIAGNOSTIC HYPOTHESIS GENERATION

Cognitive scientists studying the thought processes of expert clinicians have observed that clinicians group data into packets or "chunks," which are stored in their memories and manipulated to generate diagnostic hypotheses. Because short-term memory can typically hold only 7 to 10 items at a time, the number of packets that can be actively integrated into hypothesis-generating activities is similarly limited. The cognitive shortcuts discussed above play a key role in the generation of diagnostic hypotheses, many of which are discarded as rapidly as they are formed.

A diagnostic hypothesis sets a context for diagnostic steps to follow and provides testable predictions. For example, if the enlarged and quite tender liver felt on physical examination is due to acute hepatitis (the hypothesis), certain specific liver function tests should be markedly elevated (the prediction). If the tests come back normal, the hypothesis may need to be discarded or substantially modified.

One of the factors that makes teaching diagnostic reasoning so difficult is that expert clinicians do not follow a fixed pattern in patient examinations. From the outset, they are generating, refining, and discarding diagnostic hypotheses. The questions they ask in the history are driven by the hypotheses they are working with at the moment. Even the physical examination is driven by specific questions rather than a preordained checklist. While the student is palpating the abdomen of the alcoholic patient, waiting for a finding to strike him, the expert clinician is on a focused search mission. Is the spleen enlarged? How big is the liver? Is it tender? Are there any palpable masses or nodules? Each question focuses the attention of the examiner to the exclusion of all other inputs until answered, allowing the examiner to move on to the next specific question.

Negative findings are often as important as positive ones in establishing and refining diagnostic hypotheses. Chest discomfort that is not provoked or worsened by exertion in an active patient reduces the likelihood that chronic ischemic heart disease is the underlying cause. The absence of a resting tachycardia and thyroid gland enlargement reduces the likelihood of hyperthyroidism in a patient with paroxysmal atrial fibrillation.

While the representativeness and availability heuristics may play the major roles in shaping early diagnostic hypotheses, the acuity of a patient's illness can also be very influential. For example, clinicians are taught to consider aortic dissection routinely as a possible cause of acute severe chest discomfort along with myocardial infarction, even though the typical history of dissection is different from myocardial infarction and dissection is far less prevalent (Chap. 247). This recommendation is based on the recognition that a relatively rare but catastrophic diagnosis like aortic dissection is very difficult to make unless it is explicitly considered. If the clinician fails to elicit any of the characteristic features of dissection by history and finds equivalent blood pressures in both arms and no pulse deficits, he or she may feel comfortable in discarding the aortic dissection hypothesis. If, however, the chest x-ray shows a widened mediastinum, the hypothesis may be reinstated and a diagnostic test ordered [e.g., thoracic computed tomography (CT) scan, transesophageal echocardiogram] to evaluate it more fully. In noncritical situations, the prevalence of potential alternative diagnoses should play a much more prominent role in diagnostic hypothesis generation. The value of conducting a rapid systematic clinical survey of symptoms and organ systems to avoid missing important but inapparent clues cannot be overstated.

Because the generation and evaluation of appropriate diagnostic hypotheses is a skill that not all clinicians possess to an equal degree, errors in this process can occur, and in the patient with serious acute illness these may lead to tragic consequences. Consider the following hypothetical example. A 45-year-old male patient with a 3-week history of a "flulike" upper respiratory infection (URI) presented to his physician with symptoms of dyspnea and a productive cough. Based on the presenting complaint, the clinician pulled out a "URI Assessment Form" to improve quality and efficiency of care. The physician quickly completed the examination components outlined on this

structured form, noting in particular the absence of fever and a clear chest examination. He then prescribed an antibiotic for presumed bronchitis, showed the patient how to breathe into a paper bag to relieve his "hyperventilation," and sent him home with the reassurance that his illness was not serious. After a sleepless night with significant dyspnea unrelieved by rebreathing into a bag, the patient developed nausea and vomiting and collapsed. He was brought into the Emergency Department in cardiac arrest and could not be resuscitated. Autopsy showed a posterior wall myocardial infarction and a fresh thrombus in an atherosclerotic right coronary artery. What went wrong? The clinician decided, even before starting the history, that the patient's complaints were not serious. He therefore felt confident that he could perform an abbreviated and focused examination using the URI assessment protocol rather than considering the full range of possibilities and performing appropriate tests to confirm or refute his initial hypotheses. In particular, by concentrating on the "URI," the clinician failed to elicit the full dyspnea history, which would have suggested a far more serious disorder, and did not even search for other symptoms that could have directed him to the correct diagnosis.

This example illustrates how patients can diverge from textbook symptoms and the potential consequences of being unable to adapt the diagnostic process to real-world challenges. The expert, while recognizing that common things occur commonly, approaches each evaluation on high alert for clues that the initial diagnosis may be wrong. Patients often provide information that "does not fit" with any of the leading diagnostic hypotheses being considered. Distinguishing real clues from false trails can only be achieved by practice and experience. A less experienced clinician who tries to be too efficient (as in the above example) can make serious judgment errors.

MAJOR INFLUENCES ON CLINICAL DECISION-MAKING

More than a decade of research on variations in clinician practice patterns has shed much light on forces that shape clinical decisions. The use of heuristic "shortcuts," as detailed above, provides a partial explanation, but several other key factors play an important role in shaping diagnostic hypotheses and management decisions. These factors can be grouped conceptually into three overlapping categories: (1) factors related to physician personal characteristics and practice style, (2) factors related to the practice setting, and (3) economic incentive factors.

Practice Style Factors One of the key roles of the physician in medical care is to serve as the patient's agent to ensure that necessary care is provided at a high level of quality. Factors that influence this role include the physician's knowledge, training, and experience. It is obvious that physicians cannot practice evidence-based medicine if they are unfamiliar with the evidence. As would be expected, specialists generally know the evidence in their field better than do generalists. Surgeons may be more enthusiastic about recommending surgery than medical doctors because their belief in the beneficial effects of surgery is stronger. For the same reason, invasive cardiologists are much more likely to refer chest pain patients for diagnostic catheterization than are noninvasive cardiologists or generalists. The physician beliefs that drive these different practice styles are based on personal experience, recollection, and interpretation of the available medical evidence. For example, heart failure specialists are much more likely than generalists to achieve target angiotensin-converting enzyme (ACE) inhibitor

therapy in their heart failure patients because they are more familiar with what the targets are (as defined by large clinical trials), have more familiarity with the specific drugs (including dosages and side effects), and are less likely to overreact to foreseeable problems in therapy such as a rise in creatinine levels or symptomatic hypotension. Other intriguing research has shown a wide distribution of acceptance times of antibiotic therapy for peptic ulcer disease following widespread dissemination of the "evidence" in the medical literature. Some gastroenterologists accepted this new therapy before the evidence was clear (reflecting, perhaps, an aggressive practice style), and some gastroenterologists lagged behind (a conservative practice style, associated in this case with older physicians). As a group, internists lagged several years behind gastroenterologists.

The opinion of influential leaders can also have an important effect on practice patterns. Such influence can occur at both the national level (e.g., expert physicians teaching at national meetings) and the local level (e.g., local educational programs, "curbside consultants"). Opinion leaders do not have to be physicians. When conducting rounds with clinical pharmacists, physicians are less likely to make medication errors and more likely to use target levels of evidence-based therapies.

The patient's welfare is not the only concern that drives clinical decisions. The physician's perception about the risk of a malpractice suit resulting from either an erroneous decision or a bad outcome creates a style of practice referred to as *defensive medicine*. This practice involves using tests and therapies with very small marginal returns to preclude future criticism in the event of an adverse outcome. For example, a 40-year-old woman who presents with a long-standing history of intermittent headache and a new severe headache along with a normal neurologic examination has a very low likelihood of structural intracranial pathology. Performance of a headCT or magnetic resonance imaging (MRI) scan in this situation would constitute defensive medicine. On the other hand, the results of the test could provide reassurance to an anxious patient.

Practice Setting Factors Factors in this category relate to the physical resources available to the physician's practice and the practice environment. *Physician-induced demand* is a term that refers to the repeated observation that physicians have a remarkable ability to accommodate to and employ the medical facilities available to them. A classic early study in this area showed that physicians in Boston had an almost 50% higher hospital admission rate than did physicians in New Haven, despite there being no obvious differences in the health of the cities' inhabitants. The physicians in New Haven were not aware of using fewer hospital beds for their patients, nor were the Boston physicians aware of using less stringent criteria to admit patients.

Other environmental factors that can influence decision-making include the local availability of specialists for consultations and procedures, "high tech" facilities such as angiography suites, a heart surgery program, andMRImachines.

Economic Incentives Economic incentives are closely related to the other two categories of practice-modifying factors. Financial issues can exert both stimulatory and inhibitory influences on clinical practice. In general, physicians are paid on a fee-for-service, capitation, or salary basis (Chap. 4). In fee-for-service, the more the physician does, the more the physician gets paid. The incentive in this case is to do

more. When fees are reduced (discounted fee-for-service), doctors tend to increase the number of services billed for. Capitation, in contrast, provides a fixed payment per patient per year, encouraging physicians to take on more patients but to provide each patient with fewer services. Expensive services are more likely to be affected by this type of incentive than inexpensive preventive services. Salary compensation plans pay physicians the same regardless of the amount of clinical work performed. The incentive here is to see fewer patients. Recognizing these powerful shapers of physician behavior, managed care plans have begun to explore combinations of the three reimbursement types with the goal of improving individual physician productivity while restraining their use of expensive tests and therapies.

In summary, expert clinical decision-making can be appreciated as a complex interplay between cognitive devices used to simplify large amounts of complex information interacting with physician biases reflecting education, training, and experience, all of which are shaped by powerful, sometimes perverse, external forces. In the next section, we will review a set of statistical tools and concepts that can assist in making clinical decisions under uncertainty.

QUANTITATIVE METHODS TO AID CLINICAL DECISION-MAKING

The process of medical decision-making can be divided into two parts: (1) defining the available courses of action and estimating the likely outcomes with each, and (2) assessing the desirability of the outcomes. The former task involves integrating key information about the patient along with relevant evidence from the medical literature to create the structure of a decision problem. The remainder of this chapter will present some quantitative tools to assist the clinician in these activities. These tools can be divided into those that assist the clinician in making better outcome predictions, which are then used to make decisions, and those that support the decision process directly. While these tools are not yet used routinely in daily clinical practice, the computerization of medicine is creating the required substrate for their future widespread dissemination.

QUANTITATIVE MEDICAL PREDICTIONS

Diagnostic Testing The purpose of performing a test on a patient is to reduce uncertainty about the patient's diagnosis or prognosis and to aid the clinician in making management decisions. Although diagnostic tests are commonly thought of as laboratory tests (e.g., measurement of serum amylase level) or procedures (e.g., colonoscopy or bronchoscopy), any technology that changes our understanding of the patient's problem qualifies as a diagnostic test. Thus, even the history and physical examination can be considered a form of diagnostic test. In clinical medicine, it is common to reduce the results of a test to a dichotomous outcome, such as positive or negative, normal or abnormal. In many cases, this simplification results in the waste of useful information. However, such simplification makes it easier to demonstrate some of the quantitative ways in which test data can be used.

To characterize the accuracy of diagnostic tests, four terms are routinely used (Table 3-1). The *true-positive rate*, i.e., the sensitivity, provides a measure of how well the test correctly identifies patients with disease. The *false-negative rate* is calculated as (1-sensitivity). The *true-negative rate*, i.e., the specificity, reflects how well the test

correctly identifies patients without disease. The *false-positive rate* is (1- specificity). A perfect test would have a sensitivity of 100% and a specificity of 100% and would completely separate patients with disease from those without it.

Calculating sensitivity and specificity require selection of a cutpoint value for the test to separate "normal" from "diseased" subjects. As the cutpoint is moved to improve sensitivity, specificity typically falls and vice versa. This dynamic tradeoff between more accurate identification of subjects with versus those without disease is often displayed graphically as a receiver operating characteristic (ROC) curve. An ROC curve plots sensitivity (*y*-axis) versus 1 -specificity (*x*-axis). Each point on the curve represents a potential cutpoint with an associated sensitivity and specificity value. The area under the ROC curve is often used as a quantitative measure of the information content of a test. Values range from 0.5 (no diagnostic information at all, test is equivalent to flipping a coin) to 1.0 (perfect test).

In the diagnostic testing literature, ROC areas are often used to compare alternative tests. The test with the highest area (i.e., closest to 1.0) is presumed to be the most accurate. However, ROC curves are not a panacea for evaluation of diagnostic test utility. Like Bayes' theorem, they are typically focused on only one possible test parameter (e.g., ST segment response in a treadmill exercise test) to the exclusion of other potentially relevant data. In addition, ROC area comparisons do not simulate the way test information is actually used in clinical practice. Finally, biases in the underlying population used to generate the ROC curves (e.g., related to an unrepresentative test sample) can bias the ROC area and the validity of a comparison among tests.

Measures of Disease Probability and Bayes' Theorem Unfortunately, there are no perfect tests; after every test is completed the true disease state of the patient remains uncertain. Quantitating this residual uncertainty can be done with Bayes' theorem. This theorem provides a simple mathematical way to calculate the posttest probability of disease from three parameters: the pretest probability of disease, the test sensitivity, and the test specificity (Table 3-2). The pretest probability is a quantitative expression of the confidence in a diagnosis before the test is performed. In the absence of more relevant information it is usually estimated from the prevalence of the disease in the underlying population. For some common conditions, such as coronary artery disease (CAD), nomograms and statistical models have been created to generate better estimates of pretest probability from elements of the history and physical examination. The posttest probability, then, is a revised statement of the confidence in the diagnosis, taking into account both what was known before and after the test.

To understand how Bayes' theorem creates this revised confidence statement, it is useful to examine a nomogram version of Bayes' theorem that uses the same three parameters to predict the posttest probability of disease (Fig. 3-1). In this nomogram, the accuracy of the diagnostic test in question is summarized by the likelihood ratio for a positive test, which is the ratio of the true-positive rate to the false-positive rate [or sensitivity/(1 - specificity)]. For example, a test with a sensitivity of 0.90 and a specificity of 0.90 has a likelihood ratio of 0.90/(1 - 0.90), or 9. Thus, for this hypothetical test, a "positive" result is 9 times more likely in a patient with the disease than in a patient without it. The more accurate the test, the higher the likelihood ratio. However, if sensitivity is excellent but specificity is less so, the likelihood ratio will be substantially

reduced (e.g., with a 90% sensitivity but a 60% specificity, the likelihood ratio is 2.25). Most tests in medicine have likelihood ratios for a positive result between 1.5 and 20.

Consider two tests commonly used in the diagnosis ofCAD, an exercise treadmill and an exercise thallium-201 single photon emission CT (SPECT) test (Chap. 244). Meta-analysis has shown the treadmill to have an average sensitivity of 66% and an average specificity of 84%, yielding a likelihood ratio of 4.1 [0.66/(1 - 0.84)]. If we use this test on a patient with a pretest probability of CAD of 10%, the posttest probability of disease following a positive result rises only to about 30%. If a patient with a pretest probability of CAD of 80% has a positive test result, the posttest probability of disease is about 95%.

The exercise thalliumSPECTtest is a more accurate test for the diagnosis ofCAD. For our purposes, assume that it has both a sensitivity and specificity of 90%, yielding a likelihood ratio of 9.0 [0.90/(1 - 0.90)]. If we again test our low pretest probability patient and he has a positive test, using Fig. 3-1 we can demonstrate that the posttest probability of CAD rises from 10 to 50%. However, from a decision-making point of view, the more accurate test has not been able to improve diagnostic confidence enough to change management. In fact, the test has moved us from being fairly certain that the patient did not have CAD to being completely undecided (a 50:50 chance of disease). In a patient with a pretest probability of 80%, using the more accurate thallium SPECT test raises the posttest probability to 97% (compared with 95% for the exercise treadmill). Again, the more accurate test does not provide enough improvement in posttest confidence to alter management, and neither test has improved much upon what was known from clinical data alone.

If the pretest probability is low (e.g., £20%), even a positive result on a very accurate test will not move the posttest probability to a range high enough to rule in disease (e.g., ³80%). Conversely, with a high pretest probability, a negative test will not adequately rule out disease. Thus, the largest gain in diagnostic confidence from a test occurs when the clinician is most uncertain before performing it (e.g., pretest probability between 30 and 70%). For example, if a patient has a pretest probability forCAD of 50%, a positive exercise treadmill test will move the posttest probability to 80% and a positive exercise thalliumSPECTtest will move it to 90% (Fig. 3-1).

Bayes' theorem, as presented above, employs a number of important simplifications that should be considered. First, few tests have only two useful outcomes, positive or negative, and many tests provide numerous pieces of data about the patient. Even if these can be integrated into a summary result, multiple levels of useful information may be present (e.g., strongly positive, positive, indeterminate, negative, strongly negative). While Bayes' theorem can be adapted to this more detailed test result format, it is computationally complex to do so. Second, Bayes' theorem assumes that the information from the test is completely unique and nonoverlapping with information used to estimate the pretest probability. This independence assumption, however, is often wrong. In many cases, test results are correlated with patient characteristics. For example, the findings of cardiomegaly and pulmonary edema on chest x-ray are correlated with the historic features of heart failure and with the physical findings of a displaced left ventricular apical impulse, an S_3gallop, and rales. The unique predictive information contributed by the test in this case (the chest x-ray) is only a fraction of its

total information because much had already been learned about the probability of heart failure before the test was done.

Finally, it has long been thought that sensitivity and specificity are prevalence-independent parameters of test accuracy, and many texts still make this assertion. This statistically useful assumption, however, is clinically wrong. For example, a treadmill exercise test has a sensitivity in a population of patients with one-vesselCAD of around 30%, whereas the sensitivity in severe three-vessel CAD approaches 80%. Thus, the best estimate of sensitivity to use in a particular decision will often vary depending on the distribution of disease stages present in the tested population. A hospitalized population typically has a higher prevalence of disease and in particular a higher prevalence of more advanced disease stages than an outpatient population. As a consequence, test sensitivity will tend to be higher in hospitalized patients, whereas test specificity will be higher in outpatients.

Statistical Prediction Models Bayes' theorem, as presented above, deals with a clinical prediction problem that is unrealistically simple relative to most problems a clinician faces. Prediction models, based on multivariable statistical models, can handle much more complex problems and substantially enhance predictive accuracy for specific situations. Their particular advantage is the ability to take into account many overlapping pieces of information and assign a relative weight to each based on its unique contribution to the prediction in question. For example, a logistic regression model to predict the probability ofCADtakes into account all of the relevant independent factors from the clinical examination and diagnostic testing instead of the small handful of data that clinicians can manage in their heads or with Bayes' theorem. However, despite this strength, the models are too complex computationally to use without a calculator or computer (although this limit may be overcome when medicine is practiced from a fully computerized platform.) To date, only a handful of prediction models have been developed and properly validated. The importance of independent validation in a population separate from the one used to develop the model cannot be overstated. Unfortunately, most published models have not been properly validated, making their utility in clinical practice uncertain at best.

When statistical models have been compared directly with expert clinicians, they have been found to be more consistent, as would be expected, but not significantly more accurate. Their biggest promise, then, would seem to be to make less-experienced clinicians more accurate predictors of outcome.

DECISION SUPPORT TOOLS

DECISION SUPPORT SYSTEMS

Over the past 30 years, many attempts have been made to develop computer systems to help clinicians make decisions and manage patients. Conceptually, computers offer a very attractive way to handle the vast information load that today's physicians face. The computer can help by making accurate predictions of outcome, simulating the whole decision process, or providing algorithmic guidance. Computer-based predictions using Bayesian or statistical regression models inform a clinical decision but do not actually reach a "conclusion" or "recommendation." Artificial intelligence systems attempt to

simulate or replace human reasoning with a computer-based analogue. To date, such approaches have achieved only limited success. Reminder or protocol-directed systems do not make predictions but use existing algorithms, such as practice guidelines, to guide clinical practice. In general, however, decision support systems have shown little impact on practice. Reminder systems, although not yet in widespread use, have shown the most promise, particularly in correcting drug dosing and in promoting guideline adherence. The full potential of these approaches will only be achieved when computers are fully integrated into medical practice.

DECISION ANALYSIS

Compared with the methods discussed above, decision analysis represents a completely different approach to decision support. Its principal application is in decision problems that are complex and involve a substantial risk, a high degree of uncertainty in some key area, or an idiosyncratic feature that does not "fit" the available evidence. Three general steps are involved. First, the decision problem must be clearly defined. Second, the elements of the decision must be made explicit. This involves specifying the alternatives being considered, their relevant outcomes, the probabilities attached to each outcome, and the relative desirability (called "utility") of each outcome. Cost can also be assigned to each branch of the decision tree, allowing calculation of cost-effectiveness (Chap. 4).

An example of a decision tree used to evaluate strategies for management of the risk of infective endocarditis after catheter-associated *Staphylococcus aureus* bacteremia is shown in Fig. 3-2. Approximately 35,000 cases of *S. aureus* bacteremia occur each year in the United States. The development of complicating endocarditis, which occurs in about 6% of cases, is associated with high morbidity (31% mortality, 21% stroke rate) and medical costs. The three choices for management of the bacteremia are (1) transesophageal echocardiography (TEE), (2) a 4-week course of intravenous antibiotics (long-course), or (3) a 2-week course of intravenous antibiotics (short-course). In the TEE strategy, a 4-week course of antibiotics is given if endocarditis is evident and a 2-week course is given if it is not. With each strategy, there is a risk that the patient will develop endocarditis with or without major complications. In this analysis, the longest quality-adjusted survival (5.47 quality-adjusted life-years) was associated with the 4-week antibiotic course strategy, which also had the highest costs ($14,136 per patient), whereas the lowest costs ($9830 per patient) and worst outcomes (5.42 quality-adjusted life-years) were associated with the 2-week antibiotic course strategy. From a clinical point of view (ignoring costs), the 4-week antibiotic course was best. From a cost-effectiveness point of view, the TEE strategy (5.46 quality-adjusted life-years and $10,051 per patient costs) provided the best balance of added benefits and costs. Thus, decision analysis can be extremely helpful in clarifying tradeoffs in outcomes and costs in difficult management areas such as the above where it is highly unlikely that an adequate randomized trial will ever be done.

The data needed to fill in a decision tree (Fig. 3-2) are typically cobbled together from a variety of sources, including the literature (randomized trials, meta-analyses, observational studies) and expert opinion. Once the decision tree is finished, the decision is "analyzed" by calculating the average value of each limb of the tree. The decision arm with the highest net value (or expected utility) is the preferred choice. The

value of this exercise, however, is not so much in developing a prescription for action as it is in exploring the key elements and pressure points of a complex or difficult decision. The process of building the decision tree forces the analyst to be explicit about the choices being considered and all their relevant outcomes. Areas of high uncertainty are readily identified. Sensitivity analyses are an integral part of decision analysis and involve systematically varying the value of each key parameter in the model alone (one-way sensitivity analysis,) in pairs (two-way), or in higher combinations (multivariable) to assess the impact on choice of preferred management strategy. In the above example, varying the incidence of endocarditis resulting from *S. aureus* bacteremia from 3% to over 50% had no impact on the choice ofTEE as the preferred strategy.

User friendly personal computer-based software packages now make the creation and analysis of decision trees much more straightforward than in the past. However, the process is still too cumbersome and time-consuming to be used on a routine basis. When medicine is practiced from a fully computerized platform, a library of prestructured decision trees with user modifiable values can be made available to support practitioners working with individual patients.

CONCLUSIONS

In this era of evidence-based medicine, it is tempting to think that all the difficult decisions practitioners face have been or soon will be solved and digested into practice guidelines and computerized reminders. For the foreseeable future, however, such is not the case. Meta-analyses cannot generate evidence where there are no adequate randomized trials, and most of what clinicians face will never be thoroughly tested in a randomized trial. Excellent clinical reasoning skills and experience supplemented by well-designed quantitative tools and a keen appreciation for individual patient preferences will continue to be of paramount importance in the professional life of medical practitioners for years to come.

(Bibliography omitted in Palm version)

Back to Table of Contents

4. ECONOMIC ISSUES IN CLINICAL MEDICINE - *Daniel B. Mark*

The United States has the distinction of having some of the best medical care of any technologically advanced country. We have many of the best hospitals and doctors in the world. The research pipeline is full of significant new therapeutic advances, with revolutionary genetic-based therapies perhaps only a decade away. Our citizens largely subscribe to the principle that excellent medical care should be available to all, regardless of ability to pay. Yet we also have over 43 million people (most of them employed and earning minimal wages) without any health insurance and many more who are inadequately insured. Since the collapse of the Clinton health care reform efforts in 1994, U.S. health policy has been directed by marketplace forces that have created powerful and sometimes perverse incentives in medicine: Health insurance companies that use every available means to avoid insuring sick people; "managed care" programs that really only manage costs; doctors who are provided incentives to provide less medical care; and pharmaceutical companies that develop powerful and expensive new drugs priced beyond the reach of many of the elderly and chronically ill who need them most.

Facing such powerful and chaotic forces, physicians tend to focus narrowly on what they are most comfortable with, taking care of individual patients and conducting academic investigations. Many doctors consider economics too arcane for them to grasp and therefore do not even try. Consequently, when presented with economic arguments and evidence they are often unable to discriminate the legitimate from the fallacious. More importantly, they are ill equipped to defend their patients' interests in the crucible of cost containment that characterizes the modern managed care era.

This chapter has two goals: first, to provide a brief introduction to some of the larger economic forces that shape modern medical practices, and second, to introduce the economic tools that are used for assessing the value of medical practices, including cost effectiveness analysis.

HEALTH CARE SPENDING AND FINANCING

HOW MUCH IS SPENT ON HEALTH CARE?

In 1997, the United States spent $1.1 trillion on its health care system, representing 13.5% of the gross domestic product (GDP) (a crude measure of national income). Most of this ($969 billion) was spent on personal health care: 34% went to hospitals, 20% to physicians, 7% to nursing homes, and 8% to outpatient pharmaceuticals. In comparison, Canada and Western European countries spend a substantially smaller portion (6 to 10%) of their national income on health care but their citizens appear to be equally healthy, at least by crude metrics such as life expectancy and infant mortality rates. Economists and politicians have for years used such data to argue that the United States spends too much on health care. The issue of how much to spend is an inherently political one, however, and the discipline of economics has little to say about it.

WHO PAYS FOR HEALTH CARE?

Two major factors are continually driving up the costs of medical care: introduction into medical practice of new medical technologies (drugs, devices, procedures) that have a high price tag, and the aging of the U.S. population (since older people require more medical care than younger ones). These costs are distributed unevenly across society. In 1997, the government paid about 47% of the total national health care bill (75% federal, 25% states), private insurance paid about 32%, and individuals paid 17%. The government, of course, gets its money from taxpayers and uses the health care segment of its budget to pay for the Medicare and Medicaid programs (discussed below). To respond to rising medical costs, the government can increase taxes or redistribute funds from other programs such as defense and education. Neither of these options are politically attractive. Alternatively, because of its size in the medical marketplace, the government can impose lower prices on providers to make the available funds go farther (see "Cost-Containment Strategies," below). Much of the private insurance bill is subsidized by employers through their employee benefits packages. As medical costs go up, health insurance costs also rise and businesses must either pass on higher premium and copayment costs to their employees, raise their prices (potentially impairing their competitive position in the marketplace), or reduce their profit margin (a very unpopular move with stockholders). Like the government, businesses may also negotiate lower prices with health care providers and/or health insurance plans.

PUBLIC FINANCING OF HEALTH CARE

The public sector (i.e., government as an agent for society) finances the Medicare and Medicaid programs as well as the Veteran's Administration Hospital system, the Department of Defense military care system, the Public Health Service, and the Indian Health Service. Of these, Medicare is by far the largest and most influential, with 39 million people receiving health insurance at a total cost in 1997 of $214.6 billion (20% of total national health expenditures). The Medicare program was enacted in 1965 by Congress as an amendment to the Social Security Act of 1935 and was envisioned by President Lyndon B. Johnson as a first step toward universal health insurance in the United States, a key part of his "great society" plan. Its impact on the evolution of the U.S. health care system has been profound. The original congressional act provided health care insurance for the elderly (defined as those 65 and older) who were eligible for social security (i.e., retired workers who had paid into the system during their working years and their dependents). Amendments in 1972 extended coverage to the disabled of all ages (currently numbering around 5 million) and to patients with chronic renal failure (who currently number about 284,000).

Medicare consists of two related insurance programs. The Medicare Hospital Insurance Trust Fund (also known as Part A) covers hospital care and skilled nursing home care and is funded by compulsory federal payroll taxes on employers and employees. Medicare Part B, the Medical Supplementary Insurance Program, covers physician fees as well as laboratory and other diagnostic tests and is funded by general federal tax revenues and patient premiums. Both programs have substantial gaps in coverage, necessitating supplemental insurance (so-called Medigap policies) for those who can afford them. Because of its compulsory income redistribution feature, taking tax money from current workers to pay for health care for elderly citizens (many of whom are on fixed income close to the poverty level), Medicare is both a health insurance program

and a social welfare program designed to combat poverty in the disabled and elderly. In exchange for their tax money, the 150 million workers funding the program are promised the same type of social security when they become elderly (paid for by future generations of workers).

Medicaid is a social insurance program for the poor that is jointly run by the federal and state governments. The federal government gives each state a grant of money for the program based on that state's per capita income (in 1997, this amount totaled $95 billion), and the states pay for the rest ($65 billion in 1997). The program, like Medicare, was enacted by Congress in 1965 as a part of President Johnson's "great society" program. It is larger than Medicare in terms of eligible beneficiaries (41 million people) but smaller in terms of budget ($160 billion, or 12% of the total national health expenditures). Because the requirements to qualify for Medicaid are stringent, many low-income individuals under age 65 (especially the working poor) do not qualify. Eligibility criteria are set by each state within general federal guidelines, and the income and asset tests individuals must meet to qualify vary widely among states. Many of the dollars in the Medicaid program actually pay for care for elderly and disabled Medicare beneficiaries who also qualify for Medicaid on the basis of poverty.

PRIVATE FINANCING OF HEALTH CARE

Approximately 70% of the non-elderly U.S. population is covered by some form of private medical insurance. The feasibility of group insurance for medical care was initially demonstrated in the 1930s by Blue Cross, a franchise of nonprofit groups providing hospitalization insurance in order to help prop up the financially strapped U.S. hospital industry. Blue Shield, a separate organization modeled after Blue Cross, started providing insurance for in-hospital physician services in 1939. During World War II, employee wages were frozen by the government and to entice workers, who were in short supply, some employers started offering health insurance as a fringe benefit. With the feasibility of employer-sponsored group health insurance demonstrated by the experience of the "Blues," commercial insurers began to enter the market. To win the support of doctors and hospitals, insurers agreed to pay "reasonable and customary charges" and to defer all medical management decisions to doctors. This "fee-for-service" reimbursement system, created in the post-World War II era, sowed the seeds of the tremendous inflation observed in the U.S. medical system during the 1970s and 1980s.

The original focus of indemnity insurance plans was to cover individuals against catastrophic financial losses from high medical care bills. Insurance is a contract for protection against specific hazards that are unpredictable for individuals but can be defined with confidence for large groups. "Major medical" health insurance was designed to provide coverage for catastrophic illness, a relatively rare event in most populations. Group coverage is less expensive than individual coverage because it allows the insurance company to diffuse the risk of a large payout among a big pool of individuals who will pay premiums but make no claims. When coverage is shifted from a focus on rare catastrophes to routine maintenance medical care (comprehensive insurance policies), health insurance becomes a means for payment of expected rather than unexpected care. The consequence is higher health insurance premiums. The early appeal of health maintenance organizations (HMOs) was that they appeared to

offer an economically efficient way to provide routine preventive care and to manage the occasional catastrophic illness.

MANAGED CARE

Managed care is a generic term that embraces a wide spectrum of systems for integrating the financing and delivery of health care. Managed care organizations (MCOs) contract with doctors and hospitals to provide comprehensive care to enrolled members for a fixed, prospectively set, premium.HMOs are a form of managed care originally organized between the 1940s and 1960s as an alternative to the prevailing fee-for-service-based private insurance. With the advent of serious medical inflation in the 1970s, the HMO model was promoted by the federal government as a way to control the growth in medical spending. Early enthusiasm for this initiative was limited; in 1984, only 5% of individuals with employer-based health insurance were in an HMO. However, by 1998 that figure had risen to 85%. The exponential growth of managed care started in the 1990s in part as an employer-driven response to the uncontrolled medical inflation of the previous two decades.

The massive increase in demand for managed care by employers and by the Medicare program produced a rapid, and sometimes bewildering, evolution in the managed care industry. One important trend has been the growth of for-profit (i.e., investor owned) managed care companies. Over half ofHMOmembers now belong to a for-profit plan. Investment dollars from Wall Street have made it easier for these plans to respond quickly to increased employer demand for managed care options. However, compared with their not-for-profit counterparts, for-profit HMOs spend a smaller proportion of each premium dollar paying for health care for members (the paradoxically named "medical loss ratio"), since stockholders also have to be paid. As a result, for-profit HMOs are less successful than not-for-profit plans in providing preventive care (a presumed strength of managed care).

Another prevalent trend of the 1990s was the move from traditionalHMOmodels to virtual HMOs, built from contractual relationships with community physicians and hospitals. The three HMO models are the staff model, the group model, and the Independent Practice Association (IPA). The staff model HMO is a vertically integrated organization. That is, it owns its own hospitals, employs all its physicians full time for a set salary, and is focused in a particular geographic area. The group model HMO, exemplified by Group Health Cooperative of Puget Sound, contracts with one or more large multispecialty group practices to care for its patients for a preset capitated reimbursement. These physicians do not care for non-HMO patients. In the IPA model, the HMO contracts with an association of self-employed physicians who maintain their own offices and see both HMO and non-HMO patients. The network model refers to a hybrid of the other three forms of HMO. IPA and network model HMOs now have the majority of HMO membership in the United States.

The other portion of the managed care industry is represented by point of service (POS) plans and preferred provider organizations (PPOs). POS plans incorporate key features of bothHMOs and traditional fee-for-service plans. A patient may choose care from a provider network or go outside the network. Care within network requires only a minimal copayment, while care outside the network requires a deductible and a large (e.g., 30%)

copayment. The goal of the plan is to offer patients a choice but to provide major financial incentives to stay within the HMO portion of the plan. PPOs use a defined provider network (physicians, hospitals) that has agreed to accept discounted fee-for-service to care for enrolled members. PPOs may incorporate various managed care features, such as physician gatekeepers and utilization review.

THE UNINSURED AND UNDERINSURED

Data from the U.S. Census Bureau indicate that 43.4 million people had no health insurance for all of 1997 and 71.5 million people were without insurance for at least part of the year. The great majority of uninsured individuals either work for small employers who do not offer a health insurance benefit or, more commonly, cannot afford the premiums of the plan(s) that are offered. Underinsurance also has a significant impact on the working poor by requiring them to pay an excessive proportion of their family's income for health insurance premiums and out-of-pocket medical costs (deductibles, copayments, and uninsured care). Outpatient prescription medications are a major source of underinsurance. Prescription drug costs are now the fastest growing segment of the national medical budget and the least likely segment to be covered by insurance. The elderly are particularly affected, since Medicare does not currently cover outpatient prescriptions and even Medigap policies have limited coverage.

Some states have experimented with expanded coverage through their Medicaid programs to help the uninsured poor (such as the Oregon Medicaid program). For the forseeable future, however, it does not appear that the federal government will address this problem comprehensively.

COST-CONTAINMENT STRATEGIES

Current projections from the federal government's Health Care Finance Administration (HCFA) are that health care expenditures will double (to $2.2 trillion, or 16.2% of theGDP) by 2008. Over the past 30 years, the U.S. health care system has experimented with a vast array of cost-containment approaches. Conceptually, there are four major ways to control medical spending: (1) control prices, (2) control volume of care provided, (3) control the total budget available to pay for care, and (4) shift costs to another payer.

Two of the most important price control initiatives in medicine have been the Medicare Hospital Prospective Payment System and the Medicare Fee Schedule for physicians. In 1983, Medicare replaced its retrospective cost-based hospital reimbursement system with a prospective payment system. In this system, all hospitalizations are classified into one of approximately 500 Diagnosis Related Groups (DRGs) based on the principal discharge diagnosis for the hospitalization and a few selected additional factors such as age, the performance of surgery, and the presence of complications. Each DRG is assigned an average reimbursement (adjusted annually). If the hospital can provide care for less than this amount, they make a profit. If they spend more than this amount, they lose money. The DRG system was designed to promote efficiency and cost containment in hospital-based care. While it has helped to control Medicare costs, it has not reduced overall U.S. health care costs, probably because of substantial cost-shifting by hospitals to the private insurance sector.

Between 1975 and 1987, Medicare payments to physicians increased at an annual rate of 18%, well above the rate of inflation. While total spending for physician services accounts for less than 25% of the Medicare budget, physicians have control over aspects of care (use of procedures, length of stay, hospital admission) that extend their direct influence to over 75% of the Medicare budget. Recognizing the importance of physicians in cost containment, Congress directed the development of a new physician payment system based on the use of a resource-based relative value scale (RBRVS). The Medicare Fee Schedule, which was first used in 1992, has three components: (1) a measure of the total work (time and complexity) involved in each physician service and standardized across all specialties, (2) a practice expense to cover the cost of running an office, and (3) an amount to cover malpractice insurance costs. The Medicare Fee Schedule classifies all physician services using the American Medical Association's Current Procedural Terminology (CPT) codes. Each CPT code has an associated relative value units (RVUs) weight. The RVU weights are multiplied by a national conversion factor to generate the actual physician fee associated with the service in question.

Price controls are attractive for cost containment because they are less expensive administratively than volume controls and don't involve micromanagement of clinical care. Price controls alone, however, don't generally achieve control of costs because of compensatory responses of providers. For example, under Medicare prospective payment, hospitals have shifted much care to the outpatient setting, whereDRGs are not used. Physicians have responded to lower fees by an increased volume and intensity of service.

Volume controls include various programs to limit the diffusion of expensive technologies (such as heart surgery) or extra hospital beds. Limits can be operationalized using either a regulatory approach [such as certificate of need (CON) programs] or a budgetary approach. Utilization review approaches attempt to discern which expensive care items are medically necessary and which are not.

Budgetary controls are simpler than either price or volume control approaches. In Canada, for example, hospitals have global annual budgets. How the money is spent is decided by each hospital. If the budget is exceeded, there are no guarantees that the shortfall will be covered.

Finally, payers can control their costs by cost-shifting to other willing payers. For example, as health insurance premiums rise, employers can choose to pass these costs on to employees. Hospitals and doctors who lose money caring for Medicare patients can try to make up their losses by charging more to private insurance patients. Insurance companies can choose to offer limited or no coverage for outpatient pharmaceuticals, shifting the full cost of expensive new medicines directly to patients.

MEDICAL ECONOMIC CONCEPTS AND TOOLS

MEDICAL COST CONCEPTS

Medical cost analysis is a field that borrows heavily from both economics and

accounting. Economics provides the theoretical structure that defines the key questions to be addressed, and accounting provides many of the measurement tools. Traditional economics has as one of its major axioms that societal resources are finite. For this reason, society must choose from among the many ways that resources can be used and not all of society's goals can be fulfilled. Economics has devised a theoretical framework and a set of tools (including cost-effectiveness analysis) to help define the major competing goals for societal resources and to assist in selecting from among the ones that most efficiently fulfill societal needs. "Cost" in economics refers not so much to money but rather to the lost opportunities that occur when the limited societal resources are expended in a particular way. For example, if our medical armamentarium is enhanced over the next decade by discovery of powerful but expensive therapies and these are incorporated into standard clinical practice, the ability of the country to invest in education, defense, or transportation may be compromised. This notion of cost as a lost opportunity to use resources in alternative ways is referred to as *opportunity cost*. While representing the purest economic notion of cost, there is no practical way to measure it.

Accountants, who are much more concerned with issues of measurement, have proposed a "gold standard" of cost measurement, *true accounting cost*, that involves enumerating all the individual resources consumed in the production of a particular medical good or service and assigning market prices for each of them. The total cost is then the sum of the dollar costs for all the component resources. Even this calculation, however, may be prohibitively difficult in "real world" applications, for several reasons. First, all medical care requires not only the easily identifiable components of personnel time and disposable supplies but also the infrastructure components such as the rent on the office building where the care is provided, the cost of utilities, and the expense of an office staff. Second, even if all the components can be identified, enumeration of exactly what is used may be prohibitively expensive. Finally, medicine does not have publicly available "market prices" that can be readily obtained for a medical cost analysis, the way one can obtain prices for automobiles or refrigerators. The reasons for this relate to the lack of a true competitive free market in medicine along with the severe price distortion created in medical charges by cost-shifting practices.

KEY COST TERMS

Several key sets of cost terms are used in medicine. As the volume of health care produced is increased or decreased, costs may exhibit either variable or fixed "behavior." *Variable costs* change with each unit shift in production volume (up or down). For example, each vaccination administered to a group of children increases costs (related to the dose of vaccine and the disposable syringe) in a predictable linear fashion. *Fixed costs* do not shift with short-term changes in the volume of care provided. For example, the rent on the clinical building and the cost of heating, lighting, and so forth do not change according to the number of individuals vaccinated per day. Some types of costs display hybrid features of both variable and fixed components. For example, clinic personnel costs (e.g., nurses, secretaries) may be fixed if these personnel are paid a salary regardless of clinic volume. If the clinic volume goes up so much that evening hours must be added, either new personnel must be hired or existing personnel must work overtime. Either of these changes would graft a variable component onto the fixed personnel costs.

Marginal cost is a concept often used by economists to refer to the cost of producing one more unit of a given health care good or service. For example, the costs of doing one more or one less diagnostic cardiac catheterization would be its marginal cost. For all practical purposes, this is the same as its variable costs (since fixed costs do not change with small changes in volume). While the concept of unit changes in volume is theoretically interesting, a more pragmatic issue is the cost effect of changing a group of patients from one strategy to another. Many experts use the term *incremental costs* to refer to this type of shift (although some use marginal and incremental synonymously). Incremental analysis is a key component of cost-effectiveness analysis (see below).

Another set of cost terms relates to the traceability of costs to the production of health care goods and services. *Direct costs*, such as nursing and physician personnel and disposable supplies, can be clearly linked to the health care provided and are under the control of the health care providers. *Indirect costs*, sometimes known as *overhead*, cannot. For example, the utility, laundry, maintenance, and administration costs of a hospital cannot be linked with the care of an individual patient and are generally not under the control of the physicians and nurses providing the medical care. The distinction of direct versus indirect is useful in cost-containment efforts, where the first step is to identify all major cost components and decide how they are to be controlled.

One common error in the evaluation of medical costs is to focus on the cost of a test or therapy in isolation. Virtually every major medical management decision creates downstream consequences. For example, if physicians order a screening diagnostic test and the result is abnormal, they will need to do a confirmatory or more definitive test. If they order a potent new antibiotic and a fraction of patients develop liver failure as an unexpected toxicity, the total cost of that course of antibiotic includes not only the cost of the drug itself but also the costs of treating the liver failure in the fraction of patients who develop it. Extra costs added as a consequence of some diagnostic or therapeutic decision are referred to as *induced costs*. Similarly, if a management decision produces downstream savings, these would be referred to as *induced savings*. For example, administration of HMG CoA reductase inhibitors to patients with hypercholesterolemia can prevent future myocardial infarctions and revascularization procedures, both of which entail expensive hospitalizations.

One final important cost concept relates to the societal costs of lost productivity (primarily lost time from work) due to illness. While economists often refer to these as indirect costs, confusion with the accounting concept of indirect costs (overhead) has led many to prefer the alternative term, *productivity costs*.

COST MEASUREMENT

Using varying degrees of simplification, medical costs can be measured using either bottom-up or top-down approaches. Bottom-up approaches build from component resources to calculate total cost for an episode or type of care. Microcosting is the gold standard approach. It involves careful enumeration of all resources consumed and detailed cost-accounting estimation of the costs for each component resource. A number of medical centers have now installed computer-based cost-accounting systems that perform a modified type of microcosting analysis. For difficult-to-obtain resource

use data (such as time required for a particular type of care by a given type of personnel), these systems use expert opinion in place of empirical data. The other extreme of the bottom-up category of approaches involves enumeration and costing for only the "big ticket" or expensive items, such as hospitalization episodes and costly tests and procedures.

The top-down methods of medical cost estimation calculate a cost estimate from aggregated data. One such approach uses hospital billing charge data and charge-to-cost conversion ratios (which each hospital produces annually in its Medicare Cost Report) to estimate hospital costs. Despite the approximations involved, this approach, which can be used for most nonfederal U.S. hospitals, has provided good agreement with bottom-up estimates in the few instances where formal comparisons have been made. The other top-down approach is the use ofDRGassignments and reimbursement rates to provide standard cost weights for hospitalization episodes.

COST-EFFECTIVENESS ANALYSIS

Given a finite budget (for health care overall or for a particular health system), how can we use the available money to provide the most health benefits for our patients? For the clinician, who is less concerned with such policy issues, a prevalent question is whether a new treatment is economically attractive. The analysis method used to address this question is dependent on how the effectiveness and costs of the new therapy compare with those of "standard care" (Fig. 4-1). *Cost-effectiveness analysis* is used when effectiveness of the new treatment is greater and its costs are higher. This analysis calculates the ratio of added (or incremental) health benefits to added costs produced by a new therapy or strategy relative to some reference standard. The general formula is:

where C = costs and E= effectiveness.

The cost-effectiveness ratio provides a quantitative statement of the amount of money required to produce a single extra unit of benefit with the new therapy relative to usual care or some other relevant reference standard. The benefit can be calculated in any meaningful clinical unit, such as added survivors or extra patients with a correct diagnosis. However, the vast majority of cost-effectiveness analyses use the epidemiologic concept of life-years to express incremental benefit. Virtually all benchmarks for cost effectiveness relate to this endpoint. Because some therapies affect quality of life but not quantity, a more generally relevant effectiveness measure combines qualify of life and life expectancy into a single composite metric, the quality-adjusted life year (QALY). Calculation of incremental dollars required to add an extra QALY is called *cost-utility analysis*. The QALY is a useful concept, but many details regarding measurement and interpretation remain controversial. The third form of economic efficiency analysis, *cost-benefit analysis*, requires conversion of health benefits into monetary equivalents. Because such conversions are controversial, this form of analysis is rarely used in medicine. In theory, the time horizon of a cost-effectiveness analysis should be long enough to capture all important cost and health consequences of the therapy or strategy being evaluated. Most often, analysts

use a lifetime time frame. Because very few empirical studies are long enough to observe lifetime outcomes (especially when chronic diseases are being studied), models are required to extrapolate from available data.

A cost-effectiveness analysis can be done from a variety of perspectives, but the most widely applicable perspective is societal. Other perspectives are often much narrower and may include unattractive qualities. For example, a managed care organization may be interested only in short-term costs and outcomes, knowing that patients tend to change their health insurance every few years.

The benchmarks for cost-effectiveness ratios are determined by comparison with other well-accepted therapies in widespread medical use. A useful benchmark is hemodialysis for chronic renal failure, since the federal government has paid for all renal failure patients to get dialysis since 1973 through the End Stage Renal Disease Program. Recent estimates are that it costs this Medicare program about $50,000 to add 1 life-year to a chronic renal failure patient. Partly for this reason, many analysts use a cost-effectiveness ratio of <$50,000 per added life-year to identify therapies that are economically attractive (i.e., have a favorable balance of extra costs to extra benefits), while therapies with ratios >$100,000 per added life-year are deemed economically unattractive and therapies between $50,000 and $100,000 per added life-year are in the economic "gray zone."

Several caveats about cost-effectiveness analysis should be noted. First, cost-effectiveness analysis is descriptive, not prescriptive. It measures value that could be produced with available health care dollars but does not mandate how these dollars are to be used. If an expensive new therapy is introduced and is found to be very economically attractive by the above benchmarks, it will still not get used if there is no money in the budget to pay for it. Second, a cost-effectiveness ratio is only as good as the data that were used to calculate it. High-quality results can be obtained if economic analysis is prospectively incorporated into the design of large-scale multicenter randomized trials. Third, although cost-effectiveness ratios are often presented as deterministic (i.e., no variability), they often incorporate large amounts of uncertainty. This should be examined either with sensitivity analyses (varying each key parameter through a plausible range to see if the results are materially changed) or calculation of confidence limits.

MEDICAL ECONOMICS AND CLINICAL PRACTICE

In evaluating new therapies, three issues must be addressed: (1) is the new therapy significantly better than what is currently available? (2) how much does it cost and is it economically attractive? and (3) how many patients will need this therapy and is it affordable? The clinician should be primarily concerned with the answer to the first question. Although cost issues are now a reality of daily clinical life and cost-containment pressures are often substantial, decisions by clinicians that are based primarily on economic rather than clinical considerations put the physician in the role of the double agent (i.e., acting on behalf of both the patient and the payer) and compromise our fiduciary obligation to patients. The second question addresses cost effectiveness and, if favorable, can be used to support an argument by clinicians for adoption of the therapy. In the ideal world, at least, therapies that have a large database

of evidence demonstrating effectiveness and economic attractiveness should be given preference over therapies that do not have such supporting data. The final question is of primary concern to payers and health policy analysts. An effective therapy that is too expensive to use is of little more value than a therapy that has yet to be discovered.

(Bibliography omitted in Palm version)

Back to Table of Contents

5. INFLUENCE OF ENVIRONMENTAL AND OCCUPATIONAL HAZARDS ON DISEASE - *Howard Hu*, *Frank E. Speizer*

Exposures to hazardous materials and processes in the home, the workplace, and the community can cause or exacerbate a multitude of diseases. Physicians commonly treat the sequelae of such diseases in the practice of medicine; however, unless the underlying connection with hazardous exposures is identified and mitigated, treatment of manifestations rather than the cause at best only ameliorates the condition. At worst, the neglect of hazardous exposures may lead to both failure of treatment and failure to recognize a public health problem with wide significance.

No existing surveillance or reporting system can estimate the total contribution of hazardous exposures to morbidity and mortality. However, careful histories have identified occupational factors as etiologic in more than 10% of all admissions to general internal medicine wards in hospitals, with even higher percentages when the primary illness is either respiratory or musculoskeletal. Estimates of the number of new cases of disease due to work in the United States range from 125,000 to 350,000 per year; these cases do not include 5.3 million work-related injuries.

Environmental exposures are increasingly associated with decrements in measures of health whose outcomes range from subclinical to clinically catastrophic. For example, exposure to lead at levels that are common in the general population has been associated with increased blood pressure and decreased creatinine clearance. Ambient air pollution with respect to levels of ozone and fine-particulate matter has been related to increased rates of hospital admission for respiratory and cardiovascular diseases and to increased mortality rates, respectively. Indoor exposure to radon and passive indoor exposure to environmental tobacco smoke have been linked with an increased risk of lung cancer. There is pressure on clinicians to be aware of and act on this type of information, which is suggestive but not necessarily conclusive with respect to causation.

Patients are becoming increasingly concerned about hazardous exposures. More than 15% of patients seen in one study conducted in a primary care clinic expressed the opinion that their health problems were work-related, and 75% of this subgroup of patients reported exposure to one or more recognized toxic agents. Patients often want answers to very specific questions, such as: Is the water in our town safe to drink? Could my breathing problem be related to the new roofing sealant used in my building at work? Physicians are consulted because they are the most trusted sources of information on health risks, including chemical risks. Unfortunately, few physicians have more than rudimentary training in environmental and occupational medicine. Therefore, it becomes important for primary care physicians to be able to recognize symptoms precipitated by exposure to environmental or occupational hazards and either to manage these cases or to make appropriate referrals.

Many manifestations of exposure-related illnesses are nonspecific (e.g., dizziness, headache) or are commonly encountered in general internal medicine (e.g., myocardial infarction, cancer). The establishment of a connection with an environmental or occupational hazard requires a high index of suspicion and the application of fundamental concepts of environmental/occupational medicine. Furthermore, early

recognition by physicians of unusual patterns of illness or of evidence of asymptomatic exposure to toxins with low-level effects (e.g., an elevated blood lead level) can alert health officials to the need for control measures. Case reports either sent to local authorities or published in the literature often prompt follow-up studies that can lead to the identification of new hazards. In many states and countries, the reporting by physicians of occupational/environmental diseases is mandatory. For instance, beginning in 1992, physicians in Massachusetts were required to report cases of pneumoconiosis, occupational asthma, carpal tunnel syndrome, and carbon monoxide poisoning, among other conditions. Identification of an environmental/occupational etiology of an illness may have important economic ramifications for the patient (e.g., the awarding of worker's compensation, which covers medical bills as well as lost wages). Finally, physicians are frequently asked to provide expert medical testimony during litigation on the causal relationship between toxic exposures and diseases. In this setting, the more knowledgeable the physician is about potential hazardous exposures, the better prepared he or she is to serve the patient.

THE ENVIRONMENTAL/OCCUPATIONAL HISTORY

For a physician, the most critical steps toward recognizing these disorders are remembering to consider them in the differential diagnosis and taking an appropriate environmental/occupational history as part of the medical workup. The level of detail that is called for depends on the clinical situation. *Information should always be obtained on current and major past occupations, and patients should be asked whether they think their health problem is related to their work or to any particular environment or exposure.* In the review of systems, patients should be asked if they have been exposed to dusts, fumes, chemicals, radiation, or loud noise. When patient and physician are confronted with an illness of uncertain etiology, these factors should be explored in more detail, with the environmental/occupational history as the point of departure. (A brief outline of a sample history is shown in Table 5-1.)

The identification of specific chemical exposures can be difficult. Household products must list chemical ingredients on their labels, and this information may prove useful. For workplace exposures, the U.S. Occupational Safety and Health Administration (OSHA) requires chemical suppliers to provide material safety data sheets with their products and requires employers to retain these sheets and make them available to employees. The data sheets can be obtained by the physician or employee by a telephoned or written request; failure of an employer to provide them within 30 days of such a request is a violation of OSHA regulations and is punishable by fines. In addition to providing information on chemical ingredients and percent composition, the material safety data sheets provide basic information on toxicity. This information is seldom adequate from a clinical perspective but may indicate the general type of toxicity to be anticipated.

EVALUATION OF POSSIBLE CHEMICAL OR ENVIRONMENTAL HAZARDS

Given the wide variety of toxic exposures that may be uncovered during a workup, a clinician should routinely consult additional reference material to evaluate whether particular hazards may be associated with the illness at hand. Many sources of information exist.OSHA and some regional poison-control centers have extensive information on hazards and brief summary documents that can be transmitted over the

Internet or by telephone or facsimile. Depending on the area, other resources may include county and state health departments; regional offices of the National Institute for Occupational Safety and Health and the Environmental Protection Agency; the Consumer Products Safety Commission in Washington, DC; academic institutions; websites of these institutions; and individual toxicologists, occupational/environmental medicine specialists, or industrial hygienists. Sophisticated computerized databases are also available, including detailed listings on CD-ROM information systems. MEDLARS, the electronic database maintained by the National Library of Medicine, is accessible by modem or the Internet and is familiar to many physicians. Files other than MEDLINE, such as the Hazardous Substances Databank, provide specific toxicity information on chemicals and include toxicologic references not covered by MEDLINE. Many of these databases can also be accessed through the Internet.

As with any other illness, laboratory investigation may be crucial. For example, tests of carboxyhemoglobin level to document carbon monoxide exposure or of serum anticholinesterase level to document organophosphate pesticide absorption should be performed within hours of exposure. As in cases of acute drug overdose, it is useful to freeze samples of urine and serum from any patient suspected of having had an acute chemical exposure; such specimens can be analyzed at a later date by sensitive methods of detection. Use of other tests must rely on knowledge of the specific hazard or illness in question.

SUSPICIOUS SCENARIOS

Some medical problems or clinical scenarios demand a particularly high degree of suspicion of occupational or environmental factors as causative or contributing agents.

Respiratory Disease The contribution of occupational/environmental factors to respiratory disease is generally underrecognized, particularly among patients who smoke and among the elderly (Chap. 254). For instance, asthma related to chemical exposure may be treated without regard to cause or may be erroneously diagnosed as acute tracheobronchitis. A study of new-onset asthma among HMO members in Massachusetts found that 21% of these individuals met criteria for clinically significant asthma attributable to occupational exposures. The types of exposures and jobs in these cases varied widely; examples include exposure to smoke in a firefighter, to welding fumes in a technical school student, to cleaning compounds in a bartender, and to epoxy in an archery repairman. No single type of job or exposure predominated. Other examples of etiologic errors include shortness of breath from asbestosis that is attributed to chronic obstructive pulmonary disease and chemical pneumonitis that is misdiagnosed as a bacterial infection.

Cancer Many cancers are thought to be causally related to occupational and environmental factors in addition to tobacco. Some are particularly likely to have a chemical etiology or another environmental cause, including cancers of the skin (solar radiation, arsenic, coal tar, soot); lung (asbestos, arsenic, nickel, radon); pleura (almost exclusively asbestos); nasal cavity and sinuses (chromium, nickel, wood and leather dusts); liver (arsenic, vinyl chloride); bone marrow (benzene, ionizing radiation); and bladder (aromatic amines).

Coronary Disease and Hypertension Carbon monoxide exposure is common, particularly in homes with malfunctioning furnaces or in workplaces close to motor vehicle exhaust. By reducing oxygen transport by hemoglobin and inhibiting mitochondrial metabolism, carbon monoxide can aggravate coronary disease. Methylene chloride, a solvent used in paint stripping, is converted to carbon monoxide and thus poses the same risk. Exposure to carbon disulfide, a chemical used in the production of rayon, accelerates the rate of atherosclerotic plaque formation. Chronic lead exposure, even at modest levels, is a risk factor for the development of hypertension as well as abnormalities of cardiac conduction.

Hepatitis/Chronic Liver Disease In the absence of evidence that a viral infection, alcohol ingestion, or drug use is the main cause of hepatitis (Chaps. 295,296, and297), the involvement of a toxin must be considered. Toxin-induced hepatic injury may be cytotoxic, cholestatic, or both. The list of hepatotoxic agents is long, including organic synthetic compounds such as carbon tetrachloride (used in solvents and cleaning fluids) and methylene diamine (a resin hardener); pesticides such as chlordecone (Kepone); metals, particularly arsenic (used in pesticides and paints and found in well water); and natural toxins such as the pyrrolidizine alkaloids.

Kidney Disease Many chemical and environmental factors can cause renal injury (Chap. 269). The etiology of much chronic kidney disease, however, remains unknown. An increasing body of evidence now links chronic renal failure with hypertension to lead exposure. One study demonstrated that chelation therapy with EDTA slowed the progression of renal insufficiency in patients with a mildly elevated body lead burden. Some studies suggest that chronic exposure to hydrocarbons (e.g., gasoline, paints, solvents) may lead to various types of glomerulonephritis, including Goodpasture's syndrome. Environmental cadmium exposure has been found to promote calcium loss via urinary excretion, which results in skeletal demineralization and thus in an increased risk of fractures.

Peripheral Neuropathy Organic solvents such as *n*-hexane, heavy metals such as lead and arsenic, and some organophosphate compounds can damage the axons of peripheral nerves. Dimethylaminopropionitrile, an industrial catalyst, causes bladder neuropathy. Nerve entrapment syndromes of the upper extremity, such as carpal tunnel syndrome, may be caused by jobs that involve repetitive motion, especially those requiring the maintenance of awkward positions.

Central Nervous System Disorders Fatigue, memory loss, difficulty in concentration, and emotional lability have been linked to chronic exposure to solvents such as toluene and perchloroethylene. Painters, metal degreasers, plastics workers, and cleaners are commonly exposed to solvents and develop these symptoms at a high rate. Among the features that distinguish these patients are characteristic patterns on formal neurobehavioral testing and stabilization of symptoms with gradual improvement after discontinuation of the exposure. Other substances associated with neurobehavioral dysfunction include metals, particularly lead, mercury, arsenic, and manganese; pesticides, such as organophosphates and organochlorines; polychlorinated biphenyls (PCBs); and gases such as carbon monoxide.

Environmental factors are also suspected of contributing to other neurologic diseases,

such as degenerative disorders, motor neuron diseases, and extrapyramidal disorders. For example, a study in monozygotic and dizygotic twin pairs found a similarity in concordance indicating that environmental (as opposed to genetic) factors play a major etiologic role in cases of typical Parkinson's disease beginning after the age of 50 years.

Teratogenesis and Reproductive Problems Toxins can impair successful reproduction at a variety of levels. Examples include insecticides and herbicides,PCBs and polybrominated biphenyls (PBBs), ethylene oxide (a sterilizing gas used in hospitals), metals (lead, arsenic, cadmium, mercury), and solvents. Dibromochloropropane, a nematocide, suppresses spermatogenesis. Some toxins, such as PCBs, PBBs, and chlorinated pesticides, are concentrated in milk. Concern has arisen over the ability of specific organic pollutants, particularly pesticides, to persist in the environment and accumulate in human tissues. Some of these chemicals may disrupt endocrine function, and these effects may be related to phenomena such as the observed increases in the incidences of testicular cancer, breast cancer, and hypospadias.

Immunosuppression, Autoimmunity, and Hypersensitivity Evidence is increasing that exposures to some chemical agents can compromise the immune system, thereby leading to a generalized increase in the incidence of tumors (e.g., exposure toPBBs) or infections (e.g., respiratory infections after exposure to common air pollutants). Mercury, dieldrin, and methylcholanthrene are known to elicit autoimmune responses. Some chemicals are potent allergic sensitizers that cause dermal and respiratory problems (Chaps. 60 and254).

BIOLOGICAL MARKERS

An increasing number of methods are available for measuring and interpreting toxic exposure, including (1) the internal dose of specific toxins and (2) markers of the biologic effects of toxins. Internal-dose markers are relevant for toxins that are sequestered in the human body, such as lead (in blood), arsenic (in hair), and other metals (Chap. 395), and for halogenated compounds (such asPCBs). Examples of markers of the biologic effects of toxins include depressed levels of acetylcholinesterase in serum after exposure to organophosphate pesticides, sister chromatid exchanges in peripheral lymphocytes after exposure to the carcinogen ethylene oxide, and DNA adducts after exposure to tobacco smoke carcinogens.

MANAGING A HAZARD-RELATED ILLNESS

Once a chemical or another environmental hazard has been identified as an important contributor to an illness, the next step is to prevent further exposure. Although for chronic diseases such as cancer this step may be irrelevant for the patient in question, prevention of further exposure may still be critical for other persons who have been similarly exposed. When prevention of further exposure is important, *the physician must be willing to become an active advocate for the patient*. This advocacy may involve writing a letter stating that the patient should no longer be exposed to a hazard or should remain out of work. Alternatively, it may involve contacting appropriate officials in government, industry, or labor or other advocates who can deal with a hazardous exposure. Treatment is dependent on the specific hazard.

In few areas of medicine does a physician deal with more scientific uncertainty. Comprehensive information on toxicants is available for only a small percentage of chemicals. In general, the physician should take a conservative approach (i.e., advise the patient to avoid a hazard likely to have contributed to illness) and should use common sense and up-to-date information to evaluate causal relationships.

LOW-LEVEL EXPOSURES AND THEIR EFFECTS

The subclinical effects of toxins that are widespread in our environment and our workplaces are of increasing concern. Given the absence of any demonstrable effect threshold, low-level exposure to carcinogens should be avoided; not only carcinogenic but also noncarcinogenic effects of chronic low-level exposure to these substances are important.

Perhaps lead provides the most important example of low-level noncarcinogenic effects that constitute a major public health problem. Multiple pathways of exposure, including the combustion of leaded gasoline, the use of lead-based paints and solder, and the presence of lead in cans containing food, have contributed to exposure of the entire population. Such low-level exposures can impair neurobehavioral development in infants and children and can raise blood pressure in adults. Furthermore, absorbed lead is stored in the skeleton and may reenter the circulation at times of heightened bone turnover (e.g., pregnancy, lactation, osteoporosis, hyperthyroidism). Subclinical toxic effects can be prevented if chronic low-level exposure is detected early and curtailed. In the case of lead, such exposure is detected by tests of blood lead level, which should be performed regularly in young children living in old housing and as a precautionary measure in adults with a history of lead exposure.

(Bibliography omitted in Palm version)

Back to Table of Contents

6. WOMEN'S HEALTH - *Anthony L. Komaroff*, *Celeste Robb-Nicholson*, *Andrea E. Dunaif*

In recent years, the medical problems and health care of women have received increasing attention. There are poorly understood differences between men and women, both in morbidity and mortality and in the expression of diseases. Many research studies of disease prevention and pathophysiology have included only male subjects; most illnesses that can affect both sexes have not been as well studied in women. It also appears that women receive different care than men for certain common health problems. Finally, an increasing number of women are seeking health care in multidisciplinary women's health units that combine the expertise of gynecology, psychiatry, and internal medicine or family medicine.

MORBIDITY AND MORTALITY IN WOMEN

Morbidity Past studies have found that women experience more days of restricted activity than men at all ages, over and above the restricted activity caused by obstetric and gynecologic conditions. However, a study in 1998 concluded differently. Women make more visits to physicians, particularly for acute self-limited illnesses.

Mortality In the developed nations, women live longer than men. In the United States, as of 1996, the projected average life expectancy from birth is 79.1 years for females, and 73.1 years for males. Although there are more male fetuses conceived than female fetuses, females have a survival advantage when compared to males, in all age groups. The longer life expectancy of women versus men in developed countries is due in large part to the difference in mortality caused by ischemic heart disease (IHD).

As shown inTable 6-1, the leading causes of death among young women in the United States are accidents, homicide, and suicide. During the middle years, breast cancer is a slightly more common cause of death thanIHD and lung cancer. In women between ages 65 and 74, IHD, lung cancer, and cerebrovascular disease supercede breast cancer as the leading causes of death. Among women of all ages, IHD is the leading cause of death by a substantial margin, with a mortality rate five to sixfold higher than the rate for either lung or breast cancer. Nevertheless, polls find that U.S. women believe breast cancer poses the greatest threat to their lives.

Social Factors Influencing Morbidity and Mortality Gender differences in morbidity and mortality may be explained in part by psychosocial factors such as socially-defined gender roles, poverty, participation in the work force, health insurance, and lifestyle.

In the past 30 years in the United States, there has been a "feminization of poverty." One-third of families headed by women currently live in poverty, and the fraction is greater than one-half for African-American and Latino women. Almost a fifth of women over age 65 live below the poverty level. People of lower socioeconomic status experience poorer health and a higher mortality rate than those in higher income groups. The poor are more likely to smoke and less likely to have recommended preventive measures, including cancer screening. Lack of adequate health insurance is a major problem for many women; in general, they are more likely than men to have low-paying, part-time, non-union jobs that do not provide health insurance. Women who

are divorced or widowed may also lose health insurance that they had through their husbands.

PREVENTION (See alsoChap. 10)

Primary prevention and screening are crucial elements in improving the health of women. Based upon available literature and the consensus of experts, various authoritative organizations have published guidelines on preventive practices in women.

Most physicians believe that a baseline history and physical examination is useful to set the stage for preventive measures appropriate to each patient. In general, authorities recommend that blood pressure be measured every other year throughout life. Counseling on diet, smoking cessation, exercise, and use of seatbelts are of demonstrated value in the primary prevention of diseases and accidents. Counseling about safe sexual practices, alcohol abuse, and violence are also recommended.

Screening for glaucoma is recommended for African-American women over age 40 and for Caucasian women over age 50. Yearly examinations to test visual acuity are recommended for women over age 70.

Regular screening for breast, cervical, and colorectal cancer is recommended, but how often tests should be performed and which tools to use are still being debated. Most authorities recommend annual clinical breast examination in all women beginning at age 35 to 40. There is strong evidence to support the efficacy of annual mammography in women age 50 to 59. For women age 60 or older, the evidence for screening is less strong. The benefits of screening for women between the ages of 40 and 49 are still being debated.

Most authorities recommend Pap smear screening beginning at age 18 or when a woman becomes sexually active. After two or three consecutive normal Pap smears, most groups recommend Pap smear testing every three years. If Pap smears have been normal for 10 years, they can be discontinued in women after age 65.

Recommendations for colorectal cancer screening vary. For patients over 50, the American Cancer Society recommends yearly fecal occult blood testing and rectal examination combined with flexible sigmoidoscopy every 5 years, colonoscopy every 10 years, or double-contrast barium enema every 5 to 10 years.

Bone mineral testing has gained rapid acceptance as a screening tool for detecting osteoporosis, as well as for predicting the likelihood of the condition in the future. With the advent of multiple preventive and therapeutic strategies for osteoporosis, many authorities now recommend bone mineral testing to screen for the condition. A bone mineral density test is recomended for all women over age 65 as well as for all postmenopausal women who are at increased risk for developing osteoporosis (Chap. 342).

Cigarette smoking, a major risk factor for cardiovascular diseases and cancers in women, has been well studied (Chap. 390). Over the past 60 years there has been a sharp decline in smoking among men, but not among women; teenage women smoke at

higher rates than their male counterparts. "Low-yield" cigarettes are marketed heavily to women. The Nurses' Health Study showed that one-third of the excess risk of ischemic heart disease was eliminated two years after smoking cessation, and that all of the excess risk was eliminated by 10 to 14 years after smoking cessation.

The National Cholesterol Education Program recommends that total cholesterol and high-density lipoprotein (HDL) levels be measured once. If both are normal, a repeat test after 5 years is recommended. A meta-analysis of several small studies of women showed an increased risk ofIHD in women with serum cholesterol greater than 265, a ratio of total cholesterol to HDL cholesterol greater than 4, or an elevated fasting triglyceride.

In various case-control and observational studies, postmenopausal estrogen therapy is associated with a 40 to 50% reduction in deaths due toIHD, but its value in a prospective, randomized trial has not yet been documented.

Calcium and estrogen, as well as alendronate and the selective estrogen receptor modulators, tamoxifen and raloxifene, slow the development of osteoporosis and reduce the frequency of hip and vertebral fracture in postmenopausal women. In randomized clinical trials, both tamoxifen and raloxifene have been shown to reduce the risk of breast cancer in postmenopausal women.

Considerable research indicates that a relatively high dietary intake of various antioxidants (including vitamins E and C) is associated with lower rates of vascular disease and malignancies. Randomized trials of supplemental antioxidants are under way. Preliminary research indicates that regular aspirin use is associated with reduced rates ofIHD and colorectal carcinoma.

GENDER DIFFERENCES IN DISEASE

Obviously, some diseases and conditions occur exclusively (or nearly exclusively) in women -- e.g., menopause and various breast and gynecological disorders. These are discussed elsewhere in this book (Chaps. 52,89,336,337). In this chapter, we seek to highlight some gender differences in diseases that occur in both women and men.

Ischemic Heart Disease (See also Chap. 244) Many persons think ofIHD as a primary problem for men rather than women, perhaps because men have more than twice the total incidence of cardiovascular morbidity and mortality between the ages of 35 and 84. However, as stated earlier, in the United States IHD is among the leading causes of death among women as well as men (Table 6-1). The curve for the IHD mortality rate in women lags behind that for men by about a decade. Nevertheless, nearly 250,000 women die annually from IHD; after age 40, one in three women will die from heart disease. Although IHD mortality has been falling in men in the United States over the past 30 years, it has been increasing in women.

Why areIHDrates lower in women? They have a more favorable risk profile in some respects: higherHDLcholesterol levels, lower triglyceride levels, and less upper-body obesity than men. But women also have a less favorable risk profile in other respects: more obesity, higher blood pressure, higher plasma cholesterol levels, higher fibrinogen

levels, and more diabetes. The simplest explanation for the sex differential in IHD is the "cardioprotective" effect of estrogen, which can be due to improvement of the lipid profile, a direct vasodilatory effect, and perhaps other factors. HDL cholesterol levels appear to be a particularly important risk factor for IHD in women. HDL levels are higher in all age groups in women compared to men, and are higher in premenopausal and estrogen-treated postmenopausal women. Smoking is the most important risk factor for IHD in women.

IHDpresents differently in men and women. In the Framingham study, angina was the most frequent initial symptom of IHD in females, occurring in 47% of women, whereas myocardial infarction was the most frequent initial symptom in males, occurring in 46% of men. The exercise electrocardiogram has a substantial false positive as well as false negative rate for women, compared to men.

Women, particularly African-American women, have a higher risk of morbidity and mortality than men following a myocardial infarction. Compared to men, women who obtain coronary artery bypass graft surgery have more advanced disease, a higher perioperative mortality rate, less relief of angina, and less graft patency; however, 5- and 10-year survival rates are similar. Women undergoing percutaneous transluminal coronary angioplasty have lower rates of clinical and angiographic success than men, but also a lower rate of restenosis and a better long-term outcome. Women may benefit less and have more frequent serious bleeding complications from thrombolytic therapy than do men. Factors such as older age, more comorbid conditions, and more severeIHD in women at the time of events or procedures appear to account for at least part of the gender differences observed. Women with IHD benefit at least as much as men, and perhaps more, from reductions in cholesterol level.

The incidence ofIHDincreases markedly at menopause, consistent with the hypothesis that estrogens are cardioprotective. A number of observational studies have supported this hypothesis by demonstrating significant decreases in IHD in women on hormone replacement therapy (HRT), both estrogen alone and estrogen-progestin combination therapy. However, the HERS, a recent clinical trial of HRT for the *secondary* prevention of IHD, showed no significant difference in cardiovascular events between therapy with combined continuous conjugated equine estrogen (0.625 mg qd) and that with medroxyprogesterone acetate (2.5 mg qd), compared to placebo over four years. Indeed, in the HRT group, there was about a 50% increase in cardiovascular events in the first year of the trial. The Women's Health Initiative is investigating directly the impact of various HRT modalities as a *primary* prevention of IHD risk. Until further data are available, caution should be exercised in prescribing HRT to women with a history of IHD, or for cardioprotection alone.

Hypertension (See alsoChap. 246) Hypertension is more common in U.S. women than men, largely owing to the high prevalence of hypertension in older age groups and the longer survival rate for women. Both the effectiveness and the adverse effects of various antihypertensive drugs appear to be comparable in women and men. Benefits of treatment for severe hypertension have been dramatic in both women and men. However, in clinical trials of the treatment of mild to moderate hypertension, women have had a smaller decrease in morbidity and mortality than men, perhaps because women have a lower risk of myocardial infarction and stroke than men to begin with.

Older women benefit at least as much as men from treatment, as demonstrated by the Systolic Hypertension in Elderly study. The incidence of hypertension (above 140/90) appears to be low (less than 5%) with the current low-dose oral contraceptives. Postmenopausal estrogen therapy is not associated with increases in blood pressure.

Immunologically Mediated Diseases Several immunologically mediated diseases -- e.g., rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, Graves' disease, and thyroiditis -- occur much more frequently in women than in men. In animal models of rheumatoid arthritis -- lupus and multiple sclerosis, for example -- it is the females of the species that are predominantly affected. On the other hand, animal studies indicate that females are less susceptible to infection.

In short, female animals appear to have more vigorous immune responses, with both beneficial and adverse consequences. Increasing evidence indicates that estrogens upregulate both cellular and humoral immunity. Also, some immunocytes contain estrogen, progestin and androgen receptors, and the uterus produces a variety of cytokines, suggesting a complex interaction between the reproductive and immune systems.

Osteoporosis (See alsoChap. 342) This condition is much more prevalent in postmenopausal women than in men of similar age. Osteoporotic hip fractures are a major cause of morbidity in elderly women. Men accumulate more bone mass and lose bone more slowly than women. Gender differences in bone mass are found as early as infancy. Calcium intake, vitamin D and estrogen all play important roles in osteoporosis; calcium intake is an important determinant of peak bone mass, particularly during adolescence. Vitamin D deficiency is surprisingly common in elderly women. Receptors for estrogens and androgens have been identified in bone. The aromatase enzyme system, which converts androgens to estrogens, is also present in bone.

Therapy withHRT, or with calcium and vitamin D, has been shown to reduce the risk of osteoporotic fractures. Newer modalities, such as bisphosphonates (alendronate), calcitonin, and raloxifene, a selective estrogen receptor modulator, prevent bone loss and reduce the risk of osteoporotic fractures.

Alzheimer's Disease (See alsoChap. 362) Alzheimer's disease (AD) affects approximately twice as many women as men, in part because women live longer. Several observational studies suggest thatHRT may decrease the risk of AD and improve cognitive function in older women. These benefits are seen in both current as well as past HRT users. In a few experimental studies, estrogen replacement has been shown to be associated with improved memory compared to placebo treatment. Estrogens enhance neuronal growth and activity, providing a biologic basis for these putative cognitive effects of HRT. Prospective clinical trials, including the Women's Health Initiative, are underway to pursue these intriguing observations.

Diabetes Mellitus (See alsoChap. 333) Estrogens enhance insulin sensitivity in women but not in men. Despite this, the prevalence of type 2 diabetes mellitus (DM) is higher in women, which is related in part to the higher prevalence of female obesity. Premenopausal women with DM lose the cardioprotective effect of female gender and have identical rates ofIHD to those in males. This is partially explained by the presence

of several IHD risk factors in women with DM: obesity, hypertension and dysplipidemia. Recent evidence suggests that vascular responses differ in women with DM, as compared to normal women. Polycystic ovary syndrome and gestational diabetes mellitus -- common conditions in premenopausal women -- are associated with a significantly increased risk for type 2 DM.

Psychological Disorders (See also Chap. 385) Depression, anxiety panic disorder and eating disorders (bulimia and anorexia nervosa) occur more often in women than in men. Epidemiologic studies from both developed and developing nations consistently find major depression to be twice as common in women as in men, with the gender disparity becoming evident in early adolescence. Depression occurs in 10% of women during pregnancy and in 10 to 15% of women during the first several months of the postpartum period. The incidence of major depression diminishes after age 45, and does not increase with the onset of menopause. Depression in women also appears to have a worse prognosis than in men; episodes of depression last longer and there is a lower rate of spontaneous remission.

Social factors may account for the greater prevalence of some disorders in women; the traditionally subordinate role of women in society may generate feelings of helplessness and frustration which contribute to psychiatric illness. In addition, it is likely that biological factors, including hormonally influenced neurochemical changes, also play a role. The limbic system and hypothalamus -- areas of the brain thought to subserve appetite, satiety and emotion -- contain estradiol and testosterone receptors.

Alcohol and Drug Abuse (See alsoChap. 387) One-third of Americans who suffer from alcoholism are women. Women alcoholics are less likely to be diagnosed than men; a greater proportion of men than women seek help for alcohol and drug abuse. Men are more likely to go to an alcohol or drug treatment facility, while women tend to approach a primary care physician or mental health professional for help under the guise of a psychosocial problem. Late-life alcoholism is more common in women than men. In 1997, an epidemiologic survey reported that, among women over age 59, an estimated 1.8 million were addicted to or abused alcohol, and over 2.8 million were addicted to or abused psychoactive or mood-altering prescription drugs.

On average, alcoholic women drink less than alcoholic men, but exhibit the same degree of impairment. Blood alcohol levels are higher in women than in men after drinking equivalent amounts of alcohol, adjusted for body weight. This greater bioavailability of alcohol in women is probably due to the higher proportion of body fat and lower total body water. Women also have a lower gastric "first-pass metabolism" of alcohol, associated with lower activity of gastric alcohol dehydrogenase. In addition, alcoholic women are more likely than alcoholic men to abuse tranquilizers, sedatives, and amphetamines. Women alcoholics have a higher mortality rate than do nonalcoholic women and alcoholic men. Compared to men, women also appear to develop alcoholic liver disease and other alcohol-related diseases with shorter drinking histories and lower levels of alcohol consumption. Alcohol abuse also poses special risks to women who are or wish to become pregnant, adversely affecting fertility and the health of the baby (fetal alcohol syndrome).

Finally, there is growing evidence that for several illicit drugs, women proceed more

rapidly to drug dependence than do men.

Human Immunodeficiency Virus Infection (See also Chap. 309) As of September 1998, the Centers for Disease Control and Prevention estimate that between 120,000 and 160,000 adolescent and adult women in the United States were living with HIV infection, including those with AIDS (Table 6-1). Between 1985 and 1998, the proportion of all U.S. AIDS cases reported among women more than tripled, from 7 to 23%. HIV infection was the fourth leading cause of death among U.S. women age 25 to 44 in 1997, and the second leading cause of death among African-American women in this age group. The CDC estimates that 30% of the approximately 40,000 new HIV infections in the United States each year are among women.

Between 1996 and 1997 the incidence of new AIDS cases in the United States decreased by 18% and that of AIDS-related deaths by 42%, largely because of advances in HIV therapies. The decline continued between 1997 and 1998, albeit at a slower rate. AIDS incidence and AIDS-related mortality fell by 11 and 20%, respectively. However, AIDS incidence and deaths are not decreasing as rapidly among women as among men. HIV and AIDS continue to affect women in racial/ethnic minorities and lower socioeconomic classes disproportionately. CDC estimates that 64% of new HIV infections in 1998 occured among African-American women, 18% among Hispanic women, and 18% among white women. Of the new HIV infections among women in the United States in 1998, CDC estimates that 75% of women were infected through heterosexual sex and 25% of women through injection drug use.

Violence Against Women Violence against women in the United States is an enormous problem. Incidents of both rape and domestic violence are vastly underreported. Sexual assault is one of the most common crimes against women. One in five adult women in the United States reports having experienced sexual assault during her lifetime. Adult women are much more likely to be raped by a spouse, ex-spouse, or acquaintance than by a stranger.

Domestic violence is defined in the American Medical Association guidelines as "an ongoing, debilitating experience of physical, psychologic, and/or sexual abuse in the home, associated with increasing isolation from the outside world and limited personal freedom and accessibility to resources." It affects women of all ages, ethnic orientations, and socioeconomic groups. Based upon national crime statistics, every year an estimated 2 million women in the United States are severely injured and more than 1000 are killed by their current or former male partner. Domestic violence is the most common cause of physical injury in women, exceeding the combined incidence of all other types of injury (such as from rape, mugging, and auto accidents). Women who are young, single, pregnant, recently separated or divorced, or who have a history of substance abuse or mental illness, or a partner with substance abuse or mental illness, are at increased risk of domestic violence.

Domestic violence and sexual assault are associated with increased rates of physical and psychologic symptoms, medical office visits, and hospitalizations. Given this indirect presentation of the consequences of violence, and the high prevalence of unreported violence, clinicians should have a low threshold for pursuing the possibility of violence in female patients, particularly those with vague symptoms and psychological disorders.

The immediate treatment of rape and domestic violence focuses on assessing and treating physical injuries; providing emotional support; assessing and dealing with the risks of sexually transmitted infection and pregnancy; evaluating the safety of the patient and other family members; and documenting the patient's history and physical examination findings. In addition to dealing with the medical and psychological issues, appropriate care includes providing information about legal services, shelters and safe houses, hotlines, support groups, and counseling services.

RESEARCH IN WOMEN'S HEALTH

The growing recognition of the importance of women's health has spawned a number of research efforts, including large observational studies and clinical trials. The U.S. National Institutes of Health has introduced guidelines to mandate the inclusion of women in clinical studies, and the reporting of gender-specific data.

Studies of Prevention Large observational studies of men and women, such as the Rancho Bernardo Study and the Framingham Study, designed to analyze data specific to women have been on the increase. The Nurses' Health Study has been following more than 200,000 women, many for more than 20 years, prospectively collecting data to study the impact of smoking, diet, physical activity, medications, prevention and screening behaviors, and some psychosocial factors on the risk of various medical disorders, including breast cancer,IHD, stroke, diabetes, and fracture, as well as causes of mortality.

These studies have set the stage for clinical trials such as the Postmenopausal Estrogens/Progestins Intervention (PEPI) Trial, the first multicenter, randomized, double-blind, placebo-control trial of the effects of three estrogen/progestin regimens on risk factors for cardiovascular disease, bone mineral density, and endometrial hyperplasia. The study found that estrogen, alone or in combination with progestin, increased serum levels ofHDL and decreased low-density lipoprotein (LDL) and fibrinogen levels. While unopposed estrogen (without progestins) resulted in the most beneficial effects on lipids, it was also associated with an increased risk of endometrial hyperplasia.

In 1992, the NIH funded the Women's Health Initiative (WHI), a study of the health of postmenopausal women. The WHI, the largest research study ever funded by the NIH, involves over 160,000 postmenopausal women participating at 45 clinical centers across the United States through the year 2002. The WHI study includes both a prospective observational study and an interventional randomized trial involving over 63,000 women, which is designed to test the effects of a low-fat diet, hormone replacement therapy, and calcium and vitamin D supplementation on the risks for cardiovascular disease, breast cancer, and osteoporotic fractures.

Many other studies currently in progress promise new insights into the health of women within the next decade.

Pharmacologic Studies Historically, women have been underrepresented in drug trials, even though the majority of pharmaceuticals sold in the United States each year

are used by women. However, this has been rapidly changing. The FDA requires information on the safety and effectiveness of experimental drugs in women, on the effects of the menstrual cycle and menopause on a drug's pharmacokinetics, and on a drug's influence on the effectiveness of oral contraceptives. The increased emphasis on entering women into drug trials is likely to yield important information. Studies that have included women indicate that there are clinically significant differences in the way women respond to a number of frequently prescribed pharmaceuticals, including sedative-hypnotics, antidepressants, antipsychotics, anticonvulsants, and b-adrenergic blocking agents. The 1992 FDA Adverse Experience Report found that women have a higher frequency of adverse drug reactions than men. Other studies suggest that the efficacy of many drugs may be different in women compared to men. For example, women require lower doses of neuroleptics to control schizophrenia than men do. Women awaken from anesthesia faster than do men who are given the same doses of anesthetics, and they have a more powerful response to certain classes of analgesics than men. The reasons for these differences are not clear. However, these observations have spurred researchers to consider separating out the effects of gender in future clinical research in an effort to define "gender-based" biologic processes.

CONCLUSION

At the same time that the health of women is undergoing more rigorous study and women's clinics are becoming increasingly common and popular, a growing fraction of health professionals are women. The number of women physicians has increased by 300% between 1970 and 1990, and more than 40% of all U.S. medical students now are women. This infusion of women into the physician work force is likely to lead to a still greater recognition of the unique aspects of health and disease in women.

(Bibliography omitted in Palm version)

Back to Table of Contents

7. MEDICAL DISORDERS DURING PREGNANCY - *Robert L. Barbieri*, *John T. Repke*

Approximately 4 million births occur in the United States each year. A significant proportion of these are complicated by one or more medical disorders. Two decades ago, many medical disorders were contraindications to pregnancy. Advances in obstetrics, neonatology, obstetric anesthesiology, and medicine have increased the expectation that pregnancy will result in an excellent outcome for both mother and fetus despite most of these conditions. Successful pregnancy requires important physiologic adaptations, such as a marked increase in cardiac output. Medical problems that interfere with the physiologic adaptations of pregnancy increase the risk for poor pregnancy outcome; conversely, in some instances pregnancy may adversely impact an underlying medical disorder.

HYPERTENSION (See also Chap. 246)

In pregnancy, cardiac output increases by 40%, most of which is due to an increase in stroke volume. Heart rate increases by approximately 10 beats per minute during the third trimester. In the second trimester of pregnancy, systemic vascular resistance decreases and this is associated with a fall in blood pressure. During pregnancy, a blood pressure of 140/90 mmHg is considered to be abnormally elevated and is associated with a marked increase in perinatal morbidity and mortality. In all pregnant women, the measurement of blood pressure should be performed in the sitting position, because for many the lateral recumbent position is associated with a blood pressure lower than that recorded in the sitting position. The diagnosis of hypertension requires the measurement of two elevated blood pressures, at least 6 h apart. Hypertension during pregnancy is usually caused by preeclampsia, chronic hypertension, gestational hypertension, or renal disease.

PREECLAMPSIA

Approximately 5 to 7% of all pregnant women develop *preeclampsia*, the new onset of hypertension (blood pressure > 140/90 mmHg), proteinuria (>300 mg per 24 h), and pathologic edema. Although the precise placental factors that cause preeclampsia are unknown, the end result is vasospasm and endothelial injury in multiple organs. Preeclampsia is associated with abnormalities of cerebral circulatory autoregulation, which increase the risk of stroke at near-normal blood pressures. Risk factors for the development of preeclampsia include nulliparity, diabetes mellitus, a history of renal disease or chronic hypertension, a prior history of preeclampsia, extremes of maternal age (>35 years or <15 years), obesity, factor V Leiden mutation, angiotensinogen gene T235, antiphospholipid antibody syndrome, and multiple gestation.

There are no well-established strategies for the prevention of preeclampsia. Clinical trials have demonstrated that low-dose aspirin treatment does *not* prevent preeclampsia in either low- or high-risk women. Two meta-analyses reported that dietary calcium supplementation appeared to be effective in reducing the risk of developing preeclampsia. Subsequently, however, a large randomized clinical trial in low-risk women did not demonstrate a protective effect of calcium supplementation. Therefore, calcium supplementation may be considered in women at high risk for preeclampsia

(see above). The observation that dietary intervention may reduce the risk of hypertension in men and nonpregnant women raises the possibility that dietary manipulations will be discovered that reduce the risk of preeclampsia.

Severe preeclampsia is the presence of new-onset hypertension and proteinuria accompanied by central nervous system dysfunction (headaches, blurred vision, seizures, coma), marked elevations of blood pressure (>160/110 mmHg), severe proteinuria (>5 g per 24 h), oliguria or renal failure, pulmonary edema, hepatocellular injury (ALT >2´ the upper limits of normal), thrombocytopenia (platelet count < 100,000/uL), or disseminated intravascular coagulation. Women with *mild preeclampsia* are those with the diagnosis of new-onset hypertension, proteinuria, and edema without evidence of severe preeclampsia. The *HELLP* (*h*emolysis, *e*levated *l*iver enzymes, *l*ow *p*latelets) syndrome is a special subgroup of severe preeclampsia and is a major cause of morbidity and mortality in this disease. The presence of platelet dysfunction and coagulation disorders further increases the risk of stroke.

TREATMENT

Preeclampsia resolves within a few weeks after delivery. For pregnant women with preeclampsia prior to 37 weeks' gestation, delivery reduces the mother's morbidity but exposes the fetus to the risk of premature delivery. The management of preeclampsia is challenging because it requires the clinician to balance the health of both mother and fetus simultaneously and to make management decisions that afford both the best opportunities for infant survival. In general, prior to term, women with *mild* preeclampsia can be managed conservatively with bed rest, close monitoring of blood pressure and renal function, and careful fetal surveillance. For women with *severe* preeclampsia, delivery is recommended after 32 weeks' gestation. This reduces maternal morbidity and slightly increases the risks associated with prematurity for the newborn. Prior to 32 weeks' gestation, the risks of prematurity for the fetus are great, and some authorities recommend conservative management to allow for continued fetal maturation. Expectant management of severe preeclampsia remote from term affords some benefits for the fetus with significant risks for the mother. Such management should be restricted to tertiary care centers where maternal-fetal medicine, neonatal medicine, and critical care medicine expertise are available.

The definitive treatment of preeclampsia is delivery of the fetus and placenta. For women with severe preeclampsia, aggressive management of blood pressures> 160/110 mmHg reduces the risk of cerebrovascular accidents.

Intravenous labetalol or hydralazine are the drugs most commonly used to manage preeclampsia. Alternative agents such as calcium channel blockers may be used. Elevated arterial pressure should be reduced slowly to avoid hypotension and a decrease in blood flow to the fetus. *Angiotensin-converting enzyme (ACE) inhibitors as well as angiotensin-receptor blockers should be avoided in the second and third trimesters of pregnancy because of their adverse effects on fetal development.* Pregnant women treated with ACE inhibitors often develop oligohydramnios, which may be caused by decreased fetal renal function.

Magnesium sulfate is the treatment of choice for the prevention and treatment of

eclamptic seizures. Two large randomized clinical trials have demonstrated the superiority of magnesium sulfate over phenytoin and diazepam. Magnesium may prevent seizures by interacting with *N*-methyl-D-asparate (NMDA) receptors in the central nervous system. Given the difficulty of predicting eclamptic seizures on the basis of disease severity, it is recommended that once the decision to proceed with delivery is made, all patients carrying a diagnosis of preeclampsia be treated with magnesium sulfate (seeGuideline).

CHRONIC ESSENTIAL HYPERTENSION

Pregnancy complicated by chronic essential hypertension is associated with intrauterine growth restriction and increased perinatal mortality. Pregnant women with chronic hypertension are at increased risk for superimposed preeclampsia and abruptio placenta. Women with chronic hypertension should have a thorough prepregnancy evaluation, both to identify remediable causes of hypertension and to ensure that the prescribed antihypertensive agents are not associated with adverse pregnancy outcome (e.g.,ACEinhibitors, angiotensin-receptor blockers).a-Methyldopa and labetalol are the most commonly used medications for the treatment of chronic hypertension in pregnancy. Baseline evaluation of renal function is necessary to help differentiate the effects of chronic hypertension versus superimposed preeclampsia should the hypertension worsen during pregnancy. There are no convincing data that demonstrate that treatment of mild chronic hypertension improves perinatal outcome.

GESTATIONAL HYPERTENSION

This is the development of elevated blood pressure during pregnancy or in the first 24 h post partum in the absence of preexisting chronic hypertension and other signs of preeclampsia. Uncomplicated gestational hypertension that does not progress to preeclampsia has not been associated with adverse pregnancy outcome or adverse long-term prognosis.

RENAL DISEASE (See also Chap. 268)

Normal pregnancy is characterized by an increase in glomerular filtration rate and creatinine clearance. This occurs secondary to a rise in renal plasma flow and increase glomerular filtration pressures. Patients with underlying renal disease and hypertension may expect a worsening of hypertension during pregnancy. If superimposed preeclampsia develops, the additional endothelial injury results in a capillary leak syndrome that may make the management of these patients challenging. In general, patients with underlying renal disease and hypertension benefit from more aggressive management of blood pressure than do those with gestational hypertension. Preconception counseling is also essential for these patients so that accurate risk assessment can occur prior to the establishment of pregnancy and important medication changes and adjustments be made. In general, a prepregnancy serum creatinine level <133 umol/L (<1.5 mg/dL) is associated with a favorable prognosis. When renal disease worsens during pregnancy, close collaboration between the nephrologist and the maternal-fetal medicine specialist is essential so that decisions regarding delivery can be weighed in the context of sequelae of prematurity for the neonate versus long-term sequelae for the mother with respect to future renal function.

Successful pregnancy after renal transplantation has been reported increasingly. Predictors for success include a normal-functioning transplanted kidney, absence of rejection for at least 2 years prior to the pregnancy, absence of hypertension, and preferably minimal doses of immunosuppressant medications. Pregnancies in women using cyclosporine are more likely to be complicated by renal insufficiency and/or the development of hypertension. Such patients require very careful maternal and fetal surveillance. Nearly half of these pregnancies deliver preterm, and 20% of neonates are small for their gestational age. Rejection occurs in approximately 10% of pregnancies, and approximately 15% of patients will have deterioration in their renal function that persists after delivery. While pregnancy is generally well tolerated in renal transplant recipients, controversy remains as to whether or not deterioration of graft function is accelerated by pregnancy. More aggressive management of blood pressure has been suggested in this group of patients in an effort to protect the grafted kidney.

Another subset of patients with chronic renal disease and hypertension are those patients whose pregnancies are complicated by systemic lupus erythematosus (SLE) (Chap. 311). In the past, SLE was considered to be a contraindication to pregnancy. With improved understanding of the effects of SLE on pregnancy, and vice versa, and with improved pharmacologic methods for managing SLE, successful pregnancy outcome is likely. Good prognostic factors for establishment of pregnancy in the presence of SLE are as follows:

1. Disease quiescence> 6 months

2. Normal blood pressure (with or without medication)

3. Normal renal function [creatinine < 133 umol/L (< 1.5 mg/dL)]

4. Absence of antiphospholipid antibodies

5. Minimal or no need for immunosuppressive drugs

6. Absence of prior adverse reproductive outcome

Previously a point of controversy, there is now increasing consensus that pregnancy and the postpartum period are times of increased lupus activity. In severe flares early in gestation, pregnancy termination is often recommended. If pregnancy termination is not an option, then medical therapy to manage the lupus flare should not be influenced by the pregnancy, provided informed consent for treatment is obtained from the patient. Pulsed glucocorticoid therapy, azathioprine, hydroxychloroquine, and cyclophosphamide have all been used successfully in pregnancy.

CARDIAC DISEASE

VALVULAR HEART DISEASE (See alsoChap. 236)

This is the most common cardiac problem complicating pregnancy.

Mitral Stenosis This is the valvular disease most likely to cause death during pregnancy. The pregnancy-induced increase in blood volume and cardiac output can cause pulmonary edema in women with mitral stenosis. Pregnancy associated with long-standing mitral stenosis may result in pulmonary hypertension. Sudden death has been reported when hypovolemia has been allowed to occur in this condition. Careful control of heart rate, especially during labor and delivery, minimizes the impact of tachycardia and reduced ventricular filling times on cardiac function. Pregnant women with mitral stenosis are at increased risk for the development of atrial fibrillation and other tachyarrythmias. Medical management of severe mitral stenosis and atrial fibrillation with digoxin and beta blockers is recommended. Balloon valvulotomy can be carried out during pregnancy.

Mitral Regurgitation and Aortic Regurgitation These are both generally well tolerated during pregnancy. The pregnancy-induced decrease in systemic vascular resistance reduces the risk of cardiac failure with these conditions. As a rule, mitral valve prolapse does not present problems for the pregnant patient and aortic stenosis, unless very severe, is also well tolerated. In the most severe cases of aortic stenosis, limitation of activity or balloon valvuloplasty may be indicated.

For women with artificial valves contemplating pregnancy, it is important that warfarin be stopped and heparin initiated prior to conception. Warfarin therapy during the first trimester of pregnancy has been associated with fetal chondrodysplasia punctata. In the second and third trimester of pregnancy, warfarin may cause fetal optic atrophy and mental retardation.

CONGENITAL HEART DISEASE (See also Chap. 234)

The presence of a congenital cardiac lesion in the mother increases the risk of congenital cardiac disease in the newborn. Prenatal screening of the fetus for congenital cardiac disease with ultrasound is recommended. Atrial or ventricular septal defect is usually well tolerated during pregnancy in the absence of pulmonary hypertension, provided that the woman's prepregnancy cardiac status is favorable. Use of air filters on intravenous sets during labor and delivery in patients with intracardiac shunts is generally recommended.

OTHER CARDIAC DISORDERS

Supraventricular tachycardia (Chap. 230) is a common cardiac complication of pregnancy. Treatment is the same as in the nonpregnant patient, and fetal tolerance of medications such as adenosine and calcium channel blockers is acceptable. When necessary, electrocardioversion may be performed and is generally well tolerated by mother and fetus.

Peripartum cardiomyopathy (Chap. 238) is a rare disorder of pregnancy associated with myocarditis, and its etiology remains unknown. Treatment is directed toward symptomatic relief and improvement of cardiac function. Many patients recover completely; others are left with a progressive dilated cardiomyopathy. Recurrence in a subsequent pregnancy has been reported, and women should be counseled to avoid pregnancy after a diagnosis of peripartum cardiomyopathy.

SPECIFIC HIGH RISK CARDIAC LESIONS

Marfan Syndrome (See alsoChap. 351) This is an autosomal dominant disease, associated with a high risk of maternal morbidity. Approximately 15% of pregnant women with Marfan syndrome develop a major cardiovascular manifestation during pregnancy, with almost all women surviving. An aortic root diameter<40 mm is considered to be associated with a favorable outcome of pregnancy. Prophylactic therapy with beta blockers has been advocated, although large-scale clinical trials in pregnancy have not been performed.

Pulmonary Hypertension (See alsoChap. 260) Maternal mortality in the setting of severe pulmonary hypertension is high, and primary pulmonary hypertension is a contraindication to pregnancy. Termination of pregnancy may be advisable in these circumstances to preserve the life of the mother. In the Eisenmenger syndrome, i.e., the combination of pulmonary hypertension with right-to-left shunting due to congenital abnormalities (Chap. 234), maternal and fetal death occur frequently. Systemic hypotension may occur after blood loss, prolonged Valsalva maneuver, or regional anesthesia; sudden death secondary to hypotension is a dreaded complication. Management of these patients is challenging, and invasive hemodynamic monitoring during labor and delivery is generally recommended.

In patients with pulmonary hypertension, vaginal delivery is less stressful hemodynamically than Cesarean section, which should be reserved for accepted obstetric indications.

DEEP VENOUS THROMBOSIS AND PULMONARY EMBOLISM (See also Chaps. 248 and 261)

A hypercoagulable state is characteristic of pregnancy, and deep venous thrombosis (DVT) is a common complication. Indeed, pulmonary embolism is the most common cause of maternal death in the United States. Activated protein C resistance caused by the factor V Leiden mutation increases the risk for DVT and pulmonary embolism during pregnancy. Approximately 25% of women with DVT during pregnancy carry the factor V Leiden allele. The presence of the factor V Leiden mutation also increases the risk for severe preeclampsia. If the fetus carries a factor V Leiden mutation, the risk of extensive placental infarction is very high. Additional genetic mutations associated with DVT during pregnancy include the prothrombin G20210A mutation (heterozygotes and homozygotes) and the methylenetetrahydrofolate reductase C677T mutation (homozygotes).

TREATMENT

Aggressive diagnosis and management ofDVT and suspected pulmonary embolism optimize the outcome for mother and fetus. In general, all diagnostic and therapeutic modalities afforded the nonpregnant patient should be utilized in pregnancy. Anticoagulant therapy with heparin is indicated in pregnant women with DVT. Warfarin therapy is contraindicated in the first trimester due to its association with fetal chondrodysplasia punctata. In the second and third trimesters, warfarin may cause fetal

optic atrophy and mental retardation. In the initial treatment of DVT, heparin, which does not cross the placenta, may be administered as an intravenous bolus of approximately 100 IU per kilogram of body weight. Continuous heparin infusion is generally initiated at 1000 IU/h and then titrated to achieve a target activated partial thromboplastin time of 50 to 80 s. After initial intravenous anticoagulation, intermittent subcutaneous heparin therapy with 10,000 IU two or three times daily may be employed. When deep venous thromboembolism occurs in the postpartum period, heparin therapy for 7 to 10 days may be followed by warfarin therapy for 3 to 6 months. Warfarin is not contraindicated in breast-feeding women.

Low-molecular-weight heparins are of sufficient size and charge that they do not cross the placenta and may be substituted for unfractionated heparin in the pregnant patient. Recent concerns about low-molecular-weight heparin use and epidural hematoma suggest that caution be used in the anesthetic management of patients who had been receiving low-molecular-weight heparin near the onset of labor.

ENDOCRINE DISORDERS

DIABETES MELLITUS (See alsoChap. 333)

In pregnancy, the fetoplacental unit induces major metabolic changes, the purpose of which is to shunt glucose and amino acids to the fetus while the mother uses ketones and triglycerides to fuel her metabolic needs. These metabolic changes are accompanied by maternal insulin resistance, caused in part by placental production of steroids, a growth hormone variant, and placental lactogen. Although pregnancy has been referred to as a state of accelerated starvation, it is better characterized as accelerated ketosis. In pregnancy, after an overnight fast, plasma glucose is lower by 0.8 to 1.1 mmol/L (15 to 20 mg/dL) than in the nonpregnant state. This is due to the use of glucose by the fetus. In early pregnancy, fasting may result in circulating glucose concentrations in the range of 2.2 mmol/L (40 mg/dL) and may be associated with symptoms of hypoglycemia. In contrast to the decrease in maternal glucose concentration, plasma hydroxybutyrate and acetoacetate levels rise to two to four times normal after a fast.

TREATMENT

Pregnancy complicated by diabetes mellitus is associated with higher maternal and perinatal morbidity and mortality rates. Preconception counseling and treatment are important for the diabetic patient contemplating pregnancy. Optimizing preconception glucose control and attention to other dietary needs such as appropriate levels of folate can significantly reduce the risk of congenital fetal malformations. Folate supplementation reduces the incidence of fetal neural tube defects, which occur with greater frequency in fetuses of diabetic mothers. In addition, optimizing glucose control during key periods of organogenesis reduces other congenital anomalies including sacral agenesis, caudal dysplasia, renal agenesis, and ventricular septal defect.

Once pregnancy is established, glucose control should be managed more aggressively than in the nonpregnant state. In addition to dietary changes, this requires more frequent blood glucose monitoring and often involves additional injections of insulin or

conversion to an insulin pump. Fasting blood glucose levels should be maintained at<5.8 mmol/L (<105 mg/dL) with no values exceeding 7.8 mmol/L (140 mg/dL). Commencing in the third trimester, regular surveillance of maternal glucose control as well as assessment of fetal growth (obstetric sonography) and fetoplacental oxygenation (fetal heart rate monitoring or biophysical profile) optimize pregnancy outcome. Pregnant diabetic patients without vascular disease are at greater risk for delivering a macrosomic fetus, and attention to fetal growth via clinical and ultrasound examinations is important. Fetal macrosomia is associated with an increased risk of maternal and fetal birth trauma. Pregnant women with diabetes have an increased risk of developing preeclampsia, and those with vascular disease are at greater risk for developing intrauterine growth restriction, which is associated with an increased risk of fetal and neonatal death. Excellent pregnancy outcomes in patients with diabetic nephropathy and proliferative retinopathy have been reported with aggressive glucose control and intensive maternal and fetal surveillance.

Glycemic control may become more difficult to achieve as pregnancy progresses. Because of delayed pulmonary maturation of the fetuses of diabetic mothers, early delivery should be avoided unless there is biochemical evidence of fetal lung maturity. In general, efforts to control glucose and maintain the pregnancy until the estimated date of delivery result in the best overall outcome for both mother and newborn.

GESTATIONAL DIABETES

All pregnant women should be screened for gestational diabetes unless they are in a low-risk group. Women at low risk for gestational diabetes are those <25 years of age; those with a body mass index < 25 kg/m_2, no maternal history of macrosomia or gestational diabetes, and no diabetes in a first-degree relative; and those not members of a high-risk ethnic group (African American, Hispanic, Native American). A typical two-step strategy for establishing the diagnosis of gestational diabetes involves administration of a 50-g oral glucose challenge with a single serum glucose measurement at 60 min. If the serum glucose is < 7.8 mmol/L (<140 mg/dL), the test is considered normal. Serum glucose > 7.8 mmol/L (>140 mg/dL) warrants administration of a 100-g oral glucose challenge with serum glucose measurements obtained in the fasting state, and at 1, 2, and 3 h. Normal values are serum glucose concentrations<5.8 mmol/L (<105 mg/dL), 10.5 mmol/L (190 mg/dL), 9.1 mmol/L (165 mg/dL), and 8.0 mmol/L (145 mg/dL), respectively.

Pregnant women with gestational diabetes are at increased risk of preeclampsia, delivering infants who are large for their gestational age, and birth lacerations. Their fetuses are at risk of hypoglycemia and birth trauma (brachial plexus) injury.

TREATMENT

Gestational diabetes is first treated with dietary measures. Inability to maintain fasting glucose concentrations <5.8 mmol/L (<105 mg/dL) or 2-h postprandial glucose concentrations <6.7 mmol/L (<120 mg/dL) should prompt initiation of insulin therapy. Oral agents should not be used to treat diabetes in pregnancy. Patients with a diagnosis of gestational diabetes will benefit from postpartum follow-up as they are at increased risk for developing type 2 diabetes.

THYROID DISEASE (See alsoChap. 330)

In pregnancy, the estrogen-induced increase in thyroxine-binding globulin causes an increase in circulating levels of total T_3 and total T_4. The normal range of circulating levels of free T_4, free T_3, and thyroid stimulating hormone (TSH) remain unaltered by pregnancy.

The thyroid gland normally enlarges during pregnancy. Maternal hyperthyroidism occurs at a rate of approximately 2 per 1000 pregnancies and is generally well tolerated by pregnant women. Clinical signs and symptoms should alert the physician to the occurrence of this disease. Many of the physiologic adaptations to pregnancy may mimic subtle signs of hyperthyroidism. Although pregnant women are able to tolerate mild hyperthyroidism without adverse sequelae, more severe hyperthyroidism can cause spontaneous abortion or premature labor, and thyroid storm is associated with a significant risk of maternal mortality.

TREATMENT

Hyperthyroidism in pregnancy should be aggressively evaluated and treated. The treatment of choice is propylthiouracil. Because it crosses the placenta, the minimum effective dose should be used to maintain free T_4 in the upper normal range. Methimazole crosses the placenta to a greater degree than propylthiouracil and has been associated with fetal aplasia cutis. Radioiodine should not be used during pregnancy, either for scanning or treatment, because of effects on the fetal thyroid. In emergent circumstances, additional treatment with beta blockers and a saturated solution of potassium iodide may be necessary. Hyperthyroidism is most difficult to control in the first trimester of pregnancy and easiest to control in the third trimester.

The goal of therapy for *hypothyroidism* is to maintain the serumTSH in the normal range, and thyroxine is the drug of choice. Children born to women with an elevated serum TSH (and a normal total thyroxine) during pregnancy have impaired performance on neuropsychologic tests. During pregnancy, the dose of thyroxine required to keep the TSH in the normal range rises. In one study, the mean replacement dose of thyroxine required to maintain the TSH in the normal range was 0.1 mg daily before pregnancy, and it increased to 0.15 mg daily during pregnancy.

DISORDERS OF CALCIUM METABOLISM (See also Chap. 340)

Serum *total* calcium concentration decreases throughout gestation due to a reduction in serum albumin concentration, while serum *ionized* calcium remains unchanged during pregnancy. Circulating parathyroid hormone concentration is slightly reduced throughout the course of pregnancy. Pregnancy has been described as a state of physiologic absorptive hypercalciuria. Estrogen and increased production of 1,25-dihydroxyvitamin D by both the kidney and the placenta mediate the increased absorption of calcium during pregnancy. Due to the fetal requirements for calcium, the National Institutes of Health has recommended that pregnant women receive 1500 mg/d of elemental calcium, slightly higher than the recommended daily intake of 1200 mg/d for nonpregnant adults.

HEMATOLOGIC DISORDERS

Pregnancy has been described as a state of physiologic anemia. Part of the reduction in hemoglobin concentration is dilutional, but iron and folate deficiencies are the major causes of correctable anemia during pregnancy. Folic acid food supplementation implemented in 1998 has reduced the risk of fetal neural tube defects.

In populations at high risk for hemoglobinopathies (Chap. 106), hemoglobin electrophoresis should be performed as part of the prenatal screen. Hemoglobinopathies can be associated with increased maternal and fetal morbidity and mortality. Management is tailored to the specific hemoglobinopathy and is generally the same for both pregnant and nonpregnant women. Prenatal diagnosis of hemoglobinopathies in the fetus is readily available and should be discussed with prospective parents either prior to or early in pregnancy.

Thrombocytopenia occurs commonly during pregnancy. The majority of cases are benign gestational thrombocytopenias, but the differential diagnosis should include immune thrombocytopenia (Chap. 116) and preeclampsia. Maternal thrombocytopenia may also be caused by catastrophic obstetric events such as retention of a dead fetus, sepsis, abruptio placenta, and amniotic fluid embolism.

NEOPLASTIC DISEASES

Maternal neoplasms are rarely, if ever, transmitted to the fetus. The three most common cancers in pregnant women are cervical cancer (~1 case per 1000 pregnancies, depending on the country), breast cancer (~2 cases per 10,000 pregnancies), and lymphomas (Hodgkin's disease or non-Hodgkin's lymphomas). Cervical cancer may be missed when its early sign, vaginal bleeding, is attributed to the pregnancy. Pregnant women with vaginal bleeding should be examined, and suspicious cervical lesions biopsied. Conization is generally performed only after the first trimester because of the abortion risk.

Breast lumps may also be attributed to change associated with pregnancy. However, women with a dominant mass should undergo diagnostic evaluation (mammogram, ultrasound, biopsy). Resection of the primary lesion is safe, but radiation therapy is unsafe at any time during pregnancy. The fetus cannot be shielded from internal scattering of radiation; therapeutic doses are associated with spontaneous abortion, increased perinatal death, and defects in central nervous system and/or cognitive function. Tamoxifen is not safe for pregnant women.

Lymphoma is usually diagnosed on the basis of adenopathy or constitutional symptoms (fever, sweats, or weight loss). Staging evaluation is not undertaken during the first trimester; women in the first trimester should be counseled about termination of the pregnancy. Single-agent chemotherapy can be used in the second or third trimester as a temporizing measure. Vinblastine or doxorubicin have been used most commonly. Early induction of labor may permit the physician to maximize the survival chances of both the fetus and the mother. Survival rates for 28-week-old fetuses are about 75% and about 90% for 32-week-old fetuses.

Cancer survivors of reproductive age may desire children. Pregnancy may increase the risk of melanoma recurrence but does not influence breast cancer recurrence. Cancer treatment may deplete oocytes. Oocyte retrieval and storage of fertilized or nonfertilized eggs before cancer treatment may permit conception after the cancer has been treated successfully.

GASTROINTESTINAL AND LIVER DISEASE

Up to 90% of pregnant women experience nausea and vomiting during the first trimester of pregnancy. Occasionally, hyperemesis gravidarum requires hospitalization to prevent dehydration, and sometimes parenteral nutrition is required.

Crohn's disease may be associated with exacerbations in the second and third trimesters. Ulcerative colitis is associated with disease exacerbations in the first trimester and during the early postpartum period. Medical management of these diseases during pregnancy is identical to the management in the nonpregnant state (Chap. 287).

Exacerbation of gall bladder disease is commonly observed during pregnancy. In part this may be due to pregnancy-induced alteration in the metabolism of bile and fatty acids. Intrahepatic cholestasis of pregnancy is generally a third-trimester event. Profound pruritus may accompany this condition and may be associated with increased fetal mortality. It has been suggested that placental bile salt deposition may contribute to progressive uteroplacental insufficiency. Therefore, regular fetal surveillance should be undertaken once the diagnosis of intrahepatic cholestasis is made. Favorable results with ursodiol have been reported.

Acute fatty liver is a rare complication of pregnancy. Frequently confused with theHELLPsyndrome (see "Preeclampsia," above) and severe preeclampsia, the diagnosis of acute fatty liver of pregnancy may be facilitated by imaging studies and laboratory evaluation. Acute fatty liver of pregnancy is generally characterized by markedly increased levels of bilirubin and ammonia and by hypoglycemia. Management of acute fatty liver of pregnancy is supportive; recurrence in subsequent pregnancies has been reported.

All pregnant women should be screened for hepatitis B. This information is important for pediatricians after delivery of the infant. All infants receive hepatitis B vaccine. Infants born to mothers who are carriers of hepatitis B surface antigen should also receive hepatitis B immune globulin as soon after birth as possible and preferably within the first 72 h.

INFECTIONS

BACTERIAL INFECTIONS

Other than bacterial vaginosis, the most common bacterial infections during pregnancy involve the urinary tract (Chap. 280). Many pregnant women have asymptomatic bacteriuria, most likely due to stasis caused by progestational effects on ureteral and

bladder smooth muscle and to compression effects of the enlarging uterus. In itself, this condition is not associated with an adverse outcome of pregnancy. However, if asymptomatic bacteriuria is left untreated, symptomatic pyelonephritis may occur. Indeed, approximately 75% of cases of pregnancy-associated pyelonephritis are the result of untreated asymptomatic bacteriuria. All pregnant women should be screened with a urine culture for asymptomatic bacteriuria at the first prenatal visit. Subsequent screening with nitrite/leukocyte esterase strips is indicated for high-risk women, such as those with sickle cell trait or a history of urinary tract infections. All women with positive screens should be treated.

Because of the association between bacterial vaginosis and preterm delivery, screening for bacterial vaginosis has been used in an effort to reduce risk. However, standard treatment for bacterial vaginosis does not reduce the risk of preterm delivery.

Abdominal pain and fever during pregnancy create a clinical dilemma. The diagnosis of greatest concern is intrauterine amniotic infection. While amniotic infection most commonly follows rupture of the membranes, this is not always the case. In general, antibiotic therapy is not recommended as a temporizing measure in these circumstances. If intrauterine infection is suspected, induced delivery with concomitant antibiotic therapy is generally indicated. Intrauterine amniotic infection is most often caused by pathogens such as *Escherichia coli* and group B streptococcus. In high-risk patients at term or in preterm patients, routine intrapartum prophylaxis of group B streptococcal disease is recommended. Penicillin G and ampicillin are the drugs of choice. In penicillin-allergic patients, clindamycin is recommended.

Postpartum infection is a significant cause of maternal morbidity and mortality. While rare after vaginal delivery, postpartum endomyometritis develops in 5% of patients having elective repeat cesarean section and in 25% of patients after emergency cesarean section following prolonged labor. Prophylactic antibiotics should be given to all patients undergoing cesarean section. As most cases of postpartum endomyometritis are polymicrobial, broad-spectrum antibiotic coverage with a penicillin, aminoglycoside, and metronidazole is recommended (Chap. 167). Most cases resolve within 72 h. Women who do not respond to antibiotic treatment for postpartum endomyometritis should be evaluated for septic pelvic thrombophlebitis. Imaging studies may be helpful in establishing the diagnosis, which is primarily a clinical diagnosis of exclusion. Patients with septic pelvic thrombophlebitis generally have tachycardia out of proportion to their fever and respond rapidly to intravenous administration of heparin.

All patients are screened prenatally for gonorrhea and chlamydial infections, and the detection of either should result in prompt treatment. Ceftriaxone and azithromycin are the agents of choice (Chaps. 147 and179).

VIRAL INFECTIONS

Cytomegalovirus Infection Viral infection in pregnancy presents a significant challenge. The most common cause of congenital viral infection in the United States is cytomegalovirus (CMV) (Chap. 185). As many as 50 to 90% of women of childbearing age have antibodies to CMV, but only rarely does CMV reactivation result in neonatal infection. More commonly, primary CMV infection during pregnancy creates a risk of

congenital CMV. No currently accepted treatment of CMV during pregnancy has been demonstrated to protect the fetus effectively. Moreover, it is impossible to predict which fetus will sustain life-threatening CMV infection. Severe CMV disease in the newborn is characterized most often by petechiae, hepatosplenomegaly, and jaundice. Chorioretinitis, microcephaly, intracranial calcifications, hepatitis, hemolytic anemia, and purpura may also develop. Central nervous system involvement resulting in the development of psychomotor, ocular, auditory, and dental abnormalities over time have been described.

Rubella (See alsoChap. 195) Rubella virus is a known teratogen; first-trimester rubella carries a high risk of fetal anomalies, though the risk decreases significantly later in pregnancy. Congenital rubella may be diagnosed by percutaneous umbilical blood sampling with the detection of IgM antibodies in fetal blood. All pregnant women should be screened for their immune status to rubella. Indeed, all women of childbearing age, regardless of pregnancy status, should have their immune status for rubella verified and be immunized if necessary. The incidence of congenital rubella in the United States is extremely low.

Herpesvirus (See alsoChap. 182) The acquisition of genital herpes during pregnancy is associated with spontaneous abortion, prematurity, and congenital and neonatal herpes. A recent cohort study of pregnant women without evidence of previous herpes infection demonstrated that approximately 2% of the women acquired a new herpes infection during the pregnancy. Approximately 60% of the newly infected women had no clinical symptoms. Infection occurred equally in all three trimesters. If herpes seroconversion occurred early in pregnancy, the risk of transmission to the newborn was very low. In women who acquired genital herpes shortly before delivery, the risk of transmission was high. The risk of active genital herpes lesions at term can be reduced by prescribing acyclovir for the last 4 weeks of pregnancy to women who have had their first episode of genital herpes during the pregnancy. However, whether or not this strategy results in less viral shedding or enhanced fetal protection at delivery remains to be determined.

Herpesvirus infection in the newborn can be devastating. Disseminated neonatal herpes carries with it high mortality and morbidity rates from central nervous system involvement. It is recommended that pregnant women with active genital herpes lesions at the time of presentation in labor be delivered by cesarean section.

Parvovirus (See alsoChap. 187) Parvovirus infection (human parvovirus B19) may occur during pregnancy. It rarely causes sequelae, but susceptible women infected during pregnancy may be at risk for fetal hydrops secondary to erythroid aplasia and profound anemia.

Toxoplasmosis (See alsoChap. 217) In the United States, approximately 70% of women of childbearing age are susceptible to *Toxoplasma*. Most primary infections of toxoplasmosis in the United States come from eating undercooked meat. The diagnosis of congenital toxoplasmosis is possible through sampling of fetal umbilical blood. If there is no evidence of placental/fetal infection, single-drug treatment with spiramycin is recommended. Triple-drug therapy with spiramycin, pyrimethamine, and sulfa is recommended if there is evidence of fetal infection and the woman does not wish to

terminate the pregnancy or cannot terminate it because of advanced gestational age. Prenatal treatment has been shown to reduce the number of infants with severe infection.

Human Immunodeficiency Virus (See also Chap. 309) The predominant cause of HIV infection in children is transmission of the virus from the mother to the newborn during the perinatal period. Exposures, which increase the risk of mother-to-child transmission, include vaginal delivery, preterm delivery, trauma to the fetal skin, and maternal bleeding. Additionally, recent infection with high maternal viral load, low maternal CD4+T cell count, prolonged labor, prolonged length of membrane rupture, and the presence of other genital tract infections, such as syphilis or herpes, increase the risk of transmission. Breast feeding may also transmit HIV to the newborn and is therefore contraindicated in most developed countries for HIV-infected mothers. There is no clear evidence to suggest that the course of HIV disease is altered by pregnancy. There is also no clear evidence to suggest that uncomplicated HIV disease adversely impacts pregnancy other than by its inherent infection risk.

TREATMENT

The majority of cases of mother-to-child (vertical) transmission of HIV-1 occur during the intrapartum period. Mechanisms of vertical transmission include infection after rupture of the membranes and direct contact of the fetus with infected secretions or blood from the maternal genital tract. In women with HIV infection who are not receiving antiretroviral therapy, the rate of vertical transmission is approximately 25%. Cesarean section and treatment with zidovudine, administered both before and during delivery, decrease the rate of vertical transmission. In a meta-analysis, zidovudine treatment of both the mother during the prenatal and intrapartum periods and of the neonate at birth reduced the risk of vertical transmission to 7.3%. The combination of elective cesarean section plus zidovudine treatment reduced the risk of vertical transmission to 2%. The role of multiple drug therapy during pregnancy has not yet been established, pending safety data for the neonate.

SUMMARY

Maternal mortality has decreased steadily during the past 60 years. The maternal death rate has decreased from nearly 600/100,000 live births in 1935 to 8.5/100,000 live births in 1996. The most common causes of maternal death in the United States today are, in decreasing order of frequency, thromboembolic disease, hypertension, ectopic pregnancy, and hemorrhage. With improved diagnostic and therapeutic modalities as well as with advances in the treatment of infertility, more patients with medical complications will be seeking, and be in need of, complex obstetric care. Improving outcome of pregnancy in these women will be best obtained by assembling a team of internists and specialists in maternal-fetal medicine (high-risk obstetrics) to counsel these patients about the risks of pregnancy and to plan their treatment prior to conception. The importance of preconception counseling cannot be overstated. It is the responsibility of all physicians caring for women in the reproductive age group to assess their patient's reproductive plans as part of their overall health evaluation.

(Bibliography omitted in Palm version)

Back to Table of Contents

8. ADOLESCENT HEALTH PROBLEMS - *Mehul T. Dattani*, *Charles G. D. Brook*

Adolescence marks the transition from childhood to adulthood. It is a time of dramatic physical and psychological change. Adolescents are particularly prone to risk-taking behaviors. In the United States, 73% of deaths among adolescents and young adults result from motor vehicle and other accidents, homicide, and suicide. As a result of sexual maturation, adolescents begin to experiment sexually and, consequently, are susceptible to sexually transmitted diseases (STDs) and unwanted pregnancies. For adolescents with underlying disease, the psychological consequences can be as important as the physical disabilities; denial or resentment of disease is common and can hamper treatment. Adolescence is also a time when many lifelong health-relevant behaviors are established, including dietary habits, exercise patterns, tobacco and alcohol use, and interactions with the health care system. The physician, working together with parents, can help to guide adolescents through this dynamic period of life.

PUBERTY

Puberty encompasses (1) the adolescent growth spurt, (2) development of secondary sexual characteristics, (3) attainment of fertility, and (4) establishment of individual sexual identity.

There is wide variation in the timing of puberty. Signs of puberty are first evident between 9 and 14 years of age (mean 11.5) in 95% of American boys (Fig. 8-1*A*). In girls, puberty begins earlier, with 95% of American girls entering puberty between 8 and 12 years of age (mean 10.5) (Fig. 8-1*B*). The age of menarche in girls from developed countries has decreased by approximately 2 to 3 months per decade over the past 100 to 150 years. This trend is likely the result of improvements in socioeconomic conditions, nutritional status, and general health and well-being. Currently in the United States, the average age of menarche is 12.8 years. Genetic factors also influence the course of puberty. Data from twin studies indicate that the average age of menarche is more similar in identical twin sisters than in nonidentical twins. Secondary sexual development occurs earlier in girls of Asian and African-Caribbean heritage than in girls of European heritage. Recognition of the progressively earlier onset of puberty, the ethnic variations, and wide age distribution in the timing of puberty is important for identifying precocious or delayed puberty and for counseling adolescents and parents about the natural course of physiologic changes.

HORMONAL CHANGES

Puberty is accompanied by dramatic changes in multiple hormonal systems including alterations in adrenal steroid production, maturation of the reproductive axis, and increased production and action of growth hormone (GH). Serum GH levels increase early in puberty as a consequence of the rise in gonadal steroids. GH in turn increases the level of insulin-like growth factor 1 (IGF-1), which enhances linear bone growth (Chap. 328). The prolonged pubertal exposure to gonadal steroids ultimately causes epiphyseal closure and limits further bone growth. Appetite increases in association with the growth spurt, and sleep patterns change with a tendency to stay up later and a desire to sleep later into the morning.

The development of secondary sexual characteristics is initiated by *adrenarche*, which usually occurs between 6 and 8 years of age and marks the time when the adrenal gland begins to produce greater amounts of androgens. Despite the search for hormonal mediators, the mechanism that controls adrenarche is unknown. It may well result from adrenal cell differentiation, characterized by the growth of the innermost zone of the adrenal cortex, the zona reticularis, which is the principal site of dehydroepiandrosterone (DHEA) production. The increase in adrenal androgen (DHEA, androstenedione) secretion precedes activation of the reproductive axis. Nonetheless, adrenarche and gonadarche are independent events: children with Addison's disease enter puberty at the normal time, and children with premature adrenarche achieve gonadarche at the expected age.

The sexual maturation process is greatly accelerated by the activation of the hypothalamic-pituitary axis, leading to gonadal stimulation and the production of sex steroids. The hypothalamic-pituitary-gonadal axis is controlled predominantly by gonadotropin-releasing hormone (GnRH), a decapeptide produced by the arcuate nucleus of the mediobasal hypothalamus. GnRH is released in a pulsatile fashion, leading in turn to the pulsatile secretion of the pituitary gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). The so-called GnRH pulse-generator in the hypothalamus is active during fetal life and early infancy but is then quiescent during early childhood. In the early stages of puberty, the sensitivity to steroid inhibition is gradually lost, causing reactivation of GnRH secretion. Leptin, a hormone produced by adipose cells, may play a permissive role in this process, as leptin-deficient individuals fail to enter puberty. Early puberty is characterized by nocturnal surges of LH and FSH. As the reproductive axis matures, the characteristic patterns of feedback regulation are acquired. In males, testosterone inhibits hypothalamic GnRH and pituitary gonadotropin production (Chap. 335); in females, estrogen and progesterone feed back to generate the characteristic hormonal patterns of the menstrual cycle (Chap. 336).

FEMALE PUBERTY

The first sign of ovarian estradiol secretion is breast development, or *thelarche*. Pubic and axillary hair growth and the onset of apocrine sweat production result from adrenal androgen secretion, though they may be facilitated by estrogen. The progression of puberty is classified according to Tanner stages for breast and pubic hair development: stage 1 represents the preadolescent appearance, and stage 5 represents the adult appearance (Fig. 8-1*A*). Each of these aspects of puberty should be staged separately, as they are controlled by different underlying endocrine mechanisms. Concurrent with these outward signs of puberty are the changes in the size and shape of the uterus. The pubertal growth spurt is dependent on estradiol secretion, which leads to increasedGHsecretion. This in turn results in a doubling of the growth rate, and peak height velocity is usually coincident with breast stage 3. After menarche, a girl usually grows only an additional 5 cm.

MALE PUBERTY

In boys, growth of the testes is usually the first sign of puberty, reflecting the effects of pulsatile gonadotropin secretion on seminiferous tubule volume and, to some degree,

Leydig cell mass. Testosterone is converted to dihydrotestosterone by 5-a reductase. Both hormones act via the androgen receptor to induce growth of the external genitalia and pubic hair (Fig. 8-1;Chap. 338). The growth spurt in boys occurs at a testicular volume of about 10 to 12 mL (as measured by a Prader orchidometer). Testosterone also deepens the voice and increases muscle growth. Dihydrotestosterone stimulates prostate growth and beard growth and initiates recession of the temporal hairline. Although boys enter puberty approximately 6 to 12 months later than girls, they are potentially fertile at an earlier stage of puberty. Aromatization of testosterone to estradiol increasesGHsecretion, which acts synergistically with testosterone to induce a greater peak height velocity in boys than in girls.

DISORDERS OF PUBERTY

Precocious puberty is usually defined as an early onset of puberty in boys younger than 9 years or in girls younger than 8 years of age. Some authorities suggest that the lower limits of normal in girls be revised downward to age 7 for Caucasians and age 6 for African-American girls. *Premature thelarche* refers to breast development in the absence of other signs of puberty. It occurs most commonly in girls between infancy and 3 years of age and usually resolves spontaneously. Causes of precocious puberty are divided into central *gonadotropin-dependent* forms and peripheral *gonadotropin-independent* forms (Table 8-1). Central precocious puberty is much more common in girls than in boys, and the majority of these cases involve idiopathic activation of spontaneousGnRHpulses. It can also be caused by a variety of central nervous system tumors, structural lesions, and inflammatory conditions.

Delayed puberty is defined as the 3% of girls and boys who have not developed the first signs of puberty by 13.2 and 14.2 years, respectively. It is most commonly due to delayed activation of the hypothalamic-pituitary-gonadal axis (Table 8-1). Most individuals who meet this definition will progress through puberty normally, but at a later age. Short stature and delayed skeletal maturation are commonly seen in association with delayed puberty. Growth delay may have been evident earlier in childhood; the diagnosis of *constitutional delay of growth and puberty* can be suspected from a delayed bone age in a short child who is otherwise well. Individuals who experience delays in puberty may be emotionally as well as physically immature relative to their peers.

The main diagnostic challenge in delayed puberty is to distinguish those with constitutional delay, who will progress through puberty at a later age, from those with an underlying pathologic process.LH andFSHresponses toGnRH do not differentiate constitutional delay from pathologic causes of *hypogonadotropic hypogonadism* (Chap. 335). Thus, constitutional delay is a diagnosis of exclusion and requires ongoing evaluation during development to assure that normal growth and development occur at a later time. Reassurance without hormonal treatment is appropriate for most individuals with presumed constitutional delay of puberty. Alternatively, an anabolic steroid (e.g., 50 to 100 mg per month testosterone enanthate, intramuscularly) in boys or estrogen (5 to 10 mg/d ethinyl estradiol, orally) in girls may be useful to induce growth and secondary sexual characteristics appropriate for age. Low-dose oral oxandrolone (2.5 mg/d), an anabolic steroid that is not aromatized to estrogen, is also used for boys because it does not accelerate skeletal maturation when used for short periods. After treatment for a year or more, hormonal treatments can be stopped and the function of the

reproductive axis can be reassessed.

PSYCHOLOGICAL CHANGES AND SOCIAL FACTORS

The adolescent years are characterized by a multitude of psychological changes, including (1) the development of abstract thinking, (2) greater independence from family, (3) the formation of a personal and sexual identity, (4) the establishment of a system of values, and (5) an increase in socialization. For most adolescents, these transitions occur relatively smoothly. For others, however, these years can be frustrating and tumultuous; parents and clinicians must be attuned to the needs of those who show signs of struggling with emotional, sexual, and social issues.

Young people tend to share their feelings openly, one of which is ambivalence. These contradictory feelings most often involve both a desire for greater autonomy and, at the same time, a need to cling to the emotional and physical security provided by the family. Adolescents are granted increasing responsibilities but still lack some of the social and legal privileges of adults. This feature of adolescence can lead to conflict and challenges to parental authority.

Adolescents have a strong desire to establish an identity that is increasingly independent of the family. This new identity is strongly influenced by peer groups, some of which are institutionalized (e.g., team sports). Role confusion is quite common in adolescence, and some young people move from one intense allegiance to another with alarming speed. These transitional arrangements are eventually replaced by more permanent attachments to individuals.

In an attempt to alleviate some of the transitions associated with adolescence, many cultures have traditionally used "rites of passage" to acknowledge and accelerate an adolescent's evolution to adulthood. Among Native American Great Plains cultures, for example, a boy was sent away from the village at the time of puberty to fast and receive a vision from a spirit; upon returning to the community, he took his place among the adult men. Similarly, it was traditional in many societies for girls to be secluded at the time of the first menstruation before returning a "full-grown woman." These rites provide a public recognition of the end of childhood, and the ritual leaves the young person with the conviction that he or she has undergone a personal transformation. A relative lack of these coming-of-age rituals in western cultures may contribute to the sense of alienation experienced by some adolescents in this part of the world.

During adolescence, gender identity must be renegotiated. Though prepubertal children have a relatively secure view of themselves as either a boy or a girl, experimentation with gender roles is a common feature of adolescence. For example, adolescents may explore, at least in fantasy, alternative gender roles (e.g., cross-dressing), homosexuality, or relationships with older men or women.

The hormonal changes of puberty influence behavior as well as causing physical changes. Rising levels of testosterone in boys and the increase in adrenal and ovarian androgens in girls increase libido. The mean age of sexual intercourse varies widely within and among cultures, but ranges between ages 15 and 18 for most groups. Boys generally report sexual intercourse about 1 year earlier than girls.

ADOLESCENT VIOLENCE

Adolescents and young adults are subject to much greater rates of violence, both as victims and perpetrators. Males are involved in violence much more commonly than females and account for>90% of homicides involving those 10 to 17 years of age. Ethnic and racial differences in rates of adolescent violence have been noted consistently. African Americans, Hispanics, and Native Americans are much more likely to be victims and perpetrators of lethal violence than are people of Asian or European ancestry. The origins of different rates of violence are complex. Higher rates of lethal aggression are associated with low socioeconomic status, high housing density, increased population turnover in neighborhoods, single-parent households, and socially disorganized communities. In many cases, these factors interact; increased violence leads to high population turnover and social disorganization.

Gangs represent a potentially volatile environment that is characterized by power struggles, initiation and detachment rituals, battles over territory, and escalating violence associated with retaliation. The increase in lethal violence has been attributed in part to easier access to firearms and a greater willingness to use firearms. A Centers for Disease Control and Prevention study in 1995 found that about one-fourth of students had carried a weapon to school during the preceding month and 8 to 10% had carried a gun. Many adolescents lack the abstract reasoning skills required to understand social mores and the consequence of gun use. Though firearms do not cause violence, handguns in particular provide a facile means to a lethal outcome; widespread reduction in access to handguns is essential to curb the current trend in adolescent homicide and serious injury.

Aggressive behavior can often be recognized in early childhood; bullying is a precursor to later antisocial behavior. Child abuse, antisocial parents, inadequate child-rearing practices, and dysfunctional interpersonal interactions between parents or among siblings are associated with aggressive behavior. The physician, along with teachers, clergy, and others in positions of authority, should be alert to a pattern of aggressive behavior or problems in the home. Though these issues are not easily remedied, appropriate interventions to improve family functioning and parenting may interrupt a pattern of violence, which is all too often perpetuated by the adolescent.

HEALTH PROBLEMS

Adolescence is generally a healthy period and is often accompanied by a feeling of immortality, which leads to risk-taking. When diseases of childhood or the consequences of their treatment extend into adolescence, or when disease strikes during adolescence, the sense of unfairness may be overwhelming. Anger and denial can lead to poor compliance with therapeutic regimens.

Relatively few diseases are unique to adolescents. Rather, diseases of childhood, including many inherited disorders and infectious diseases, extend into the adolescent period. Similarly, many of the disorders that affect teenagers are also seen in the adult population. The presentation and management of asthma, for example, is similar in adolescents and adults. Some of the diseases with relatively increased prevalence

during adolescence are summarized inTable 8-2. These diseases should be borne in mind when considering the differential diagnosis. For example, when an adolescent presents with exertional chest pain, dyspnea, and syncope, hypertrophic cardiomyopathy or congenital heart disease should be considered as likely diagnoses, whereas coronary artery disease would be more likely in an adult.

SEXUALLY TRANSMITTED DISEASES

Sexually active adolescents are at greater risk of acquiringSTDsthan their adult counterparts (Chap. 132). Prevention of STDs in adolescence depends on adequate sexual education coupled with access to appropriate clinical services. Early age of first sexual intercourse is associated with (1) an increased number of lifetime sexual partners; (2) an increased risk of acquiring chronic STDs, such as herpes simplex, HIV, and hepatitis B; and (3) cervical cancer in women. In addition, pelvic inflammatory disease in adolescent females increases the likelihood of future ectopic pregnancy, tubal infertility, and chronic pelvic inflammation. A low rate of barrier contraceptive use, combined with ignorance about the acquisition and prevention of infectious diseases, also contributes to the increased risk of STDs among adolescents. Screening for STDs is recommended in sexually active teens (Table 8-3). Adolescents with sexually transmitted infections, particularly those who deny sexual activity, may be victims of sexual abuse.

CHILD SEXUAL ABUSE

Child sexual abuse is defined as the involvement of developmentally immature children and adolescents in sexual activities they do not comprehend, to which they are unable to give consent, or that violate social taboos or family roles. In a U.S. study in 1985, sexual abuse during childhood was reported by 27% of adult females and 16% of adult males. Females are more likely than males to have been sexually abused by a family member. Although there is a paucity of literature on male sexual abuse, it is probably more common than generally recognized. The psychological trauma appears to be similar for boys and girls. Sexual abuse during adolescence may merge with peer sexual assault, or "date rape." Sexual abuse in adolescent girls can be associated with a constant fear of pregnancy. Teenage pregnancy orSTD may, in fact, be the first indication of ongoing abuse.

Psychological consequences of child sexual abuse often involve behavioral problems, psychiatric disturbances, or adjustment difficulties at the onset of adolescence, even though the actual abuse may have taken place at a younger age. Child sexual abuse may lead to low self-esteem and/or a degree of sexual disinhibition. The cognitive maturation that occurs with adolescence may bring about the realization and expression of these feelings. Young women who have been sexually abused have significantly higher rates of early-onset consensual sexual activity, teenage pregnancy, multiple sexual partners, unprotected intercourse,STDs, and later sexual assault. Poor psychological outcome is related to the duration of abuse, the extent to which the abuse involves violence or coercion, and the perception that the child has cooperated with the abuser, with ensuing feelings of guilt. The impact of these sequelae can be reduced by supportive peer and family relationships. Disclosure of the abuse may help to ameliorate some of the psychological traumas associated with abuse.

SUBSTANCE ABUSE

Substance abuse and drug misuse among adolescents is a significant cause of morbidity and mortality (Chaps. 386 to 389). The prevalence rates vary widely by region, ethnic group, age, and gender. The age of initiation into substance abuse has gradually declined. In 1997, rates among American teenagers for substance use or abuse, at some stage during their lifetimes, were: cigarettes smoking (70%), alcohol use (79%), marijuana use (47%), cocaine use (8%), anabolic steroids (4%), injected illegal drugs (2%), and other illegal drugs (17%), e.g., lysergic acid (LSD), phencyclidine (PCP), methylenedioxymethamphetamine (ecstasy), methamphetamine (ice), or heroin.

The forms of substance abuse change continuously. Anabolic steroids, for example, are now used by 3 to 5% of male high school seniors, with a 10% prevalence rate among male adolescent athletes. In addition to their use by athletes in an effort to increase muscle strength, nonathletes use anabolic steroids with a goal of achieving a more virile appearance. In contrast to popular views, anabolic steroids do not appear to enhance performance except at very high doses, which are associated with significant side effects (Chap. 335). Other performance-enhancing agents include human growth hormone and erythropoietin (EPO), but the high cost of these hormones limits their use.

In addition to the direct effect on health, substance abuse is associated with other risk-taking behaviors. The relationship of alcohol use and motor vehicle accidents, for example, is well documented. However, drug and alcohol use are also correlated with many other problems during adolescence including violence, suicide, depression,STD, and unwanted pregnancies. Therefore, the presence of one form of risky behavior should prompt consideration of others.

SUICIDAL BEHAVIOR AND DEPRESSION

After motor vehicle accidents and homicide, suicide is the third leading cause of death in adolescents, and the rate has risen almost fourfold over the past 50 years. In 1988, the suicide rate among 15- to 19-year olds was 11.3 in 100,000. The causes for increased rates of suicide are not well understood, but one theory holds that modern society fosters increased social isolation and alienation. Nearly one-fourth of adolescents acknowledge seriously considering suicide, and 8% have actually attempted it. Attempted suicide is three times more common in females than males, with drug overdose or wrist-cutting being the most common means of suicide attempt. Completed suicide is three to five times more common in teenage boys than girls and usually involves firearms, hanging, or jumping from heights. Suicide is rare before puberty. Risk factors for suicide among adolescents include prior attempt of suicide, a history of depression or other major psychiatric disorder, history of substance abuse, medical illness, family history of suicidal behavior, and knowing someone who has committed suicide. Unfortunately, these and other risk factors are relatively common among nonsuicidal youth as well, making suicide difficult to predict in individual cases. Stressful events can precipitate depression and increase risk of suicide; these can include the death of a relative or friend, disciplinary crisis, rejection or humiliation, school difficulty, and anxiety about homosexuality. Apparently impulsive actions may be harbingers of more serious underlying mood disturbances, personality disorders, or substance abuse.

Major depression occurs in 4 to 6% of adolescents, and the *DSM-IV* criteria for diagnosis are the same as in adults (Chap. 385). Every depressed or suicidal adolescent should undergo psychiatric examination, whether hospitalized or not. Comprehensive evaluation requires exploration of the adolescent's history of mental health problems, symptoms of depression, level of functioning in school, interactions with friends and family, and evaluation for comorbid disorders. Indications for hospitalization include imminent risk of suicide as evidenced by an identified plan and access to lethal means, recurrent suicide attempts, the presence of severe depression or psychosis, substance abuse, and the need to remove the individual from an overwhelmingly stressful environment.

ADOLESCENT EATING DISORDERS

Many adolescents have voracious appetites in response to the increased energy and caloric requirements generated by the growth spurt. The unique physical, psychological, and social transitions of adolescence provide a context for the development and perpetuation of eating patterns. Adolescents with a body mass index (BMI), measured as weight (kg)/height (m_2), greater than the 95th percentile for age and gender are overweight, and those between the 85th and 94th percentiles are at risk for becoming overweight. Based on the NHANES III survey for 1988 to 1994, there was evidence for a 6% increase in the prevalence of overweight adolescents compared to the previous decade. The increasing prevalence of obesity is multifactorial and involves patterns of eating behavior as well as alterations in activity level (Chap. 77). Physical activity among both girls and boys tends to decline steadily during adolescence. Regular involvement in enjoyable forms of exercise should be encouraged to help promote lifelong habits that involve physical activity.

Eating disorders such as anorexia nervosa or bulimia nervosa often have their onset during adolescence (Chap. 78). Control over dietary intake is perhaps one of the first mechanisms that adolescents use to establish autonomy and achieve independence from family. The majority of female adolescents and young adults in western cultures report feeling discontented with their body shape. Surveys of normal adolescent populations disclose a surprisingly high frequency of dieting and abnormal eating patterns. For instance, up to 79% binge, 70% consider themselves fat, 11% induce vomiting, 5% abuse laxatives, and about 3% meet diagnostic criteria for anorexia or bulimia nervosa. Eating disorders also occur in males, but much less frequently than in females.

PHYSICIAN-ADOLESCENT RELATIONSHIP

The transition from the pediatrician to an adult medical practice can be difficult for adolescents, their parents, and their physicians. The emergence of adolescent medicine as a specialty practice has helped to facilitate this transition and to focus on the special needs of this group. When adolescents transfer to an adult-based practice, the physician should first establish a relationship with the patient and his or her parents. Previous medical history should be reviewed and medical records obtained. The need for the teenager to be seen alone, and office policies concerning confidentiality, should be discussed and agreed to with the parent(s) and adolescent together.

Legal issues related to the medical care of minors arise frequently, and laws vary in different countries and from state to state (Chap. 2). As a general rule, anyone who has reached the age of majority (usually 18 years) may consent to treatment. Under this age, a parent or legal guardian must consent for medical intervention. However, there are several exceptions to this requirement. The delivery of medical care is generally accepted in an emergency, but it is important to document the nature of the emergency and any efforts to notify parents. Emancipated minors may also provide consent. This group includes those fulfilling adult roles (e.g., military service), married teens, and those who are financially independent and living separately from their parents. In addition, when the health of a minor is potentially endangered by disorders for which they may be reluctant to seek parental consent, such as substance abuse, pregnancy, orSTD, mature minors may generally provide consent. In these circumstances, the caregiver must assess the minor's maturity, ability to understand the risks and benefits of treatment, and capacity to provide informed consent. Mature or emancipated minors do not need to reveal consent or treatment to their parents.

Obtaining a medical history from an adolescent includes many elements that are distinct from an adult history. It should include, for example, schoolwork, home environment, and relationships with parents, siblings, and peers. Adolescents often lack knowledge about medical issues and may be reluctant to discuss sensitive topics with authority figures. Most will be nervous, even when these issues do not pertain. It can be useful, therefore, to provide printed forms or questionnaires. These not only serve to gather information in a relatively nonthreatening manner but also provide an indication of the kinds of issues that might be discussed with the physician. It is difficult to predict the topics that are paramount to the adolescent. Some may be preoccupied with concerns about the onset of acne, whereas others fear HIV or pregnancy. Adolescents may harbor guilt about sexual abuse or feel overwhelmed by peer pressure to engage in certain activities. The physician is well positioned to assist with many of these issues, if there is trust and an indication of interest and understanding. Because of these types of questions, it is important to interview the adolescent in private. Some parents will resist this approach, but it is necessary if the adolescent is to volunteer information that he or she is unwilling to discuss in the presence of parents. It is useful to reinforce the fact that conversations will be kept confidential. Adolescents are sometimes willing to bring concerns to the attention of nurses or other caregivers before raising these issues with a physician. It is helpful, therefore, to have another health care provider interact with the patient, if only briefly. In addition to direct questioning, general conversation about topical issues or inquiries about school or peers may provide insight into an adolescent's interests, activities, and potential risk factors. Because of the prevalence of substance abuse, risk-taking behavior, suicide, sexual orientation crises,STDs, unwanted pregnancies, sexual abuse, depression, and eating disorders in the teenage years, these topics warrant specific inquiry as part of routine health assessment. The interview should also include adequate time for education, health care guidance, and counseling. It is also useful to provide written information about topics that are pertinent to the care of the adolescent.

The physical examination of the adolescent, while incorporating many elements of the adult examination, has several unique features. Foremost among these is the assessment of growth and sexual development. In addition to questionnaires that allow

the adolescent an opportunity to self-assess stages of pubertal development, the examination can be made less stressful by using it as opportunity to explain normal physiology. The issue of when to perform a pelvic examination as part of the routine health maintenance is controversial. Some advocate pelvic examinations in all sexually active young women as a means to detectSTDs and for Pap smears. With the advent of urinary screening tests for chlamydia and gonorrhea, others suggest that pelvic examinations are not routinely necessary in the absence of specific indications. When a pelvic examination is performed, the patient should be asked whether she prefers her mother or a member of the health care team as an observer. The physical examination should also focus on diseases that tend to present during adolescence (Table 8-2).

Disorders such as hypertension, hyperlipidemia, and obesity are often first detected during adolescence. Strategies for disease prevention also include immunization, avoidance of cigarette smoking or excessive alcohol use, establishing good dietary habits, and engaging in regular exercise. General guidelines for adolescent preventive services (GAPS) are summarized inTable 8-3.

SUMMARY

The term *adolescent* is derived from a Latin phrase meaning "to grow up." Adolescence is, in many ways, the culmination of development, with the achievement of identity and reproductive competence. Though these processes are triggered by internal physiologic events, they are intimately intertwined with the family and social environment. Physicians have an important role to facilitate these transitions by providing information and managing the diseases of adolescents. Moreover, it should be remembered that many adolescents view physicians as role models and will seek objective and informed advice about issues that reach beyond medicine.

(Bibliography omitted in Palm version)

Back to Table of Contents

9. GERIATRIC MEDICINE - *Neil M. Resnick*

Of all the people who have ever lived to age 65, more than half are now alive. This statistic has important demographic and economic implications, and its impact on medical care is also substantial.

BIOLOGY OF AGING

Numerous molecular concomitants of aging have been described. For instance, there is an increase in chromosome structural abnormalities, DNA cross-linking, and frequency of single-strand breaks; a decline in DNA methylation; and loss of DNA telomeric sequences. The primary structure of proteins is unaltered, but posttranslational changes, such as deamidation, oxidation, cross-linking, and nonenzymatic glycation, increase. Mitochondrial structure also deteriorates, albeit not universally.

However, the biologic changes are clearer than the mechanisms that mediate them. In fact, although the senescent phenotype appears to be ubiquitous, biologists disagree about whether senescence even exists beyond zoos and civilized societies and whether it occurs at all in many species. There is little evolutionary rationale for a process that happens after reproduction is complete, particularly one associated with such a long and complex course. In nature, senescence is most notable for its absence; nearly all animals die of predation, disease, or environmental hazards rather than aging. The argument that different species have different maximum life spans can be explained without invoking a specific aging process: while growth and development are based on a genetic template, aging may reflect merely the accumulation of random damage rather than a specific mechanism.

If aging exists as a distinct process, there is consensus that the mechanisms are likely multifactorial, environmentally influenced, and species-specific, if not organ- and cell-specific, making the paucity of available human data particularly problematic. As a result, there are nearly as many theories of aging as investigators. Most theories overlap or are not mutually exclusive, and none is completely compatible with the dearth of data. As a group, the theories can be divided into two broad categories, based on whether they attribute aging to a genetic program or to progressive and random damage to homeostatic systems.

Enthusiasm for genetic theories of aging is fueled by several observations, including the dramatic species-specific differences in maximal life span, the strong correlation with survival among monozygotic compared to dizygotic twins, and the fact that single mutations can prolong life span by more than 50% in some nematodes and mice. However, all genetic theories must account for the fact that evolutionary selection pressure is minimal following completion of reproduction. Three genetic theories have recently been advanced, but few relevant data have yet been accrued. The first theory suggests that, since animals usually succumb to natural forces long before reaching their maximal life span, aging might reflect mutations that impair long-term survival. These mutations would accumulate in the genome because there is no selection pressure to delete them. A second theory, "pleiotropic antagonism," proposes that aging may be caused by the late and deleterious effects of genes that are conserved because of the survival advantages they confer prior to reproduction. The third theory applies to

ecological niches where extrinsic hazards are relatively low. In such an environment, evolution might select for mutations that retard the aging process since these might allow an animal to produce and protect many more litters. In support of this theory, the rate of aging in an isolated clan of Virginia opossums was calculated to be roughly half of that seen in their less fortunate cousins.

The "random damage" theories are based on the possibility that the balance between ongoing damage and repair is disrupted. The theories differ in the emphasis placed on increased damage (e.g., by free radicals, oxidation, or glycation) versus deficient repair, as well as in the mechanisms that might mediate each. However, all share the observation that cell and organ repair capacity declines with age. Some 40 years ago, Hayflick and Moorehead observed that the number of replications among cultured cells is finite. Subsequent research revealed that this replicative senescence was due to arrest of the cell cycle at the G_1/S phase, the point at which DNA synthesis begins. Recently, cell replication has also been linked to the length of telomeric DNA. Present at the termini of chromosomes, telomeric DNA prevents chromosomal instability, fragmentation, and rearrangement; anchors chromosomes to nuclear matrix; and provides a buffer between coding regions of DNA and the ends of the chromosomes. In addition, telomeric DNA is necessary for cell division. With each cell division, however, roughly 50 of the total 2000 base pairs of the telomere are lost. Telomeric shortening might thus result in loss of gene accessibility, which is necessary to repair ongoing cell damage caused by metabolism. Together with cytoplasmic factors mediating arrest of DNA synthesis, telomeric shortening could also limit the cell's ability to divide and thereby replace cells lost to apoptosis.

Many mechanisms previously postulated to mediate aging have not been borne out, including the somatic mutation theory (in which aging would result from cumulative spontaneous mutations), the error catastrophe theory (in which aging would result from errors in the synthesis of proteins critical to the synthesis of genetic material or protein-synthesizing machinery), and the intrinsic mutagenesis theory (in which aging is the result of ongoing intrinsic DNA rearrangements).

To date, the only intervention known to delay aging is caloric restriction. The salutary effect of restricting caloric intake by 30 to 40% has been documented in multiple species, from single-cell organisms to rodents. In rodents, it not only increases average life expectancy and maximum life span but also delays the onset of some typical age-associated diseases as well as deterioration of physiologic systems (e.g., immune responsiveness, glucose metabolism, muscle atrophy). Moreover, its impact is evident in both mitotic and postmitotic cells, in gene expression, and in protein turnover and cross-linking. Although the mechanism is still not determined, it is specific to caloric restriction rather than to reduction of any dietary component (e.g., fat intake) or supplements with vitamins or antioxidants. Unfortunately, adequate data from primates are not yet available, and the effect of caloric restriction in humans is still unknown.

PRINCIPLES OF GERIATRIC MEDICINE

Despite the biologic controversy, from a physiologic standpoint human aging is characterized by progressive constriction of the homeostatic reserve of every organ system. This decline, often referred to as *homeostenosis*, is evident by the third decade

and is gradual and progressive, although the rate and extent of decline vary. The decline of each organ system (Table 9-1) appears to occur independently of changes in other organ systems and is influenced by diet, environment, and personal habits as well as by genetic factors.

Several important principles follow from these facts: (1) Individuals become more dissimilar as they age, belying any stereotype of aging; (2) an *abrupt* decline in any system or function is always due to disease and not to "normal aging"; (3) "normal aging" can be attenuated by modification of risk factors (e.g., increased blood pressure, smoking, sedentary lifestyle); and (4) "healthy old age" is not an oxymoron. In fact, *in the absence of disease, the decline in homeostatic reserve causes no symptoms and imposes few restrictions on activities of daily living regardless of age*.

Appreciation of these facts may make it easier to understand the striking increases that have occurred in life expectancy. Average life expectancy is now 17 years at age 65, 11 years at age 75, 6 years at age 85, 4 years at age 90, and 2 years at age 100. Moreover, the bulk of these years is characterized by a lack of significant impairment (Table 9-2). Even beyond age 85, only 30% of people are impaired in any activity required for daily living and only 20% reside in a nursing home. Yet, as individuals age they are more likely to suffer from disease, disability, and the side effects of drugs, all of which, when combined with the decrease in physiologic reserve, make the older person more vulnerable to environmental, pathologic, and pharmacologic challenges.

The following concepts underlie the remainder of the chapter:

1. Disease presentation is often atypical in the elderly, especially in those more than 75 to 80 years old. Homeostatic strain caused by onset of a new disease often leads to symptoms associated with a different organ system, particularly one compromised by preexisting disease. For example, fewer than one-fourth of older patients with hyperthyroidism present with goiter, tremor, and exophthalmos; more likely are atrial fibrillation, confusion, depression, syncope, and weakness. Significantly, because the "weakest link" is so often the brain, the lower urinary tract, or the cardiovascular or musculoskeletal system, a limited number of presenting symptoms predominate -- acute confusion, depression, incontinence, falling, and syncope -- no matter what the underlying disease. Thus for the most common geriatric syndromes, regardless of the presenting symptom, the differential diagnosis is often largely similar. The corollary is equally important: The organ system usually associated with a particular symptom is less likely to be the source of that symptom in older individuals than in younger ones. Compared with middle-aged individuals, for example, acute confusion in older patients is less often due to a new brain lesion, depression to a psychiatric disorder, incontinence to bladder dysfunction, falling to a neuropathy, or syncope to heart disease.

2. Because of decreased physiologic reserve, older patients often develop symptoms at an earlier stage of their disease (Fig. 9-1). For example, heart failure may be precipitated by mild hyperthyroidism, cognitive dysfunction by mild hyperparathyroidism, urinary retention by mild prostatic enlargement, and nonketotic hyperosmolar coma by mild glucose intolerance. Paradoxically, therefore, treatment of the underlying disease may be easier because it is frequently less advanced at the time of presentation. A

corollary is that drug side effects can occur with drugs and drug doses unlikely to produce side effects in younger people (Chap. 71). For instance, an antihistamine (e.g., diphenhydramine) may cause confusion, loop diuretics may precipitate urinary incontinence, digoxin may induce depression even with normal serum levels, and over-the-counter sympathomimetics may precipitate urinary retention in men with mild prostatic obstruction.

Unfortunately, the predisposition to develop symptoms at an earlier stage of disease is often offset by the change in illness behavior that occurs with age. Raised at a time when symptoms and debility were accepted as normal consequences of aging, the elderly are less likely to seek attention until symptoms become disabling. Thus, any symptom, particularly those associated with a change in functional status, must be taken seriously and evaluated promptly.

3. Since many homeostatic mechanisms may be compromised concurrently, there are usually multiple abnormalities amenable to treatment, and small improvements in each may yield dramatic benefits overall. For instance, cognitive impairment in patients with Alzheimer's disease may respond much better to interventions that alleviate comorbidity than to prescription of donepezil (Fig. 9-2). Similar approaches apply to most other geriatric syndromes, including falls, incontinence, depression, delirium, syncope, and fracture. In each case, substantial functional improvement can result from treating the contributing factors even if -- as in Alzheimer's disease -- the disease itself is largely untreatable.

4. Many findings that are abnormal in younger patients are relatively common in older people -- e.g., bacteriuria, premature ventricular contractions, low bone mineral density, impaired glucose tolerance, and uninhibited bladder contractions. However, they may not be responsible for a particular symptom but only be incidental findings that result in missed diagnoses and misdirected therapy. For instance, the finding of bacteriuria should not end the search for a source of fever in an acutely ill older patient, nor should an elevated random blood sugar -- especially in an acutely ill patient -- be incriminated as the cause of neuropathy. On the other hand, certain other abnormalities must not be dismissed as due to old age -- e.g., there is no anemia, impotence, depression, or confusion of old age.

5. Because symptoms in older people are often due to multiple causes, the diagnostic "law of parsimony" often does not apply. For instance, fever, anemia, retinal embolus, and a heart murmur prompt almost a reflex diagnosis of infective endocarditis in a younger patient but may reflect aspirin-induced blood loss, a cholesterol embolus, insignificant aortic sclerosis, and a viral illness in an older patient. Moreover, even when the diagnosis is correct, treatment of a single disease in an older patient is unlikely to result in cure. For instance, in a younger patient, incontinence due to involuntary bladder contractions is treated effectively with a bladder relaxant medication. However, in an older patient with the same condition but who also has fecal impaction, takes medications that cloud the sensorium, and suffers from arthritis-associated impairments of mobility and manual dexterity, treatment of the bladder spasms alone is unlikely to restore continence. On the other hand, disimpaction, discontinuation of the offending medications, and treatment of the arthritis are likely to restore continence without the need for a bladder relaxant. Failure to recognize these principles often leads to

prescribing "ineffective" therapy and to unjustified therapeutic nihilism towards older patients.

6. Because the older patient is more likely to suffer the adverse consequences of disease, treatment -- and even prevention -- may be equally or even more effective. For instance, the survival benefits of exercise, as well as thrombolysis and beta-blocker therapy after a myocardial infarction, are as impressive in older patients as in younger ones; and treatment of hypertension and transient ischemic attacks, as well as immunization against influenza and pneumococcal pneumonia, are more effective in older patients. In addition, prevention in older patients must often be seen in a broader context. For instance, although interventions to increase bone density may be limited in older patients, fracture may still be prevented by efforts to improve balance, strengthen legs, reduce peripheral edema, treat other contributing medical conditions, replete nutritional deficits, eliminate environmental hazards, and remove adverse medications -- not so much those that affect bone metabolism, but rather those that induce orthostasis, confusion, and extrapyramidal stiffness.

In summary, optimal treatment of the older patient generally requires treating much more than the organ system usually associated with the disease or symptom, and often permits ignoring that system entirely.

EVALUATION

Evaluation of the older patient can be time-consuming, even when it is tailored to the problem. Yet, such initial investment can reduce subsequent morbidity and resource utilization and enhance patient and physician satisfaction. Additionally, the assessment can often be accomplished over several visits. Moreover, much can be gleaned from questionnaires filled out by the patient or caregiver in advance as well as from observation. For instance, greeting the patient in the waiting room allows the physician to note affective and cognitive response, the strength of the handshake, the ease of rising from a chair without using the arms, the length and steadiness of the stride, and the ability to follow directions to the examining room and to sit down safely in the examining room chair. Observing the patient dress or undress can also enhance detection of impaired cognition, fine motor skills, balance, and judgment. Such observations often provide more information than standard examinations and can shorten the clinical evaluation.

HISTORY TAKING IN ELDERLY PATIENTS

Most older patients are able to provide a reliable medical history; however, a multitude of complaints may make obtaining a history more difficult. If the patient is unable to comprehend or communicate, data should be sought from family, friends, and caregivers. The history should also include drug ingestion; dietary patterns; falling, incontinence, sexual dysfunction, depression and anxiety.

Advance Directives All older patients should be asked whether they have drafted advance health care directives, and, if they have, a copy should be placed in the record. Such directives may consist of a health care proxy or durable power of attorney for health care, in which patients designate a surrogate decision-maker who makes health

care decisions if the patient cannot, and/or a living will or medical directive, in which patients specify their desires for treatment in specific situations if they cannot communicate at the critical time.

Whether or not the patient has formally drafted these directives, it is useful to indicate in the record who should make health care decisions if the patient is no longer able to do so. Patients should then be encouraged to discuss their thoughts with the physician as well as the designated proxy. It is not feasible to cover all possible future complications in such discussions. Ascertaining patients' perspectives on specific interventions, such as resuscitation or intubation, is also difficult because preferences will likely differ depending on prognosis. For instance, a patient may not be interested in feeding tube placement following a massive stroke with little chance of recovery but would prefer the same intervention if it is short-term and helps ensure more rapid and complete recovery from an intercurrent illness such as pneumonia. More useful is a discussion that uses open-ended questions and empathic comments to elicit the patient's values and goals. Moreover, for any given condition, preferences may differ depending on baseline clinical status. For robust elderly individuals, recovery is a realistic goal, albeit the odds of complications are higher than for younger individuals. For the frail elderly patient with comorbidity that impairs functional status, reduction or alleviation of symptoms may be the goal. For patients with advanced dementia or terminal illness, palliation may be the most appropriate strategy. In each situation, however, early elicitation of a patient's preferences and values -- when the patient can still state them -- can often help both physicians and families in subsequent difficult decisions by giving surrogate decision-makers the sense that they are doing as the patient would have wanted.

PHYSICAL EXAMINATION

Certain features of the examination should receive special attention, depending in part on clues from the history. Weight and postural blood pressure should be measured at most visits. Vision and hearing should be checked; if hearing is impaired, excess cerumen should be removed from the external auditory canals prior to audiologic referral. Denture fit should be assessed, and the oral cavity should be inspected with the dentures removed. Although thyroid disease becomes more common with age, the sensitivity and specificity of related findings are substantially lower than in younger individuals; consequently, the physical examination can rarely corroborate or exclude thyroid dysfunction in older patients. The breasts should not be overlooked, since older women are more likely to have breast cancer and less likely to do breast self-examination. The systolic murmur of aortic sclerosis is common and may be difficult to differentiate from aortic stenosis, especially since the presence of a fourth heart sound in an elderly person does not imply significant cardiac disease, and the carotid upstroke normally increases owing to age-related arterial stiffening.

In inactive patients and those with fecal or urinary incontinence, one should check for fecal impaction. In patients with urinary incontinence -- especially men -- a distended bladder must be looked for, since it may be the only finding in urinary retention; perineal sensation and the bulbocavernosus reflex should also be tested. Patients who fall should be observed standing up from a chair, bending down, reaching up, walking 10 feet, turning, returning, and sitting again; abnormalities of gait and balance should be evaluated with the patient's eyes open and closed and in response to a sternal push. It

should be appreciated that "frontal release signs" (e.g., "snout," "glabellar," or palmomental reflexes) and absent ankle jerks and vibratory sense in the feet may be normal in the elderly.

MENTAL STATUS EXAMINATION

In addition to evaluating mood and affect, some form of cognitive testing is essential in all elderly patients, even if it involves only checking different components of the history for consistency. People with mild degrees of dementia usually retain their social graces and may mask intellectual impairment by a cheerful and cooperative manner. Thus, the examiner should always probe for content. For patients who follow the news, one can ask what stories they are particularly interested in and why; the same applies to reading, social events -- even the soap operas on television.

If there is any suspicion of a cognitive deficit after this kind of conversational probing, further questioning is indicated. An examination that tests only orientation as to person, place, and time is insufficient to detect mild or moderate intellectual impairment. As a quick screen, simply assessing orientation and asking the patient to draw a clock with the hands at a set time (e.g., 10 min before 2:00) can be very informative regarding cognitive status, visuospatial deficits, ability to comprehend and execute instructions in logical sequence, and presence or absence of perseveration. For slightly more detailed examinations, many practical mental status tests are available. The most widely used is the Mini-Mental Status Examination of Folstein (Chap. 24), which provides a numerical score that can be obtained in 5 to 10 min. Regardless of the test employed, the total score is less useful diagnostically than is knowledge of the specific domain of the deficit. As a general rule, disproportionate difficulty with immediate recall (e.g., of a list of three items) suggests depression, while predominant difficulty with recalling the items 5 min later suggests dementia. For patients with deficits of attention -- recognized by inability to spell simple words backwards, repeat five digits, or recite the months of the year backwards -- delirium is probably present, and the accuracy of the remainder of the test is dubious. However, the test can be interpreted accurately only in the context of a comprehensive evaluation.

EVALUATION OF FUNCTIONAL CAPACITY

Medical problem lists, a standard tool for assessing and following younger patients, often prove inadequate for older patients. Heart failure, stroke, and prostate cancer can describe a bedbound institutionalized person as well as a Supreme Court justice. Thus, it is essential to ascertain the patient's degree of functional incapacity owing to both medical and psychosocial problems. The functional assessment includes determination of the patient's ability to perform basic activities of daily life (ADL), which are those needed for personal self-care, as well as the ability to perform more complex tasks required for independent living, the instrumental activities of daily living (IADL). ADLs include bathing, dressing, toileting, feeding, getting in and out of chairs and bed, and walking. IADLs include shopping, cooking, money management, housework, using a telephone, and traveling outside the home. For frail patients, an assessment in the home by a trained observer may be required, but for most patients a questionnaire dealing with these activities can be completed by the family or patient. In either case, the physician must determine the cause of any impairment and whether it can be

treated. Assessment should conclude with determination of the socioeconomic circumstances and social support systems.

MANAGEMENT OF COMMON GERIATRIC CONDITIONS

Diseases more common in the elderly are covered elsewhere in the text. The medical problems discussed below do not usually present as clear-cut organ-specific diagnoses and are most common in the frail elderly, especially those over 80 years of age.

INTELLECTUAL IMPAIRMENT

The predominant causes of impaired mentation in older patients are delirium, dementia, and depression. Each condition is covered elsewhere in the text in detail (Chaps. 24 and362), but their management in the elderly is discussed here.

Differentiating the causes of impaired mentation is important, but in older patients they frequently coexist. Thus, the most important first step is to search for and correct all factors that may contribute to cognitive impairment, even in patients with dementia (Fig. 9-2). Evidence of dangerous behavior should also be sought (e.g., leaving the stove on, wandering, and getting lost), and plans should be devised to deal with it. Although there is no specific pharmacologic treatment for Alzheimer's disease and agents such as donepezil are of limited efficacy, this does not mean that the physician has no further role in treating the patient and family. In addition to discontinuing all nonessential medications and treating new intercurrent illness, the physician should help the family and patient predict and deal with the disease; indeed, the family often needs the physician's support more than the patient does.

TREATMENT

Community services should be suggested as needed, including a visiting nurse, a home health aide to assist with personal hygiene, a homemaker to assist with housework, meal delivery, transportation services, day health centers, and respite care to ease the burden on family members. Support groups such as the Alzheimer's Association are often of value to the family and help them to anticipate problems. Signs of patient abuse by an overstressed caregiver should be watched for. Legal counsel should be recommended to help the patient and family devise plans for ongoing management and ultimate disposition of assets not already obtained; advance directives should be sought as soon as possible while the patient can still participate.

Finally, abrupt worsening of mentation or the onset of disruptive behavior should always prompt a search for new illness or medication. Exacerbation of cognitive dysfunction may occur with mild infections (e.g., subungual toe abscess, vaginitis, or pressure ulcer); with "therapeutic" levels of many drugs; with use of nonprescribed drugs or alcohol; with modest abnormalities of serum sodium, calcium, glucose, or thyroxine; with mild hypoxia; with borderline nutritional deficiencies; with subdural hematoma or "minor" stroke; and with the development of fecal impaction, urinary retention, pain, or change in environment, particularly in frail older patients. However, if a cause is not found and behavior does not respond to environmental manipulation (e.g., ignoring the behavior, distracting the patient, addressing situational "triggers," and providing a calm

environment), low doses of an antipsychotic medication may be helpful (e.g., haloperidol 0.25 to 2 mg/d orally; see below).

DEPRESSION

Depression of significant degree occurs in 5 to 10% of community-dwelling elderly but is often overlooked. At highest risk are individuals with recent medical illness (e.g., stroke or fracture), bereavement, lack of social supports, recent nursing home admission, or psychiatric history (including alcohol abuse). The diagnosis requires the presence of a depressed mood for at least two consecutive weeks plus at least four of the following eight symptoms: sleep disturbance, lack of interest, feelings of guilt, decreased energy, decreased concentration, decreased appetite, psychomotor agitation/retardation, and suicidal ideation. Also helpful diagnostically are a personal or family history of depression, anhedonia (loss of pleasure), and past response to an antidepressant. It is essential to bear in mind that depression in older patients is often caused or contributed to by drugs or a systemic illness. Although "subsyndromal" depression (fewer than four of the above symptoms) also causes substantial morbidity and health resource utilization, it appears to be less responsive than major depression to therapy.

TREATMENT

For the hospitalized patient in whom acute depression delays recovery or rehabilitation -- when correction of medical and pharmacologic contributing factors is ineffective and there is no prior history of mania or major depression -- methylphenidate, 5 to 10 mg at 8 A.M. and noon (to avoid insomnia) is often very effective, with benefits discernible within a few days. For patients with major depression, there is no ideal antidepressant drug. All are about equally effective, but the side effects differ (see below and Chap. 385). Consequently, one should become familiar with one or two agents for patients with psychomotor retardation (e.g., sertaline, desipramine) and for those with agitation (e.g., nortriptyline or nefazodone). Because of its potent anticholinergic and orthostatic side effects, amitriptyline should be avoided whenever possible in older patients. Initial low dosages should be increased slowly to avoid serious side effects; low doses of each medication (e.g., nortriptyline, 10 to 50 mg daily; desipramine, 25 to 75 mg daily; or sertraline 50 to 150 mg daily) are often effective in the elderly. Careful follow-up is required to anticipate and minimize anticholinergic side effects, orthostatic hypotension, sedating effects, confusion, bizarre mental symptoms, cardiovascular complications, and drug overdose with suicidal intent. Adverse drug reactions should not be assumed to be due to the aging process.

Cautious use of the monoamine oxidase inhibitors is sometimes of benefit when other antidepressants are ineffective. Neither monoamine oxidase inhibitors nor selective serotonin reuptake inhibitors should be used in combination with the cyclic compounds. Electroconvulsive therapy has been successful and is usually well tolerated by elderly patients who remain severely depressed despite drug treatment, particularly if they also have delusions.

URINARY INCONTINENCE

Transient Incontinence (Table 9-3) Because urinary continence requires adequate

mobility, mentation, motivation, and manual dexterity -- in addition to integrated control of the lower urinary tract -- problems outside the bladder can result in incontinence.

1. *Delirium*. A clouded sensorium impedes recognition of both the need to void and the location of the nearest toilet; once delirium clears, incontinence resolves.

2. *Infection*. Symptomatic urinary tract infection commonly causes or contributes to incontinence; asymptomatic infection does not.

3. *Atrophic urethritis/vaginitis*. Atrophic urethritis/vaginitis, characterized by the presence of vaginal telangiectasia, petechiae, erythema, or friability, commonly contributes to incontinence in women and responds to a several-month course of low-dose estrogen or vaginal estrogen creams.

4. *Pharmaceutical*. The drugs most commonly causing transient incontinence are listed inTable 9-4.

5. *Psychologic*. Depression and psychosis are uncommon but treatable causes.

6. *Excess urine output*. Excess urine output may overwhelm the ability to reach a toilet in time. Causes include diuretics, alcohol, excess fluid intake, and metabolic abnormalities (e.g., hyperglycemia, hypercalcemia, diabetes insipidus); nocturnal incontinence may also result from mobilization of peripheral edema.

7. *Restricted mobility*. If mobility cannot be improved, access to a urinal or commode may restore continence. (See "Immobility," below.)

8. *Stool impaction*. This is a common cause of urinary incontinence, especially in hospitalized or immobile patients. Although the mechanism is unknown, a clue to its presence is the coexistence of both urinary and fecal incontinence. Disimpaction restores continence.

Established Incontinence (Table 9-3) The causes of established incontinence include irreversible functional deficits, such as *end-stage* Alzheimer's disease, and intrinsic lower urinary tract dysfunction. Lower urinary tract dysfunction should be sought after transient causes have been excluded.

Detrusor Overactivity This disorder (involuntary bladder contraction) accounts for two-thirds of geriatric incontinence in both sexes, regardless of whether patients are demented. Detrusor overactivity can be diagnosed presumptively in a woman when leakage occurs in the absence of stress maneuvers or urinary retention and is preceded by the abrupt onset of an intense urge to urinate that cannot be forestalled. In men, the symptoms are similar, but since detrusor overactivity often coexists with urethral obstruction, urodynamic testing should be done if prescription of a bladder relaxant is planned. Because detrusor overactivity may also be due to bladder stones or tumor, the abrupt onset of otherwise unexplained urge incontinence -- especially if accompanied by perineal/suprapubic discomfort or sterile hematuria -- should prompt cystoscopy and cytologic examination.

TREATMENT

The cornerstone of treatment is behavioral therapy with or without biofeedback. Patients without dementia are instructed to void every 1 to 2 h (while awake only) and to suppress urgency in between; once daytime continence is restored, the interval between voiding can be progressively increased. Demented patients are "prompted" to void at similar intervals. When drugs are necessary, they should be added to these regimens and monitored to avoid inducing urinary retention. Effective drugs include oxybutynin (2.5 to 5 mg three or four times daily, or sustained release, 5 to 20 mg once daily), dicyclomine (10 to 30 mg three times daily), tolterodine (1 to 2 mg twice daily), and imipramine or doxepin (25 to 100 mg at bedtime). If prescribed for older patients, DDAVP should be used cautiously -- especially in the setting of renal insufficiency or heart failure -- and it probably should not be given to patients with hyponatremia or urine output >2500 mL/d. Alternative treatments, such as neuromodulation, are under investigation.

Indwelling catheterization is rarely indicated for detrusor overactivity. If all measures fail, an external collection device or protective pad or undergarment may be required.

Stress Incontinence This disorder, the second most common cause of established incontinence in older women (it is rare in men), is characterized by symptoms and evidence of *instantaneous* leakage of urine in response to stress. Leakage is worse or occurs only during the day unless another abnormality (e.g., detrusor overactivity) is also present. On examination, with the bladder full and the perineum relaxed, instantaneous leakage upon coughing strongly suggests stress incontinence, especially if it reproduces symptoms and if urinary retention has been excluded by a postvoiding residual determination; a several-second delay suggests that leakage is instead caused by an involuntary bladder contraction induced by coughing.

TREATMENT

Surgery is the most effective treatment. For women who can comply indefinitely, pelvic muscle exercises are an option for mild to moderate stress incontinence, but they often require specialized training using vaginal cones or biofeedback. If not contraindicated, an a-adrenergic agonist (e.g., phenylpropanolamine) is also helpful in such cases, especially if combined with estrogen. Occasionally, a pessary or even a tampon (for women with vaginal stenosis) provides some relief.

Urethral Obstruction Rarely present in women, urethral obstruction (due to prostatic enlargement, urethral stricture, bladder neck contracture, or prostate cancer) is the second most common cause of established incontinence in older men. It can present as dribbling incontinence after voiding, urge incontinence due to detrusor overactivity (which coexists in two-thirds of cases), or overflow incontinence due to urinary retention. Renal ultrasound is recommended to exclude hydronephrosis in men whose postvoiding residual volume exceeds 100 to 200 mL; in older men for whom surgery is planned, urodynamic confirmation of obstruction is strongly advised.

TREATMENT

Surgical decompression is the most effective treatment for obstruction, especially if there is urinary retention. For a nonoperative candidate, intermittent or indwelling catheterization is used; a condom catheter is contraindicated when urinary retention is present. For a man with prostatic obstruction who is not in retention, treatment with an a-adrenergic antagonist (e.g., terazosin 5 to 10 mg daily) may lessen symptoms in a few weeks. The 5a-reductase inhibitor finasteride may also ameliorate symptoms in a third or more of patients, but its impact is modest and not apparent for many months. Combined treatment with both agents has proved no better than treatment with an alpha blocker alone in most men.

Detrusor Underactivity Whether idiopathic or due to sacral lower motor nerve dysfunction, this is the least common cause of incontinence (<10% of cases). When it causes incontinence, detrusor underactivity is associated with urinary frequency, nocturia, and frequent leakage of small amounts. The elevated postvoiding residual volume (generally >450 mL) distinguishes it from detrusor overactivity and stress incontinence, but only urodynamic testing (rather than cystoscopy or intravenous urography) differentiates it from urethral obstruction in men; such testing is not usually required in women, in whom obstruction is rare.

TREATMENT

For the patient with a poorly contractile bladder, augmented voiding techniques (e.g., double voiding or applying suprapubic pressure) are often effective; pharmacologic agents (e.g., bethanechol) are rarely effective. If further emptying is needed or for the patient with an acontractile bladder, intermittent or indwelling catheterization is the only option. Antibiotics should be used for symptomatic upper tract infection, or as prophylaxis for recurrent symptomatic infections only in a patient using intermittent catheterization; they should not be used as prophylaxis with an indwelling catheter.

FALLS

Falls are a major problem for elderly people, especially women. Some 30% of community-dwelling elderly individuals fall each year, and the proportion increases with age. Nonetheless, falling must *not* be viewed as accidental, inevitable, or untreatable.

Causes of Falls Balance and ambulation require a complex interplay of cognitive, neuromuscular, and cardiovascular function and the ability to adapt rapidly to an environmental challenge. With age, balance becomes impaired and sway increases. The resulting vulnerability predisposes the older person to fall when challenged by an additional insult to *any* of these systems. Thus, a seemingly minor fall may be due to a serious problem, such as pneumonia or a myocardial infarction.

Much more commonly, however, falls are due to the complex interaction between a variably impaired patient and an environmental challenge. While a warped floorboard may pose little problem for a vigorous, unmedicated, alert person, it may be sufficient to precipitate a fall and hip fracture in the patient with impaired vision, strength, balance, or cognition. Thus, falls in older people are rarely due to a single cause, and effective prevention entails a comprehensive assessment of the patient's intrinsic deficits (usually diseases and medications), the routine activities, and the environmental obstacles.

Intrinsic deficits are those that impair sensory input, judgment, blood pressure regulation, reaction time, and balance and gait (Table 9-5). Medications and alcohol use are among the most common, significant, and reversible causes of falling. Other treatable contributors include postprandial hypotension (which peaks 30 to 60 min after a meal), insomnia, urinary urgency, foot problems, and peripheral edema [which can burden impaired leg strength and gait with an additional 2 to 5 kg (5 to 10 lb)].

Environmental obstacles are listed inTable 9-6. Since most falls occur in or around the home, a visit by a visiting nurse, physical therapist, or physician often reaps substantial dividends.

Complications of Falls and Treatment One out of four people who fall suffers serious injury. About 5% of falls result in fractures, and an equal proportion cause serious soft tissue damage. Falls are the sixth leading cause of death for older people and a contributing factor in 40% of admissions to nursing homes. Resultant hip problems and fear of falls are major causes of loss of independence.

Subdural hematoma is a treatable but easily overlooked complication of falls that must be considered in any elderly patient presenting with new neurologic signs, including confusion alone, even in the absence of a headache. Dehydration, electrolyte imbalance, pressure sores, rhabdomyolysis, and hypothermia may also occur and endanger the patient's life following a fall.

The risk of falling is related to the number of contributory conditions. Because the relationship is multiplicative rather than additive, however, even minor improvement in a number of these factors will reduce the risk substantially. In addition, gait training by a physical therapist often alleviates fear of falling. Ensuring the availability of phones at floor level, a portable phone, or a lightweight radio call system is also important, as is detection and treatment of osteoporosis.

IMMOBILITY

The main causes of immobility are weakness, stiffness, pain, imbalance, and psychological problems. Weakness may result from disuse of muscles, malnutrition, electrolyte disturbances, anemia, neurologic disorders, or myopathies. The most common cause of stiffness in the elderly is osteoarthritis; however, Parkinson's disease, rheumatoid arthritis, gout, pseudogout, and antipsychotic drugs such as haloperidol may also contribute. Pain, whether from bone (e.g., osteoporosis, osteomalacia, Paget's disease, metastatic bone cancer, trauma), joints (e.g., osteoarthritis, rheumatoid arthritis, gout), bursa, muscle (e.g., polymyalgia rheumatica, intermittent claudication, or "pseudoclaudication"), or foot problems may immobilize the patient.

Imbalance and fear of falling are major causes of immobilization. Imbalance may result from general debility, neurologic causes (e.g., stroke; loss of postural reflexes; peripheral neuropathy due to diabetes mellitus, alcohol, or malnutrition; and vestibulocerebellar abnormalities), orthostatic or postprandial hypotension, or drugs (e.g., diuretics, antihypertensives, neuroleptics, and antidepressants) or may occur following prolonged bed rest. Psychological conditions such as severe anxiety or

depression may also contribute to immobilization.

Consequences In addition to thrombophlebitis and pulmonary embolus, there are multiple hazards of bed rest in the elderly. Deconditioning of the cardiovascular system occurs within days and involves fluid shifts, fluid loss, decreased cardiac output, decreased peak oxygen uptake, and increased resting heart rate. Striking changes also occur in skeletal muscle. At the cellular level, intracellular ATP and glycogen concentrations decrease, rates of protein degradation increase, and contractile velocity and strength decline, while at the whole-muscle level, atrophy, weakness, and shortening are seen. Pressure sores are another serious complication; mechanical pressure, moisture, friction, and shearing forces all predispose to their development. As a result, within days of being confined to bed, the risk of postural hypotension, falls, and skin breakdown rises. Moreover, these changes usually take weeks to months to reverse.

TREATMENT

The most important step is preventive -- to avoid bedrest whenever possible. When it cannot be avoided, several measures can be employed to minimize its consequences. Patients should be positioned as close to the upright position as possible several times daily. Range-of-motion exercises should begin immediately, and the skin over pressure points should be inspected frequently. Isometric and isotonic exercises should be performed while the patient is in bed, and whenever possible patients should assist their own positioning, transferring, and self-care. As mobility becomes feasible, graduated ambulation should begin. For individuals confined to a wheelchair, ring-shaped devices ("donuts") should not be used to prevent pressure ulcers since they cause venous congestion and edema and actually increase the risk.

If a pressure ulcer develops, therapy depends on its stage. Stage 1 ulcers are characterized by nonblanchable erythema of intact skin; stage 2 lesions involve an ulcer of the epidermis, dermis, or both; stage 3 ulcers extend to the subcutaneous tissue; and stage 4 lesions involve muscle, bone, and/or the supporting tissues. For stage 1 lesions, eliminating excess pressure and ensuring adequate nutrition and hygiene are sufficient. For the remaining types, the caregiver must also ensure that the wound stays clean and moist; thus, if saline dressings are used they should be changed when they are damp rather than dry. Synthetic dressings are more expensive than saline but are more effective because they require fewer changes (with less disruption of reepithelialization) and protect against contamination. Because bacterial colonization of pressure ulcers is universal, swab cultures should not be performed and topical treatment should be considered only for patients whose ulcers have not healed after 2 weeks of therapy. By contrast, associated cellulitis, osteomyelitis, or sepsis requires systemic therapy after cultures of blood and the wound border (by needle aspiration or biopsy) have been obtained. Surgical or enzymatic debridement is required for stage 3 and 4 lesions. In addition to a daily multivitamin, prescribing vitamin C (500 mg twice daily) is also useful. For debilitated patients, special mattresses are beneficial, including those that reduce pressure (e.g., static air mattress or foam) and those that relieve it (e.g., dynamic units that sequentially inflate and deflate).

In addition to treating all identified factors that contribute to immobility, consultation with

a physical therapist should be sought. Installing handrails, lowering the bed, and providing chairs of proper height with arms and rubber skid guards may allow the patient to be safely mobile in the home. A properly fitted cane or walker may be helpful.

IATROGENIC DRUG REACTIONS

For several reasons, older patients are two or three times more likely to have adverse drug reactions (Chap. 71). Drug clearance is often markedly reduced. This is due to a decrease in renal plasma flow and glomerular filtration rate and a reduced hepatic clearance. The last is due to a decrease in activity of the drug-metabolizing microsomal enzymes and an overall decline in blood flow to the liver with aging. The volume of distribution of drugs is also affected, since the elderly have a decrease in total-body water and a relative increase in body fat. Thus, water-soluble drugs become more concentrated, and fat-soluble drugs have longer half-lives. In addition, serum albumin levels decline, particularly in sick patients, so that there is a decrease in protein binding of some drugs (e.g., warfarin, phenytoin), leaving more free (active) drug available.

In addition to impaired drug clearance, which alters pharmacokinetics, older patients have altered responses to similar serum drug levels, a phenomenon known as *altered pharmacodynamics*. They are more sensitive to some drugs (e.g., opiates, anticoagulants) and less sensitive to others (e.g.,b-adrenergic agents). Finally, the older patient with multiple chronic conditions is likely to be taking several drugs, including nonprescribed agents. Thus, adverse drug reactions and dosage errors are more likely to occur, especially if the patient has visual, hearing, or memory deficits.

Precautions to Avoid Drug Toxicity

Drug Selection and Administration Before initiating treatment, the physician should first ensure that the symptom requiring treatment is not itself due to another drug. For example, antipsychotic agents can cause symptoms that mimic depression (flat affect, restlessness, and pacing); such symptoms should prompt lowering of the dose rather than initiation of an antidepressant. In addition, drug therapy should be employed only after nonpharmacologic means have been considered or tried and only when the benefit clearly outweighs the risk.

Once pharmacotherapy has been decided upon, it should begin at less than the usual adult dosage and the dose should be increased slowly. However, given the marked variability in pharmacokinetics and pharmacodynamics in the elderly, dose escalation should continue until either a successful endpoint is reached or an intolerable side effect is encountered. The final dosage schedule should be kept as simple as possible, and the number of pills should be kept as low as possible. Serum drug levels are often useful in older patients, especially for monitoring drugs with narrow therapeutic indices such as phenytoin, theophylline, quinidine, aminoglycosides, lithium, and psychotropic agents such as nortriptyline. However, toxicity can occur even with "normal" therapeutic levels of some drugs (e.g., digoxin, phenytoin).

Over-the-Counter Agents Nearly three-quarters of the elderly regularly use nonprescribed drugs, many of which cause significant symptoms and/or interact with other medications. Frequent offenders include nonprescribed agents for insomnia (all of

which are anticholinergics), and nonsteroidal anti-inflammatory drugs (NSAIDs), which can hamper control of hypertension in addition to causing renal dysfunction and gastrointestinal bleeding. Gingko biloba, increasingly used as a "memory booster," may interfere with previously stable anticoagulation regimens. Because older patients often consider such agents "nostrums" rather than drugs, the physician must ask about them directly.

Sedative-Hypnotics If nonpharmacologic treatment of insomnia is unsuccessful, low-dose and short-term or intermittent use of an intermediate-acting agent whose metabolism is not affected by age (e.g., oxazepam, 10 to 30 mg/d) may be useful. Because of the increased risk of confusion and other adverse effects, benzodiazepines with either short (e.g., triazolam) or long duration of action (e.g., flurazepam and diazepam) should be avoided. Barbiturates should be avoided for the same reasons. An antidepressant should not be prescribed for insomnia unless the patient is depressed.

Antibiotics Serum creatinine is not a good index of renal function in old people; however, when it is elevated, special care must be taken with the administration of drugs normally excreted by the kidneys. Concentrations of relevant antibiotics should be measured directly.

Cardiac Drugs In older patients, digitalis, procainamide, and quinidine have prolonged half-lives and narrow therapeutic windows; toxicity is common at the usual dosages. For example, digoxin toxicity -- especially anorexia, confusion, or depression -- can occur even with therapeutic digoxin levels.

H_2Receptor Antagonists Most of these agents interfere with hepatic metabolism of other drugs, and all can produce confusion in the elderly. Because they are renally excreted, lower doses should be used to minimize the risk of toxicity in older individuals.

Antipsychotics and Tricyclic Antidepressants These drugs can produce anticholinergic side effects in old people (e.g., confusion, urinary retention, constipation, dry mouth). These can be minimized by switching to a nonanticholinergic agent (e.g., sertraline or nefazodone) or one with less anticholinergic effect (e.g., olanzapine, desipramine). In general, the least potent agents for psychosis (e.g., chlorpromazine) have the most sedating and anticholinergic effects and are the most likely to induce postural hypotension. By contrast, the most potent antipsychotic agents (e.g., haloperidol) have the least sedating, anticholinergic, and hypotensive side effects but cause extrapyramidal side effects, including dystonia, akathisia, rigidity, and tardive dyskinesia. The newer potent antipsychotics (e.g., risperidone, olanzapine, quetiapine, and clozapine) are relative exceptions to this rule. More specific for serotonin than dopamine D_2receptors, these medications may be safer for older demented patients, especially those with hallucinations associated with Lewy body dementia or in those receiving therapy for Parkinson's disease. Unfortunately, even these newer drugs lose their specificity at the higher doses that are commonly required in clinical practice. Thus all of these agents are potentially toxic. Moreover, since both depression and agitation often remit spontaneously, cautious discontinuation of these drugs should be considered periodically.

Glaucoma Medications Both topical beta blockers and carbonic anhydrase inhibitors can

cause systemic side effects. The latter can cause malaise and anorexia independent of the induced metabolic acidosis.

Anticoagulants Elderly patients benefit from anticoagulation as much as do younger individuals but are more vulnerable to serious bleeding and drug interactions. Hence, more careful monitoring and less aggressive anticoagulation are advisable.

Analgesics Both propoxyphene and meperidine are associated with a disproportionate risk of delirium, and propoxyphene also increases the risk of hip fracture. Of theNSAIDs, indomethacin is most likely to induce confusion, fluid retention, and gastrointestinal bleeding. Each of these agents should be avoided in the elderly.

Avoidance of Overtreatment Drugs are frequently not indicated in some common clinical situations. For instance, antibiotics need not be given for asymptomatic bacteriuria unless obstructive uropathy, other anatomic abnormalities, or stones are also present. Ankle edema is often due to venous insufficiency, drugs such asNSAIDsor some calcium antagonists, or even inactivity or malnutrition in chairbound patients. Diuretics are usually not indicated unless edema is associated with heart failure. Fitted, pressure gradient stockings are often helpful. Regular exercise is much more useful for claudication than is pentoxifylline. Finally, since older patients generally tolerate aspirin and other NSAIDs less well than do younger patients, localized pain should be treated when possible with local measures such as injection, physical therapy, heat, ultrasound, or transcutaneous electrical stimulation (Chap. 12).

PREVENTION

Much can be done to prevent the progression and even the onset of disease in older people. Dietary inadequacies should be corrected. Daily calcium intake should approximate 1500 mg, and most elderly people should take 400 to 800 IU of vitamin D daily (contained in one to two multivitamin tablets). Tobacco and alcohol use should be minimized, since the benefits of discontinuing these accrue even to individuals over age 65. The importance of reviewing all of a patient's medications and discontinuing them whenever feasible cannot be overemphasized.

Hypertension, whether isolated systolic hypertension or combined systolic and diastolic hypertension, should be treated. Treatment reduces the risk of stroke and the risk of death due to cardiovascular causes substantially in this age group and may also reduce the risk of cognitive impairment. These benefits have been achieved using *low doses* of a thiazide-like diuretic (e.g., chlorthalidone, 12.5 to 25 mg/d) as the first step (alone effective in almost half of patients) and adding low-dose reserpine (0.05 to 0.1 mg/d) or atenolol (25 to 50 mg/d) only as needed. Benefits are dramatic, side effects are minimal, cost is trivial, and concerns about potential toxicity have not been borne out.

Because of the prevalence, functional impact, and ease of treatment, glaucoma should be screened for, and visual and auditory impairment should be corrected. Dentures should be assessed for their fit, and oral lesions beneath them should be detected.

Because thyroid dysfunction is more prevalent in the elderly, difficult to detect clinically, and treatable, serum levels of thyroid-stimulating hormone should be measured at least

once in asymptomatic older people and probably every 3 to 5 years thereafter. Serum cholesterol is worth measuring in patients with established coronary heart disease, but in those without apparent disease, screening for hypercholesterolemia is controversial. It seems reasonable to screen those who would be willing to comply with therapy, whose quality of life is good (from the patient's viewpoint), whose life expectancy exceeds several years (long enough to potentially benefit from therapy), and whose other risk factors -- for which benefit of treatment has been definitely established -- have already been addressed. A Papanicolaou test should be done in women who have not had one before, since the incidence of both preventable cervical carcinoma and associated death increases with age, especially in this group; it should be repeated triennially in all older women unless two previous tests have been normal. Screening for colon cancer is warranted until a minimum age of 80 to 85, at least in the community-dwelling elderly, although the optimal method is unclear. Immunizations for influenza, pneumococcal pneumonia, and tetanus should be current. Purified protein derivative (PPD) testing should be done on residents of chronic care facilities and on others at high risk of tuberculosis; those who have recently converted probably should be treated. Since responsiveness wanes with age, the test, if negative, should be repeated in a week to increase the chances of detecting all exposed patients. Because older women with breast cancer are more likely to die *of* it than *with* it, screening mammography is indicated every 1 to 2 years at least until age 75 and thereafter if a positive finding would result in therapeutic intervention. The relative risks and benefits of low-dose aspirin and (for women) estrogen replacement therapy have not yet been elucidated sufficiently in the elderly to warrant routine use, but they should be considered on an individual basis.

Exercise should be encouraged not only because of its beneficial effects on blood pressure, cardiovascular conditioning, glucose homeostasis, bone density, insomnia, functional status, and even longevity, but also because it may improve mood and social interaction, reduce constipation, and prevent falls. Resistance training should be encouraged as much as a walking program. Spinal flexion exercises should be avoided in patients with osteopenia; consultation with a physical therapist may be helpful.

Measures should be taken to prevent falling, as outlined in Tables 9-5 and9-6. Now that alendronate has proved effective in preventing vertebral and hip fractures in older women, bone density should be measured in women who are willing to take the drug and who do not already take estrogen. Counseling about driving is important, especially for patients with cognitive impairment.

Perhaps the most valuable preventive measure in old people is to take a careful history, focusing not only on the "chief complaint" but also on common and often hidden conditions such as falls, confusion, depression, alcohol abuse, sexual dysfunction, and incontinence. In addition, one should always identify the complications for which the specific patient is at risk and take steps to avert them. For instance, a patient with cognitive impairment who smokes is at risk not only for lung cancer but also for starting a fire, and a patient who requires narcotics is at risk for fecal impaction, delirium, urinary retention, and confusion. Community-dwelling patients who are at highest risk of rapid deterioration and institutionalization and who should be monitored more closely include those over age 80, those who live alone, those who are bereaved or depressed, and those who are intellectually impaired.

(Bibliography omitted in Palm version)

Back to Table of Contents

10. PRINCIPLES OF DISEASE PREVENTION - *Maureen T. Connelly*, *Thomas S. Inui*

PERSPECTIVES ON PREVENTION

The primary goals of prevention in medicine are to prolong life, to decrease morbidity, and to improve quality of life -- all with the available resources. Working in partnership with patients, physicians play critical roles as educators, managers of access to screening and intervention services, and interpreters of divergent recommendations for promoting health. Despite evidence of the effectiveness of many preventive services in prolonging healthy life and decreasing medical costs, physicians frequently do not integrate appropriate preventive practices into their care. Obstacles to providing optimal preventive care include lack of appropriate training, doubt about the effectiveness of preventive interventions, skepticism about patients' commitment to change, limited reimbursement and time, and conflicting professional recommendations. Success achieved for populations may not be visible to individuals, and physicians may not appreciate the cumulative benefit of their efforts. Despite considerable success in some areas, such as the reduction of smoking by U.S. adults from 40 to 25% in the last 35 years, effective behavior change in other domains is often elusive, challenging and frustrating physicians and patients alike.

DEFINITIONS

This chapter will be devoted to a discussion of *primary* and *secondary* prevention. *Primary prevention*, including various forms of health promotion and vaccination, is care intended to minimize risk factors and the subsequent incidence of disease. *Secondary prevention* is screening for detection of early disease, for example the use of mammography to detect preclinical breast cancer. While the term secondary prevention is also sometimes used for the prevention of recurrent episodes of an existing illness, most would consider this activity to be *tertiary prevention*, care intended to ameliorate the course of established disease.

Deciding what types of primary and secondary preventive care clinicians should offer to their patients is not a trivial matter. The United States Preventive Services Task Force (USPSTF), The Canadian Task Force on the Periodic Health Examination, and the American College of Physicians, among other organizations, have critically reviewed the strength of available evidence for preventive practices and have made recommendations. Adopting an evidence-based approach to the development of preventive practices policy is an essential step to assuaging provider concerns about the validity of particular recommendations, to identifying the specific basis of controversies in prevention, and to reassuring patients that certain interventions will do more good than harm.

PRIMARY PREVENTION

RISK MODIFICATION

Of the more than 2 million deaths that occur in the United States each year, as many as half may be due to preventable causes (Table 10-1). Life-style and behavior play a

central role in the primary causes of morbidity and mortality for adults -- coronary heart disease, cancer, and injuries.

Tobacco The largest potentially modifiable risk to health is the abuse of tobacco products. Responsible for more than 400,000 deaths each year and an estimated annual cost to society as high as $50 billion, tobacco abuse accounts for a substantial fraction of cardiovascular, cancer, and pulmonary morbidity and mortality. Recent evidence also suggests that passive exposure to tobacco smoke results in chronic pulmonary disease, cardiovascular disease, and lung cancer for some adults. Because of the addictive properties of nicotine, preventing the initiation of tobacco abuse is the tobacco control intervention of choice. Most adult smokers acquire their habit as teenagers, and primary efforts to discourage initial tobacco use must engage younger audiences. However, smoking cessation extends life, even in individuals who quit after age 65 or those with established disease.

Counseling regarding the health risks of tobacco and methods for quitting is advised by all prevention advisory panels. Because 70% of smokers come into contact with health professionals each year, the medical encounter provides an opportunity to address the health implications of tobacco abuse. Although 70% of smokers say that they want to stop smoking, many are not ready to make an immediate change. The role of the provider is to motivate smokers to attempt cessation, reduce barriers to smoking cessation, and advise effective methods for cessation, including an expanding array of pharmacotherapeutic agents. Ninety percent of successful quitters will stop smoking without the aid of programmatic interventions. Setting a date to quit, arranging follow-up visits or phone calls during the initial quitting period, providing literature, and considering the use of nicotine replacement systems and other effective medications, such as bupropion, are all interventions that may improve the quitting success rate. Compared to placebo, nicotine replacement systems and bupropion have approximately twice the success rate at 6 months.

Diet Mounting evidence suggests that modification of caloric intake, particularly the quality of calories, can result in decreased morbidity and mortality from cardiovascular disease, cancer, and diabetes. Excess weight is an independent risk factor for coronary disease, in addition to its contribution to the incidence of diabetes, hyperlipidemia, and hypertension. Between 20 and 30% of Americans are overweight, defined as 20% above the acceptable body-mass index (kg/m_2), and more than 40% of certain subpopulations, such as black, Native American, and Mexican-American women, are overweight. Despite concern about the risk of weight cycling, the health hazards of obesity appear to outweigh the potential harm of repeated weight loss and gain.

Americans derive excess calories from fats, particularly saturated fats, rather than from more beneficial sources such as complex carbohydrates, monounsaturated fats, and fiber. Since intake of saturated fat correlates with cholesterol level, and coronary heart disease is reduced by 2 to 3% for every 1% reduction in plasma cholesterol level, dietary modification will play a central role in decreasing the primary cause of mortality in America. Excess dietary fat intake has also been associated with breast, colon, prostate, and lung cancer in epidemiologic studies. The once widely accepted goals of reducing calories from all fats to 30% and from saturated fat to 10% have been challenged as the impact of types of fat (not simply fat itself) on morbidity and mortality

is further elucidated. Increasing the intake of dietary fiber, such as from plant, legume, and grain sources, may contribute specifically to a decrease in colon cancer incidence.

Dietary sodium restriction may benefit those who have salt-sensitive hypertension, although the need for such restriction in the general population is unclear. Calcium and vitamin D are protective against osteoporosis, particularly in young women prior to reaching menopause, and evidence suggests that females at all ages have an inadequate intake. Menstruating women are at risk for iron-deficiency anemia. To achieve the recommended daily intake of vitamins and minerals, a varied diet including fish, lean meats, dairy products, whole grains, and five to six servings of fruits and vegetables daily is recommended, rather than the use of vitamin supplements. However, certain nutrients, such as adequate folate to prevent neural tube defects in developing fetuses, are not readily obtained from the typical American diet and may be best found in supplements. While evidence supporting the use of antioxidants such as vitamins E and C is still incomplete, the recommended quantities of these micronutrients can be obtained from a balanced diet.

Alcohol and Drugs The use of alcohol and drugs accounts for more than 100,000 deaths annually. While the ability of health care providers to prevent the initiation of such behaviors has not been proven, screening for exposure and addiction could potentially direct medical effort to the prevention of alcohol and drug-associated problems such as injury, violence, and medical complications of drug abuse. Although instruments such as the CAGE questionnaire have proven to be valuable for detection of alcohol abuse, no comparable brief screening strategy is available for the routine identification of illicit drug abuse. Health care providers screen inadequately for both disorders, despite evidence for effective early treatment of addictions and their complications. Reviewing recent data that moderate alcohol consumption may lower the risk of heart disease may open a discussion with patients about appropriate use. Legal implications of identifying illicit drug use may hinder detection of this problem. When screening for these disorders is feasible, interventions that have proven effective include brief counseling, referral to ambulatory and in-patient treatment programs, use of 12-step and other community organizations, and appropriate use of medications such as methadone for heroin abuse.

Physical Activity Not only can increased physical activity decrease obesity, but avoiding a sedentary life-style can also decrease the incidence of cardiac disease, hypertension, diabetes, and osteoporotic fracture. It is estimated that only 22% of U.S. adults engage in at least light to moderate physical activity, such as walking for 30 min three to five times per week. A full quarter of the population pursues no vigorous physical activity at all. The magnitude of benefit derived from physical activity may be as great as a 35% reduction in coronary heart disease, and even light exercise is preferable to no exercise. At present, the intensity, frequency, duration, and type of physical activity required to achieve optimal cardiac benefit remain unclear. While earlier studies suggested that vigorous exercise was needed to achieve maximal risk reduction, recent studies suggest that regular moderate-intensity activity, such as walking for exercise on most days, is associated with a reduced risk of cardiac events. A sudden onset of vigorous activity in the unfit may increase the risk for myocardial infarction and sudden death. Patients should be informed that, despite previous physical inactivity, the incremental adoption of a regular fitness program can decrease their risk

of cardiovascular and other diseases to the level of those who have remained fit throughout their lives. Successful exercise programs are integrated into daily routines, self-directed, and injury-free.

Sexual Behavior Because of the substantial risks of infectious diseases and unwanted pregnancy from unprotected sexual activity, patients should be strongly advised to use barrier methods for all high-risk practices such as oral, anal, and vaginal intercourse as well as additional contraceptive methods when pregnancy would not be welcome.

Environment Physicians should adopt a broad construction of environmental risks to health, considering the physical, social, and occupational environments of their patients. Taking a complete exposure history, focusing on home, work, neighborhood, hobbies, and dietary habits, can help direct interventions and recommendations. While local circumstances will dictate specific risks to which patients should be alerted, such as regional infectious diseases or particular toxic exposures produced by local industry, certain general recommendations should be adopted universally for health promotion.

Since skin cancers, the vast majority of them secondary to sun exposure, constitute the most common form of malignancy, all patients should be counseled to avoid sun overexposure and to use sunscreens. Patients should be encouraged to consider potential toxin exposures, such as those due to air pollution, household smoking, or carbon monoxide and radon gases, and be informed of the medical symptoms and consequences of such exposures. Proper food preparation and storage decrease the incidence of food-borne infectious disease.

Unintended injury constitutes a significant preventable burden of morbidity and mortality and is the leading cause of death for the general population under 40. Automobile accidents are the leading cause of unintentional injuries. The risk of being involved in a disabling traffic accident may be as high as 30% in the course of an individual's lifetime, and 50% of deaths from automobile accidents could be prevented with regular seatbelt use. Physicians should recommend seatbelt use, as well as helmet use for motorcycle and bicycle riders, since evidence supports a higher likelihood of use among patients who receive such advice. Clinicians should also recommend against operating a motor vehicle after drinking, since alcohol (and illicit drugs) is a clear-cut risk cofactor.

Smoke detectors are underused, being found in only 80% of homes. Since most deaths due to fire occur in the residential setting, patients should be encouraged to install at least one on each floor of their home.

Attention to health hazards in the workplace can identify those at risk and prevent long-term consequences of exposure. Evaluation of the work environment should include questions about exposure to metals, dusts, fibers, chemicals, fumes, radiation, loud noises, extreme temperatures, and biologic agents.

Community and family violence, particularly through the misuse of firearms, is the second leading cause of death from unintentional injury. Firearms, especially handguns, are far more likely to injure a family member than an intruder and are associated with increased rates of suicide and harm to children. Patients should be encouraged to remove their weapons from the home and should be informed of the risks associated

with improper security and storage of firearms. At a minimum, trigger locks may prevent accidental injury from firearms. While community and family violence are epidemic in the United States, interventions to curtail violent behavior are not well established. Screening for exposure to relationship violence, developing plans for safe havens, and referrals to appropriate community and government agencies can prevent continued abuse.

IMMUNIZATION

As many as 70,000 deaths due to influenza, pneumococcal infections, and hepatitis B occur in the United States annually. Despite good availability and evidence for the cost-effectiveness of recommended vaccinations for adults, only 40% or fewer members of target populations are immunized. Factors explaining poor adherence to adult immunization guidelines include lack of confidence in vaccine efficacy among providers and patients, underestimation of the severity of the target diseases, incomplete reimbursement, lack of systems to identify and vaccinate high-risk populations, and the absence of an adult requirement for vaccination equivalent to our vaccination policies for school-age children.Table 10-2 lists recommended adult immunizations.

CHEMOPROPHYLAXIS

There is significant supportive evidence for the use of certain medications in primary prevention. Therapy of this nature in the otherwise healthy person, however, is not risk-free. The use of aspirin for the prevention of cardiovascular disease or colorectal cancer, for example, is supported by evidence from cohort and, in the case of cardiovascular disease, randomized controlled trials. The potential for cerebral bleeds and gastrointestinal intolerance, however, must be balanced against a patient's individual risk for the target diseases. Although no randomized trials have measured the impact on mortality, postmenopausal hormone replacement therapy is another therapy given to healthy women for the prevention of future disease (coronary heart disease and osteoporosis), as well as to control menopausal symptoms. These benefits must be weighed against the risks of possible breast and endometrial cacinoma. Patient involvement in the decision-making process, perhaps even informed consent, is recommended to ensure compliance, proper use of medication, and sustained monitoring for side effects.

SECONDARY PREVENTION

SCREENING

Widespread screening for the presence of existing diseases should meet the following criteria:

1. The targeted disease must be sufficiently burdensome to the population that a screening program is warranted. Minor changes in relative risk should have a substantial impact on the absolute risk within the population.

2. The target disease must have a well-understood natural history with a long preclinical latent period.

3. The screening method must have acceptable technical performance parameters, detecting the disease at an earlier stage than would be possible without screening and minimizing false-positive and false-negative results.

4. Efficacious treatment for the target illness must be available.

5. Early detection must improve disease outcome.

6. Cost, feasibility, and acceptability of screening and early treatment should be established.

While physicians under-provide certain screening services that have met these criteria (for example, regular mammograms for women over age 50 years), it is also the case that some prevalent screening practices today are not solidly rooted in evidence. Screening tests such as mammography in women under 50 and measurement of prostate-specific antigen have been adopted for use by many clinicians despite lack of complete current evidence that these services will decrease the risk of morbidity or mortality or improve the quality of life. SeeTable 10-2 recommendations of theUSPSTFfor screening of adults who are at average risk for target conditions. Recommendations for special-risk and vulnerable populations are available in the USPSTF *Guide*.

COMMUNITY HEALTH ADVOCACY

In addition to the direct clinical provision of preventive and health-promoting services, physicians can bring their knowledge, expertise, clinical experience, and influence to bear at the community level to promote health. Whether arguing for the denormalization of tobacco use or providing data about the health risks of local incinerators, physicians are important sources of information and support for improving health beyond the clinical office. Such activities are consistent with the overall objective of caring for patients and may have a substantial impact on decreasing the prevalence of the root causes of disease.

(Bibliography omitted in Palm version)

Back to Table of Contents

11. ALTERNATIVE MEDICINE - *Adriane Fugh-Berman*

Alternatives to conventional medicine will always exist. Defined by its outsider status, alternative therapies of one era may be conventional therapies in another. Radiation treatment and the use of transcutaneous electrical nerve stimulation (TENS), now widely used in medicine, were once unconventional therapies. Even leeches have made a minor comeback; hirudin, a potent anticoagulant secreted by leeches, has been approved by the U.S. Food and Drug Administration (FDA) as an antithrombotic agent. Alternative medicine is any approach to a health problem that is different from those used by conventional medical practitioners; many alternative therapies complement rather than supplant conventional medicine.

Alternative medicine ranges from systems with distinct disease theories, diagnostic methods, and multiple treatment options (including traditional Chinese medicine and Ayurvedic medicine) to single-component panaceas, such as bee pollen (Table 11-1). While conventional western medicine is primarily based upon physiology and pathophysiology, alternative therapies may be based on alternative paradigms (e.g., eastern concepts of energy, called "qi" in Chinese medicine or "prana" in Ayurvedic medicine) or may be based on unproven biochemical hypotheses (e.g., high doses of vitamin C). Although many unconventional therapies are not supported by rigorous prospective clinical trials, "alternative" is not synonymous with "unproven." Data from well-designed and well-executed clinical studies support the use of certain alternative medicines in some settings (see below).

The use of unconventional therapies is widespread. In 1997, a telephone survey found that 42% of English-speaking adults in the United States used some sort of alternative therapy. Usage is especially high among people of color, reflecting the fact that, for many, alternative therapies *are* traditional medicine. Patients who use alternative medicine do so most often at opposite ends of the disease spectrum: either for symptom relief in mild or chronic illnesses or for life-threatening conditions (typically as adjuncts to conventional medicine).

PLACEBO OR NONSPECIFIC EFFECTS

To what extent does the *placebo effect* explain the popularity of alternative medicine? Now more commonly called *nonspecific effects*, this phenomenon encompasses numerous factors including the environment, the relationship between the patient and the practitioner, beliefs and expectations (of both the patient and practitioner), natural history of the disease, and individual variability. The nature of the placebo effect is itself under investigation.

It may be assumed that nonspecific effects enhance positive outcomes of both unconventional and conventional therapies. However, scientific evaluation must be the primary determinant of the physician's attitude toward alternative therapies.

HERBAL MEDICINE

Herbalism (also called *phytomedicine*, *phytotherapy*, or *botanical medicine*) is the medicinal use of plants or plant constituents. The use of plants as medicine predates

even human evolution; great apes have been noted to consume specific medicinal plants when they are ill. Many of the drugs in clinical practice are derived from plants, as are the majority of analgesics: lidocaine and novocaine from the coca plant (*Erythroxylum coca*); opioids from the poppy (*Papaver somniferum*); aspirin from meadowsweet (*Spirea ulmaria*), whence the "spir" part of its name derives. The progestin component of oral contraceptives comes from the Mexican yam (*Diascorea villosa*); digoxin comes from foxglove (*Digitalis lanata*); cromolyn sodium is a khellin derivative from the Ayurvedic herb *Ammi visnaga*; and warfarin is a derivative of dicoumarin, from sweet clover (*Melilotus officinalis*). The ipecac that is kept in the medicine cabinet for poisonings comes from the root of a South American shrub (*Cephaelis ipecacuanha*); the benzoin that attaches bandages to skin is a gum resin from *Styrax benzoin*; and the witch hazel used to soothe hemorrhoids is an extract of *Hamamelis virginiana*. Fungi also have contributed to pharmaceuticals: penicillin was isolated from the fungus *Penicillium notatum*, cephalosporins were derived from a marine fungus (*Cephalosporium acremonium*), and lovastatin was derived from the fungus *Aspergillus terreus*.

Essentially every culture has a tradition of herbal medicine. In western herbalism, herbs are often used singly but may sometimes be combined. Chinese herbal medicine utilizes complex mixtures of many herbs, sometimes combined with animal materials. Herbs may be used to treat or prevent disease. As preventives, "tonics" support the function of specific organs, and "adaptogens" are nonspecific treatments that facilitate a return to homeostasis.

Some clinical trial data support the use of Saint John's wort (SJW) for depression, kava for anxiety, saw palmetto for benign prostatic hyperplasia (BPH), and ginkgo for increasing cerebral blood flow. Evidence also supports the use of garlic for lowering cholesterol, hawthorn to improve cardiac function, and echinacea for treating (but not preventing) upper respiratory infections.

Saint John's Wort (*Hypericum perforatum*) A meta-analysis of the use ofSJW for mild to moderate depression examined 23 randomized trials (20 were double-blind) with a total of 1757 outpatients. In 15 placebo-controlled trials, SJW was found to be significantly more effective than placebo. In eight treatment-controlled trials, clinical improvement in those receiving SJW did not differ significantly from those receiving tricyclic antidepressants. The trials in this meta-analysis were heterogeneous and used varying diagnostic criteria and dosages of SJW.

Most clinical trials have been done with extracts of the flowering tops standardized to 0.3% hypericin (long thought to be the most active compound inSJW). However, hyperforin is most likely the prime active agent, and extracts standardized to 3% hyperforin have been tested in clinical trials and are available. The usual dose of both of these preparations is 300 mg tid.

In vitro, a high concentration ofSJWinhibits the uptake of serotonin, norepinephrine, and dopamine. Among neurotransmitters, its strongest binding affinity is tog-aminobutyric acid (GABA) A and B receptors. Although SJW demonstrates monoamine oxidase (MAO) inhibition in vitro, this effect has not been demonstrated in vivo, nor have there been any reported cases of MAO inhibitor-associated hypertensive crises in humans

using SJW (the herb may, however, potentiate serotonin reuptake inhibitors). Side effects of SJW include gastrointestinal symptoms, fatigue, and photosensitization.

Ginkgo (*Ginkgo biloba*) Ginkgo leaf extracts have been proposed for treating Alzheimer's or multi-infarct dementia. In several randomized controlled trials, a small but statistically significant effect after 3 to 6 months of treatment with 40 to 80 mg tid of standardized *G. biloba* extract (containing 22 to 27% flavonoid glycosides and 5 to 7% terpene lactones) was noted on objective measures of cognitive function in patients with Alzheimer's disease. Ginkgo appears to have vasoregulatory and antioxidant effects as well as inhibiting platelet-activating factor (PAF). Serious intracerebral bleeding associated with ginkgo use has been reported, including two subdural hematomas, one intracerebral hemorrhage, one subarachnoid hemorrhage, and one case of spontaneous hyphema. In most cases, these patients were receiving concurrent anticoagulant drugs.

Kava (*Piper methysticum*) Used in Polynesia as a ceremonial beverage, the roots and rhizomes of kava are used medicinally for anxiety and insomnia in Europe and in North America. Several placebo-controlled trials have shown significant anxiolytic activity of kava products (standardized to 70% kavalactones), usually in a dose of 70 mg tid. *Kavalactones* (also called *kavapyrones*) are muscle relaxants; they include kawain, dihydrokawain, methysticin, and dihydromethysticin (the latter two are potent inhibitors of norepinephrine uptake).

Therapeutic doses may result in mild gastrointestinal complaints or allergic skin reactions (incidence ~1.5%). Chronic use of high-dose kava may result in *kava dermopathy*, a reversible ichthyosiform eruption that is often accompanied by eye irritation.

Ginseng (*Panax ginseng* and other *Panax* species) Ginseng is a popular herb in both western and eastern medicine. The place of ginseng root in the treatment of specific conditions remains to be shown in clinical trials. Ginseng contains ginsenosides, polyacetylenes, and sesquiterpenes. It appears to have some glucocorticoid-like actions and hypoglycemic activity and also affects neurotransmitter activity. Glucocorticoid administration blocks the effect of ginsenosides both in vitro and in vivo; ginsenosides increase adrenal cyclic AMP in intact but not hypophysectomized rats, so effects on adrenal secretion appear to be through the pituitary gland. Several cases of postmenopausal uterine bleeding have been reported from ginseng use, although ginseng does not contain known phytoestrogens.

Saw Palmetto (*Serenoa repens*) Saw palmetto fruits were used as food by Native Americans; their most common use today is to treatBPH. A systemic review of randomized controlled trials of extracts of the fruit *S. repens* (alone or in combination with other herbs) identified 18 randomized trials (16 double-blind) that included 2939 men and lasted 4 to 48 weeks. Compared with placebo, *S. repens* improved urinary symptom scores, nocturia, and peak urine flow. In two studies that compared finasteride with *S. repens*, improvements in urinary symptoms scores were similar. Adverse effects due to *S. repens* were mild and infrequent.

Saw palmetto is usually administered in liposterolic extracts standardized to 70 to 95%

free fatty acids; the usual dose is 160 mg bid. Saw palmetto appears to have multiple mechanisms of action including inhibition of 5a-reductase and inhibition of dihydrotestosterone binding to cytosolic androgen receptors.

Pollen/Bee Pollen Pollen may be collected directly from plants or their pollinators; bee pollen is flower pollen collected from bees. A double-blind placebo-controlled trial of a mixed-pollen extract in 60 patients withBPHfound that subjective improvement was significantly better in the treated group; there was a significant decrease in residual urine and in diameter of the prostate on ultrasound. However, flow rate and volume were unchanged. Allergic reactions, hypereosinophilia, and eosinophilic gastroenteritis have been associated with bee pollen intake.

Echinacea Species Echinacea roots are used to treat or prevent infections. The three species used commercially are *Echinacea purpurea*, *E. angustifolia*, and *E. pallida*. A systemic review of 16 trials (8 prevention and 8 treatment trials on upper respiratory tract infections) with a total of 3396 participants found a wide variation in preparations and methodologic quality of trials. Although many available studies reported positive results of echinacea compared to placebo, reviewers concluded that the evidence is not strong enough to recommend a specific dose, product, or preparation.

Immunomodulatory effects are attributed to five classes of compounds in echinacea preparations: caffeic acid derivatives, alkylamides, polyacetylenes, glycoproteins, and polysaccharides. The alkylamides are regarded as the most active chemical constituent. Echinacea stimulates both humoral and cellular immunity; theoretically, it could worsen symptoms in atopic individuals or those with autoimmune disease.

Adverse Effects of Herbs Herbs have pharmacologic effects and can be associated with adverse effects or interactions. Many medicinal herbs (and pharmaceutical drugs) are therapeutic at one dose and toxic at another. The relative dearth of reports of adverse events and interactions attributed to herbal products reflects a combination of underreporting and the relatively nontoxic nature of most herbal usage.

The most dangerous plants used medicinally are aconite and any herbs containing unsaturated pyrrolizidine alkaloids (saturated pyrrolizidine alkaloids lack toxicity). Several herbs that do not contain pyrrolizidine alkaloids have also shown hepatotoxicity.

Aconite (*Aconitum* spp.), sometimes used in Chinese herb mixtures to treat pain or heart failure, contains aconitine and other C_{19}diterpenoid alkaloids. Proper curing of aconite reduces alkaloids by 90%, but even appropriately cured aconite can result in serious, sometimes fatal, cardiac arrhythmias. The first symptoms of aconite poisoning occur within 90 min of ingestion; the majority of patients present with neurologic symptoms (most commonly oral numbness or burning), progressing to peripheral paresthesia and generalized muscle weakness. Nausea and vomiting are also common.

Cardiovascular effects include bradycardia, hypotension, and arrhythmias (including ventricular or supraventricular tachycardia, bidirectional tachycardia, sinus bradycardia with first-degree heart block, bundle branch block with junctional escape rhythm, or torsade de pointes). Other symptoms may include chest pain, abdominal pain, diarrhea, hyperventilation, respiratory distress, dizziness, sweating, confusion, headache, and

excessive lacrimation. No specific antidote for aconite is known, and treatment is mainly supportive. Atropine may be given if symptoms of cholinergic excess are apparent. Antiarrhythmics are often helpful, but characteristically, electrical cardioversion is markedly unsuccessful in aconite poisoning.

Unsaturated pyrrolizidine alkaloids are hepatotoxic; children may be especially sensitive. Unsaturated pyrrolizidine alkaloids occur in comfrey (*Symphytum*), borage (*Borago officinalis*) leaf (seed oils are safe), coltsfoot (*Tussilago farfara*), and species of *Crotalaria* and *Senecio*. Liver toxicity has also been associated with chaparral (*Larrea divaricata*), germander (*Teucrium chamaedrys*), and a Chinese medicine called *jin bu huan* (which contains 36% levo-tetrahydropalmitine, a chemical present in *Stephania* and *Corydalis* genera).

Drug Interactions One of the most serious herb-drug interactions is increased risk of bleeding when warfarin is combined with anticoagulant herbs: cases of bleeding have been reported with ginkgo (*G. biloba*), garlic (*Allium sativum*), and the Chinese herbs danshen (*Salvia miltiorrhiza*) and dong quai (*Angelica sinensis*). The soluble fibers guar gum and psyllium can slow or reduce the absorption of many drugs, and anthranoid-containing laxatives, including senna (*Cassia senna* and *C. angustifolia*) and cascara sagrada (*Rhamnus purshiana*), can also reduce the absorption of many drugs. An Ayurvedic syrup, shankhapushpi, has been associated with reduced levels of phenytoin. Licorice (*Glycyrrhiza glabra*) can potentiate both oral and topical glucocorticoids.

Several herbs can interact with psychotropic drugs; the herb yohimbe (*Pausinystalia yohimbe*) (also available as the drug yohimbine; both forms are used to treat impotence) increases the risk of hypertension when combined with tricyclic antidepressants. Extrapyramidal effects have occurred in patients ingesting neuroleptics and betel nut (*Areca catechu*); mania has been induced in depressed patients who mix antidepressants and *P. ginseng*; andSJW(*H. perforatum*) combined with a serotonin reuptake inhibitor may produce a mild "serotonin syndrome" (with symptoms of nausea, vomiting, and confusion); the full-blown syndrome may include myoclonus, agitation, fever, abdominal cramping, and hypertension (Chap. 385).

Adulterants and Contaminants Herbal products may be contaminated, mislabeled, or contain misidentified plants. Medicinal plants from India and Sri Lanka can be contaminated with toxigenic fungi, including *Aspergillus* and *Fusarium*. Heavy metals have been detected in some Asian herbal products (metals are sometimes deliberately used in the preparation of Ayurvedic herbal medicines). Without any mention on the label, pharmaceutical drugs may also be incorporated into herbal products, a particular problem in Chinese herbal preparations imported from Hong Kong and Taiwan. Nonsteroidal anti-inflammatory drugs and benzodiazepines have been found in Chinese herbal products, including Miracle Herb, Tung Shueh, and Chuifong Toukuwan (since 1974 this notorious brand has incorporated at least 10 different drugs into the preparation). The absence of standard manufacturing practices creates risks that are difficult to quantify.

ACUPUNCTURE

Acupuncture was recognized in western medical texts a century ago; Sir William Osler's *Principles and Practice of Medicine*, first published in 1892, recommended acupuncture for both sciatica and lumbago; and the 1901 edition of *Gray's Anatomy* noted the use of acupuncture for sciatica.

Stimulation of acupuncture points may be done by needles, finger pressure, electrical stimulation, or heat (usually applied by a smoldering cone or rod of "moxa," made of the herb mugwort, *Artemisia vulgaris*). Acupuncture is effective in the treatment of nausea and vomiting. In 27 of 33 controlled trials, superiority of acupuncture point stimulation over placebo for nausea and vomiting of various etiologies was demonstrated. In substance abusers, the therapy may reduce withdrawal symptoms, but evidence is lacking about whether acupuncture has any long-term effect in preventing recidivism. An analysis of 16 randomized controlled trials of acupuncture for smoking cessation showed no beneficial effect of acupuncture over sham or no treatment. Limited preliminary data suggest a possible beneficial effect for acupuncture in stroke rehabilitation. Although acupuncture is known to stimulate endorphin release, and its use for pain is better accepted than for other conditions, controlled clinical trials of pain treatment have had mixed results.

Risks Inadequately sterilized acupuncture needles have been linked to infections, including HIV infection and an epidemic of hepatitis B. Two cases of fatal *Staphylococcus* sepsis have been reported. More than 100 cases of pneumothorax have been reported. Rare cases both of spinal trauma and cardiac tamponade (caused by penetration of a congenital sternal foramen) have been reported.

HOMEOPATHY

Originated in the early nineteenth century by Samuel Hahnemann, a German physician, homeopathy is based on the "doctrine of similars"; animal, vegetable, or mineral substances that cause symptoms in a well person are used to treat those same symptoms in a sick person. For example, poison ivy (*Rhus toxicodendron*) is used to treat varicella (chickenpox). Because conventional treatments aim to counter rather than reproduce symptoms, practitioners of homeopathy refer to conventional medicine as "allopathy."

Usually, remedies are used in highly dilute concentrations, and homeopaths believe that the most dilute remedies are the most potent. If analyzed chemically, many homeopathic remedies contain no detectable levels of the original substance. It is difficult to conceive of a scientifically testable hypothesis that could explain the putative effects of homeopathic medicine where a preparation containing few or no molecules of an active agent are said to have pharmacologic effects. A meta-analysis of 89 placebo-controlled trials of homeopathy found that the odds ratio was 2.45 (CI, 2.05 to 2.93) in favor of homeopathy over placebo. The quality of these studies was not uniform, and the studies with the best methodologic quality yielded significantly less positive results.

Risks A case of pancreatitis associated with intake of homeopathic medication has been reported. Potentially toxic levels of arsenic and cadmium have been found in "low potency" (less dilute) homeopathic preparations.

SPINAL MANIPULATION

Therapeutic manipulation of the body has ancient roots; Hippocrates, Aesculapius, and Galen all used some form of it. A physician, Andrew Taylor Still, originated osteopathy in 1892. Daniel David Palmer invented chiropractic in 1895. A meta-analysis of nine methodologically acceptable studies of spinal manipulation for low-back pain found a definite improvement at 3 weeks for patients with uncomplicated, acute back pain. For patients with chronic pain or sciatic nerve irritation, chiropractic was not helpful.

A meta-analysis of cervical manipulation for neck pain concluded that manipulation, in combination with other treatment, may produce short-term pain relief.

Risks Complications of spinal manipulation include vertebrobasilar accidents, disc herniations, vertebral fracture, spinal cord compression, and cauda equina syndrome. More than 80% of serious complications from chiropractic occur after cervical manipulation. It is impossible to determine the true rate of complications from chiropractic manipulations, but estimates vary from 1 in 400,000 to between 3 and 6 per 10 million. The incidence of cauda equina syndrome is thought to be less than 1 per 10 million manipulations.

MASSAGE

Several studies support the use of massage for reducing lymphedema; the technique matches the effectiveness of uniform-pressure pneumatic devices.

Numerous studies in hospital settings describe the use of infant massage to decrease hospital stays of premature babies. A meta-analysis of randomized trials found that massage interventions improved daily weight gain by 5 g, while gentle, still touch did not show a benefit. Methodologic concerns about the blinding undermine the conclusions.

Risks Massage has been used in an effort to prevent pressure sores, but it is not clearly effective and may increase tissue trauma when done over bony prominences.

MIND/BODY THERAPIES

Biofeedback Biofeedback uses instruments to translate information on physiologic function into audio or visual signals that patients use as cues to help them learn to affect functions not normally thought to be under voluntary control. Commonly used modalities include electromyographic (EMG) feedback of skeletal muscle contraction, thermal feedback of skin temperature (an indirect measure of peripheral blood flow), electroencephalographic (EEG) feedback, electrodermal response (EDR) (feedback of sweat gland activity on the fingers), and perineometry (feedback of contraction of pelvic floor muscles and anal sphincter).

Biofeedback treatment modalities may be combined. For example, for urinary incontinence, biofeedback may be used to measure pelvic muscle activity through urethral sphincter pressure and electromyography, circumvaginal muscle manometry and electromyography, and anorectal manometry and electromyography. Detrusor

pressure feedback may be measured by cystometry, and feedback on intraabdominal pressure may be used to help patients learn to simultaneously contract pelvic muscles while relaxing abdominal muscles (to avoid putting excess pressure on the bladder). Clinical trials support the use of biofeedback in the treatment of urinary incontinence (stress, urge, or mixed), fecal incontinence, migraine, tension headaches, and in stroke rehabilitation.

Hypnosis Traditional hypnosis utilizes the induction of a deep trance state to enhance suggestibility. Several clinical trials indicate that hypnosis is effective in chemotherapy-associated nausea and may be helpful in the treatment of irritable bowel syndrome and pain syndromes. Numerous uncontrolled trials suggest a beneficial effect of hypnosis on smoking cessation, but controlled trials are less impressive. A meta-analysis of hypnosis in nine randomized controlled trials of hypnotherapy found significant heterogeneity among the results of the individual studies, with conflicting results for the effectiveness of hypnotherapy compared to no treatment or to advice. Hypnotherapy was not more effective than rapid smoking (an aversive therapy in which cigarettes are smoked in quick succession) or psychological treatment.

DIETARY SUPPLEMENTS

The fundamental principles of good nutrition are to eat a balanced variety of foods and to maintain a balance of calories taken in with calories burned through activity. Dietary supplements have become a large and lucrative business that promotes the idea that our food is somehow lacking in specific nutrients and that additional intake of any number of food components will treat or prevent specific diseases, improve athletic and sexual performance, and make us live longer. These claims are largely either untested or unproven (Table 11-2). Unfortunately, the *Dietary Supplement Health and Education Act* (DSHEA) of 1994 prevents theFDA from monitoring the quality or safety of dietary supplements before marketing. Supplements do not even have to be dietary components to be sold as dietary supplements (the availability of over-the-counter hormones, including DHEA, progesterone topical creams, and organ extracts is particularly worrisome).

Even vitamin and mineral supplementation may have unexpected results. For example, high dietary consumption of carrots, sweet potatoes, greens, and other foods rich inb-carotene is associated with decreased risk of cardiovascular disease and cancer. However, two large prospective randomized controlled trials of b-carotene [the Alpha Tocopherol Beta Carotene (ATBC) Cancer Prevention Study and the Beta-Carotene and Retinal Efficacy Trial (CARET)] found that b-carotene increased rates of lung cancer in supplemented groups. The ATBC trial also failed to show a benefit in terms of cardiovascular disease, andb-carotene supplementation has been disappointing in other trials. b-carotene may well be a dietary marker for more beneficial carotenoids (including lycopene, lutein,a-carotene, andb-cryptoxanthin) that typically occur in mixtures in foods.

Similarly, while dietary intake of vitamin E is associated with a reduced risk of coronary heart disease and several observational studies (including the Health Professionals Follow-Up Study and the Nurses Health Study) found a protective effect of supplemental vitamin E, prospective placebo-controlled trials have had mixed results. In the

Cambridge Heart Antioxidant Study (CHAOS), a placebo-controlled trial of vitamin E supplementation in patients with coronary artery disease, vitamin E supplementation reduced nonfatal myocardial infarction but did not appear to decrease cardiovascular deaths or all-cause mortality. In theATBCstudy, vitamin E failed to protect male smokers with a previous myocardial infarction from major coronary events and significantly increased risk of death from hemorrhagic stroke. Although vitamin E improves endothelium-dependent vasodilation significantly and attenuates the development of nitrate tolerance, prospective trials have found no benefit of vitamin E for angina.

Vitamin E is not a single compound; the term applies to eight related compounds in two groups, the tocopherols and the tocotrienols. Most North American dietary sources of vitamin E are composed of two-thirds g-tocopherol and one-thirda-tocopherol. g-Tocopherol neutralizes both oxygen and nitrogen free radicals, whilea-tocopherol is selective for oxygen free radicals. Most vitamin E supplements, however, contain only D- or D,L-a-tocopherol, and large doses ofa-tocopherol displaceg-tocopherol in plasma and tissues.

While it is possible that tomorrow's trendy carotenoid or tocopherol will be more successful than yesterday's, it is more likely that carotenoids and tocopherols are most beneficial when consumed in the naturally occurring mixtures of nutrients found in foods.

Dietary supplements that are safe under one set of conditions may be harmful under other conditions. For example, nearly all physiologic reactions that generate free radicals are oxidation-reduction reactions, capable of either generating free radicals or reducing free radicals. While low doses of vitamin C and other substances have antioxidant effects, high levels may actually have a prooxidant effect.

There is clinical trial evidence supporting the supplemental use of some vitamins, minerals, and amino acids (Table 11-2); but at high doses these supplements should be treated with the same respect given pharmaceuticals.

SUMMARY

The incorporation of science into medicine began only in the middle of the nineteenth century; *evidence-based medicine* is a recent term and still not necessarily standard practice. There is no such thing as objective care of a patient. There are many ways in which we physicians communicate goals and our own beliefs in our therapies; it is likely that a substantial proportion of benefit from even rational interventions is due to nonspecific effects.

Many of our patients take alternative medicines; physicians need to adopt an open-minded, nonjudgmental attitude toward the practice. Inquire about all the medications and supplements a patient is taking. Increasingly, data are available to help guide decision-making about alternative medicines. A thorough knowledge of all therapies a patient is utilizing may help explain unexpected findings and is an important component of a holistic approach to patient care.

Conventional medicine often casts a wary eye on therapies outside its boundaries, but it

is incumbent upon physicians to evaluate evidence regarding alternative therapies with the same rigor with which we evaluate conventional therapies. Scientific evidence from controlled clinical trials supports specific applications of alternative medicine, and some therapies considered alternative today may well be incorporated into conventional medicine in the future.

(Bibliography omitted in Palm version)

Back to Table of Contents

PART TWO -CARDINAL MANIFESTATIONS AND PRESENTATION OF DISEASE

SECTION 1 -PAIN

12. PAIN: PATHOPHYSIOLOGY AND MANAGEMENT - *Howard L. Fields*, *Joseph B. Martin*

The task of medicine is to preserve and restore health and to relieve suffering. Understanding pain is essential to both these goals. Because pain is universally understood as a signal of disease, it is the most common symptom that brings a patient to a physician's attention. The function of the pain sensory system is to detect, localize, and identify tissue-damaging processes. Since different diseases produce characteristic patterns of tissue damage, the quality, time course, and location of a patient's pain complaint and the location of tenderness provide important diagnostic clues and are used to evaluate the response to treatment.

THE PAIN SENSORY SYSTEM

Pain is an unpleasant sensation localized to a part of the body. It is often described in terms of a penetrating or tissue-destructive process (e.g., stabbing, burning, twisting, tearing, squeezing) and/or of a bodily or emotional reaction (e.g., terrifying, nauseating, sickening). Furthermore, any pain of moderate or higher intensity is accompanied by anxiety and the urge to escape or terminate the feeling. These properties illustrate the duality of pain: it is both sensation and emotion. When acute, pain is characteristically associated with behavioral arousal and a stress response consisting of increased blood pressure, heart rate, pupil diameter, and plasma cortisol levels. In addition, local muscle contraction (e.g., limb flexion, abdominal wall rigidity) is often present.

THE PRIMARY AFFERENT NOCICEPTOR

A peripheral nerve consists of the axons of three different types of neurons: primary sensory afferents, motor neurons, and sympathetic postganglionic neurons (Fig. 12-1). The cell bodies of primary afferents are located in the dorsal root ganglia in the vertebral foramina. The primary afferent axon bifurcates to send one process into the spinal cord and the other to innervate tissues. Primary afferents are classified by their diameter, degree of myelination, and conduction velocity. The largest-diameter fibers, A-beta (Ab), respond maximally to light touch and/or moving stimuli; they are present primarily in nerves that innervate the skin. In normal individuals, the activity of these fibers does not produce pain. There are two other classes of primary afferents: the small-diameter myelinated A-delta (Ad) and the unmyelinated (C fiber) axons (Fig. 12-1). These fibers are present in nerves to the skin and to deep somatic and visceral structures. Some tissues, such as the cornea, are innervated only by Ad and C afferents. Most Ad and C afferents respond maximally only to intense (painful) stimuli and produce the subjective experience of pain when they are electrically stimulated; this defines them as *primary afferent nociceptors* (*pain receptors*). The ability to detect painful stimuli is completely abolished when Ad and C axons are blocked.

Individual primary afferent nociceptors can respond to several different types of noxious stimuli. For example, most nociceptors respond to heating, intense mechanical stimuli

such as a pinch, and application of irritating chemicals.

Sensitization When intense, repeated, or prolonged stimuli are applied in the presence of damaged tissue or inflammation, the threshold for activating primary afferent nociceptors is lowered and the frequency of firing is higher for all stimulus intensities. Inflammatory mediators such as bradykinin, some prostaglandins, and leukotrienes contribute to this process, which is called *sensitization*. In sensitized tissues normally innocuous stimuli can produce pain. Sensitization is a clinically important process that contributes to tenderness, soreness, and hyperalgesia. A striking example of sensitization is sunburned skin, in which severe pain can be produced by a gentle slap on the back or a warm shower.

Under normal conditions, viscera are relatively insensitive to noxious mechanical and thermal stimuli. Hollow viscera do generate significant discomfort when distended. Furthermore, when affected by a disease process with an inflammatory component, deep structures such as joints or hollow viscera characteristically become exquisitely sensitive to mechanical stimulation.

A large proportion of Ad and C afferents innervating viscera are completely insensitive in normal noninjured, noninflamed tissue. That is, they cannot be activated by known mechanical or thermal stimuli and are not spontaneously active. However, in the presence of inflammatory mediators, these afferents become sensitive to mechanical stimuli. Such afferents have been termed *silent nociceptors*, and their characteristic properties may explain how under pathologic conditions the relatively insensitive deep structures can become the source of severe and debilitating pain and tenderness.

Nociceptor-Induced Inflammation One important concept to emerge in recent years is that afferent nociceptors also have a neuroeffector function. Most nociceptors contain polypeptide mediators that are released from their peripheral terminals when they are activated (Fig. 12-2). An example is substance P, an 11-amino-acid peptide. Substance P is released from primary afferent nociceptors and has multiple biologic activities. It is a potent vasodilator, degranulates mast cells, is a chemoattractant for leukocytes, and increases the production and release of inflammatory mediators. Interestingly, depletion of substance P from joints reduces the severity of experimental arthritis. Primary afferent nociceptors are not simply passive messengers of threats to tissue injury but also play an active role in tissue protection through these neuroeffector functions.

CENTRAL PATHWAYS FOR PAIN

The Spinal Cord and Referred Pain The axons of primary afferent nociceptors enter the spinal cord via the dorsal root. They terminate in the dorsal horn of the spinal gray matter (Fig. 12-3). The terminals of primary afferent axons contact spinal neurons that transmit the pain signal to brain sites involved in pain perception. The axon of each primary afferent contacts many spinal neurons, and each spinal neuron receives convergent inputs from many primary afferents.

From a clinical standpoint, the convergence of many sensory inputs to a single spinal pain-transmission neuron is of great importance because it underlies the phenomenon of referred pain. All spinal neurons that receive input from the viscera and deep

musculoskeletal structures also receive input from the skin. The convergence patterns are determined by the spinal segment of the dorsal root ganglion that supplies the afferent innervation of a structure. For example, the afferents that supply the central diaphragm are derived from the third and fourth cervical dorsal root ganglia. Primary afferents with cell bodies in these same ganglia supply the skin of the shoulder and lower neck. Thus sensory inputs from both the shoulder skin and the central diaphragm converge on pain-transmission neurons in the third and fourth cervical spinal segments. *Because of this convergence and the fact that the spinal neurons are most often activated by inputs from the skin, activity evoked in spinal neurons by input from deep structures is mislocalized by the patient to a place that is roughly coextensive with the region of skin innervated by the same spinal segment*. Thus inflammation near the central diaphragm is usually reported as discomfort near the shoulder. This spatial displacement of pain sensation from the site of the injury that produces it is known as *referred pain*.

Ascending Pathways for Pain A majority of spinal neurons contacted by primary afferent nociceptors send their axons to the contralateral thalamus. These axons form the contralateral spinothalamic tract which lies in the anterolateral white matter of the spinal cord, the lateral edge of the medulla, and the lateral pons and midbrain. The spinothalamic pathway is crucial for pain sensation in humans. Interruption of this pathway produces permanent deficits in pain and temperature discrimination.

Spinothalamic tract axons connect to thalamic neurons that project to somatosensory cortex (Fig. 12-4). This pathway from spinal cord to thalamus to somatosensory cortex appears to be particularly important for the sensory aspects of pain, i.e., its location, intensity, and quality. Spinothalamic tract axons also connect to thalamic and cortical regions linked to emotional responses, such as the cingulate gyrus and frontal lobe. This pathway is thought to subserve the affective or unpleasant emotional dimension of pain.

PAIN MODULATION

The pain produced by similar injuries is remarkably variable in different situations and in different people. For example, athletes have been known to sustain serious fractures with only minor pain, and Beecher's classic World War II survey revealed that many men were unbothered by battle injuries that would have produced agonizing pain in civilian patients. Furthermore, even the suggestion of relief can have a significant analgesic effect (placebo). On the other hand, many patients find even minor injuries (such as venipuncture) unbearable, and the expectation of pain has been demonstrated to induce pain *without a noxious stimulus*.

The powerful effect of expectation and other psychological variables on the perceived intensity of pain implies the existence of brain circuits that can modulate the activity of the pain-transmission pathways. Although there are probably several circuits that can modulate pain, only one has been studied extensively. This circuit has links in the hypothalamus, midbrain, and medulla, and it selectively controls spinal pain-transmission neurons through a descending pathway (Fig. 12-4).

There is good evidence that this pain-modulating circuit contributes to the pain-relieving

effect of opioid analgesic medications. Each of the component structures of the pathway contains opioid receptors and is sensitive to the direct application of opioid drugs. Furthermore, lesions of the system reduce the analgesic effect of systemically administered opioids such as morphine. Along with the opioid receptor, the component nuclei of this pain-modulating circuit contain endogenous opioid peptides such as the enkephalins andb-endorphin.

The most reliable way to activate this endogenous opioid-mediated modulating system is by prolonged pain and/or fear. There is evidence that pain-relieving endogenous opioids are released following operative procedures and in patients given a placebo for pain relief.

Pain modulation is bidirectional. Pain-modulating circuits not only produce analgesia but are also capable of increasing pain. Both pain-inhibiting and pain-facilitating neurons in the medulla project to and control spinal pain-transmission neurons. Since pain-transmission neurons can be activated by modulatory neurons, it is theoretically possible to generate a pain signal with no peripheral noxious stimulus. Some such mechanism could account for the finding that pain can be induced by suggestion alone and may provide a framework for understanding how psychological factors can contribute to chronic pain.

NEUROPATHIC PAIN

The normal nervous system transmits coded signals that result in pain. Thus lesions of the peripheral or central nervous system may result in a loss or impairment of pain sensation. Paradoxically, damage or dysfunction of the nervous system can produce pain. For example, damage to peripheral nerves, as occurs in diabetic neuropathy, or to primary afferents, as in herpes zoster, can result in pain that is referred to the body region innervated by the damaged nerves. Though rare, pain may also be produced by damage to the central nervous system, particularly the spinothalamic pathway or thalamus. Such neuropathic pains are often severe and are notoriously intractable to standard treatments for pain.

Neuropathic pains typically have an unusual burning, tingling, or electric shock-like quality and may be triggered by very light touch. These features are rare in other types of pain. On examination, a sensory deficit is characteristically present in the area of the patient's pain.

A variety of mechanisms contribute to neuropathic pain. As with sensitized primary afferent nociceptors, damaged primary afferents, including nociceptors, become highly sensitive to mechanical stimulation and begin to generate impulses in the absence of stimulation. There is evidence that this increased sensitivity and spontaneous activity is due to an increased concentration of sodium channels. Damaged primary afferents may also develop sensitivity to norepinephrine. Interestingly, spinal pain-transmission neurons cut off from their normal input may also become spontaneously active. Thus both central and peripheral nervous system changes may contribute to neuropathic pain.

Sympathetically Maintained Pain A certain percentage of patients with peripheral

nerve injury develop a severe burning pain (causalgia) in the region innervated by the nerve. The pain typically begins after a delay of hours to days or even weeks. The pain is accompanied by swelling of the extremity, periarticular osteoporosis, and arthritic changes in the distal joints. A similar syndrome called *reflex sympathetic dystrophy* can be produced without obvious nerve damage by a variety of injuries, including fractures of bone, soft tissue trauma, myocardial infarction, and stroke (Chap. 366). Although the pathophysiology of this condition is poorly understood, the pain can be relieved within minutes by blocking the sympathetic nervous system. This implies that sympathetic activity activates nociceptors even if they are not obviously damaged. These results also suggest that the sympathetic nervous system can, under some circumstances, play an active role in inflammation.

TREATMENT

The ideal treatment for any pain is to remove the cause. Sometimes this is possible, but more often after diagnosis and initiation of appropriate treatments for the cause, there is a lag period before the pain subsides. Furthermore, some conditions are so painful that rapid and effective analgesia is essential (e.g., the postoperative state, burns, trauma, cancer, sickle cell crisis). Analgesic medications are a first line of treatment in these cases, and their use should be familiar to all practitioners.

Aspirin, Acetaminophen, and Nonsteroidal Anti-Inflammatory Agents (NSAIDS)

These drugs are considered together because they are used for similar problems and may have a similar mechanism of action (Table 12-1). All these compounds inhibit cyclooxygenase (COX), and, except for acetaminophen, all have anti-inflammatory actions, especially at higher dosages. They are particularly effective for mild to moderate headache and for pain of musculoskeletal origin.

Since they are effective for these common types of pains and are available without prescription,COXinhibitors are by far the most commonly used analgesics. They are absorbed well from the gastrointestinal tract and, with occasional use, side effects are minimal. With chronic use, gastric irritation is a common side effect of aspirin andNSAIDsand is the problem that most frequently limits the dose that can be given. Gastric irritation is most severe with aspirin, which may cause erosion of the gastric mucosa, and because aspirin irreversibly acetylates platelets and interferes with coagulation of the blood, gastrointestinal bleeding is a risk. The NSAIDs are less problematic in this regard. Although toxic to the liver when taken in a high dose, acetaminophen rarely produces gastric irritation and does not interfere with platelet function.Table 12-1 lists the dosages and durations of action of the commonly used drugs of this class.

The introduction of a parenteral form ofNSAID, ketorolac, extends the usefulness of this class of compounds in the management of acute severe pain. Ketorolac is sufficiently potent and rapid in onset to supplant opioids for many patients with acute severe headache and musculoskeletal pain.

There are two major classes ofCOX: COX 1 is constitutively expressed, and COX 2 is induced in the inflammatory state. COX 2-selective drugs have recently been introduced for the treatment of arthritis and are associated with a significant reduction of gastric

irritation. Whether COX 2-selective drugs have analgesic actions equivalent to otherNSAIDsremains to be demonstrated.

Opioid Analgesics Opioids are the most potent pain-relieving drugs currently available. Furthermore, of all analgesics, they have the broadest range of efficacy, providing the most reliable method for rapidly relieving pain. Although side effects are common, they are usually not serious except for respiratory depression and can be reversed rapidly with the narcotic antagonist naloxone. The physician should not hesitate to use opioid analgesics in patients with acute severe pain. Table 12-1 lists the most commonly used opioid analgesics.

Opioids produce analgesia by actions in the central nervous system. They activate pain-inhibitory neurons and directly inhibit pain-transmission neurons. Most of the commercially available opioid analgesics act at the same opioid receptor (mu receptor), differing mainly in potency, speed of onset, duration of action, and optimal route of administration. Although the dose-related side effects (sedation, respiratory depression, pruritus, constipation) are similar among the different opioids, some side effects are due to accumulation of nonopioid metabolites that are unique to individual drugs. One striking example of this is normeperidine, a metabolite of meperidine. Normeperidine produces hyperexcitability and seizures that are not reversible with naloxone. Normeperidine accumulation is much greater in patients with renal failure.

The most rapid relief with opioids is obtained by intravenous administration; relief with oral administration is significantly slower. Common acute side effects include nausea, vomiting, and sedation. The most serious side effect is respiratory depression. Patients with any form of respiratory compromise must be kept under close observation following opioid administration; an oxygen saturation monitor may be useful. The opioid antagonist, naloxone, should be readily available. These effects are dose-related, and there is great variability among patients in the doses that relieve pain and produce side effects. Because of this, initiation of therapy requires titration to optimal dose and interval. The most important principle is to provide adequate pain relief. This requires asking the patient whether the drug has relieved the pain and, if so, when the relief wears off. *The most common error made by physicians in managing severe pain with opioids is to prescribe an inadequate dose. Since many patients are reluctant to complain, this practice leads to needless suffering*. In the absence of sedation at the expected time of peak effect, a physician should not hesitate to repeat the initial dose to achieve satisfactory pain relief.

An innovative approach to the problem of achieving adequate pain relief is the use of patient-controlled analgesia (PCA). PCA requires a device that delivers a baseline continuous dose of an opioid drug, and preprogrammed additional doses whenever the patient pushes a button. The device can be programmed to limit the total hourly dose so that overdosing is impossible. The patient can then titrate the dose to the optimal level. This approach is used most extensively for the management of postoperative pain, but there is no reason why it should not be used for any hospitalized patient with persistent severe pain. PCA is also used for short-term home care of patients with intractable pain, such as is caused by metastatic cancer.

Many physicians, nurses, and patients have a certain trepidation about using opioids

that is based on an exaggerated fear of patients becoming addicted. In fact, there is a vanishingly small chance of patients becoming addicted to narcotics as a result of their appropriate medical use.

The availability of new routes of administration has extended the usefulness of opioid analgesics. Most important is the availability of spinal administration. Opioids can be infused through a spinal catheter placed either intrathecally or epidurally. By applying opioids directly to the spinal cord, regional analgesia can be obtained using a relatively low total dose. In this way, such side effects as sedation, nausea, and respiratory depression can be minimized. This approach has been used extensively in obstetric procedures and for lower-body postoperative pain. Opioids can also be given intranasally (butorphanol), rectally, and transdermally (fentanyl), thus avoiding the discomfort of frequent injections in patients who cannot be given oral medication.

Opioid and Cyclooxygenase Inhibitor Combinations When used in combination, opioids andCOXinhibitors have additive effects. Because a lower dose of each can be used to achieve the same degree of pain relief and their side effects are nonadditive, such combinations can be used to lower the severity of dose-related side effects. Fixed-ratio combinations of an opioid with acetaminophen carry a special risk. Dose escalation as a result of increased severity of pain or decreased opioid effect as a result of tolerance may lead to levels of acetaminophen that are toxic to the liver.

CHRONIC PAIN

PATIENT EVALUATION

Managing patients with chronic pain is intellectually and emotionally challenging. The patient's problem is often difficult to diagnose: such patients are demanding of the physician's time and often appear emotionally distraught. The traditional medical approach of seeking an obscure organic pathology is usually unhelpful. On the other hand, psychological evaluation and behaviorally based treatment paradigms are frequently helpful, particularly in the setting of a multidisciplinary pain-management center.

There are several factors that can cause, perpetuate, or exacerbate chronic pain. First, of course, the patient may simply have a disease that is characteristically painful for which there is presently no cure. Arthritis, cancer, migraine headaches, fibromyalgia, and diabetic neuropathy are examples of this. Second, there may be secondary perpetuating factors that are initiated by disease and persist after that disease has resolved. Examples include damaged sensory nerves, sympathetic efferent activity, and painful reflex muscle contraction. Finally, a variety of psychological conditions can exacerbate or even cause pain.

There are certain areas to which special attention should be paid in the medical history. Because depression is the most common emotional disturbance in patients with chronic pain, patients should be questioned about their mood, appetite, sleep patterns, and daily activity. A simple standardized questionnaire, such as the Beck Depression Inventory, can be a useful screening device. It is important to remember that major depression is a common, treatable, and potentially fatal illness.

Other clues that a significant emotional disturbance is contributing to a patient's chronic pain complaint include: pain that occurs in multiple unrelated sites; a pattern of recurrent, but separate, pain problems beginning in childhood or adolescence; pain beginning at a time of emotional trauma, such as the loss of a parent or spouse; a history of physical or sexual abuse; and past or present substance abuse.

On examination, special attention should be paid to whether the patient guards the painful area and whether certain movements or postures are avoided because of pain. Discovering a mechanical component to the pain can be useful both diagnostically and therapeutically. Painful areas should be examined for deep tenderness, noting whether this is localized to muscle, ligamentous structures, or joints. Chronic myofascial pain is very common, and in these patients deep palpation may reveal highly localized trigger points that are firm bands or knots in muscle. If injection of local anesthetic into these trigger points relieves the pain, it supports the diagnosis. A neuropathic component to the pain is indicated by evidence of nerve damage, such as sensory impairment, exquisitely sensitive skin, weakness and muscle atrophy, or loss of deep tendon reflexes. Evidence suggesting sympathetic nervous system involvement includes the presence of diffuse swelling, changes in skin color and temperature, and hypersensitive skin and joint tenderness compared with the normal side. Relief of the pain with a sympathetic block is diagnostic.

A guiding principle in evaluating patients with chronic pain is to assess both emotional and organic factors before initiating therapy. Addressing these issues together, rather than waiting to "rule out" organic causes of the pain, improves compliance in part because it assures patients that a psychological evaluation does not mean that the physician is questioning the validity of their complaint. Even when an organic cause for a patient's pain can be found, it is still wise to look for other factors. For example, cancer patients with painful bony metastases may also have pain due to nerve damage and significant depression. Optimal therapy requires that each of these factors be looked for and treated.

TREATMENT

Once the evaluation process has been completed and the likely causative and exacerbating factors identified, an explicit treatment plan should be developed. An important part of this process is to identify specific and realistic functional goals for therapy, such as getting a good night's sleep, being able to go shopping, or returning to work. A multidisciplinary approach that utilizes medications, counseling, physical therapy, nerve blocks, and even surgery may be required to improve the patient's quality of life. This may require referral to a pain clinic; however, this is not necessary for all chronic pain patients. For some, pharmacologic management alone can provide significant help.

Antidepressant Medications The tricyclic antidepressants (Table 12-1) are extremely useful for the management of patients with chronic pain. Although developed for the treatment of depression, the tricyclics have a spectrum of dose-related biologic activities that include the production of analgesia in a variety of clinical conditions. Although the mechanism is unknown, the analgesic effect of tricyclics has a more rapid onset and

occurs at a lower dose than is typically required for the treatment of depression. Furthermore, patients with chronic pain who are not depressed obtain pain relief with antidepressants. There is evidence that tricyclic drugs potentiate opioid analgesia, so they are useful adjuncts for the treatment of severe persistent pain such as occurs with malignant tumors.Table 12-2 lists some of the painful conditions that respond to tricyclics. Tricyclics are of particular value in the management of neuropathic pain such as occurs in diabetic neuropathy and postherpetic neuralgia, for which there are few other therapeutic options.

The tricyclics that have been shown to relieve pain have significant side effects (Table 12-1;Chap 385). Unfortunately, some of the serotonin-selective reuptake inhibitors such as fluoxetine (Prozac) that have fewer and less serious side effects have not been shown to provide pain relief. On the other hand, venlafaxine (Effexor), a nontricyclic antidepressant that blocks both serotonin and norepinephrine reuptake, appears to be useful in patients who cannot tolerate tricyclics.

Anticonvulsants and Antiarrhythmics (Table 12-1) These drugs are useful primarily for patients with neuropathic pain. Phenytoin (Dilantin) and carbamazepine (Tegretol) were first shown to relieve the pain of trigeminal neuralgia. This pain has a characteristic brief, shooting, electric shock-like quality. In fact, anticonvulsants seem to be helpful largely for pains that have such a lancinating quality. A new-generation anticonvulsant, gabapentin (Neurontin), which increases brain g-aminobutyric acid levels, is effective for a broad range of neuropathic pains.

Antiarrhythmic drugs such as low-dose lidocaine and mexiletine (Mexitil) are also effective for neuropathic pains. These drugs block the spontaneous activity of primary afferent nociceptors that appears when they are damaged.

Chronic Opioid Medication The long-term use of opioids is accepted for patients with pain due to malignant disease. Although its use for chronic pain of nonmalignant origin is controversial, it is clear that for many such patients opioid analgesics are the only option available for obtaining effective relief. This is understandable since opioids are the most potent and have the broadest range of efficacy of any analgesic medications. Although addiction is rare in patients who first use opioids for pain relief, some degree of tolerance and physical dependence are likely to occur with long-term use. Therefore, before embarking on opioid therapy, other options should be explored, and the limitations and risks of opioids should be explained to the patient. It is also important to point out that some opioid analgesic medications have mixed agonist-antagonist properties (e.g., pentazocine and butorphanol). From a practical standpoint, this means that they may worsen pain by inducing an abstinence syndrome in patients who are physically dependent on other opioid analgesics.

With long-term outpatient use of orally administered opioids it is desirable to use long-acting compounds such as levorphanol, methadone, or sustained-release morphine (Table 12-1). The pharmacokinetic profile of these drugs enables prolonged pain relief, minimizes side effects such as sedation that are associated with high peak plasma levels, and, perhaps, reduces the likelihood of rebound pain associated with a rapid fall in plasma opioid concentration. Constipation is a virtually universal side effect of opioid use and should be treated expectantly.

It is worth emphasizing, in conclusion, that many patients, especially those with chronic pain, seek medical attention primarily because they are suffering and because only physicians can provide the medications required for their relief. A primary responsibility of all physicians is to minimize the physical and emotional discomfort of their patients. Familiarity with pain mechanisms and analgesic medications is an important step toward accomplishing this aim.

(Bibliography omitted in Palm version)

Back to Table of Contents

13. CHEST DISCOMFORT AND PALPITATIONS - *Thomas H. Lee*

CHEST DISCOMFORT

Chest discomfort is one of the most common challenges for clinicians in the office or emergency department. The differential diagnosis includes conditions affecting organs throughout the thorax and abdomen, with prognostic implications that vary from benign to life-threatening (Table 13-1). Failure to recognize potentially serious conditions such as acute ischemic heart disease, aortic dissection, or pulmonary embolism can lead to serious complications, including death. Conversely, overly conservative management of low-risk patients leads to unnecessary hospital admissions, tests, and procedures.

CAUSES OF CHEST DISCOMFORT

Myocardial Ischemia and Injury (See also Chap. 244) Myocardial ischemia occurs when the oxygen supply to the heart is not sufficient to meet metabolic needs. This mismatch can result from a decrease in oxygen supply, a rise in demand, or both. The most common underlying cause of myocardial ischemia is obstruction of coronary arteries by atherosclerosis; in the presence of such obstruction, transient ischemic episodes are usually precipitated by an increase in oxygen demand as a result of physical exertion. However, ischemia can also result from psychological stress, fever, or large meals or from compromised oxygen delivery due to anemia, hypoxia, or hypotension. Ventricular hypertrophy due to valvular heart disease, hypertrophic cardiomyopathy, or hypertension can predispose the myocardium to ischemia because of impaired penetration of blood flow from epicardial coronary arteries to the endocardium.

Angina Pectoris The chest discomfort of myocardial ischemia is a visceral discomfort that is usually described as a heaviness, pressure, or squeezing (Table 13-2). Other common adjectives for anginal pain are burning and aching. Some patients deny any "pain" but may admit to dyspnea or a vague sense of anxiety. The word "sharp" is sometimes used by patients to describe intensity rather than quality.

The location of angina pectoris is usually retrosternal; most patients do not localize the pain to any small area. The discomfort may radiate to the neck, jaw, teeth, arms, or shoulders, reflecting the common origin in the posterior horn of the spinal cord of sensory neurons supplying the heart and these areas. Some patients present with aching in sites of radiated pain as their only symptoms of ischemia. Occasional patients report epigastric distress with ischemic episodes. Less common is radiation to below the umbilicus or to the back.

Stable angina pectoris usually develops gradually with exertion, emotional excitement, or after heavy meals. Rest or treatment with sublingual nitroglycerin typically leads to relief within several minutes. In contrast, pain that is fleeting (lasting only a few seconds) is rarely ischemic in origin. Similarly, pain that lasts for several hours is unlikely to represent angina, particularly if the patient's electrocardiogram does not show evidence of ischemia.

Anginal episodes can be precipitated by any physiologic or psychological stress that

induces tachycardia. Most myocardial perfusion occurs during diastole, when there is minimal pressure opposing coronary artery flow from within the left ventricle. Since tachycardia decreases the percentage of the time in which the heart is in diastole, it decreases myocardial perfusion.

Unstable angina and myocardial infarction (See also Chaps. 243 and244) Patients with these acute ischemic syndromes usually complain of symptoms similar in quality to angina pectoris, but more prolonged and severe. The onset of these syndromes may occur with the patient at rest, and sublingual nitroglycerin may lead to transient or no relief. Accompanying symptoms may include diaphoresis, dyspnea, nausea, and light-headedness.

The physical examination may be completely normal in patients with chest discomfort due to ischemic heart disease. Careful auscultation during ischemic episodes may reveal a third or fourth heart sound, reflecting myocardial systolic or diastolic dysfunction. A transient murmur of mitral regurgitation suggests ischemic papillary muscle dysfunction. Severe episodes of ischemia can lead to pulmonary congestion and even pulmonary edema.

Other Cardiac Causes Myocardial ischemia caused by hypertrophic cardiomyopathy, aortic stenosis, or other conditions leads to angina pectoris similar to that caused by coronary atherosclerosis. In such cases, a systolic murmur or other findings usually suggest abnormalities other than coronary atherosclerosis that may be contributing to the patient's symptoms.

Pericarditis (See alsoChap. 239) The pain in pericarditis is believed to be due to inflammation of the adjacent parietal pleura, since most of the pericardium is believed to be insensitive to pain. Thus, infectious pericarditis, which usually involves adjoining pleura surfaces, tends to be associated with pain, while conditions that cause only local inflammation (e.g., myocardial infarction or uremia) and cardiac tamponade tend to result in mild or no chest pain.

The adjacent parietal pleura receives its sensory supply from several sources, so the pain of pericarditis can be experienced in areas ranging from the shoulder and neck to the abdomen and back. Most typically, the pain is retrosternal and is aggravated by coughing, deep breaths, or changes in position -- all of which lead to movements of pleural surfaces. The pain is often worse in the supine position and relieved by sitting upright and leaning forward. Less common is a steady aching discomfort that mimics acute myocardial infarction.

Diseases of the Aorta (See alsoChap. 247) *Aortic dissection* is a potentially catastrophic condition that is due to spread within the wall of the aorta of a subintimal hematoma. The hematoma may begin with a tear in the intima of the aorta or with rupture of the vasa vasorum within the aortic media. This syndrome can occur with trauma to the aorta, including motor vehicle accidents or medical procedures in which catheters or intraaortic balloon pumps damage the intima of the aorta. Nontraumatic aortic dissections are rare in the absence of hypertension and/or conditions associated with deterioration of the elastic or muscular components of the media within the aorta's wall. Cystic medial degeneration is a feature of several inherited connective tissue

diseases, including Marfan and Ehlers-Danlos syndromes. About half of all aortic dissections in women under 40 years of age occur during pregnancy.

Almost all patients with acute dissections present with severe chest pain, although some patients with chronic dissections are identified without associated symptoms. Unlike the pain of ischemic heart disease, symptoms of aortic dissection tend to reach peak severity immediately, often causing the patient to collapse from its intensity. The adjectives used to describe the pain reflect the process occurring within the wall of the aorta -- "ripping" and "tearing" -- and the location usually correlates with the site and extent of the dissection. Thus, dissections that begin in the ascending aorta and extend to the descending aorta tend to cause pain in the front of the chest that extends into the back, between the shoulder blades.

Physical findings may also reflect extension of the aortic dissection that compromises flow into arteries branching off the aorta. Thus, loss of a pulse in one or both arms, cerebrovascular accident, or paraplegia can all be catastrophic consequences of aortic dissection. Hematomas that extend proximally and undermine the coronary arteries or aortic valve apparatus may lead to acute myocardial infarction or acute aortic insufficiency. Rupture of the hematoma into the pericardial space leads to pericardial tamponade.

Another abnormality of the aorta that can cause chest pain is a *thoracic aortic aneurysm*. Aortic aneurysms are frequently asymptomatic but can cause chest pain and other symptoms by compressing adjacent structures. This pain tends to be steady, deep, and sometimes severe.

Pulmonary Embolism (See also Chap. 261) Chest pain due to pulmonary embolism is believed to be due to distention of the pulmonary artery or infarction of a segment of the lung adjacent to the pleura. Massive pulmonary emboli may lead to substernal pain that is suggestive of acute myocardial infarction. More commonly, smaller emboli lead to focal pulmonary infarctions that cause pain that is lateral and pleuritic. Associated symptoms include dyspnea and, occasionally, hemoptysis. Tachycardia is usually present.

Pneumonia or Pleuritis Lung diseases that damage and cause inflammation of the pleura of the lung usually cause a sharp, knifelike pain that is aggravated by inspiration or coughing.

Gastrointestinal Conditions Esophageal pain from acid reflux from the stomach, spasm, obstruction, or injury can be difficult to discern from myocardial syndromes. Acid reflux typically causes a deep burning discomfort that may be exacerbated by alcohol, aspirin, or some foods; this discomfort is often relieved by antacid or other acid-reducing therapies. Acid reflux tends to be exacerbated by lying down and may be worse in early morning when the stomach is empty of food that might otherwise absorb gastric acid.

Esophageal spasm may occur in the presence or absence of acid reflux, and leads to a squeezing pain indistinguishable from angina. Prompt relief of esophageal spasm is often provided by antianginal therapies such as sublingual nifedipine, further promoting confusion between these syndromes. Chest pain can also result from injury to the

esophagus, such as a Mallory-Weiss tear caused by severe vomiting.

Chest pain can result from diseases of the gastrointestinal tract below the diaphragm, including *peptic ulcer disease*, *biliary disease*, and *pancreatitis*. These conditions usually cause abdominal pain as well as chest discomfort; symptoms are not likely to be associated with exertion. The pain of ulcer disease typically occurs 60 to 90 min after meals, when postprandial acid production is no longer neutralized by food in the stomach. Cholecystitis usually causes a pain that is described as aching, occurring an hour or more after meals.

Neuromusculoskeletal Conditions *Cervical disk disease* can cause chest pain by compression of nerve roots. Pain in a dermatomal distribution can also be caused by *intercostal muscle cramps* or by *herpes zoster*. Chest pain symptoms due to herpes zoster may occur before skin lesions are apparent.

Costochondral and chondrosternal syndromes are the most common causes of anterior chest musculoskeletal pain. Only occasionally are physical signs of costochondritis such as swelling, redness, and warmth (Tietze's syndrome) present. The pain of such syndromes is usually fleeting and sharp, but some patients experience a dull ache that lasts for hours. Direct pressure on the chondrosternal and costochondral junctions may reproduce the pain from these and other musculoskeletal syndromes. Arthritis of the shoulder and spine and bursitis may also cause chest pain. Some patients who have these conditions and myocardial ischemia blur and confuse symptoms of these syndromes.

Emotional and Psychiatric Conditions As many as 10% of patients who present to emergency departments with acute chest pain have panic disorder or other emotional conditions. The symptoms in these populations are highly variable, but frequently the discomfort is described as visceral tightness or aching that lasts more than 30 min. Some patients offer other atypical descriptions, such as pain that is fleeting, sharp, and/or localized to a small region. The electrocardiogram in patients with emotional conditions may be difficult to interpret if hyperventilation causes ST-T-wave abnormalities. A careful history may elicit clues of depression, prior panic attacks, somatization, agoraphobia, or other phobias.

Approach to the Patient

The evaluation of the patient with chest discomfort must accommodate two goals -- determining the diagnosis and assessing the safety of the immediate management plan. The latter issue is often dominant when the patient has acute chest discomfort, such as patients seen in the emergency department. In such settings, the clinician must focus on questions such as the safety of discharge to home, admission to a non-coronary care unit facility, or immediate exercise testing. Table 13-3 displays a sequence of questions that can be used in the evaluation of the patient with chest discomfort, with the diagnostic entities that are most important for consideration at each stage of the evaluation.

Acute Chest Discomfort In patients with acute chest discomfort, the clinician must first assess the patient's respiratory and hemodynamic status. If either is compromised,

initial management should focus on stabilizing the patient before the diagnostic evaluation is pursued. If, however, the patient does not require emergent interventions, then a focused history, physical examination, and laboratory evaluation should be performed to assess the patient's risk of life-threatening conditions, including acute ischemic heart disease, aortic dissection, and pulmonary embolism.

The *history* should include questions about the quality and location of the chest discomfort (Table 13-2). The patient should also be asked about the nature of onset of the pain and its duration. Myocardial ischemia is usually associated with a gradual intensification of symptoms over a period of minutes. Pain that is fleeting or that lasts hours without being associated with electrocardiographic changes is not likely to be ischemic in origin.

The *physical examination* should include evaluation of blood pressure in both arms and of pulses in both legs. Poor perfusion of a limb may be due to an aortic dissection that has compromised flow to an artery branching from the aorta. Chest auscultation may reveal diminished breath sounds; a pleural rub; or evidence of pneumothorax, pulmonary embolism, pneumonia, or pleurisy. The cardiac examination should seek pericardial rubs, systolic and diastolic murmurs, and third or fourth heart sounds.

An *electrocardiogram* is an essential test for adults with chest discomfort that is not due to an obvious traumatic cause. The presence of electrocardiographic changes consistent with ischemia or infarction (Chap. 226) is associated with high risks of acute myocardial infarction or unstable angina (Table 13-4); such patients should be admitted to a unit with electrocardiographic monitoring and the capacity to respond to a cardiac arrest. The absence of such changes does not exclude acute ischemic heart disease, but the risk of life-threatening complications is low for patients with normal electrocardiograms or only nonspecific ST-T-wave changes. If these patients are not considered appropriate for immediate discharge, they are often candidates for early or immediate exercise testing.

Markers of myocardial injury are often obtained in the emergency department evaluation of acute chest discomfort. The most commonly used markers are creatine kinase (CK), CK-MB, and the cardiac troponins (I and T). Single values of these markers do not have high sensitivity for acute myocardial infarction or for prediction of complications. Hence, decisions to discharge patients home should not be made on the basis of single negative values of these tests.

Provocative tests for coronary artery disease are not appropriate for patients with ongoing chest pain. In such patients, rest myocardial perfusion scans can be considered; a normal scan reduces the likelihood of coronary artery disease. Clinicians frequently employ therapeutic trials with sublingual nitroglycerin or antacids, and a common error is to assume that a response to either of these interventions clarifies the diagnosis. While such information is often helpful, the patient's response may be due to the placebo effect. Hence, myocardial ischemia should never be considered excluded solely because of a response to antacid therapy. Similarly, failure of nitroglycerin to relieve pain does not exclude the diagnosis of coronary disease.

If the patient's history or examination is consistent with aortic dissection, imaging studies

to evaluate the aorta must be pursued promptly because of the high risk of catastrophic complications with this condition. A chest x-ray is not sufficient to exclude this diagnosis. Appropriate tests include a chest computed tomography scan with contrast or a magnetic resonance imaging scan in patients who are hemodynamically stable, or a transesophageal echocardiogram in patients who are less stable. Aortic angiography is no longer a first test at most institutions.

Acute pulmonary embolism should be considered in patients with respiratory symptoms, pleuritic chest pain, hemoptysis, or a history of venous thromboembolism or coagulation abnormalities. Initial tests usually include a lung scan and/or pulmonary arteriography.

If patients with acute chest discomfort show no evidence of life-threatening conditions, the clinician should then focus on serious chronic conditions with the potential to cause major complications, the most common of which is stable angina. Early use of treadmill exercise testing for such patients, whether in the office or the emergency department, is now an accepted management strategy for low-risk patients. Exercise testing is not appropriate, however, for patients who (1) report pain that is believed to be ischemic occurring at rest or (2) have electrocardiographic changes consistent with ischemia not known to be old.

Patients with sustained chest discomfort who do not have evidence for life-threatening conditions should be evaluated for evidence of conditions likely to benefit from acute treatment (Table 13-3). Pericarditis may be suggested by the history, physical examination, and electrocardiogram (Table 13-2). Clinicians should carefully assess blood pressure patterns and consider echocardiography in such patients to detect evidence of impending pericardial tamponade. Chest x-rays can be used to evaluate the possibility of pulmonary disease.

GUIDELINES AND CRITICAL PATHWAYS FOR ACUTE CHEST PAIN

Guidelines for the initial evaluation for patients with acute chest pain have been developed by the American College of Emergency Physicians (ACEP) and other organizations. The ACEP statement describes *rules* and *guidelines* about the data that should be recorded as part of the evaluation, and the actions that should follow from certain findings (Table 13-5). In the ACEP framework, *rules* are actions that are general principles of good practice, while *guidelines* are actions that should be considered but are not always followed. Hence, failure to follow a guideline is not necessarily improper care.

Other organizations, including the Agency for Health Care Policy and Research (AHCPR) and the National Heart Attack Alert Program, have also issued guidelines for management of patients with a high probability of acute ischemic heart disease. In these and other guidelines, patients with possible or probable acute myocardial infarction as suggested by the description of their pain or electrocardiographic findings are expected to be admitted to the hospital. The AHCPR guidelines for unstable angina note that not all patients with that syndrome require admission but recommend that patients with unstable angina be monitored electrocardiographically during their evaluation; that those with ongoing rest pain should be placed at bed rest during the initial phase of stabilization. TheACEPpolicy statement indicates that patients who are discharged

should be given a referral for follow-up care and instructions regarding treatment and circumstances that require a return to the emergency department.

Many medical centers have adopted critical pathways and other forms of guidelines to increase efficiency. These guidelines emphasize two strategies:

· Triage to non-coronary care unit monitored facilities such as intermediate care units or chest pain units of patients with a low risk for complications, such as patients without new ischemic changes on their electrocardiograms and without ongoing chest pain. Such patients can usually be safely observed in non-coronary care unit settings, undergo early exercise testing, or be discharged home. Risk stratification can be assisted through use of prospectively validated multivariate algorithms that have been published for acute ischemic heart disease and its complications.

· Shortening lengths of stay in the coronary care unit and hospital. Recommendations regarding the minimum length of stay in a monitored bed for a patient who has no further symptoms have decreased in recent years to 12 h or less if exercise testing or other risk stratification technologies are available.

NONACUTE CHEST DISCOMFORT

The management of patients who do not require admission to the hospital or who no longer require inpatient observation should seek to identify the cause of the symptoms and the likelihood of major complications. Cost-effectiveness analyses support use of noninvasive testing for coronary disease, such as exercise electrocardiography and stress echocardiography. These tests serve both to diagnose coronary disease and to identify patients with high-risk forms of coronary disease who may benefit from revascularization. Gastrointestinal causes of chest pain can be evaluated via endoscopy or radiology studies. Emotional and psychiatric conditions warrant appropriate evaluation and treatment; randomized trial data indicate that cognitive therapy and group interventions lead to decreases in symptoms for such patients.

PALPITATIONS

Palpitations are characterized by an awareness of the beating of the heart. Patients commonly describe "pounding" or "fluttering" heart beats or report a sensation that the heart is stopping or skipping beats. These symptoms may be caused by a change in the heart's rhythm or rate or by an increase in the force of its contractions. In many cases, this awareness reflects lack of competing sensory stimuli, such as when a person is lying in bed, unable to sleep.

Palpitations are often manifestations of psychiatric conditions, the most common of which are depression and panic disorder. For example, in one study of outpatients referred for ambulatory electrocardiographic monitoring to evaluate palpitations, 19% were found to have a psychiatric disorder. Patients with psychiatric disorders were more likely than other patients to report that their palpitations lasted longer than 15 min or were accompanied by ancillary symptoms. In this study, physicians usually recognized the emotional basis of the patients' symptoms but frequently did not refer the patient for specific therapy.

Palpitations can also be caused by virtually any cardiac arrhythmia as well as by other cardiac and noncardiac conditions. A markedly enlarged left ventricle can cause awareness of the heart beat by contact with the chest wall. Any condition associated with increased catecholamine levels can lead to palpitations both by increasing the forcefulness of cardiac contractions and by increasing the rate of premature beats.

Palpitations can be intermittent or sustained and regular or irregular. Patients with this complaint should be asked to describe their palpitations' onset, duration, associated symptoms and the circumstances in which they occur. Abrupt onset and termination after several minutes may reflect a sustained ventricular or supraventricular tachyarrhythmia. Gradual onset and termination of a pounding heart beat is more consistent with sinus tachycardia. Patients should try to replicate the rhythm of their palpitations by tapping on a table. This maneuver can help the physician determine the nature of any cardiac arrhythmia. Patients should also be taught to take their pulse so that they can more accurately report their approximate heart rate and whether the rhythm was regular.

DIFFERENTIAL DIAGNOSIS

Patients who report "skipped" beats or a "flopping" sensation often have atrial or ventricular extrasystoles (Chap. 230). These premature beats are followed by a compensatory pause, and the first heart beat after the pause may be unusually strong due to increased left ventricular volume and enhanced contractility (a phenomenon called *postextrasystolic potentiation*). Sustained bursts of rapid heart beats may be due to ventricular or supraventricular tachyarrhythmias. A sustained irregular rhythm suggests atrial fibrillation.

Conditions that cause marked left ventricular enlargement such as aortic regurgitation can cause an awareness of the heart beat that is sometimes positional. Presumably because of associated arrhythmias, hypertrophic cardiomyopathy, mitral valve prolapse, and other cardiac structural abnormalities are also associated with palpitations.

Palpitations can also be a prominent symptom in noncardiac conditions, including thyrotoxicosis, hypoglycemia, pheochromocytoma, and fever. The physiologic basis of palpitations with these conditions is either arrhythmia or increased catecholamine levels leading to greater myocardial contractility. Drugs that can precipitate arrhythmias and palpitations include tobacco, coffee, tea, alcohol, epinephrine, ephedrine, aminophylline, and atropine.

Approach to the Patient

The first goal in the evaluation of patients with palpitations is to exclude the possibility of life-threatening arrhythmias. The risk for such arrhythmias is highest in patients with coronary artery disease, congestive heart failure, or other structural cardiac abnormalities. The history, physical examination, and electrocardiogram should therefore be focused on stratifying patients according to the risk of such conditions. Palpitations are also more likely to reflect serious arrhythmias if they are associated with symptoms that suggest hemodynamic compromise, such as syncope, light-headedness,

dizziness, or shortness of breath.

The most common first test after the initial evaluation of palpitations is continuous electrocardiographic (Holter) monitoring. This test is especially useful if patients have palpitations on a daily basis. For patients with more sporadic palpitations, a variety of new technologies have become available to allow capture of electrocardiographic tracings at the time of their symptoms. These technologies include loop recorders, that can freeze the last several minutes of data when the patient presses a button, and telephonic monitors, which can be used to "call in" tracings when symptoms occur. If episodes are associated with physical stress, exercise electrocardiography can be used in an attempt to elicit an arrhythmia.

Most patients with palpitations do not have evidence of major arrhythmias or abnormal physiologic conditions associated with increased catecholamine levels. Patients with emotional or psychological causes of palpitations should be evaluated for possible cognitive and pharmaceutical therapy. Drugs and medications that may precipitate palpitations should be eliminated or reduced. A trial of beta blockers is often successful in reducing premature beats and symptoms. Regardless of the cause and treatment, the clinician should remain aware that palpitations are extremely bothersome symptoms for patients. Reassurance that a comprehensive evaluation has been performed and that the palpitations do not adversely affect the patient's prognosis is a critical part of the patient's care.

(Bibliography omitted in Palm version)

Back to Table of Contents

14. ABDOMINAL PAIN - *William Silen*

The correct interpretation of acute abdominal pain is challenging. Since proper therapy may require urgent action, the unhurried approach suitable for the study of other conditions is sometimes denied. Few other clinical situations demand greater judgment, because the most catastrophic of events may be forecast by the subtlest of symptoms and signs. A meticulously executed, detailed history and physical examination is of great importance. The etiologic classification inTable 14-1, although not complete, forms a useful basis for the evaluation of patients with abdominal pain.

The diagnosis of "acute or surgical abdomen" is not an acceptable one because of its often misleading and erroneous connotation. The most obvious of "acute abdomens" may not require operative intervention, and the mildest of abdominal pains may herald an urgently correctable lesion. Any patient with abdominal pain of recent onset requires early and thorough evaluation and accurate diagnosis.

SOME MECHANISMS OF PAIN ORIGINATING IN THE ABDOMEN

Inflammation of the Parietal Peritoneum The pain of parietal peritoneal inflammation is steady and aching in character and is located directly over the inflamed area, its exact reference being possible because it is transmitted by somatic nerves supplying the parietal peritoneum. The intensity of the pain is dependent on the type and amount of material to which the peritoneal surfaces are exposed in a given time period. For example, the sudden release into the peritoneal cavity of a small quantity of *sterile* acid gastric juice causes much more pain than the same amount of grossly contaminated neutral feces. Enzymatically active pancreatic juice incites more pain and inflammation than does the same amount of sterile bile containing no potent enzymes. Blood and urine are often so bland as to go undetected if exposure of the peritoneum has not been sudden and massive. In the case of bacterial contamination, such as in pelvic inflammatory disease, the pain is frequently of low intensity early in the illness until bacterial multiplication has caused the elaboration of irritating substances.

The rate at which the irritating material is applied to the peritoneum is important. Perforated peptic ulcer may be associated with entirely different clinical pictures dependent only on the rapidity with which the gastric juice enters the peritoneal cavity.

The pain of peritoneal inflammation is invariably accentuated by pressure or changes in tension of the peritoneum, whether produced by palpation or by movement, as in coughing or sneezing. The patient with peritonitis lies quietly in bed, preferring to avoid motion, in contrast to the patient with colic, who may writhe incessantly.

Another characteristic feature of peritoneal irritation is tonic reflex spasm of the abdominal musculature, localized to the involved body segment. The intensity of the tonic muscle spasm accompanying peritoneal inflammation is dependent on the location of the inflammatory process, the rate at which it develops, and the integrity of the nervous system. Spasm over a perforated retrocecal appendix or perforated ulcer into the lesser peritoneal sac may be minimal or absent because of the protective effect of overlying viscera. A slowly developing process often greatly attenuates the degree of muscle spasm. Catastrophic abdominal emergencies such as a perforated ulcer may be

associated with minimal or no detectable pain or muscle spasm in obtunded, seriously ill, debilitated elderly patients or in psychotic patients.

Obstruction of Hollow Viscera The pain of obstruction of hollow abdominal viscera is classically described as intermittent, or colicky. Yet the lack of a truly cramping character should not be misleading, because distention of a hollow viscus may produce steady pain with only very occasional exacerbations. It is not nearly as well localized as the pain of parietal peritoneal inflammation.

The colicky pain of obstruction of the small intestine is usually periumbilical or supraumbilical and is poorly localized. As the intestine becomes progressively dilated with loss of muscular tone, the colicky nature of the pain may diminish. With superimposed strangulating obstruction, pain may spread to the lower lumbar region if there is traction on the root of the mesentery. The colicky pain of colonic obstruction is of lesser intensity than that of the small intestine and is often located in the infraumbilical area. Lumbar radiation of pain is common in colonic obstruction.

Sudden distention of the biliary tree produces a steady rather than colicky type of pain; hence the term *biliary colic* is misleading. Acute distention of the gallbladder usually causes pain in the right upper quadrant with radiation to the right posterior region of the thorax or to the tip of the right scapula, and distention of the common bile duct is often associated with pain in the epigastrium radiating to the upper part of the lumbar region. Considerable variation is common, however, so that differentiation between these may be impossible. The typical subscapular pain or lumbar radiation is frequently absent. Gradual dilatation of the biliary tree, as in carcinoma of the head of the pancreas, may cause no pain or only a mild aching sensation in the epigastrium or right upper quadrant. The pain of distention of the pancreatic ducts is similar to that described for distention of the common bile duct but, in addition, is very frequently accentuated by recumbency and relieved by the upright position.

Obstruction of the urinary bladder results in dull suprapubic pain, usually low in intensity. Restlessness without specific complaint of pain may be the only sign of a distended bladder in an obtunded patient. In contrast, acute obstruction of the intravesicular portion of the ureter is characterized by severe suprapubic and flank pain that radiates to the penis, scrotum, or inner aspect of the upper thigh. Obstruction of the ureteropelvic junction is felt as pain in the costovertebral angle, whereas obstruction of the remainder of the ureter is associated with flank pain that often extends into the same side of the abdomen.

Vascular Disturbances A frequent misconception, despite abundant experience to the contrary, is that pain associated with intraabdominal vascular disturbances is sudden and catastrophic in nature. The pain of embolism or thrombosis of the superior mesenteric artery or that of impending rupture of an abdominal aortic aneurysm certainly may be severe and diffuse. Yet, just as frequently, the patient with occlusion of the superior mesenteric artery has only mild continuous diffuse pain for 2 or 3 days before vascular collapse or findings of peritoneal inflammation appear. The early, seemingly insignificant discomfort is caused by hyperperistalsis rather than peritoneal inflammation. Indeed, absence of tenderness and rigidity in the presence of continuous, diffuse pain in a patient likely to have vascular disease is quite characteristic of

occlusion of the superior mesenteric artery. Abdominal pain with radiation to the sacral region, flank, or genitalia should always signal the possible presence of a rupturing abdominal aortic aneurysm. This pain may persist over a period of several days before rupture and collapse occur.

Abdominal Wall Pain arising from the abdominal wall is usually constant and aching. Movement, prolonged standing, and pressure accentuate the discomfort and muscle spasm. In the case of hematoma of the rectus sheath, now most frequently encountered in association with anticoagulant therapy, a mass may be present in the lower quadrants of the abdomen. Simultaneous involvement of muscles in other parts of the body usually serves to differentiate myositis of the abdominal wall from an intraabdominal process that might cause pain in the same region.

REFERRED PAIN IN ABDOMINAL DISEASES

Pain referred to the abdomen from the thorax, spine, or genitalia may prove a vexing diagnostic problem, because diseases of the upper part of the abdominal cavity such as acute cholecystitis or perforated ulcer are frequently associated with intrathoracic complications. A most important, yet often forgotten, dictum is that the possibility of intrathoracic disease must be considered in every patient with abdominal pain, especially if the pain is in the upper part of the abdomen. Systematic questioning and examination directed toward detecting myocardial or pulmonary infarction, pneumonia, pericarditis, or esophageal disease (the intrathoracic diseases that most often masquerade as abdominal emergencies) will often provide sufficient clues to establish the proper diagnosis. Diaphragmatic pleuritis resulting from pneumonia or pulmonary infarction may cause pain in the right upper quadrant and pain in the supraclavicular area, the latter radiation to be distinguished from the referred subscapular pain caused by acute distention of the extrahepatic biliary tree. The ultimate decision as to the origin of abdominal pain may require deliberate and planned observation over a period of several hours, during which repeated questioning and examination will provide the diagnosis.

Referred pain of thoracic origin is often accompanied by splinting of the involved hemithorax with respiratory lag and decrease in excursion more marked than that seen in the presence of intraabdominal disease. In addition, apparent abdominal muscle spasm caused by referred pain will diminish during the inspiratory phase of respiration, whereas it is persistent throughout both respiratory phases if it is of abdominal origin. Palpation over the area of referred pain in the abdomen also does not usually accentuate the pain and in many instances actually seems to relieve it. Thoracic and abdominal disease frequently coexist and may be difficult or impossible to differentiate. For example, the patient with known biliary tract disease often has epigastric pain during myocardial infarction, or biliary colic may be referred to the precordium or left shoulder in a patient who has suffered previously from angina pectoris.**For an explanation of the radiation of pain to a previously diseased area, see Chap. 12.*

Referred pain from the spine, which usually involves compression or irritation of nerve roots, is characteristically intensified by certain motions such as cough, sneeze, or strain and is associated with hyperesthesia over the involved dermatomes. Pain referred to the abdomen from the testicles or seminal vesicles is generally accentuated by the

slightest pressure on either of these organs. The abdominal discomfort is of dull aching character and is poorly localized.

METABOLIC ABDOMINAL CRISES

Pain of metabolic origin may simulate almost any other type of intraabdominal disease. Several mechanisms may be at work. In certain instances, such as hyperlipidemia, the metabolic disease itself may be accompanied by an intraabdominal process such as pancreatitis, which can lead to unnecessary laparotomy unless recognized. C¢I esterase deficiency associated with angioneurotic edema is often associated with episodes of severe abdominal pain. Whenever the cause of abdominal pain is obscure, a metabolic origin always must be considered. Abdominal pain is also the hallmark of familial Mediterranean fever (Chap. 289).

The problem of differential diagnosis is often not readily resolved. The pain of porphyria and of lead colic is usually difficult to distinguish from that of intestinal obstruction, because severe hyperperistalsis is a prominent feature of both. The pain of uremia or diabetes is nonspecific, and the pain and tenderness frequently shift in location and intensity. Diabetic acidosis may be precipitated by acute appendicitis or intestinal obstruction, so if prompt resolution of the abdominal pain does not result from correction of the metabolic abnormalities, an underlying organic problem should be suspected. Black widow spider bites produce intense pain and rigidity of the abdominal muscles and back, an area infrequently involved in intraabdominal disease.

NEUROGENIC CAUSES

Causalgic pain may occur in diseases that injure sensory nerves. It has a burning character and is usually limited to the distribution of a given peripheral nerve. Normal stimuli such as touch or change in temperature may be transformed into this type of pain, which is frequently present in a patient at rest. The demonstration of irregularly spaced cutaneous pain spots may be the only indication of an old nerve lesion underlying causalgic pain. Even though the pain may be precipitated by gentle palpation, rigidity of the abdominal muscles is absent, and the respirations are not disturbed. Distention of the abdomen is uncommon, and the pain has no relationship to the intake of food.

Pain arising from spinal nerves or roots comes and goes suddenly and is of a lancinating type (Chap. 16). It may be caused by herpes zoster, impingement by arthritis, tumors, herniated nucleus pulposus, diabetes, or syphilis. It is not associated with food intake, abdominal distention, or changes in respiration. Severe muscle spasm, as in the gastric crises of tabes dorsalis, is common but is either relieved or is not accentuated by abdominal palpation. The pain is made worse by movement of the spine and is usually confined to a few dermatomes. Hyperesthesia is very common.

Psychogenic pain conforms to none of the aforementioned patterns. Mechanism is hard to define. The most common problem is the hysterical adolescent or young person who develops abdominal pain and who frequently loses an appendix or other organs because of it. Ovulation or some other natural event that causes brief mild abdominal discomfort may be experienced as an abdominal catastrophe.

Psychogenic pain varies enormously in type and location but usually has no relation to meals. It is often markedly accentuated during the night. Nausea and vomiting are rarely observed. Spasm is seldom induced in the abdominal musculature and, if present, does not persist, especially if the attention of the patient can be distracted. Persistent localized tenderness is rare, and if found, the muscle spasm in the area is inconsistent or absent. Shallow respiration is the most common breathing abnormality; anxiety may produce a smothering or choking sensation. It occurs in the absence of thoracic splinting or change in the respiratory rate.

Approach to the Patient

Few abdominal conditions require such urgent operative intervention that an orderly approach need be abandoned, no matter how ill the patient. Only those patients with exsanguinating hemorrhage must be rushed to the operating room immediately, but in such instances, only a few minutes are required to assess the critical nature of the problem. Under these circumstances, all obstacles must be swept aside, adequate venous access for fluid replacement obtained, and the operation begun. Many patients of this type have died in the radiology department or the emergency room while awaiting such unnecessary examinations as electrocardiograms or abdominal films. *There are no contraindications to operation when massive hemorrhage is present*. This situation fortunately is relatively rare.

Nothing will supplant an orderly, painstakingly *detailed history*, which is far more valuable than any laboratory or radiographic examination. This kind of history is laborious and time-consuming, making it not especially popular, even though a reasonably accurate diagnosis can be made on the basis of the history alone in the majority of cases. Computer-aided diagnosis of abdominal pain provides no advantage over clinical assessment alone. In cases of *acute* abdominal pain, a diagnosis is readily established in most instances, whereas success is not so frequent in patients with *chronic* pain. Irritable bowel syndrome is one of the most common causes of abdominal pain and must always be kept in mind (Chap. 288). The *chronological sequence of events* in the patient's history is often more important than emphasis on the location of pain. If the examiner is sufficiently open-minded and unhurried, asks the proper questions, and listens, the patient will usually provide the diagnosis. Careful attention should be paid to the extraabdominal regions that may be responsible for abdominal pain. An accurate menstrual history in a female patient is essential. Narcotics or analgesics should *not* be withheld until a definitive diagnosis or a definitive plan has been formulated; obfuscation of the diagnosis by adequate analgesia is unlikely.

In the examination, simple critical inspection of the patient, e.g., of facies, position in bed, and respiratory activity, may provide valuable clues. The amount of information to be gleaned is directly proportional to the *gentleness* and thoroughness of the examiner. Once a patient with peritoneal inflammation has been examined brusquely, accurate assessment by the next examiner becomes almost impossible. Eliciting rebound tenderness by sudden release of a deeply palpating hand in a patient with suspected peritonitis is cruel and unnecessary. The same information can be obtained by gentle percussion of the abdomen (rebound tenderness on a miniature scale), a maneuver that can be far more precise and localizing. Asking the patient to cough will elicit true

rebound tenderness without the need for placing a hand on the abdomen. Furthermore, the forceful demonstration of rebound tenderness will startle and induce protective spasm in a nervous or worried patient in whom true rebound tenderness is not present. A palpable gallbladder will be missed if palpation is so brusque that voluntary muscle spasm becomes superimposed on involuntary muscular rigidity.

As in history taking, there is no substitute for sufficient time spent in the examination. Abdominal signs may be minimal but nevertheless, if accompanied by consistent symptoms, may be exceptionally meaningful. Abdominal signs may be virtually or totally absent in cases of pelvic peritonitis, so careful *pelvic and rectal examinations are mandatory in every patient with abdominal pain*. Tenderness on pelvic or rectal examination in the absence of other abdominal signs can be caused by operative indications such as perforated appendicitis, diverticulitis, twisted ovarian cyst, and many others.

Much attention has been paid to the presence or absence of peristaltic sounds, their quality, and their frequency. Auscultation of the abdomen is one of the least revealing aspects of the physical examination of a patient with abdominal pain. Catastrophes such as strangulating small intestinal obstruction or perforated appendicitis may occur in the presence of normal peristalsis. Conversely, when the proximal part of the intestine above an obstruction becomes markedly distended and edematous, peristaltic sounds may lose the characteristics of borborygmi and become weak or absent, even when peritonitis is not present. It is usually the severe chemical peritonitis of sudden onset that is associated with the truly silent abdomen. Assessment of the patient's state of hydration is important.

Laboratory examinations may be of great value in assessment of the patient with abdominal pain, yet with few exceptions they rarely establish a diagnosis. Leukocytosis should never be the single deciding factor as to whether or not operation is indicated. A white blood cell count greater than 20,000/uL may be observed with perforation of a viscus, but pancreatitis, acute cholecystitis, pelvic inflammatory disease, and intestinal infarction may be associated with marked leukocytosis. A normal white blood cell count is not rare in cases of perforation of abdominal viscera. The diagnosis of anemia may be more helpful than the white blood cell count, especially when combined with the history.

The urinalysis may reveal the state of hydration or rule out severe renal disease, diabetes, or urinary infection. Blood urea nitrogen, glucose, and serum bilirubin levels may be helpful. Serum amylase levels may be increased by many diseases other than pancreatitis, e.g., perforated ulcer, strangulating intestinal obstruction, and acute cholecystitis; thus, elevations of serum amylase do not rule out the need for an operation. The determination of the serum lipase may have greater accuracy than that of the serum amylase.

Plain and upright or lateral decubitus radiographs of the abdomen may be of value in cases of intestinal obstruction, perforated ulcer, and a variety of other conditions. They are usually unnecessary in patients with acute appendicitis or strangulated external hernias. In rare instances, barium or water-soluble contrast study of the upper part of the gastrointestinal tract may demonstrate partial intestinal obstruction that may elude diagnosis by other means. If there is any question of obstruction of the colon, oral

administration of barium sulfate should be avoided. On the other hand, in cases of suspected colonic obstruction (with perforation), contrast enema may be diagnostic.

In the absence of trauma, peritoneal lavage has been replaced as a diagnostic tool by ultrasound, computed tomography (CT), and laparoscopy. Ultrasonography has proved to be useful in detecting an enlarged gallbladder or pancreas, the presence of gallstones, an enlarged ovary, or a tubal pregnancy. Laparoscopy is especially helpful in diagnosing pelvic conditions, such as ovarian cysts, tubal pregnancies, salpingitis, and acute appendicitis. Radioisotopic scans (HIDA) may help differentiate acute cholecystitis from acute pancreatitis. A CT scan may demonstrate an enlarged pancreas, ruptured spleen, or thickened colonic or appendiceal wall and streaking of the mesocolon or mesoappendix characteristic of diverticulitis or appendicitis.

Sometimes, even under the best circumstances with all available aids and with the greatest of clinical skill, a definitive diagnosis cannot be established at the time of the initial examination. Nevertheless, despite lack of a clear anatomic diagnosis, it may be abundantly clear to an experienced and thoughtful physician and surgeon that on clinical grounds alone operation is indicated. Should that decision be questionable, watchful waiting with repeated questioning and examination will often elucidate the true nature of the illness and indicate the proper course of action.

(Bibliography omitted in Palm version)

Back to Table of Contents

15. HEADACHE, INCLUDING MIGRAINE AND CLUSTER HEADACHE - *Neil H. Raskin, Stephen J. Peroutka*

Few of us are spared the experience of head pain. As many as 90% of individuals have at least one headache per year. Severe, disabling headache is reported to occur at least annually by 40% of individuals worldwide. A useful classification of the many causes of headache is shown inTable 15-1. Headache is usually a benign symptom, but occasionally it is the manifestation of a serious illness such as brain tumor, subarachnoid hemorrhage, meningitis, or giant cell arteritis. In emergency settings, approximately 5% of patients with headache are found to have a serious underlying neurologic disorder. Therefore, it is imperative that the serious causes of headache be diagnosed rapidly and accurately.

PAIN-SENSITIVE STRUCTURES OF THE HEAD

Pain is most commonly due to tissue injury resulting in stimulation of peripheral nociceptors in an intact nervous system. Pain can also result from damage to or anomalous activation of pain-sensitive pathways of the peripheral or central nervous system. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain-sensitive: the scalp, middle meningeal artery, dural sinuses, falx cerebri, and the proximal segments of the large pial arteries. The ventricular ependyma, choroid plexus, pial veins, and much of the brain parenchyma are pain-insensitive. Electrical stimulation of the midbrain in the region of the dorsal raphe has resulted in migraine-like headaches. Thus, whereas most of the brain is insensitive to electrode probing, a site in the midbrain represents a possible source of headache generation. Sensory stimuli from the head are conveyed to the central nervous system via the trigeminal nerves for structures above the tentorium in the anterior and middle fossae of the skull and via the first three cervical nerves for those in the posterior fossa and the inferior surface of the tentorium.

Headache can occur as the result of (1) distention, traction, or dilation of intracranial or extracranial arteries; (2) traction or displacement of large intracranial veins or their dural envelope; (3) compression, traction, or inflammation of cranial and spinal nerves; (4) spasm, inflammation, or trauma to cranial and cervical muscles; (5) meningeal irritation and raised intracranial pressure; or (6) other possible mechanisms such as activation of brainstem structures.

GENERAL CLINICAL CONSIDERATIONS

The quality, location, duration, and time course of the headache and the conditions that produce, exacerbate, or relieve it should be carefully reviewed. Ascertaining the *quality* of cephalic pain is occasionally helpful for diagnosis. Most tension-type headaches are described as tight "bandlike" pain or as dull, deeply located, and aching pain. Jabbing, brief, sharp cephalic pain, often occurring multifocally (ice pick-like pain), is the signature of a benign, nondescript disorder. A throbbing quality and tight muscles about the head, neck, and shoulder girdle are common nonspecific accompaniments of vascular headaches.

Pain *intensity* rarely has diagnostic value, although from the patient's perspective, it is

the single aspect of pain that is most important. Although meningitis, subarachnoid hemorrhage, and cluster headache produce intense cranial pain, most patients entering emergency departments with the most severe headache of their lives usually have migraine. Contrary to common belief, the headache produced by a brain tumor is not usually distinctive or severe.

Data regarding *location* of headache may be informative. If the source is an extracranial structure, as in giant cell arteritis, the correspondence with the site of pain is fairly precise. Inflammation of an extracranial artery causes pain and exquisite tenderness localized to the site of the vessel. Lesions of paranasal sinuses, teeth, eyes, and upper cervical vertebrae induce less sharply localized pain, but pain that is still referred in a regional distribution. Intracranial lesions in the posterior fossa cause pain that is usually occipitonuchal, and supratentorial lesions most often induce frontotemporal pain.

Duration and *time-intensity* curves of headaches are diagnostically useful. A ruptured aneurysm results in head pain that peaks in an instant, thunderclap-like; much less often, unruptured aneurysms may signal their presence in the same way. Cluster headache attacks reach their peak over 3 to 5 min, remain at maximal levels for about 45 min, and then taper off. Migraine attacks build up over hours, are maintained for several hours to days, and are characteristically relieved by sleep. Sleep disruption and early morning headaches that improve during the day are characteristics of headaches produced by brain tumors.

The analysis of facial pain requires a different approach. Trigeminal and, less commonly, glossopharyngeal neuralgia are frequent causes of facial pain (Chap. 367). "Neuralgias" are painful disorders characterized by paroxysmal, fleeting, often electric shock-like episodes that are frequently caused by demyelinating lesions of nerves (the trigeminal or glossopharyngeal nerves in cranial neuralgias). Certain maneuvers characteristically trigger paroxysms of pain. However, the most common cause of facial pain by far is dental; provocation by hot, cold, or sweet foods is typical. The application of a cold stimulus will repeatedly induce dental pain, whereas in neuralgic disorders, a refractory period usually occurs after the initial response so that pain cannot be repeatedly induced.

The effect of eating on facial pain may provide insight into its cause. Is it the chewing, swallowing, or taste of the food that elicits pain? Chewing points toward trigeminal neuralgia, temporomandibular joint dysfunction, or giant cell arteritis ("jaw claudication"), whereas swallowing *and* taste provocation point toward glossopharyngeal neuralgia. Pain upon swallowing is common among patients with carotidynia (see below) because the inflamed, tender carotid artery abuts the esophagus during deglutition.

Many patients with facial pain do not experience stereotypic neuralgias; the term *atypical facial pain* has been used in this setting. Vague, poorly localized, continuous facial pain is characteristic of nasopharyngeal carcinoma; a burning pain often develops as deafferentation occurs and evidence of cranial neuropathy appears. Burning facial pain may also occur with tumors of the fifth cranial nerve (meningioma or schwannoma) or with lesions of the pons that interrupt the dorsal root entry zone of the nerve (multiple sclerosis). In patients with facial pain, the finding of objective sensory loss is an important clue to a serious underlying disorder. Occasionally, the cause of a pain

problem cannot be resolved promptly, necessitating periodic follow- up until further signs appear.

CLINICAL EVALUATION OF ACUTE, NEW-ONSET HEADACHE

Patients who present with their first severe headache raise entirely different diagnostic possibilities than those with recurrent headaches over many years. In new-onset and severe headaches, the probability of finding a potentially serious cause is considerably greater than in recurrent headache. When a patient complains of an acute, new-onset headache, a number of causes should be considered including meningitis, subarachnoid hemorrhage, epidural or subdural hematoma, glaucoma, and purulent sinusitis. Clinical features of acute, new-onset headache caused by serious underlying conditions are summarized inTable 15-2.

A complete neurologic examination is an essential first step in the evaluation. In most cases, an abnormal examination should be followed by a computed tomography (CT) or a magnetic resonance imaging (MRI) study. As a screening procedure for intracranial pathology in this setting, CT and MRI methods appear to be equally sensitive. A general evaluation of acute headache might include the investigation of cardiovascular and renal status by blood pressure monitoring and urine examination; eyes by fundoscopy, intraocular pressure measurement, and refraction; cranial arteries by palpation; and cervical spine by the effect of passive movement of the head and imaging.

The psychological state of the patient should also be evaluated since a relationship exists between head pain and depression. Many patients in chronic daily pain cycles become depressed; moreover, there is a greater-than-chance coincidence of migraine with both bipolar (manic depressive) and unipolar major depressive disorders. Drugs with antidepressant actions are also effective in the prophylactic treatment of both tension-type headache and migraine.

Underlying recurrent headache disorders may be activated by pain that follows otologic or endodontic surgical procedures. Treatment of the headache problem is largely ineffective until the cause of the primary problem is addressed. Thus, pain about the head as the result of diseased tissue or trauma may reawaken an otherwise quiescent migrainous syndrome.

Serious underlying conditions that are associated with headache are described below and inTable 15-3.

MENINGITIS

In general, acute, severe headache with stiff neck and fever suggests meningitis. Lumbar puncture is mandatory. Often there is striking accentuation of pain with eye movement. Meningitis is particularly easy to mistake for migraine in that the cardinal symptoms of pounding headache, photophobia, nausea, and vomiting are present.*A detailed discussion of meningitis can be found in Chaps. 372 to 374.*

INTRACRANIAL HEMORRHAGE

In general, acute, severe headache with stiff neck but without fever suggests subarachnoid hemorrhage. A ruptured aneurysm, arteriovenous malformation, or intraparenchymal hemorrhage may also present with only headache. Rarely, if the hemorrhage is small or below the foramen magnum, the headCT scan can be normal. Therefore, a lumbar puncture may be required to make the definitive diagnosis of a subarachnoid hemorrhage.*A detailed discussion of intracranial hemorrhage can be found in Chap. 361.*

BRAIN TUMOR

Approximately 30% of patients with brain tumors consider headache to be their chief complaint. The head pain is usually nondescript -- an intermittent deep, dull aching of moderate intensity, which may worsen with exertion or change in position and may be associated with nausea and vomiting. This pattern of symptoms results from migraine far more often than from brain tumor. Headache of brain tumor disturbs sleep in about 10% of patients. Vomiting that precedes the appearance of headache by weeks is highly characteristic of posterior fossa brain tumors. A history of amenorrhea or galactorrhea should lead one to question whether a prolactin-secreting pituitary adenoma (or the polycystic ovary syndrome) is the source of headache. Headache arising de novo in a patient with known malignancy suggests either cerebral metastases and/or carcinomatous meningitis. Head pain appearing abruptly after bending, lifting, or coughing can be the clue to a posterior fossa mass (or a Chiari malformation).*A detailed discussion of brain tumors can be found in Chap. 370.*

TEMPORAL ARTERITIS (See alsoChaps. 28 and317)

Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation. This is a common disorder of the elderly; its annual incidence is 77:100,000 in individuals aged 50 and older. The average age of onset is 70 years, and women account for 65% of cases. About half of patients with untreated temporal arteritis develop blindness due to involvement of the ophthalmic artery and its branches; indeed, the ischemic optic neuropathy induced by giant cell arteritis is the major cause of rapidly developing bilateral blindness in patients over 60 years of age. Because treatment with glucocorticoids is effective in preventing this complication, prompt recognition of this disorder is important.

Typical presenting symptoms include headache, polymyalgia rheumatica (Chap. 317), jaw claudication, fever, and weight loss. Headache is the dominant symptom and often appears in association with malaise and muscle aches. Head pain may be unilateral or bilateral and is located temporally in 50% of patients but may involve any and all aspects of the cranium. Pain usually appears gradually over a few hours before peak intensity is reached; occasionally, it is explosive in onset. The quality of pain is only seldom throbbing; it is almost invariably described as dull and boring with superimposed episodic ice pick-like lancinating pains similar to the sharp pains that appear in migraine. Most patients can recognize that the origin of their head pain is superficial, external to the skull, rather than originating deep within the cranium (the pain site for migraineurs). Scalp tenderness is present, often to a marked degree; brushing the hair or resting the head on a pillow may be impossible because of pain. Headache is usually worse at night and is often aggravated by exposure to cold. Reddened, tender nodules or red

streaking of the skin overlying the temporal arteries may be found in patients with headache, as is tenderness of the temporal or, less commonly, the occipital arteries.

The erythrocyte sedimentation rate (ESR) is often, though not always, elevated; a normal ESR does not exclude giant cell arteritis. A temporal artery biopsy and the initiation of prednisone at 80 mg daily for the first 4 to 6 weeks should be instituted when clinical suspicion is high. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headaches with prednisone, so that one must be cautious about interpreting the therapeutic response.

GLAUCOMA

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The history will usually reveal that the headache started with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.*A detailed discussion of glaucoma can be found in Chap. 28.*

OTHER CAUSES OF HEADACHE

Systemic Illness There is hardly any illness that is never manifested by headache; however, some illnesses are frequently associated with headache. These include infectious mononucleosis, systemic lupus erythematosus, chronic pulmonary failure with hypercapnia (early morning headaches), Hashimoto's thyroiditis, inflammatory bowel disease, many of the illnesses associated with HIV, and the acute blood pressure elevations that occur in pheochromocytoma and in malignant hypertension. The last two examples are the exceptions to the generalization that hypertension per se is a very uncommon cause of headache; diastolic pressures of at least 120 mmHg are requisite for hypertension to cause headache. Persistent headache and fever are often the manifestations of an acute systemic viral infection; if the neck is supple in such a patient, lumbar puncture may be deferred. Some drugs and drug-withdrawal states, e.g., oral contraceptives, ovulation-promoting medications, and glucocorticoid withdrawal, are also associated with headache in some individuals.

Idiopathic Intracranial Hypertension (Pseudotumor Cerebri) Headache, clinically resembling that of brain tumor, is a common presenting symptom of pseudotumor cerebri, a disorder of raised intracranial pressure probably resulting from impaired cerebrospinal fluid CSF absorption by the arachnoid villi. Transient visual obscurations and papilledema with enlarged blind spots and loss of peripheral visual fields are additional manifestations. Most patients are young, female, and obese. They often have a history of exposure to provoking agents such as vitamin A and glucocorticoids.**Treatment of idiopathic intracranial hypertension is discussed in Chap. 28.*

Cough A male-dominated (4:1) syndrome, cough headache is characterized by transient, severe head pain upon coughing, bending, lifting, sneezing, or stooping. Head pain persists for seconds to a few minutes. Many patients date the origins of the syndrome to a lower respiratory infection accompanied by severe coughing or to strenuous weight-lifting programs. Headache is usually diffuse but is lateralized in about

one-third of patients. The incidence of serious intracranial structural anomalies causing this condition is about 25%; the Chiari malformation (Chap. 368) is a common cause. Thus,MRI is indicated for most patients with cough headache. The benign disorder may persist for a few years; it responds dramatically to indomethacin at doses ranging from 50 to 200 mg daily. Approximately half of patients will also show a response to therapeutic lumbar puncture with removal of 40 mL ofCSF.

Many patients with migraine note that attacks of headache may be provoked by *sustained* physical exertion, such as during the third mile of a 5-mile run. Such headaches build up over hours, in contrast to cough headache. The term *effort migraine* has been used for this syndrome to avoid the ambiguous term *exertional headache*.

Lumbar Puncture Headache following lumbar puncture (Chap. 356) usually begins within 48 h but may be delayed for up to 12 days. Its incidence is between 10 and 30%. Head pain is dramatically positional; it begins when the patient sits or stands upright; there is relief upon reclining or with abdominal compression. The longer the patient is upright, the longer the latency before head pain subsides. It is worsened by head shaking and jugular vein compression. The pain is usually a dull ache but may be throbbing; its location is occipitofrontal. Nausea and stiff neck often accompany headache, and occasional patients report blurred vision, photophobia, tinnitus, and vertigo. The symptoms resolve over a few days but may on occasion persist for weeks to months.

Loss ofCSFvolume decreases the brain's supportive cushion, so that when a patient is upright there is probably dilation and tension placed on the brain's anchoring structures, the pain-sensitive dural sinuses, resulting in pain. Intracranial hypotension often occurs, but severe lumbar puncture headache may be present even in patients who have normal CSF pressure.

Treatment with intravenous caffeine sodium benzoate given over a few minutes as a 500-mg dose will promptly terminate headache in 75% of patients; a second dose given in 1 h brings the total success rate to 85%. An epidural blood patch accomplished by injection of 15 mL of autologous whole blood rarely fails for those who do not respond to caffeine. The mechanism for these treatment effects is not straightforward. The blood patch has an *immediate* effect, making it unlikely that sealing off a dural hole with blood clot is its mechanism of action.

Postconcussion Following seemingly trivial head injuries and particularly after rear-end motor vehicle collisions, many patients report varying combinations of headache, dizziness, vertigo, and impaired memory. Anxiety, irritability and difficulty with concentration are other hallmarks of this syndrome. Symptoms may remit after several weeks or persist for months and even years after the injury. Postconcussion headaches may occur whether or not a person was rendered unconscious by head trauma. Typically, the neurologic examination is normal with the exception of the behavioral abnormalities, andCT orMRIstudies are unrevealing. Chronic subdural hematoma may on occasion mimic this disorder. Although the cause of postconcussive headache disorder is not known, it should not in general be viewed as a primary psychological disturbance. It often persists long after the settlement of pending lawsuits. The treatment is symptomatic support. Repeated encouragement that the syndrome

eventually remits is important.

Coital Headache This is another male-dominated (4:1) syndrome. Attacks occur periorgasmically, are very abrupt in onset, and subside in a few minutes if coitus is interrupted. These are nearly always benign events and usually occur sporadically; if they persist for hours or are accompanied by vomiting, subarachnoid hemorrhage must be excluded (Chap. 361).

PRINCIPAL CLINICAL VARIETIES OF RECURRENT HEADACHE

There is usually little difficulty in diagnosing the serious types of headaches listed above because of the clues provided by the associated symptoms and signs. It is when headache is chronic, recurrent, and unattended by other important signs of disease that the physician faces a challenging and unique medical problem. The following sections describe a variety of headache types, ranging from the most common (e.g., tension-type headache) to rare causes of recurrent headache.

TENSION-TYPE HEADACHE

The term *tension-type headache* is still commonly used to describe a chronic head pain syndrome characterized by bilateral tight, bandlike discomfort. Patients may report that the head feels as if it is in a vise or that the posterior neck muscles are tight. The pain typically builds slowly, fluctuates in severity, and may persist more or less continuously for many days. Exertion does not usually worsen the headache. The headache may be episodic or chronic (i.e., present more than 15 days per month). Tension-type headache is common in all age groups, and females tend to predominate. In some patients, anxiety or depression coexist with tension headache.

The pathophysiologic basis of tension-type headache remains unknown. Some investigators believe that periodic tension headache is biologically indistinguishable from migraine, whereas others believe that tension-type headache and migraine are two distinct clinical entities. Abnormalities of cervical and temporal muscle contraction are likely to exist, but the exact nature of the dysfunction has not yet been elucidated.

Relaxation almost always relieves tension-type headaches. Patients should be encouraged to find a means of relaxation, which, for a given individual, could include bed rest, massage, and/or formal biofeedback training. Pharmacologic treatment consists of either simple analgesics and/or muscle relaxants. Ibuprofen and naproxen sodium are useful treatments for most individuals. When simple over-the-counter analgesics such as acetaminophen, aspirin, ibuprofen, and/or other nonsteroidal anti-inflammatory drugs (NSAIDs) alone fail, the addition of butalbital and caffeine (in a combination compound such as Fiorinal, Fioricet) to these analgesics may be effective. A list of commonly used analgesics for tension-type headaches is presented in Table 15-4. For chronic tension-type headache, prophylactic therapy is recommended. Low doses of amitriptyline (10 to 50 mg at bedtime) can provide effective prophylaxis.

MIGRAINE

Migraine, the most common cause of vascular headache, afflicts approximately 15% of

women and 6% of men. A useful definition of migraine is a benign and recurring syndrome of headache, nausea, vomiting, and/or other symptoms of neurologic dysfunction in varying admixtures (Table 15-5). Migraine can often be recognized by its activators (red wine, menses, hunger, lack of sleep, glare, estrogen, worry, perfumes, let-down periods) and its deactivators (sleep, pregnancy, exhiliration, sumatriptan). A classification of the many subtypes of migraine, as defined by the International Headache Society, is shown inTable 15-1.

Severe headache attacks, regardless of cause, are more likely to be described as throbbing and associated with vomiting and scalp tenderness. Milder headaches tend to be nondescript -- tight, bandlike discomfort often involving the entire head -- the profile of tension-type headache.

Pathogenesis

Genetic Basis of Migraine Migraine has a definite genetic predisposition. Specific mutations leading to *rare* causes of vascular headache have been identified (Table 15-6). For example, the MELAS syndrome consists of a *m*itochondrial *e*ncephalomyopathy, *l*actic *a*cidosis, and *s*troke-like episodes and is caused by an A® G point mutation in the mitochondrial gene encoding for $tRNA_{Leu(UUR)}$at nucleotide position 3243. Episodic migraine-like headaches are another common clinical feature of this syndrome, especially early in the course of the disease. The genetic pattern of mitochondrial disorders is unique, since only mothers transmit mitochondrial DNA. Thus, all children of mothers with MELAS syndrome are affected with the disorder.

Familial hemiplegic migraine (FHM) is characterized by episodes of recurrent hemiparesis or hemiplegia during the aura phase of a migraine headache. Other associated symptoms may include hemianesthesia or paresthesia; hemianopic visual field disturbances; dysphasia; and variable degrees of drowsiness, confusion, and/or coma. In severe attacks, these symptoms can be quite prolonged and persist for days or weeks, but characteristically they last for only 30 to 60 min and are followed by a unilateral throbbing headache.

Approximately 50% of cases ofFHMappear to be caused by mutations within the CACNL1A4 gene on chromosome 19, which encodes a P/Q type calcium channel subunit expressed only in the central nervous system. The gene is very large (>300 kb in length) and consists of 47 exons. Four distinct point mutations have been identified within the gene (in five different families) that cosegregate with the clinical diagnosis of FHM. Analysis of haplotypes in the two families with the same mutation suggest that each mutation arose independently rather than representing a founder effect. Thus, certain subtypes of FHM are caused by mutations in the CACNL1A4 gene. The function of the CACNL1A4 gene remains unknown, but it is likely to play a role in calcium-induced neurotransmitter release and/or contraction of smooth muscle. Different mutations within this gene are the cause of another neurogenetic disorder, episodic ataxia type 2 (Chap. 364).

In a genetic association study, a *Nco*I polymorphism in the gene encoding the D_2dopamine receptor (DRD2) was overrepresented in a population of patients with migraine with aura compared to a control group of nonmigraineurs, suggesting that

susceptibility to migraine with aura is modified by certain DRD2 alleles. In a Sardinian population, an association between different DRD2 alleles and migraine has also been demonstrated. Therefore, these initial studies suggest that variations in dopamine receptor regulation and/or function may alter susceptibility to migraine since molecular variations within the DRD2 gene have been associated with variations in dopaminergic function. However, since not all individuals with certain DRD2 genotypes suffer from migraine with aura, additional genes or factors must also be involved. Migraine is likely to be a complex disorder with polygenic inheritance and a strong environmental component.

The Vascular Theory of Migraine It was widely held for many years that the headache phase of migrainous attacks was caused by extracranial vasodilatation and that the neurologic symptoms were produced by intracranial vasoconstriction (i.e., the "vascular" hypothesis of migraine). Regional cerebral blood flow studies have shown that in patients with classic migraine there is, during attacks, a modest cortical hypoperfusion that begins in the visual cortex and spreads forward at a rate of 2 to 3 mm/min. The decrease in blood flow averages 25 to 30% (insufficient to explain symptoms on the basis of ischemia) and progresses anteriorly in a wavelike fashion independent of the topography of cerebral arteries. The wave of hypoperfusion persists for 4 to 6 h, appears to follow the convolutions of the cortex, and does not cross the central or lateral sulcus, progressing to the frontal lobe via the insula. Perfusion of subcortical structures is normal. Contralateral neurologic symptoms appear during temporoparietal hypoperfusion; at times, hypoperfusion persists in these regions after symptoms cease. More often, frontal spread continues as the headache phase begins. A few patients with classic migraine show no flow abnormalities; an occasional patient has developed focal ischemia sufficient to cause symptoms. However, focal ischemia does not appear to be *necessary* for focal symptoms to occur.

The ability of these changes to induce the symptoms of migraine has been questioned. Specifically, the decrease in blood flow that is observed does not appear to be significant enough to cause focal neurologic symptoms. Second, the increase in blood flow per se is not painful, and vasodilatation alone cannot account for the local edema and focal tenderness often observed in migraineurs. Moreover, in migraine without aura, no flow abnormalities are usually seen. Thus, it is unlikely that simple vasoconstriction and vasodilatation are the fundamental pathophysiologic abnormalities in migraine. However, it is clear that cerebral blood flow is altered during certain migraine attacks, and these changes may explain some, but clearly not all, of the clinical syndrome of migraine.

The Neuronal Theory of Migraine In 1941, the psychologist KS Lashley charted his own *fortification spectrum*, which is a migraine aura characterized by a slowly enlarging visual scotoma with luminous edges (see below). He was able to estimate that the evolution of his own scotoma proceeded across the occipital cortex at a rate of 3 mm/min. He speculated that a wavefront of intense excitation followed by a wave of complete inhibition of activity were propagated across the visual cortex. In 1944, the phenomenon that has come to be known as *spreading depression* was described by the Brazilian physiologist Leao in the cerebral cortex of laboratory animals. It is a slowly moving (2 to 3 mm/min), potassium-liberating depression of cortical activity, preceded by a wavefront of increased metabolic activity that can be produced by a variety of

experimental stimuli, including hypoxia, mechanical trauma, and the topical application of potassium. These observations suggest that neuronal abnormalities, most likely initiated in the brainstem, could be the cause of a migraine attack. More recently, both cortical and brainstem changes have been observed in positron emission tomography (PET) scan studies of migraine. Thus, the existence of a specific "brainstem generator" for migraine remains an intriguing possibility that might represent the pathophysiologic basis of migraine.

The Trigeminovascular System in Migraine Activation of cells in the trigeminal nucleus caudalis in the medulla (a pain-processing center for the head and face region) results in the release of vasoactive neuropeptides, including substance P and calcitonin gene-related peptide (CGRP), at vascular terminations of the trigeminal nerve. These peptide neurotransmitters have been proposed to induce a sterile inflammation that activates trigeminal nociceptive afferents originating on the vessel wall, further contributing to the production of pain. This mechanism also provides a potential mechanism for the soft tissue swelling and tenderness of blood vessels that attend migraine attacks. However, numerous pharmacologic agents that are effective in preventing or reducing inflammation in this animal model (e.g., selective 5-HT_{1D}agonists, NK-1 antagonists, endothelin antagonists) have failed to demonstrate any clinical efficacy in recent migraine trials.

5-Hydroxytryptamine in Migraine Pharmacologic and other data point to the involvement of the neurotransmitter 5-hydroxytryptamine (5-HT; also know as serotonin) in migraine. Approximately 40 years ago, methysergide was found to antagonize certain peripheral actions of 5-HT and was introduced as the first drug capable of preventing migraine attacks. Subsequently, it was found that platelet levels of 5-HT fall consistently at the onset of headache and that drugs that cause 5-HT to be released may trigger migrainous episodes. Such changes in circulating 5-HT levels proved to be pharmacologically trivial, however, and interest in the humoral role of 5-HT in migraine declined.

More recently, interest in the role of5-HT in migraine has been renewed due to the introduction of the triptan class of antimigraine drugs. The triptans are designed to stimulate selectively a particular subpopulation of 5-HT receptors. Molecular cloning studies have demonstrated that at least 14 specific 5-HT receptors exist in humans. The triptans (e.g., naratriptan, rizatriptan, sumatriptan, and zolmitriptan) are potent agonists of 5-HT_{1B}, 5-HT_{1D}, and 5-HT_{1F}receptors and are less potent at 5-$HT_{1A}$and 5-HT_{1E}receptors. A growing body of data indicates that the antimigraine efficacy of the triptans relates to their ability to stimulate 5-HT_{1B}receptors, which are located both on blood vessels and nerve terminals. Selective 5-HT_{1D}receptor agonists have, thus far, failed to demonstrate clinical efficacy in migraine. Triptans that are weak 5-HT_{1F}agonists are also effective in migraine; however, only 5-HT_{1B}efficacy is currently thought to be essential for antimigraine efficacy.

Physiologically, electrical stimulation near dorsal raphe neurons can result in migraine-like headaches. Blood flow in the pons and midbrain increases focally during migraine headache episodes; this alteration probably results from increased activity of cells in the dorsal raphe and locus caeruleus. There are projections from the dorsal raphe that terminate on cerebral arteries and alter cerebral blood flow. There are also

major projections from the dorsal raphe to important visual centers, including the lateral geniculate body, superior colliculus, retina, and visual cortex. These various serotonergic projections may represent the neural substrate for the circulatory and visual characteristics of migraine. The dorsal raphe cells stop firing during deep sleep, and sleep is known to ameliorate migraine; the antimigraine prophylactic drugs also inhibit activity of the dorsal raphe cells through a direct or indirect agonist effect.

RecentPET scan studies have demonstrated that midbrain structures near the dorsal raphe are differentially activated during a migraine attack. In one study of acute migraine, an injection of sumatriptan relieved the headache, but did not alter the brainstem changes noted on the PET scan. These data suggest that a "brainstem generator" may be the cause of migraine and that certain antimigraine medications may not interfere with the underlying pathologic process in migraine.

Dopamine in Migraine A growing body of biologic, pharmacologic, and genetic data support a role for dopamine in the pathophysiology of certain subtypes of migraine. Most migraine symptoms can be induced by dopaminergic stimulation. Moreover, there is dopamine receptor hypersensitivity in migraineurs, as demonstrated by the induction of yawning, nausea, vomiting, hypotension, and other symptoms of a migraine attack by dopaminergic agonists at doses that do not affect nonmigraineurs. Conversely, dopamine receptor antagonists are effective therapeutic agents in migraine, especially when given parenterally or concurrently with other antimigraine agents. As noted above, recent genetic data also suggest that molecular variations within dopamine receptor genes play a modifying role in the pathophysiology of migraine with aura. Therefore, modulation of dopaminergic neurotransmission should be considered in the therapeutic management of migraine.

The Sympathetic Nervous System in Migraine Biochemical changes occur within the sympathetic nervous system (SNS) of migraineurs before, during, and between migraine attacks. Factors that activate the SNS are all trigger factors for migraine. Specific examples include environmental changes (e.g., stress, sleep patterns, hormonal shifts, hypoglycemia) and agents that cause release and a secondary depletion of peripheral catecholamines [e.g., tyramine, phenylethylamine, fenfluramine, m-chlorophenylpiperazine (mCPP) and reserpine]. By contrast, effective therapeutic approaches to migraine share an ability to mimic and/or enhance the effects of norepinephrine in the peripheral SNS. For example, norepinephrine itself, sympathomimetics (e.g., isometheptene), monoamine oxidase inhibitors (MAOIs) and reuptake blockers alleviate migraine. Dopamine antagonists, prostaglandin synthesis inhibitors, and adenosine antagonists are pharmacologic agents effective in the acute treatment of migraine. These drugs block the negative feedback inhibition or norepinephrine release induced by endogenous dopamine, prostaglandins, and adenosine. Therefore, migraine susceptibility may relate to genetically based variations in the ability to maintain adequate concentrations of certain neurotransmitters within postganglionic sympathetic nerve terminals. This hypothesis has been called the *empty neuron theory* of migraine.

Clinical Features

Migraine without Aura (Common Migraine) In this syndrome no focal neurologic

disturbance precedes the recurrent headaches. Migraine without aura is by far the more frequent type of vascular headache. The International Headache Society criteria for migraine include moderate to severe head pain, pulsating quality, unilateral location, aggravation by walking stairs or similar routine activity, attendant nausea and/or vomiting, photophobia and phonophobia, and multiple attacks, each lasting 4 to 72 h.

Migraine with Aura (Classic Migraine) In this syndrome headache is associated with characteristic premonitory sensory, motor, or visual symptoms. Focal neurologic disturbances are more common during headache attacks than as prodromal symptoms. Focal neurologic disturbances without headache or vomiting have come to be known as *migraine equivalents* or *migraine accompaniments* and appear to occur more commonly in patients between the ages of 40 and 70 years. The term *complicated migraine* has generally been used to describe migraine with dramatic transient focal neurologic features or a migraine attack that leaves a persisting residual neurologic deficit.

The most common premonitory symptoms reported by migraineurs are visual, arising from dysfunction of occipital lobe neurons. Scotomas and/or hallucinations occur in about one-third of migraineurs and usually appear in the central portions of the visual fields. A highly characteristic syndrome occurs in about 10% of patients; it usually begins as a small paracentral scotoma, which slowly expands into a "C" shape. Luminous angles appear at the enlarging outer edge, becoming colored as the scintillating scotoma expands and moves toward the periphery of the involved half of the visual field, eventually disappearing over the horizon of peripheral vision. The entire process lasts 20 to 25 min. This phenomenon is pathognomonic for migraine, and has never been described in association with a cerebral structural anomaly. It is commonly referred to as a *fortification spectrum* because the serrated edges of the hallucinated "C" seemed to resemble a "fortified town with bastions all round it"; "spectrum" is used in the sense of an apparition or specter.

Basilar Migraine Symptoms referable to a disturbance in brainstem function, such as vertigo, dysarthria, or diplopia, occur as the only neurologic symptoms of the attack in about 25% of patients. A dramatic form of basilar migraine (Bickerstaff's migraine) occurs primarily in adolescent females. Episodes begin with total blindness accompanied or followed by admixtures of vertigo, ataxia, dysarthria, tinnitus, and distal and perioral paresthesia. In about one-quarter of patients, a confusional state supervenes. The neurologic symptoms usually persist for 20 to 30 min and are generally followed by a throbbing occipital headache. This basilar migraine syndrome is now known also to occur in children and in adults over age 50. An altered sensorium may persist for as long as 5 days and may take the form of confusional states superficially resembling psychotic reactions. Full recovery after the episode is the rule.

Carotidynia The carotidynia syndrome, sometimes called *lower- half headache* or *facial migraine*, is most common among older patients, with the incidence peaking in the fourth through sixth decades. Pain is usually located at the jaw or neck, although sometimes periorbital or maxillary pain occurs; it may be continuous, deep, dull, and aching, and it becomes pounding or throbbing episodically. There are often superimposed sharp, ice pick-like jabs. Attacks occur one to several times per week, each lasting several minutes to hours. Tenderness and prominent pulsations of the cervical carotid artery and soft tissue swelling overlying the carotid are usually present

ipsilateral to the pain; many patients also report throbbing ipsilateral headache concurrent with carotidynia attacks as well as between attacks. Dental trauma is a common precipitant of this syndrome. Carotid artery involvement also appears to be common in the more traditional forms of migraine; over 50% of patients with frequent migraine attacks are found to have carotid tenderness at several points on the side most often involved during hemicranial migraine attacks.

TREATMENT

Nonpharmacologic Approaches for All Migraineurs Migraine can often be managed to some degree by a variety of nonpharmacologic approaches (Table 15-7). The measures that apply to a given individual should be used routinely since they provide a simple, cost-effective approach to migraine management. Patients with migraine do not encounter more stress than headache-free individuals; overresponsiveness to stress appears to be the issue. Since the stresses of everyday living cannot be eliminated, lessening one's response to stress by various techniques is helpful for many patients. These include yoga, transcendental meditation, hypnosis, and conditioning techniques such as biofeedback. For most patients, this approach is, at best, an adjunct to pharmacotherapy. Avoidance of migraine trigger factors may also provide significant prophylactic benefits (Table 15-7). Unfortunately, these measures are unlikely to prevent all migraine attacks. When these measures fail to prevent an attack, then pharmacologic approaches are needed to abort an attack.

Pharmacologic Treatment of Acute Migraine The mainstay of pharmacologic therapy is the judicious use of one or more of the many drugs that are effective in migraine. The selection of the optimal regimen for a given patient depends on a number of factors, the most important of which is the severity of the attack (Table 15-8). Mild migraine attacks can usually be managed by oral agents; the average efficacy rate is 50-70%. Severe migraine attacks may require parenteral therapy. Most drugs effective in the treatment of migraine are members of one of three major pharmacologic classes: anti-inflammatory agents, 5-HT1 agonists, and dopamine antagonists.

Table 15-9 lists specific drugs effective in migraine. In general, an adequate dose of whichever agent is chosen should be used as soon as possible after the onset of an attack. If additional medication is required within 60 min because symptoms return or have not abated, the initial dose should be increased for subsequent attacks. Migraine therapy must be individualized for each patient; a standard approach for all patients is not possible. A therapeutic regimen may need to be constantly refined and personalized until one is identified that provides the patient with rapid, complete, and consistent relief with minimal side effects.

Nonsteroidal anti-inflammatory agents Both the severity and duration of a migraine attack can be reduced significantly by anti- inflammatory agents. Indeed, many undiagnosed migraineurs are self- treated with nonprescription anti-inflammatory agents (Table 15-4). A general consensus is thatNSAIDsare most effective when taken early in the migraine attack. However, the effectiveness of anti-inflammatory agents in migraine is usually less than optimal in moderate or severe migraine attacks. The combination of acetaminophen, aspirin, and caffeine (Excedrin Migraine) has been approved for use by the U.S. Food and Drug Administration (FDA) for the treatment of mild to moderate

migraine. The combination of aspirin and metoclopramide has been show to be equivalent to a single dose of sumatriptan. Major side effects of NSAIDs include dyspepsia and gastrointestinal irritation.

5-HT$_1$agonists

ORAL Stimulation of5-HT$_1$receptors can stop an acute migraine attack. Ergotamine and dihydroergotamine are nonselective receptor agonists, while the series of drugs known as triptans are selective 5-HT$_1$receptor agonists. A variety of triptans (e.g., naratriptan, rizatriptan, sumatriptan, zolmitriptan) are now available for the treatment of migraine (Table 15-9).

Each of the triptan class of drugs has similar pharmacologic properties, but varies slightly in terms of clinical efficacy. Rizatriptan appears to be the fastest acting and most efficacious of the triptans currently available in the United States. Sumatriptan and zolmitriptan have similar rates of efficacy as well as time to onset, whereas naratriptan is the slowest acting and the least efficacious. Clinical efficacy appears to be related more to the t_{max}(time to peak plasma level) than to the potency, half-life, or bioavailability (Table 15-10). This observation is in keeping with a significant body of data indicating that faster-acting analgesics are more efficacious than slower-acting agents.

Unfortunately, monotherapy with a selective oral5-HT$_1$agonist does not result in rapid, consistent, and complete relief of migraine in all patients. Triptans are not effective in migraine with aura unless given after the aura is completed and the headache initiated. Side effects, although often mild and transient, occur in up to 89% of patients. Moreover, 5-HT$_1$agonists are contraindicated in individuals with a history of cardiovascular disease. Recurrence of headache is a major limitation of triptan use, and occurs at least occasionally in 40 to 78% of patients.

Ergotamine preparations offer a nonselective means of stimulating5-HT$_1$receptors. A nonnauseating dose of ergotamine should be sought since a dose that provokes nausea is too high and may intensify head pain. Except for a sublingual formulation of ergotamine (Ergomar), oral formulations of ergotamine also contain 100 mg caffeine (theoretically to enhance ergotamine absorption and possibly to add additional vasoconstrictor activity). The average oral ergotamine dose for a migraine attack is 2 mg. Since the clinical studies demonstrating the efficacy of ergotamine in migraine predated the clinical trial methodologies used with the triptans, it is difficult to assess the clinical efficacy of ergotamine versus the triptans. In general, ergotamine appears to have a much higher incidence of nausea than triptans, but less headache recurrence.

NASAL The fastest acting nonparenteral antimigraine therapies that can be self-administered include nasal formulations of dihydroergotamine (Migranal) or sumatriptan (Imitrex Nasal). The nasal sprays result in substantial blood levels within 30 to 60 min. However, the nasal formulations suffer from inconsistent dosing, poor taste, and variable efficacy. Although in theory the nasal sprays might provide faster and more effective relief of a migraine attack than oral formulations, their reported efficacy is only approximately 50 to 60%.

PARENTERAL Parenteral administration of drugs such as dihydroergotamine (DHE-45

Injectable) and sumatriptan (Imitrex SC) is approved by the FDA for the rapid relief of a migraine attack. Peak plasma levels of dihydroergotamine are achieved 3 min after intravenous dosing, 30 min after intramuscular dosing, and 45 min after subcutaneous dosing. If an attack has not already peaked, subcutaneous or intramuscular administration of 1 mg dihydroergotamine suffices for about 80 to 90% of patients. Sumatriptan, 6 mg subcutaneously is effective in approximately 70 to 80% of patients.

Dopamine Antagonists

ORAL Oral dopamine antagonists should be considered as adjunctive therapy in migraine. Drug absorption is impaired during migrainous attacks because of reduced gastrointestinal motility. Delayed absorption occurs in the absence of nausea and is related to the severity of the attack and not its duration. Therefore, when oral NSAIDs and/or triptan agents fail, the addition of a dopamine antagonist such as metoclopramide, 10 mg, should be considered to enhance gastric absorption. In addition, dopamine antagonists decrease nausea/vomiting and restore normal gastric motility.

PARENTERAL Parenteral dopamine antagonists (e.g., chlorpromazine, prochlorperazine, metoclopramide) can also provide significant acute relief of migraine; they can be used in combination with parenteral 5-HT_1 agonists. A common intravenous protocol used for the treatment of severe migraine is the administration over 2 min of a mixture of 5 mg of prochlorperazine and 0.5 mg of dihydroergotamine.

Other Medications for Acute Migraine

ORAL The combination of acetaminophen, dichloralphenazone, and isometheptene (i.e., Midrin, Duradrin, generic), one to two capsules, has been classified by the FDA as "possibly" effective in the treatment of migraine. Since the clinical studies demonstrating the efficacy of this combination analgesic in migraine predated the clinical trial methodologies used with the triptans, it is difficult to assess the clinical efficacy of this sympathomimetic compound in comparison to other agents.

NASAL A nasal preparation of butorphanol is available for the treatment of acute pain. As with all narcotics, the use of nasal butorphanol should be limited to a select group of migraineurs, as described below.

PARENTERAL Narcotics are effective in the acute treatment of migraine. For example, intravenous meperidene (Demerol), 50 to 100 mg, is given frequently in the emergency room. This regimen "works" in the sense that the pain of migraine is eliminated. However, this regimen is clearly suboptimal in patients with recurrent headache for two major reasons. First, narcotics do not treat the underlying headache mechanism; rather, they act at the thalamic level to alter pain sensation. Second, the recurrent use of narcotics can lead to significant problems. In patients taking oral narcotics such as oxycodone (Percodan) or hydrocodone (Vicoden), narcotic addiction can greatly confuse the treatment of migraine. The headache that results from narcotic craving and/or withdrawal can be difficult to distinguish from chronic migraine. Therefore, it is recommended that narcotic use in migraine be limited to patients with severe, but infrequent, headaches that are unresponsive to other pharmacologic approaches.

Prophylactic Treatment of Migraine A substantial number of drugs are now available that have the capacity to stabilize migraine (Table 15-11). The decision of whether to use this approach depends on the frequency of attacks and on how well acute treatment is working. The occurrence of at least three attacks per month could be an indication for this approach. Drugs must be taken daily and there is usually a lag of at least 2 to 6 weeks before an effect is seen. The drugs that have been approved by theFDA for the prophylactic treatment of migraine include propranolol, timolol, sodium valproate, and methysergide. In addition, a number of other drugs appear to display prophylactic efficacy. This group of drugs includes amitriptyline, nortriptyline, verapamil, phenelzine, isocarbazid, and cyproheptadine. Phenelzine and methysergide are usually reserved for recalcitrant cases because of their serious potential side effects. Phenelzine is anMAOI; therefore, tyramine-containing foods, decongestants, and meperidine are contraindicated. Methysergide may cause retroperitoneal or cardiac valvular fibrosis when it is used for more than 8 months, thus monitoring is required for patients using this drug; the risk of the fibrotic complication is about 1:1500 and is likely to reverse after the drug is stopped.

The probability of success with any one of the antimigraine drugs is 50 to 75%; thus, if one drug is assessed each month, there is a good chance that effective stabilization will be achieved within a few months. Many patients are managed adequately with low-dose amitriptyline, propranolol, or valproate. If these agents fail or lead to unacceptable side effects, then methysergide or phenelzine can be used. Once effective stabilization is achieved, the drug is continued for 5 to 6 months and then slowly tapered to assess the continued need. Many patients are able to discontinue medication and experience fewer and milder attacks for long periods, suggesting that these drugs may alter the natural history of migraine.

CLUSTER HEADACHE

A variety of names have been used for this condition, including *Raeder's syndrome*, *histamine cephalalgia*, and *sphenopalatine neuralgia*. *Cluster headache* is a distinctive and treatable vascular headache syndrome. The episodic type is most common and is characterized by one to three short-lived attacks of periorbital pain per day over a 4- to 8-week period, followed by a pain- free interval that averages 1 year. The chronic form, which may begin de novo or several years after an episodic pattern has become established, is characterized by the absence of sustained periods of remission. Each type may transform into the other. Men are affected seven to eight times more often than women; hereditary factors are usually absent. Although the onset is generally between ages 20 and 50, it may occur as early as the first decade of life. Propranolol and amitriptyline are largely ineffective. Lithium is beneficial for cluster headache and ineffective in migraine. The cluster syndrome is thus clinically, genetically, and therapeutically different from migraine. Nevertheless, mixed features of the two disorders are occasionally present, suggesting some common elements to their pathogenesis.

Clinical Features Periorbital or, less commonly, temporal pain begins without warning and reaches a crescendo within 5 min. It is often excruciating in intensity and is deep, nonfluctuating, and explosive in quality; only rarely is it pulsatile. Pain is strictly unilateral

and usually affects the same side in subsequent months. Attacks last from 30 min to 2 h; there are often associated symptoms of homolateral lacrimation, reddening of the eye, nasal stuffiness, lid ptosis, and nausea. Alcohol provokes attacks in about 70% of patients but ceases to be provocative when the bout remits; this on-off vulnerability to alcohol is pathognomonic of cluster headache. Only rarely do foods or emotional factors precipitate pain, in contrast to migraine.

There is a striking periodicity of attacks in at least 85% of patients. At least one of the daily attacks of pain recurs at about the same hour each day for the duration of a cluster bout. Onset is nocturnal in about 50% of the cases, and then the pain usually awakens the patient within 2 h of falling asleep.

Pathogenesis No consistent cerebral blood flow changes accompany attacks of pain. Perhaps the strongest evidence for a central mechanism is the periodicity of attacks; the existence of a central mechanism is also suggested by the observation that autonomic symptoms that accompany the pain are bilateral and are more severe on the painful side. The hypothalamus may be the site of activation in this disorder. The posterior hypothalamus contains cells that regulate autonomic functions, and the anterior hypothalamus contains cells (in the suprachiasmatic nuclei) that constitute the principal circadian pacemaker in mammals. Activation of both is necessary to explain the symptoms of cluster headache. The pacemaker is modulated via serotonergic dorsal raphe projections. It can be concluded tentatively that both migraine and cluster headache result from abnormal serotonergic neurotransmission, albeit at different loci.

TREATMENT

The most satisfactory treatment is the administration of drugs to prevent cluster attacks until the bout is over. Effective prophylactic drugs are prednisone, lithium, methysergide, ergotamine, sodium valproate, and verapamil. Lithium (600 to 900 mg daily) appears to be particularly useful for the chronic form of the disorder. A 10-day course of prednisone, beginning at 60 mg daily for 7 days followed by a rapid taper, may interrupt the pain bout for many patients. When ergotamine is used, it is most effective when given 1 to 2 h before an expected attack. Patients must be educated regarding the early symptoms of ergotism when ergotamine is used daily; a weekly limit of 14 mg should be adhered to.

For the attacks themselves, oxygen inhalation (9 L/min via a loose mask) is the most effective modality; 15 min of inhalation of 100% oxygen is often necessary. Sumatriptan, 6 mg subcutaneously, will usually shorten an attack to 10 to 15 min.

(Bibliography omitted in Palm version)

Back to Table of Contents

16. BACK AND NECK PAIN - *John W. Engstrom*

The importance of back and neck pain in our society is underscored by the following: (1) the annual societal cost of back pain in the United States is estimated to be between $20 and $50 billion; (2) back symptoms are the most common cause of disability in patients under 45 years of age; (3) 50% of working adults, in one survey, admitted to having a back injury each year; and (4) approximately 1% of the U.S. population is chronically disabled because of back pain.

The enormous economic pressure to provide rational and efficient care of patients with back pain has resulted in clinical practice guidelines (CPGs) for these patients. CPGs are algorithms which guide evaluation or treatment at specific steps in patient care. CPGs for *acute low back pain* (ALBP) are based upon incomplete evidence (see algorithms,Fig. 16-6) but represent an attempt to standardize common medical practice. Major revisions in CPGs for back pain can be anticipated in the future. Management of patients with *chronic low back pain* (CLBP) is complex and not amenable to a simple algorithmic approach at this time.

ANATOMY OF THE SPINE

The anterior portion of the spine consists of cylindrical vertebral bodies separated by intervertebral disks and held together by the anterior and posterior longitudinal ligaments. The intervertebral disks are composed of a central gelatinous nucleus pulposus surrounded by a tough cartilagenous ring, the annulus fibrosis; disks are responsible for 25% of spinal column length (Figs. 16-1 and16-2). The disks are largest in the cervical and lumbar regions where movements of the spine are greatest. The disks are elastic in youth and allow the bony vertebrae to move easily upon each other. Elasticity is lost with age. The function of the anterior spine is to absorb the shock of typical body movements such as walking and running.

The posterior portion of the spine consists of the vertebral arches and seven processes. Each arch consists of paired cylindrical pedicles anteriorly and paired laminae posteriorly (Fig. 16-1). The vertebral arch gives rise to two transverse processes laterally, one spinous process posteriorly, plus two superior and two inferior articular facets. The functions of the posterior spine are to protect the spinal cord and nerves within the spinal canal and to stabilize the spine by providing sites for the attachment of muscles and ligaments. The contraction of muscles attached to the spinous and transverse processes produces a system of pulleys and levers that results in flexion, extension, and lateral bending movements of the spine. Normal upright posture in humans places the center of gravity anterior to the spine. The graded contraction of well-developed paraspinal muscles attached to the laminae, transverse processes, and spinous processes is necessary to maintain normal upright posture.

The nerve roots exit at a level above their respective vertebral bodies in the cervical region (the C7 nerve root exits at the C6-C7 level) and below their respective vertebral bodies in the thoracic and lumbar regions (the T1 nerve root exits at the T1-T2 level). The spinal cord ends at the L1 or L2 level of the bony spine. Consequently, the lumbar nerve roots follow a long intraspinal course and can be injured anywhere from the upper lumbar spine to their exit at the intervertebral foramen. For example, it is common for

disk herniation at the L4-L5 level to produce compression of the S1 nerve root (Fig. 16-3). In contrast, cervical nerve roots follow a short intraspinal course and exit at the level of their respective spinal cord segments (upper cervical) or one segment below the corresponding levels (lower cervical cord). Cervical spine pathology can result in spinal cord compression, but lumbar spine pathology cannot.

Pain-sensitive structures in the spine include the vertebral body periosteum, dura, facet joints, annulus fibrosus of the intervertebral disk, epidural veins, and the posterior longitudinal ligament. Damage to these nonneural structures may cause pain. The nucleus pulposus of the intervertebral disk is not pain-sensitive under normal circumstances. Pain sensation is conveyed by the sinuvertebral nerve that arises from the spinal nerve at each spine segment and reenters the spinal canal through the intervertebral foramen at the same level. Disease of these diverse pain-sensitive spine structures may explain many cases of back pain without nerve root compression. The lumbar and cervical spine possess the greatest potential for movement and injury.

Approach to the Patient

Types of Back Pain An understanding of the nature of the pain as described by the patient is the essential first step in evaluation. Attention is also focused on identification of risk factors for serious underlying diseases that require specific evaluation.

Local pain is caused by stretching of pain-sensitive structures that compress or irritate sensory nerve endings. The site of the pain is near the affected part of the back.

Pain referred to the back may arise from abdominal or pelvic viscera. The pain is usually described as primarily abdominal or pelvic but is accompanied by back pain and usually unaffected by posture. The patient may occasionally complain of back pain only.

Pain of spine origin may be located in the back or referred to the buttocks or legs. Diseases affecting the upper lumbar spine tend to refer pain to the lumbar region, groin, or anterior thighs. Diseases affecting the lower lumbar spine tend to produce pain referred to the buttocks, posterior thighs, or rarely the calves or feet. Provocative injections into pain-sensitive structures of the spine (diskography) may produce leg pain that does not follow a dermatomal distribution. The exact pathogenesis of this "sclerotomal" pain is unclear, but it may explain many instances in which combined back and leg pain is unaccompanied by evidence of nerve root compression.

Radicular back pain is typically sharp and radiates from the spine to the leg within the territory of a nerve root (see "Lumbar Disk Disease," below). Coughing, sneezing, or voluntary contraction of abdominal muscles (lifting heavy objects or straining at stool) may elicit the radiating pain. The pain may increase in postures that stretch the nerves and nerve roots. Sitting stretches the sciatic nerve (L5 and S1 roots) because the nerve passes posterior to the hip. The femoral nerve (L2, L3, and L4 roots) passes anterior to the hip and is not stretched by sitting. The description of the pain alone often fails to distinguish clearly between sclerotomal pain and radiculopathy.

Pain associated with muscle spasm, although of obscure origin, is commonly associated with many spine disorders. The spasms are accompanied by abnormal posture, taut

paraspinal muscles, and dull pain.

Back pain at rest or unassociated with specific postures should raise the index of suspicion for an underlying serious cause (e.g., spine tumor, fracture, infection, or referred pain from visceral structures). Knowledge of the circumstances associated with the onset of back pain is important when weighing possible serious underlying causes for the pain. Some patients involved in accidents or work-related injuries may exaggerate their pain for the purpose of compensation or for psychological reasons.

Examination of the Back A physical examination that includes the abdomen and rectum is advisable. Back pain referred from visceral organs may be reproduced during palpation of the abdomen (pancreatitis, abdominal aortic aneurysm) or percussion over the costovertebral angles (pyelonephritis, adrenal disease, L1-L2 transverse process fracture).

The normal spine (Fig. 16-2) displays a thoracic kyphosis, lumbar lordosis, and cervical lordosis. Exaggeration of these normal alignments may result in hyperkyphosis (lameback) of the thoracic spine or hyperlordosis (swayback) of the lumbar spine. Spasm of lumbar paraspinal muscles results in flattening of the usual lumbar lordosis. Inspection may reveal lateral curvature of the spine (scoliosis) or an asymmetry in the appearance of the paraspinal muscles, suggesting muscle spasm. Taut paraspinal muscles limit the motion of the lumbar spine. Back pain of bony spine origin is often reproduced by palpation or percussion over the spinous process of the affected vertebrae.

Forward bending is frequently limited by paraspinal muscle spasm. Flexion of the hips is normal in patients with lumbar spine disease, but flexion of the lumbar spine is limited and sometimes painful. Lateral bending to the side opposite the injured spinal element may stretch the damaged tissues, worsen pain, and limit motion. Hyperextension of the spine (with the patient prone or standing) is limited when nerve root compression or bony spine disease is present.

Pain from hip disease may mimic the pain of lumbar spine disease. The first movement is typically internal rotation of the hip. Manual internal and external rotation at the hip with the knee and hip in flexion (Patrick sign) may reproduce the pain, as may percussion of the heel (of an outstretched leg) with the palm of the examiner's hand.

In the supine position passive flexion of the thigh on the abdomen while the knee is extended produces stretching of the L5 and S1 nerve roots and the sciatic nerve because the nerve passes posterior to the hip. Passive dorsiflexion of the foot during the maneuver adds to the stretch. While flexion to at least 80° is normally possible without causing pain, tight hamstrings commonly limit motion, may result in pain, and are readily identified by the patient. This *straight leg-raising* (SLR) *sign* is positive if the maneuver reproduces the patient's usual back or limb pain. Eliciting the SLR sign in the sitting position may help determine if the finding is reproducible. The patient may describe pain in the low back, buttocks, posterior thigh, or lower leg, but the key feature is reproduction of the patient's usual pain. The *crossed* SLR sign is positive when performance of the maneuver on one leg reproduces the patient's pain symptoms in the opposite leg or buttocks. The nerve or nerve root lesion is always on the side of the

pain. The *reverse* SLR sign is elicited by standing the patient next to the examination table and passively extending each leg while the patient continues to stand. This maneuver stretches the L2-L4 nerve roots and the femoral nerve because the nerves pass anterior to the hip. The reverse SLR test is positive if the maneuver reproduces the patient's usual back or limb pain.

The neurologic examination includes a search for weakness, muscle atrophy, focal reflex changes, diminished sensation in the legs, and signs of spinal cord injury. Findings with specific nerve root lesions are shown inTable 16-1 and are discussed below.

Laboratory Studies Routine laboratory studies such as a complete blood count, erythrocyte sedimentation rate, chemistry panel, and urinalysis are rarely needed for the initial evaluation of acute (<3 months), nonspecific, low back pain. If risk factors for a serious underlying disease are present, then laboratory studies (guided by the history and examination) are indicated (Fig. 16-6*B*).

Plain films of the lumbar or cervical spine are helpful when risk factors for vertebral fracture (trauma, chronic steroid use) are present. *In the absence of risk factors, routine x-rays of the lumbar spine in the setting of acute, nonspecific, low back pain are expensive and rarely helpful*. Magnetic resonance imaging (MRI) and computed tomography (CT)-myelography have emerged as the radiologic tests of choice for evaluation of most serious diseases involving the spine. In general, the definition of soft tissue structures by MRI is superior, whereas CT-myelography provides optimal imaging of bony lesions in the region of the lateral recess and intervertebral foramen and is tolerated by claustrophobic patients. With rare exceptions, conventional myelography and bone scan are inferior to MRI and CT-myelography.

Electromyography (EMG) can be used to assess the functional integrity of the peripheral nervous system (Chap. 357) in the setting of back pain. Sensory nerve conduction studies are normal when focal sensory loss is due to nerve root damage because the nerve roots are proximal to the nerve cell bodies in the dorsal root ganglia. The diagnostic yield of needle EMG is higher than that of nerve conduction studies for radiculopathy. Denervation changes in a myotomal (segmental) distribution are detected by sampling multiple muscles supplied by different nerve roots and nerves; the pattern of muscle involvement indicates the nerve root(s) responsible for the injury. Needle EMG provides objective information about motor nerve fiber injury when the clinical evaluation of weakness is limited by pain or poor effort. EMG and nerve conduction studies will be normal when only limb pain or sensory nerve root injury or irritation is present. Mixed nerve somatosensory evoked potentials and F-wave studies are of uncertain value in the evaluation of radiculopathy.

CAUSES OF BACK PAIN

CONGENITAL ANOMALIES OF THE LUMBAR SPINE

Spondylolysis is a bony defect in the pars interarticularis (a segment near the junction of the pedicle with the lamina) of the vertebra; the etiology of the defect may be a stress fracture in a congenitally abnormal segment. The defect (usually bilateral) is best

visualized on oblique projections in plain x-rays or byCT scan and occurs in the setting of a single injury, repeated minor injuries, or growth.

Spondylolisthesis is the anterior slippage of the vertebral body, pedicles, and superior articular facets, leaving the posterior elements behind. Spondylolisthesis is associated with spondylolysis and degenerative spine disease and occurs more frequently in women. The slippage may be asymptomatic but may also cause low back pain, nerve root injury (the L5 root most frequently), or symptomatic spinal stenosis. Tenderness may be elicited near the segment that has "slipped" forward (most often L4 on L5 or occasionally L5 on S1). A "step" may be present on deep palpation of the posterior elements of the segment above the spondylolisthetic joint. The trunk may be shortened and the abdomen protuberant as a result of extreme forward displacement of L4 on L5 in severe degrees of spondylolisthesis. In these cases, cauda equina syndrome may occur (Chap. 368).

TRAUMA

Trauma is an important cause of acute low back pain. A patient complaining of back pain and inability to move the legs may have a spinal fracture or dislocation, and, with fractures above L1, spinal cord compression. In such cases care must be taken to avoid further damage to the spinal cord or nerve roots. The back should be immobilized pending results of plain x-rays.

Sprains and Strains The terms *low back sprain*, *strain*, or *mechanically induced muscle spasm* are used for minor, self-limited injuries associated with lifting a heavy object, a fall, or a sudden deceleration such as occurs in an automobile accident. These terms are used loosely and do not clearly describe a specific anatomic lesion. The pain is usually confined to the lower back, and there is no radiation to the buttocks or legs. Patients with low back pain and paraspinal muscle spasm often assume unusual postures.

Vertebral Fractures Most traumatic fractures of the lumbar vertebral bodies result from compression or flexion injuries producing anterior wedging or compression. With more severe trauma, the patient may sustain a fracture-dislocation or a "burst" fracture involving not only the vertebral body but posterior elements as well. Traumatic vertebral fractures are caused by falls from a height (a pars interarticularis fracture of the L5 vertebra is common), sudden deceleration in an automobile accident, or direct injury. Neurologic impairment is commonly associated with these injuries, and early surgical treatment is indicated (Chap. 369).

When fractures are atraumatic, the bone is presumed to be weakened by a pathologic process. The cause is usually postmenopausal (type 1) or senile (type 2) osteoporosis (Chap. 342). Underlying systemic disorders such as osteomalacia, hyperparathyroidism, hyperthyroidism, multiple myeloma, metastatic carcinoma, or glucocorticoid use may also weaken the vertebral body. The clinical context, neurologic signs, and x-ray appearance of the spine establish the diagnosis. Antiresorptive drugs including biphosphatonates, alendronate, transdermal estrogen, and tamoxifen have been shown to reduce the risk of osteoporotic fractures.

LUMBAR DISK DISEASE

This disorder is a common cause of chronic or recurrent low back and leg pain. Disk disease is most likely to occur at the L4-L5 and L5-S1 levels, but upper lumbar levels are involved occasionally. The cause of the disk injury is often unknown; the risk is increased in overweight individuals. Degeneration of the nucleus pulposus and the annulus fibrosus increases with age and may be asymptomatic or painful. A sneeze, cough, or trivial movement may cause the nucleus pulposus to prolapse, pushing the frayed and weakened annulus posteriorly. In severe disk disease, the nucleus may protrude through the annulus (herniation) or become extruded to lie as a free fragment in the spinal canal.

The mechanism by which intervertebral disk injury causes back pain is controversial. The inner annulus fibrosus and nucleus pulposis are normally devoid of innervation. Inflammation and production of proinflammatory cytokines within the protruding or ruptured disk may trigger or perpetuate back pain. Ingrowth of nociceptive (pain) nerve fibers into inner portions of diseased intervertebral disk may be responsible for chronic "diskogenic" pain. Nerve root injury (*radiculopathy*) from disk herniation may be due to compression, inflammation, or both; pathologically, varying degrees of demyelination and axonal loss are usually present.

The symptoms of a ruptured intervertebral disk include back pain, abnormal posture, limitation of spine motion (particularly flexion), or radicular pain. A dermatomal pattern of sensory loss or a reduction in or loss of a deep tendon reflex is more suggestive of a specific root lesion than the pattern of pain. Motor findings (focal weakness, muscle atrophy, or fasciculations) occur less frequently than sensory or reflex changes, but a myotomal pattern of involvement can suggest specific nerve root injury. Lumbar disk disease is usually unilateral (Fig. 16-4), but bilateral involvement does occur with large central disk herniations that compress several nerve roots at the same level. Clinical manifestations of specific lumbosacral nerve root lesions are summarized in Table 16-1. There is evidence to suggest that lumbar disk herniation with a nonprogressive nerve root deficit can be managed conservatively (i.e., nonsurgically) with a successful outcome. The size of the disk protrusion may naturally decrease over time.

Degeneration of the intervertebral disk without frank extrusion of disk tissue may give rise to low back pain only. There may be referred pain in the leg, buttock, or hip with little or no discomfort in the back and no signs of nerve root involvement. Lumbar disk syndromes are usually unilateral, but large central disk herniations can cause bilateral symptoms and signs and may produce a cauda equina syndrome.

Breakaway weakness describes a variable power of muscle contraction by a patient who is asked to provide maximal effort. The weakness may be due to pain or a combination of pain and underlying true weakness. Breakaway weakness without pain is due to lack of effort; patients who exhibit breakaway weakness should be asked if testing a specific muscle is painful. In uncertain cases,EMG can determine whether or not true weakness is present.

The differential diagnosis of lumbar disk disease includes a variety of serious and treatable conditions, including epidural abscess, hematoma, or tumor. Fever, constant

pain uninfluenced by position, sphincter abnormalities, or signs of spinal cord disease suggest an etiology other than lumbar disk disease. Bilateral absence of ankle reflexes can be a normal finding in old age or a sign of bilateral S1 radiculopathy. An absent deep tendon reflex or focal sensory loss may reflect injury to a nerve root, but other sites of injury along the nerve must also be considered. For example, an absent knee reflex may be due to a femoral neuropathy rather than an L4 nerve root injury. A focal decrease in sensation over the foot and distal lateral calf may result from a peroneal or lateral sciatic neuropathy rather than an L5 nerve root injury. Focal muscle atrophy may reflect loss of motor axons from a nerve root or peripheral nerve injury, an anterior horn cell disease, or disuse.

AnMRI scan orCT-myelogram is necessary to establish the location and type of pathology. Simple MRI yields exquisite views of intraspinal and adjacent soft tissue anatomy and is more likely to establish a specific anatomic diagnosis than plain films or myelography. Bony lesions of the lateral recess or intervertebral foramen may be seen with optimal clarity on CT-myelographic studies.

The correlation of neuroradiologic findings to symptoms, particularly pain, is often problematic. As examples, contrast-enhancing tears in the annulus fibrosus or disk protrusions are widely accepted as common sources of back pain. However, one recent study found that over half of asymptomatic adults have annular tears on lumbar spine MR imaging, nearly all of which demonstrate contrast enhancement. Furthermore, asymptomatic disk protrusions are common in adults, and many of these abnormalities enhance with contrast. These observations strongly suggest that MRI findings of disk protrusion, tears in the annulus fibrosus, or contrast enhancement are common incidental findings that by themselves should not dictate management decisions for patients with back pain. The presence or absence of persistent disk herniation 10 years after surgical or conservative treatment has no bearing on a successful clinical outcome.

There are four indications for intervertebral disk surgery: (1) progressive motor weakness from nerve root injury demonstrated on clinical examination orEMG, (2) bowel or bladder disturbance or other signs of spinal cord disease, (3) incapacitating nerve root pain despite conservative treatment for at least 4 weeks, and (4) recurrent incapacitating pain despite conservative treatment. The latter two criteria are more subjective and less well established than the others. Surgical treatment should also be considered if the pain and/or neurologic findings do not substantially improve over 4 to 12 weeks.

Surgery is preceded by MRI scan or CT-myelogram to define the location and type of pathology. The usual surgical procedure is a partial hemilaminectomy with excision of the involved and prolapsed intervertebral disk. Arthrodesis of the involved lumbar segments is considered only in the presence of significant spinal instability (i.e., degenerative spondylolisthesis or isthmic spondylolysis).

OTHER CAUSES OF LOW BACK PAIN

Spinal stenosis is an anatomic diagnosis reflecting a narrowed lumbar or cervical spinal canal. Classic *neurogenic claudication* occurs in the setting of moderate to severe spinal stenosis and typically consists of back and buttock or leg pain induced by walking or

standing. The pain is relieved by sitting. Symptoms in the legs are usually bilateral. Focal weakness, sensory loss, or reflex changes may occur when associated with radiculopathy. Unlike vascular claudication, the symptoms are often provoked by standing without walking. Unlike lumbar disk disease, the symptoms are usually relieved by sitting. Severe neurologic deficits, including paralysis and urinary incontinence, occur rarely. Spinal stenosis usually results from acquired (75%), congenital, or mixed acquired/congenital factors. Congenital forms (achondroplasia, idiopathic) are characterized by short, thick pedicles that produce both spinal canal and lateral recess stenosis. Acquired factors that may contribute to spinal stenosis include degenerative diseases (spondylosis, spondylolisthesis, scoliosis), trauma, spine surgery (postlaminectomy, fusion), metabolic or endocrine disorders (epidural lipomatosis, osteoporosis, acromegaly, renal osteodystrophy, hypoparathyroidism), and Paget's disease.MRI orCT-myelography provide the best definition of the abnormal anatomy (Fig. 16-5).

Conservative treatment includes nonsteroidal anti-inflammatory drugs (NSAIDs), exercise programs, and symptomatic treatment of acute pain exacerbations. Surgical therapy is considered when medical therapy does not relieve pain sufficiently to allow for activities of daily living or when significant focal neurologic signs are present. Between 65 and 80% of properly selected patients treated surgically experience >75% relief of back and leg pain. Up to 25% develop recurrent stenosis at the same spinal level or an adjacent level 5 years after the initial surgery; recurrent symptoms usually respond to a second surgical decompression.

Facet joint hypertrophy can produce unilateral radicular symptoms, due to bony compression, that are indistinguishable from disk-related radiculopathy. Patients may exhibit stretch signs, focal motor weakness, hyporeflexia, or sensory loss. Hypertrophic superior or inferior facets can often be visualized radiologically. Foraminotomy results in long-term relief of leg and back pain in 80 to 90% of patients.

Lumbar adhesive arachnoiditis with radiculopathy is the result of a fibrotic process following an inflammatory response to local tissue injury within the subarachnoid space. The fibrosis results in nerve root adhesions, producing back and leg pain associated with motor, sensory, and reflex changes. Myelography-induced arachnoiditis has become rare with the abandonment of oil-based contrast. Other causes of arachnoiditis include multiple lumbar operations, chronic spinal infections, spinal cord injury, intrathecal hemorrhage, intrathecal injection of steroids and anesthetics, and foreign bodies. The spineMRIappearance of arachnoiditis includes nerve roots clumping together centrally and adherent to the dura peripherally, or loculations of cerebrospinal fluid (CSF) within the thecal sac that obscure nerve root visualization. Treatment is often unsatisfactory. Microsurgical lysis of adhesions, dorsal rhizotomy, and dorsal root ganglionectomy have resulted in poor outcomes. Dorsal column stimulation for pain relief has produced varying results. Epidural steroid injections have been of limited value.

ARTHRITIS

Arthritis is a major cause of spine pain.

Spondylosis Osteoarthritic spine disease typically occurs in later life and primarily involves the cervical and lumbosacral spine. Patients often complain of back pain that is increased by motion and associated with stiffness or limitation of motion. The relationship between clinical symptoms and radiologic findings is usually not straightforward. Pain may be prominent when x-ray findings are minimal; alternatively, large osteophytes can be seen in asymptomatic patients in middle and later life. Hypertrophied facets and osteophytes may compress nerve roots in the lateral recess or intervertebral foramen. Osteophytes arising from the vertebral body may cause or contribute to central spinal canal stenosis. Loss of intervertebral disk height reduces the vertical dimensions of the intervertebral foramen; the descending pedicle may compress the nerve root exiting at that level. Osteoarthritic changes in the lumbar spine may rarely compress the cauda equina.

Ankylosing Spondylitis (See alsoChap. 315) This distinctive arthritic spine disease typically presents with the insidious onset of low back and buttock pain. Patients are often males below age 40. Associated features include morning back stiffness, nocturnal pain, pain unrelieved by rest, an elevated sedimentation rate, and the histocompatibility antigen HLA-B27. The differential diagnosis includes tumor and infection. Onset at a young age and back pain characteristically improving with exercise suggest ankylosing spondylitis. Loss of the normal lumbar lordosis and exaggeration of thoracic kyphosis are seen as the disease progresses. Inflammation and erosion of the outer fibers of the annulus fibrosus at the point of contact with the vertebral body are followed by ossification and bone growth. Bony growth (syndesmophyte) bridges adjacent vertebral bodies and results in reduced spine mobility in all planes. The radiologic hallmarks of the disease are periarticular destructive changes, sclerosis of the sacroiliac joints, and bridging of vertebral bodies by bone to produce the fused "bamboo spine." Similar restricted movement may accompany Reiter's syndrome, psoriatic arthritis, and chronic inflammatory bowel disease. Stress fractures through the spontaneously ankylosed posterior bony elements of the rigid, osteoporotic spine may result in focal spine pain, spinal cord compression or cauda equina syndrome. Occasional atlantoaxial subluxation with spinal cord compression occurs. Bilateral ankylosis of the ribs to the spine and a decrease in the height of axial thoracic structures may cause marked impairment of respiratory function.

OTHER DESTRUCTIVE DISEASES

Neoplasm (See alsoChap. 370) Back pain is the most common neurologic symptom among patients with systemic cancer. One-third of patients with undiagnosed back or neck pain and known systemic cancer have epidural extension or metastasis of tumor, and one-third have pain associated with vertebral metastases alone. About 11% have back pain unrelated to metastatic disease. Metastatic carcinoma (breast, lung, prostate, thyroid, kidney, gastrointestinal tract), multiple myeloma, and non-Hodgkin's and Hodgkin's lymphomas frequently involve the spine. Back pain may be the presenting symptom because the primary tumor site may be overlooked or asymptomatic. The pain tends to be constant, dull, unrelieved by rest, and worse at night. In contrast, mechanical low back pain is usually improved with rest. Plain x-rays usually, though not always, show destructive lesions in one or several vertebral bodies without disk space involvement.MRI orCT-myelography are the studies of choice in the setting of suspected spinal metastasis, but the trend of evidence favors the use of MRI. The procedure of

choice is the study most rapidly available because the patient may worsen during a diagnostic delay.

Infection *Vertebral osteomyelitis* is usually caused by staphylococci, but other bacteria or the tubercle bacillus (Pott's disease) may be the responsible organism. A primary source of infection, most often from the urinary tract, skin, or lungs, can be identified in 40% of patients. Intravenous drug use is a well-recognized risk factor. Back pain exacerbated by motion and unrelieved by rest, spine tenderness over the involved spine segment, and an elevated erythrocyte sedimentation rate are the most common findings. Fever or elevated white blood cell count are found in a minority of patients. Plain radiographs may show a narrowed disk space with erosion of adjacent vertebrae; these diagnostic changes may take weeks or months to appear.MRI andCT are sensitive and specific for osteomyelitis; MRI definition of soft tissue detail is exquisite. CT scan may be more readily available and better tolerated by some patients with severe back pain.

Spinal epidural abscess (Chap. 368) presents with back pain (aggravated by palpation or movement) and fever. The patient may exhibit nerve root injury or spinal cord compression accompanied by a sensory level, incontinence, or paraplegia. The abscess may track over multiple spinal levels and is best delineated by spineMRI.

Osteoporosis and Osteosclerosis Considerable loss of bone may occur with or without symptoms in association with medical disorders, including hyperparathyroidism, chronic glucocorticoid use, or immobilization. Compression fractures occur in up to half of patients with severe osteoporosis. The risk of osteoporotic vertebral fracture is 4.5 times greater over 3 years among patients with a baseline fracture compared with osteoporotic controls. The sole manifestation of a compression fracture may be focal lumbar or thoracic aching (often after a trivial injury) that is exacerbated by movement. Other patients experience thoracic or upper lumbar radicular pain. Focal spine tenderness is common. When compression fractures are found, treatable risk factors should be sought. Compression fractures above the midthoracic region suggest malignancy.

Osteosclerosis is readily identifiable on routine x-ray studies (e.g., Paget's disease) and may or may not produce back pain. Spinal cord or nerve root compression may result from bony encroachment on the spinal canal or intervertebral foramina. Single dual-beam photon absorptiometry or quantitativeCT can be used to detect small changes in bone mineral density.**For further discussion of these bone disorders, see Chaps. 341 to 343.*

REFERRED PAIN FROM VISCERAL DISEASE

Diseases of the pelvis, abdomen, or thorax may produce referred pain to the posterior portion of the spinal segment that innervates the diseased organ. Occasionally, back pain may be the first and only sign. In general, pelvic diseases refer pain to the sacral region, lower abdominal diseases to the lumbar region (around the second to fourth lumbar vertebrae), and upper abdominal diseases to the lower thoracic or upper lumbar region (eighth thoracic to the first and second lumbar vertebrae). Local signs (pain with spine palpation, paraspinal muscle spasm) are absent, and minimal or no pain

accompanies normal spine movements.

Low Thoracic and Upper Lumbar Pain in Abdominal Disease Peptic ulcer or tumor of the posterior stomach or duodenum typically produces epigastric pain (Chaps. 285 and90), but midline back or paraspinal pain may occur if retroperitoneal extension is present. Back pain due to peptic ulcer may be precipitated by ingestion of an orange, alcohol, or coffee and relieved by food or antacids. Fatty foods are more likely to induce back pain associated with biliary disease. Diseases of the pancreas may produce back pain to the right of the spine (head of the pancreas involved) or to the left (body or tail involved). Pathology in retroperitoneal structures (hemorrhage, tumors, pyelonephritis) may produce paraspinal pain with radiation to the lower abdomen, groin, or anterior thighs. A mass in the iliopsoas region often produces unilateral lumbar pain with radiation toward the groin, labia, or testicle. The sudden appearance of lumbar pain in a patient receiving anticoagulants suggests retroperitoneal hemorrhage.

Isolated low back pain occurs in 15 to 20% of patients with a contained rupture of an abdominal aortic aneurysm (AAA). The classic clinical triad of abdominal pain, shock, and back pain in an elderly man occurs in fewer than 20% of patients. Two of these three features are present in two-thirds of patients, and hypotension is present in half. Ruptured AAA has a high mortality rate; the typical patient is an elderly male smoker with back pain. The diagnosis is initially missed in at least one-third of patients because the symptoms and signs can be nonspecific. Common misdiagnoses include nonspecific back pain, diverticulitis, renal colic, sepsis, and myocardial infarction. A careful abdominal examination revealing a pulsatile mass (present in 50 to 75% of patients) is an important physical finding.

Lumbar Pain with Lower Abdominal Diseases Inflammatory bowel disorders (colitis, diverticulitis) or colonic neoplasms may produce lower abdominal pain, midlumbar back pain, or both. The pain may have a beltlike distribution around the body. A lesion in the transverse or initial descending colon may refer pain to the middle or left back at the L2-L3 level. Sigmoid colon disease may refer pain to the upper sacral or midline suprapubic regions or left lower quadrant of the abdomen.

Sacral Pain in Gynecologic and Urologic Disease Pelvic organs rarely cause low back pain, except for gynecologic disorders involving the uterosacral ligaments. The pain is referred to the sacral region. Endometriosis or uterine carcinoma may invade the uterosacral ligaments; malposition of the uterus may cause uterosacral ligament traction. The pain associated with endometriosis begins during the premenstrual phase and often continues until it merges with menstrual pain. Malposition of the uterus (retroversion, descensus, and prolapse) may lead to sacral pain after standing for several hours.

Menstrual pain may be felt in the sacral region. The poorly localized, cramping pain can radiate down the legs. Other pelvic sources of low back pain include neoplastic invasion of pelvic nerves, radiation necrosis, and pregnancy. Pain due to neoplastic infiltration of nerves is typically continuous, progressive in severity, and unrelieved by rest at night. Radiation therapy of pelvic tumors may produce sacral pain from late radiation necrosis of tissue or nerves. Low back pain with radiation into one or both thighs is common in the last weeks of pregnancy.

Urologic sources of lumbosacral back pain include chronic prostatitis, prostate carcinoma with spinal metastasis, and diseases of the kidney and ureter. Lesions of the bladder and testes do not usually produce back pain. The diagnosis of metastatic prostate carcinoma is established by rectal examination, spine imaging studies (MRI orCT), and measurement of prostate-specific antigen (PSA) (Chap. 95). Infectious, inflammatory, or neoplastic renal diseases may result in ipsilateral lumbosacral pain, as can renal artery or vein thrombosis. Ureteral obstruction due to renal stones may produce paraspinal lumbar pain.

Postural Back Pain There is a group of patients with chronic, nonspecific low back pain in whom no anatomic or pathologic lesion can be found despite exhaustive investigation. These individuals complain of vague, diffuse back pain with prolonged sitting or standing that is relieved by rest. The physical examination is unrevealing except for "poor posture." Imaging studies and laboratory evaluations are normal. Exercises to strengthen the paraspinal and abdominal muscles are sometimes therapeutic.

Psychiatric Disease Chronic low back pain (CLBP) may be encountered in patients with compensation hysteria, malingering, substance abuse, chronic anxiety states, or depression. Many patients with CLBP have a history of psychiatric illness (depression, anxiety, substance abuse) or childhood trauma (physical or sexual abuse) that antedates the onset of back pain. Preoperative psychological assessment has been used to exclude patients with marked psychological impairment who are at high risk for a poor surgical outcome. It is important to be certain that the back pain in these patients does not represent serious spine or visceral pathology in addition to the impaired psychological state.

Unidentified The cause of low back pain occasionally remains unclear. Some patients have had multiple operations for disk disease but have persistent pain and disability. The original indications for surgery may have been questionable with back pain only, no definite neurologic signs, or a minor disk bulge noted onCT orMRI. Scoring systems based upon neurologic signs, psychological factors, physiologic studies, and imaging studies have been devised to minimize the likelihood of unsuccessful surgical explorations and to avoid selection of patients with psychological profiles that predict poor functional outcomes.

TREATMENT

Acute Low Back Pain A practical approach to the management of low back pain is to consider acute and chronic presentations separately.ALBP is defined as pain of less than 3 months' duration. Full recovery can be expected in 85% of adults with ALBP unaccompanied by leg pain. Most of these patients exhibit "mechanical" symptoms -- pain that is aggravated by motion and relieved by rest.

Observational, population-based studies have been used to justify a minimalist approach to individual patient care. These studies share a number of limitations: (1) a true placebo control group is often lacking; (2) patients who consult different provider groups (generalists, orthopedists, neurologists) are assumed to have similar etiologies

for their back pain; (3) no information is provided about the details of treatment within each provider group or between provider groups; and (4) no attempt to tabulate serious causes ofALBP is made. The appropriateness of specific diagnostic procedures or therapeutic interventions for low back pain cannot be assessed from these studies.

The proposed algorithms (Fig. 16-6) for management ofALBP in adults draw considerably from published guidelines. However, it must be emphasized that currentCPGs for the treatment of low back pain are based on incomplete evidence -- for example, there is a paucity of well-designed studies documenting the natural history of disk lesions associated with a focal neurologic deficit. Guidelines should not substitute for sound clinical judgment.

The initial assessment excludes serious causes of spine pathology that require urgent intervention, including infection, cancer, and trauma. Risks factors for a possible serious underlying cause of back pain include: age > 50 years, prior diagnosis of cancer or other serious medical illness, bed rest without relief, duration of pain >1 month, urinary incontinence or recent nocturia, focal leg weakness or numbness, pain radiating into the leg(s) from the back, intravenous drug use, chronic infection (pulmonary or urinary), pain increasing with standing and relieved by sitting, history of spine trauma, and glucocorticoid use. Clinical signs associated with a possible serious etiology include unexplained fever, well-documented and unexplained weight loss, positiveSLR sign or reverse SLR sign, crossed SLR sign, percussion tenderness over the spine or costovertebral angle, an abdominal mass (pulsatile or nonpulsatile), a rectal mass, focal sensory loss (saddle anesthesia or focal limb sensory loss), true leg weakness, spasticity, and asymmetric leg reflexes. Laboratory studies are unnecessary unless a serious underlying cause (Fig. 16-6, Algorithms *A* and *B*) is suspected. Plain spine films are rarely indicated in the first month of symptoms unless a spine fracture is suspected.

The roles of bed rest, early exercise, and traction in the treatment of acute uncomplicated low back pain have been the subject of recent prospective studies. Clinical trials fail to demonstrate any benefit of prolonged (>2 days) bed rest forALBP. There is evidence that bed rest is also ineffective for patients with sciatica or for acute back pain with findings of nerve root injury. Theoretical advantages of early ambulation for ALBP include maintenance of cardiovascular conditioning, improved disk and cartilage nutrition, improved bone and muscle strength, and increased endorphin levels. A recent trial did not show benefit from an early vigorous exercise program, but the benefits of less vigorous exercise or other exercise programs remain unknown. The early resumption of normal physical activity (without heavy manual labor) is likely to be beneficial. Well-designed clinical studies of traction that include a sham traction group have failed to show a benefit of traction for ALBP. Despite this knowledge, one survey of physicians' perceptions of effective treatment identified strict bed rest for >3 days, trigger point injections (see below), and physical therapy (PT) as beneficial for more than 50% of patients with ALBP. In many instances, the behavior of treating physicians does not reflect the current medical literature.

Proof is lacking to support the treatment of acute back and neck pain with acupuncture, transcutaneous electrical nerve stimulation, massage, ultrasound, diathermy, or electrical stimulation. Cervical collars can be modestly helpful by limiting spontaneous and reflex neck movements that exacerbate pain. Evidence regarding the efficacy of ice

or heat is lacking, but these interventions are optional given the lack of negative evidence, low cost, and low risk. Biofeedback has not been studied rigorously. Facet joint, trigger point, and ligament injections are not recommended in the treatment ofALBP.

A beneficial role for specific exercises or modification of posture has not been validated by rigorous clinical studies. As a practical matter, temporary suspension of activity known to increase mechanical stress on the spine (heavy lifting, prolonged sitting, bending or twisting, straining at stool) may be helpful.

Patient education is an important part of treatment. Studies reveal that patient satisfaction and the likelihood of follow-up increase when patients are educated about prognosis, treatment methods, activity modifications, and strategies to prevent future exacerbations. In one study, patients who felt they did not receive an adequate explanation for their symptoms wanted more diagnostic tests. Evidence for the efficacy of structured education programs ("back school") is inconclusive; in one controlled study, patients attending back school had a shorter duration of sick leave during the initial episode but not during subsequent episodes. Recent large, controlled, randomized studies of back school for primary prevention of low back injury and pain have failed to demonstrate a benefit.

Medications used in the treatment ofALBPincludeNSAIDs, acetaminophen, muscle relaxants, and opioids. NSAIDs are superior to placebo for back pain relief. Acetaminophen is superior to placebo in the treatment of other types of pain but has not been compared against placebo for low back pain. Muscle relaxants provide short-term (4 to 7 days) benefit compared with placebo, but drowsiness often limits their daytime use. The efficacy of muscle relaxants compared to NSAIDs or in combination with NSAIDs is unclear. Opioid analgesics have not been shown to be more effective than NSAIDs or acetaminophen for relief of ALBP or likelihood of return to work. Short-term use of opioids in selected patients unresponsive to or intolerant of acetaminophen or NSAIDs may be helpful. There is no evidence to support the use of oral glucocorticoids or tricyclic antidepressants in treatment of ALBP.

The role of diagnostic and therapeutic nerve root blocks for patients with acute back or neck pain remains controversial. Equivocal data suggests that epidural steroids may occasionally produce short-term pain relief in patients withALBP and radiculopathy, but proof is lacking for pain relief beyond 1 month. Epidural anesthetics, steroids, or opioids are not indicated as initial treatment for ALBP without radiculopathy. Diagnostic selective nerve root blocks have been advocated to determine if pain originates from a nerve root. However, these studies may be falsely positive due to a placebo effect, in patients with a painful lesion located distally along the peripheral nerve, or from anesthesia of the sinuvertebral nerve. Therapeutic selective nerve root blocks are an option after brief conservative measures fail, particularly when temporary relief of pain may be important for patient function. Needle position is confirmed under fluoroscopic guidance with nonionic contrast before injection of glucocorticoid and local anesthetic.

A short course of spinal manipulation orPT for symptomatic relief of uncomplicatedALBP is an option. A prospective, randomized study comparing PT, chiropractic manipulation, and education interventions for patients with ALBP found

modest trends toward benefit with both PT and chiropractic manipulation at 1 year. Costs per year were equivalent in the PT/chiropractic group and ~$280 less for the group treated with the education booklet alone. The extent to which this modest improvement in symptoms and outcome is worth the cost must be determined for each patient. Extended duration of treatment or treatment of patients with radiculopathy is of unknown value and carries potential risk. The appropriate frequency or duration of spinal manipulation has not been addressed adequately.

Chronic Low Back Pain CLBP is defined as pain lasting longer than 12 weeks. Patients with CLBP account for 50% of back pain costs. Overweight individuals appear to be at particular risk. Other risk factors include: female gender, older age, prior history of back pain, restricted spinal mobility, pain radiating into a leg, high levels of psychological distress, poor self-rated health, minimal physical activity, smoking, job dissatisfaction, and widespread pain. Combinations of these premorbid factors have been used to predict which individuals withALBP are likely to develop CLBP. The initial approach to these patients is similar to that for ALBP, and the differential diagnosis of CLBP includes most of the conditions described in this chapter. Treatment of this heterogeneous group of patients is directed toward the underlying cause when possible; the ultimate goal is to restore function to the greatest extent possible.

Many conditions that produceCLBP can be identified by the combination of neuroimaging and electrophysiologic studies. SpineMRI orCT-myelography are the techniques of choice but are generally not indicated within the first month after initial evaluation in the absence of risk factors for a serious underlying cause. Imaging studies should be performed only in circumstances where the results are likely to influence surgical or medical treatment.

Diskography is of questionable value in the evaluation of back pain. No additional anatomic information is provided beyond what is available byMRI. Reproduction of the patient's typical pain with the injection is often used as evidence that a specific disk is the pain generator, but it is not known whether this information has any value in selecting candidates for surgery. There is no proven role for thermography in the assessment of radiculopathy.

The diagnosis of nerve root injury is most secure when the history, examination, results of imaging studies, and theEMG are concordant. The correlation betweenCT and EMG for localization of nerve root injury is between 65 and 73%. Up to one-third of asymptomatic adults have a disk protrusion detected by CT orMRIscans. Thus, surgical intervention based solely upon radiologic findings and pain increases the likelihood of an unsuccessful outcome.

CLBPcan be treated with a variety of conservative measures. Acute and subacute exacerbations are managed withNSAIDsand comfort measures. There is no good evidence to suggest that one NSAID is more effective than another. Bed rest should not exceed 2 days. Activity tolerance is the primary goal, while pain relief is secondary. Exercise programs can reverse type II muscle fiber atrophy in paraspinal muscles and strengthen trunk extension. Supervised, intensive physical exercise or "work hardening" regimens (under the guidance of a physical therapist) have been effective in returning some patients to work, improving walking distances, and diminishing pain. The benefit

can be sustained with home exercise regimens; compliance with the exercise regimen strongly influences outcome. The role of manipulation, back school, or epidural steroid injections in the treatment of CLBP is unclear. Up to 30% of "blind" epidural steroid injections miss the epidural space even when performed by an experienced anesthesiologist. There is no strong evidence to support the use of acupuncture or traction in this setting. A reduction in sick leave days, long-term health care utilization, and pension expenditures may offset the initial expense of multidisciplinary treatment programs. In one study comparing 3 weeks of hydrotherapy versus routine ambulatory care, hydrotherapy resulted in diminished duration and intensity of back pain, reduced analgesic drug consumption, improved spine mobility, and improved functional score. Functional score returned to baseline at the 9-month follow-up, but all other beneficial effects were sustained. Percutaneous electrical nerve stimulation (PENS) has been shown to provide significant short-term relief of CLBP, but additional studies regarding long-term efficacy and cost are necessary.

PAIN IN THE NECK AND SHOULDER

Approach to the Patient

In one recent epidemiologic survey, the 6-month prevalence of disabling neck pain was 4.6% among adults. Neck pain commonly arises from diseases of the cervical spine and soft tissues of the neck. Neck pain arising from the cervical spine is typically precipitated by neck movements and may be accompanied by focal spine tenderness and limitation of motion. Pain arising from the brachial plexus, shoulder, or peripheral nerves can be confused with cervical spine disease, but the history and examination usually identify a more distal origin for the pain. Cervical spine trauma, disk disease, or spondylosis may be asymptomatic or painful and can produce a myelopathy, radiculopathy, or both. The nerve roots most commonly affected are C7 and C6.

TRAUMA TO THE CERVICAL SPINE

Unlike injury to the low back, trauma to the cervical spine (fractures, subluxation) places the spinal cord at risk for compression. Motor vehicle accidents, violent crimes, or falls account for 87% of spinal cord injuries, which can have devastating consequences (Chap. 369). Emergency immobilization of the neck prior to complete assessment is mandatory to minimize further spinal cord injury from movement of unstable cervical spine segments.

Whiplash injury is due to trauma (usually automobile accidents) causing cervical musculoligamental sprain or strain due to hyperflexion or hyperextension. This diagnosis should not be applied to patients with fractures, disk herniation, head injury, or altered consciousness. One prospective study found that 18% of patients with whiplash injury had persistent injury-related symptoms 2 years after the car accident. Such patients were older, had a higher incidence of inclined or rotated head position at impact, greater intensity of initial neck and head pain, greater number of initial symptoms, and more osteoarthritic changes on cervical spine x-rays at baseline compared to patients who ultimately recovered. Objective data on the pathology of neck soft tissue injuries is lacking. Patients with severe initial injury are at increased risk for poor long-term outcome.

CERVICAL DISK DISEASE

Herniation of a lower cervical disk is a common cause of neck, shoulder, arm, or hand pain. Neck pain (worse with movement), stiffness, and limited range of neck motion are common. With nerve root compression, pain may radiate into a shoulder or arm. Extension and lateral rotation of the neck narrows the intervertebral foramen and may reproduce radicular symptoms (Spurling's sign). In young individuals, acute cervical nerve root compression from a ruptured disk is often due to trauma. Subacute radiculopathy is less likely to be related to a specific traumatic incident and may involve both disk disease and spondylosis. Cervical disk herniations are usually posterolateral near the lateral recess and intervertebral foramen. The usual patterns of reflex, sensory, and motor changes that accompany specific cervical nerve root lesions are listed inTable 16-2. When evaluating patients with suspected cervical radiculopathy it is important to consider the following: (1) overlap in function between adjacent nerve roots is common, (2) the anatomic pattern of pain is the most variable of the clinical features, and (3) the distribution of symptoms and signs may be evident in only part of the injured nerve root territory.

Surgical management of cervical herniated disks usually consists of an anterior approach with diskectomy followed by anterior interbody fusion. A simple posterior partial laminectomy with diskectomy is an alternative approach. The risk of subsequent radiculopathy or myelopathy at cervical segments adjacent to the fusion is 3% per year and 26% at 10 years. Although the risk is sometimes portrayed as a late complication of cervical surgery, it may also reflect the natural history of degenerative cervical spine disease in this subpopulation of patients.

CERVICAL SPONDYLOSIS

Osteoarthritis of the cervical spine may produce neck pain that radiates into the back of the head, shoulders, or arms. Arthritic or other pathologic conditions of the upper cervical spine may be the source of headaches in the posterior occipital region (supplied by the C2-C4 nerve roots). Cervical spondylosis with osteophyte formation in the lateral recess or hypertrophic facet joints may produce a monoradiculopathy (Fig. 16-7). Narrowing of the spinal canal by osteophytes, ossification of the posterior longitudinal ligament, or a large central disk may compress the cervical spinal cord. In some patients, a combination of radiculopathy and myelopathy occur. An electrical sensation elicited by neck flexion and radiating down the spine from the neck (Lhermitte's symptom) usually indicates cervical or upper thoracic (T1-T2) spinal cord involvement. When little or no neck pain accompanies the cord compression, the diagnosis may be confused with amyotrophic lateral sclerosis (Chap. 365), multiple sclerosis (Chap. 371), spinal cord tumors (Chap. 368), or syringomyelia (Chap. 368). The possibility of this treatable cervical spinal cord disease must be considered even when the patient presents with leg complaints only. Furthermore, lumbar radiculopathy or polyneuropathy may mask an associated cervical myelopathy.MRI orCT-myelography can define the anatomic abnormalities, andEMG and nerve conduction studies can quantify the severity and localize the levels of motor nerve root injury.

OTHER CAUSES OF NECK PAIN

Rheumatoid arthritis (RA) (Chap. 312) of the cervical apophyseal joints results in neck pain, stiffness, and limitation of motion. In typical cases with symmetric inflammatory polyarthritis, the diagnosis of RA is straightforward. In advanced RA, synovitis of the atlantoaxial joint (C1-C2;Fig. 16-2) may damage the transverse ligament of the atlas, producing forward displacement of the atlas on the axis (atlantoaxial subluxation). Radiologic evidence of atlantoaxial subluxation occurs in 30% of patients with RA. Not surprisingly, the degree of subluxation correlates with the severity of erosive disease. When subluxation is present, careful neurologic assessment is important to identify early signs of myelopathy. Occasional patients develop high spinal cord compression leading to quadriparesis, respiratory insufficiency, and death. Although low back pain is common among RA patients, the frequency of facet disease, fracture, and spondylolisthesis is no greater than among age- and sex-matched controls with mechanical low back pain.

Ankylosing spondylitis can cause neck pain and on occasion atlantoaxial subluxation; when spinal cord compression is present or threatened, surgical intervention is indicated. Herpes zoster produces neck and posterior occipital pain in a C2-C3 distribution prior to the outbreak of vesicles. Neoplasms metastatic to the cervical spine, infections (osteomyelitis and epidural abscess), and metabolic bone diseases may also be the cause of neck pain. Neck pain may also be referred from the heart in the setting of coronary artery ischemia (cervical angina syndrome).

THORACIC OUTLET

The thoracic outlet is an anatomic region containing the first rib, the subclavian artery and vein, the brachial plexus, the clavicle, and the lung apex. Injury to these structures may result in posture or task-related pain around the shoulder and supraclavicular region. There are at least three subtypes of thoracic outlet syndrome (TOS). *True neurogenic TOS* results from compression of the lower trunk of the brachial plexus by an anomalous band of tissue connecting an elongate transverse process at C7 with the first rib. Neurologic deficits include weakness of intrinsic muscles of the hand and diminished sensation on the palmar aspect of the fourth and fifth digits.EMG and nerve conduction studies confirm the diagnosis. Definitive treatment consists of surgical division of the anomalous band compressing either the lower trunk of the brachial plexus or ventral rami of the C8 or T1 nerve roots. The weakness and wasting of intrinsic hand muscles typically does not improve, but surgery halts the insidious progression of weakness. The *arterial TOS* results from compression of the subclavian artery by a cervical rib; the compression results in poststenotic dilatation of the artery and thrombus formation. Blood pressure is reduced in the affected limb, and signs of emboli may be present in the hand; neurologic signs are absent. Noninvasive ultrasound techniques confirm the diagnosis. Treatment is with thrombolysis or anticoagulation (with or without embolectomy) and surgical excision of the cervical rib compressing the subclavian artery or vein. The *disputed TOS* includes a large number of patients with chronic arm and shoulder pain of unclear cause. The lack of sensitive and specific findings on physical examination or laboratory markers for this condition frequently results in diagnostic uncertainty. The role of surgery in disputed TOS is controversial; conservative approaches often include multidisciplinary pain management. Treatment is often unsuccessful.

BRACHIAL PLEXUS AND NERVES

Pain from injury to the brachial plexus or arm peripheral nerves can occasionally be confused with pain of cervical spine origin. Neoplastic infiltration of the lower trunk of the brachial plexus may produce shoulder pain radiating down the arm, numbness of the fourth and fifth fingers, and weakness of intrinsic hand muscles innervated by the ulnar and median nerves. Postradiation fibrosis (breast carcinoma is the most common setting) or a Pancoast tumor of the lung (Chap. 88) may produce similar findings. A Horner's syndrome is present in two-thirds of patients with a Pancoast tumor. Suprascapular neuropathy may produce severe shoulder pain, weakness, and wasting of the supraspinatous and infraspinatous muscles. *Acute brachial neuritis* is often confused with radiculopathy. It consists of the acute onset of severe shoulder or scapular pain followed over days to weeks by weakness of the proximal arm and shoulder girdle muscles innervated by the upper or middle trunks or cords of the brachial plexus. The onset is often preceeded by an infection or immunization. Separation of this syndrome from cervical radiculopathy is important because slow, complete recovery of brachial neuritis occurs in 75% of patients after 2 years and in 89% after 3 years. Occasional cases of carpal tunnel syndrome produce pain and paresthesia extending into the forearm, arm, and shoulder resembling a C5 or C6 root lesion. Lesions of the radial or ulnar nerve can mimic a radiculopathy at C7 or C8, respectively.EMG and nerve conduction studies can accurately localize lesions to the nerve roots, brachial plexus, or nerves.**For further discussion of peripheral nerve disorders, see Chap. 377.*

SHOULDER

Pain in the shoulder region can be difficult to separate clearly from neck pain. If the symptoms and signs of radiculopathy are absent, then the differential diagnosis includes mechanical shoulder pain (tendonitis, bursitis, rotator cuff tear, dislocation, adhesive capsulitis, and cuff impingement under the acromion) and referred pain (subdiaphragmatic irritation, angina, Pancoast tumor). Mechanical pain is often worse at night, associated with local shoulder tenderness, and aggravated by abduction, internal rotation, or extension of the arm. The pain of shoulder disease may at times radiate into the arm or hand, but the sensory, motor, and reflex changes that indicate disease of the nerve roots, plexus, or peripheral nerves are absent.

TREATMENT

A paucity of well-designed clinical trials exists for the treatment of neck pain. Symptomatic treatment of neck pain can include the use of analgesic medications and/or a soft cervical collar. Current indications for cervical disk surgery are similar to those for lumbar disk surgery; because of the risk of spinal cord injury with cervical spine disease, an aggressive approach is generally indicated whenever spinal cord injury is threatened. Surgical management of cervical herniated disks usually consists of an anterior approach with diskectomy followed by anterior interbody fusion. A simple posterior partial laminectomy with diskectomy is an acceptable alternative approach. The cumulative risk of subsequent radiculopathy or myelopathy at cervical segments adjacent to the fusion is approximately 3% per year and 26% per decade. Although this

risk is sometimes portrayed as a late complication of surgery, it may also reflect the natural history of degenerative cervical spine disease. Nonprogressive cervical radiculopathy (associated with a focal neurologic deficit) due to a herniated cervical disk may be treated conservatively with a high rate of success. Cervical spondylosis with bony, compressive cervical radiculopathy is generally treated with surgical decompression to interrupt the progression of neurologic signs. Cervical spondylotic myelopathy is typically managed with either anterior decompression and fusion or laminectomy. Outcomes in both surgical groups vary, but late functional deterioration occurs in 20 to 30% of patients; a prospective, controlled study comparing different surgical interventions is sorely needed.

(Bibliography omitted in Palm version)

Back to Table of Contents

SECTION 2 -ALTERATIONS IN BODY TEMPERATURE

17. FEVER AND HYPERTHERMIA - *Charles A. Dinarello*, *Jeffrey A. Gelfand*

Body temperature is controlled by the hypothalamus. Neurons in both the preoptic anterior hypothalamus and the posterior hypothalamus receive two kinds of signals: one from peripheral nerves that reflect warmth/cold receptors and the other from the temperature of the blood bathing the region. These two types of signals are integrated by the thermoregulatory center of the hypothalamus to maintain normal temperature. In a neutral environment, the metabolic rate of humans consistently produces more heat than is necessary to maintain the core body temperature at 37°C. Therefore, the hypothalamus controls temperature by mechanisms of heat loss.

A normal body temperature is ordinarily maintained, despite environmental variations, because the hypothalamic thermoregulatory center balances the excess heat production derived from metabolic activity in muscle and the liver with heat dissipation from the skin and lungs. According to recent studies of healthy individuals 18 to 40 years of age, the mean oral temperature is 36.8°± 0.4°C (98.2° ±0.7°F), with low levels at 6 A.M. and higher levels at 4 to 6 P.M. The maximum normal oral temperature is 37.2°C (98.9°F) at 6 A.M. and 37.7°C (99.9°F) at 4 P.M.; these values define the 99th percentile for healthy individuals. In light of these studies, *an A.M. temperature of>37.2°C (98.9°F) or a P.M. temperature of>37.7°C (99.9°F) would define a fever*. The normal daily temperature variation is typically 0.5°C (0.9°F). However, in some individuals recovering from a febrile illness, this daily variation can be as great as 1.0°C. During a febrile illness, diurnal variations are usually maintained but at higher levels. Daily temperature swings do not occur in patients with hyperthermia (see below). Rectal temperatures are generally 0.4°C (0.7°F) higher than oral readings. The lower oral readings are probably attributable to mouth breathing, which is a particularly important factor in patients with respiratory infections and rapid breathing. Lower esophageal temperatures closely reflect core temperature. Tympanic membrane (TM) thermometers measure radiant heat energy from the tympanic membrane and nearby ear canal and display that absolute value (unadjusted mode) or a value automatically calculated from the absolute reading on the basis of nomograms relating the radiant temperature measured to actual core temperatures obtained in clinical studies (adjusted mode). These measurements, although convenient, may be more variable than directly determined oral or rectal values. Studies in adults show that readings are lower with unadjusted-mode than with adjusted-mode TM thermometers and that unadjusted-mode TM values are 0.8°C (1.6°F) lower than rectal temperatures.

In women who menstruate, the A.M. temperature is generally lower in the 2 weeks before ovulation; it then rises by about 0.6°C (1°F) with ovulation and remains at that level until menses occur. Seasonal variation in body temperature has been described but may reflect a metabolic change and is not common. Body temperature is elevated in the postprandial state, but this elevation does not represent fever. Pregnancy and endocrinologic dysfunction also affect body temperature. The daily temperature variation appears to be fixed in early childhood; in contrast, elderly individuals can exhibit a reduced ability to develop fever, with only a modest fever even in severe infections.

FEVER VERSUS HYPERTHERMIA

FEVER

Fever is an elevation of body temperature that exceeds the normal daily variation and occurs *in conjunction with an increase in the hypothalamic set point* -- for example, from 37°C to 39°C. This shift of the set point from "normothermic" to febrile levels very much resembles the resetting of the home thermostat to a higher level in order to raise the ambient temperature in a room. Once the hypothalamic set point is raised, neurons in the vasomotor center are activated and vasoconstriction commences. The individual first notices vasoconstriction in the hands and feet. Shunting of blood away from the periphery to the internal organs essentially decreases heat loss from the skin, and the person feels cold. For most fevers, body temperature increases by 1 to 2°C. Shivering, which increases heat production from the muscles, may begin at this time; however, shivering is not required if heat conservation mechanisms raise blood temperature sufficiently. Heat production from the liver also increases. In humans, behavioral instincts (e.g., putting on more clothing or bedding) lead to a reduction of exposed surfaces, which helps raise body temperature.

The processes of heat conservation (vasoconstriction) and heat production (shivering and increased metabolic activity) continue until the temperature of the blood bathing the hypothalamic neurons matches the new thermostat setting. Once that point is reached, the hypothalamus maintains the temperature at the febrile level by the same mechanisms of heat balance that are operative in the afebrile state. When the hypothalamic set point is again reset downward (due to either a reduction in the concentration of pyrogens or the use of antipyretics), the processes of heat loss through vasodilation and sweating are initiated. Behavioral changes triggered at this time include the removal of insulating clothing or bedding. Loss of heat by sweating and vasodilation continues until the blood temperature at the hypothalamic level matches the lower setting.

A fever of>41.5°C (106.7°F) is called *hyperpyrexia*. This extraordinarily high fever can develop in patients with severe infections but most commonly occurs in patients with central nervous system hemorrhages. In the preantibiotic era, fever due to a variety of infectious diseases rarely exceeded 106°F, and there has been speculation that this natural "thermal ceiling" is mediated by neuropeptides functioning as central antipyretics.

In some rare cases, the hypothalamic set point is elevated as a result of local trauma, hemorrhage, tumor, or intrinsic hypothalamic malfunction. The term *hypothalamic fever* is sometimes used to describe elevated temperature caused by abnormal hypothalamic function. However, most patients with hypothalamic damage have *sub*normal, not *supra*normal, body temperatures. These patients do not respond properly to mild environmental temperature changes. For example, when exposed to only mildly cold conditions, their core temperature falls quickly rather than over the normal period of a few hours. In the very few patients in whom elevated core temperature is suspected to be due to hypothalamic damage, diagnosis depends on the demonstration of other abnormalities in hypothalamic function, such as the production of hypothalamic releasing factors, abnormal response to cold, and absence of circadian temperature and

hormonal rhythms.

HYPERTHERMIA

Hyperthermia is characterized by *an unchanged (normothermic) setting of the thermoregulatory center* in conjunction with an uncontrolled increase in body temperature that exceeds the body's ability to lose heat. Exogenous heat exposure and endogenous heat production are two mechanisms by which hyperthermia can result in dangerously high internal temperatures. Excessive heat production can easily cause hyperthermia despite physiologic and behavioral control of body temperature. For example, over-insulating clothing can result in an elevated core temperature, and work or exercise in hot environments can produce heat faster than peripheral mechanisms can lose it.

Although most patients with elevated body temperature have fever, there are a few circumstances in which elevated temperature represents not fever but hyperthermia (Table 17-1). *Heat stroke*, caused by thermoregulatory failure in association with a warm environment, may be categorized as exertional or nonexertional. *Exertional heat stroke* typically occurs in younger individuals exercising at ambient temperatures and/or humidities that are higher than normal. Even in normal individuals, dehydration or the use of common medications (e.g., over-the-counter antihistamines with anticholinergic side effects) may help to precipitate exertional heat stroke. *Nonexertional* or *classic heat stroke* typically occurs in elderly individuals, particularly during heat waves. For example, in Chicago in July 1995, 465 deaths were certified as heat related. The elderly, the bedridden, persons taking anticholinergic or antiparkinsonian drugs or diuretics, and individuals confined to poorly ventilated and non-air-conditioned environments are most susceptible.

Drug-induced hyperthermia has become increasingly common as a result of the increased use of prescription psychotropic drugs and illicit drugs. Drug-induced hyperthermia may be caused by monoamine oxidase inhibitors, tricyclic antidepressants, and amphetamines and by the illicit use of phencyclidine, lysergic acid diethylamide (LSD), or cocaine.

Malignant hyperthermia occurs in individuals with an inherited abnormality of skeletal-muscle sarcoplasmic reticulum that causes a rapid increase in intracellular calcium levels in response to halothane and other inhalational anesthetics or to succinylcholine. Elevated temperature, increased muscle metabolism, rigidity, rhabdomyolysis, acidosis, and cardiovascular instability develop rapidly. This condition is often fatal. The *neuroleptic malignant syndrome* can occur with phenothiazines and other drugs such as haloperidol and is characterized by muscle rigidity, autonomic dysregulation, and hyperthermia. This disorder appears to be caused by the inhibition of central dopamine receptors in the hypothalamus, which results in increased heat generation and decreased heat dissipation. Thyrotoxicosis and pheochromocytoma can also cause increased thermogenesis.

It is important to distinguish between fever and hyperthermia since hyperthermia can be rapidly fatal and characteristically does not respond to antipyretics. However, there is no rapid way to make this distinction. Hyperthermia is often diagnosed on the basis of the

events immediately preceding the elevation of core temperature -- e.g., heat exposure or treatment with drugs that interfere with thermoregulation. However, in addition to the clinical history of the patient, the physical aspects of some forms of hyperthermia may alert the clinician. For example, in patients with heat stroke syndromes and in those taking drugs that block sweating, the skin is hot but dry. Moreover, antipyretics do not reduce the elevated temperature in hyperthermia, whereas in fever -- and even in hyperpyrexia -- adequate doses of either aspirin or acetaminophen usually result in some decrease in body temperature.

PYROGENS

The term *pyrogen* is used to describe any substance that causes fever. *Exogenous* pyrogens are derived from outside the patient; most are microbial products, microbial toxins, or whole microorganisms. The classic example of an exogenous pyrogen is the lipopolysaccharide endotoxin produced by all gram-negative bacteria. Endotoxins are potent not only as pyrogens but also as inducers of various pathologic changes in gram-negative infections. Another group of potent bacterial pyrogens is produced by gram-positive organisms and includes the enterotoxins of *Staphylococcus aureus* and the group A and B streptococcal toxins, also called *superantigens*. One staphylococcal toxin of clinical importance is the toxic shock syndrome toxin associated with isolates of *S. aureus* from patients with toxic shock syndrome. Like the endotoxins of gram-negative bacteria, the toxins produced by staphylococci and streptococci cause fever in experimental animals when injected intravenously at concentrations of <1 ug/kg of body weight. Endotoxin is a highly pyrogenic molecule in humans: a dose of 2 to 3 ng/kg produces fever and generalized symptoms of malaise in volunteers.

PYROGENIC CYTOKINES

Cytokines are small proteins (molecular mass, 10,000 to 20,000 Da) that regulate immune, inflammatory, and hematopoietic processes. For example, stimulation of lymphocyte proliferation during an immune response to vaccination is the result of the cytokines interleukin (IL) 2, IL-4, and IL-6. Another cytokine, granulocyte colony-stimulating factor, stimulates granulocytopoiesis in the bone marrow. Some cytokines cause fever and hence are called *pyrogenic cytokines*. From a historic point of view, the field of cytokine biology began in the 1940s with laboratory investigations into fever induction by products of activated leukocytes. These fever-producing molecules were called *endogenous pyrogens*. When endogenous pyrogens were purified from activated leukocytes, they were shown to possess various biologic activities, which are now recognized as the properties of the various cytokines.

The known pyrogenic cytokines includeIL-1, IL-6, tumor necrosis factor (TNF), ciliary neurotropic factor (CNTF), and interferon (IFN) a. Others probably exist. Each cytokine is encoded by a separate gene, and each pyrogenic cytokine has been shown to cause fever in laboratory animals and in humans. When injected into humans, IL-1, IL-6, and TNF produce fever at low doses (10 to 100 ng/kg).

The synthesis and release of endogenous pyrogenic cytokines are induced by a wide spectrum of exogenous pyrogens, most of which have recognizable bacterial or fungal sources. Viruses also induce pyrogenic cytokines by infecting cells. However, in the

absence of microbial infection, inflammation, trauma, tissue necrosis, or antigen-antibody complexes can induce the production ofIL-1,TNF, and/or IL-6, which -- individually or in combination -- trigger the hypothalamus to raise the set point to febrile levels. The cellular sources of pyrogenic cytokines are primarily monocytes, neutrophils, and lymphocytes, although many other types of cells can synthesize these molecules when stimulated.

ELEVATION OF THE HYPOTHALAMIC SET POINT BY CYTOKINES

During fever, levels of prostaglandin E_2(PGE_2) are elevated in hypothalamic tissue and the third cerebral ventricle. The concentrations of PGE_2 are highest near the circumventricular vascular organs (organum vasculosum of lamina terminalis) -- networks of enlarged capillaries surrounding the hypothalamic regulatory centers. Destruction of these organs reduces the ability of pyrogens to produce fever. Most studies in animals have failed to show, however, that pyrogenic cytokines pass from the circulation into the brain itself. Thus, it appears that both exogenous and endogenous pyrogens interact with the endothelium of these capillaries and that this interaction is the first step in initiating fever -- i.e., in raising the set point to febrile levels.

The key events in the production of fever are illustrated inFig. 17-1. As has been mentioned, several cell types can produce pyrogenic cytokines. Pyrogenic cytokines such asIL-1, IL-6, andTNF are released from the cells and enter the systemic circulation. Although the systemic effects of these circulating cytokines lead to fever by inducing the synthesis ofPGE_2, they also induce PGE_2 in peripheral tissues. The increase in PGE_2 in the periphery accounts for the nonspecific myalgias and arthralgias that often accompany fever. However, it is the induction of PGE_2 in the brain that starts the process of raising the hypothalamic set point for core temperature.

There are four receptors forPGE_2, and each signals the cell in different ways. Of the four receptors, the third (EP-3) is essential for fever: when the gene for this receptor is deleted in mice, no fever follows the injection ofIL-1 or endotoxin. Deletion of the other PGE_2receptor genes leaves the fever mechanism intact. Although PGE_2 is essential for fever, it is not a neurotransmitter. Rather, the release of $PGE_2$from the brain side of the hypothalamic endothelium triggers the PGE_2receptor on glial cells, and this stimulation results in the rapid release of cyclic adenosine 5¢-monophosphate (cyclic AMP), which is a neurotransmitter. As shown inFig. 17-1, the release of cyclic AMP from the glial cells activates neuronal endings from the thermoregulatory center that extend into the area. The elevation of cyclic AMP is thought to account for changes in the hypothalamic set point either directly or indirectly by inducing the release of monoamine neurotransmitters. Since receptors for endotoxin are in many ways similar to IL-1 receptors, the activation of endotoxin receptors on the hypothalamic endothelium also results in PGE_2production and fever.

PRODUCTION OF CYTOKINES IN THE CENTRAL NERVOUS SYSTEM

Several viral diseases produce active infection in the brain. Glial and possibly neuronal cells synthesizeIL-1,TNF, and IL-6.CNTF is also synthesized by neural as well as neuronal cells. What role in the production of fever is played by these cytokines produced in the brain itself? In experimental animals, the concentrations of cytokine

required to cause fever are several orders of magnitude lower with direct injection into the brain than with intravenous injection. Therefore, central nervous system production of these cytokines apparently can raise the hypothalamic set point, bypassing the circumventricular organs involved in fever caused by circulating cytokines. Central nervous system cytokines may account for the hyperpyrexia of central nervous system hemorrhage, trauma, or infection.

Approach to the Patient

History It is in the diagnosis of a febrile illness that the science and art of medicine come together. In no other clinical situation is a meticulous history more important. Painstaking attention must be paid to the chronology of symptoms in relation to the use of prescription drugs (including drugs or herbs taken without a physician's supervision) or treatments such as surgical or dental procedures. The exact nature of any prosthetic materials and/or implanted devices should be ascertained. A careful occupational history should include exposures to animals; toxic fumes; potential infectious agents; possible antigens; or other febrile or infected individuals in the home, workplace, or school. A history of the geographic areas in which the patient has lived and a travel history should include locations during military service. Information on unusual hobbies, dietary proclivities (such as raw or poorly cooked meat, raw fish, and unpasteurized milk or cheeses), and household pets should be elicited, as should that on sexual orientation and practices, including precautions taken or omitted. Attention should be directed to the use of tobacco, marijuana, intravenous drugs, or alcohol; trauma; animal bites; tick or other insect bites; and prior transfusions, immunizations, drug allergies, or hypersensitivities. A careful family history should include information on family members with tuberculosis, other febrile or infectious diseases, arthritis or collagen vascular disease, or unusual familial symptomatology such as deafness, urticaria, fevers and polyserositis, bone pain, or anemia. Ethnic origin may be critical. For example, blacks are more likely than persons in other groups to have hemoglobinopathies. Turks, Arabs, Armenians, and Sephardic Jews are especially likely to have familial Mediterranean fever.

Physical Examination A meticulous physical examination should be repeated on a regular basis. All the vital signs are relevant. The temperature may be taken orally or rectally, but the site used should be consistent. Axillary temperatures are notoriously unreliable. Particular attention should be paid to daily (or sometimes more frequent) physical examination, which should continue until the diagnosis is certain and the anticipated response has been achieved. Special attention should be paid to the skin, lymph nodes, eyes, nail beds, cardiovascular system, chest, abdomen, musculoskeletal system, and nervous system. Rectal examination is imperative. The penis, prostate, scrotum, and testes should be examined carefully and the foreskin, if present, retracted. Pelvic examination must be part of every complete physical examination of a woman, with a search for such causes of fever as pelvic inflammatory disease and tubo-ovarian abscess.

Laboratory Tests Few signs and symptoms in medicine have as many diagnostic possibilities as fever. If the history, epidemiologic situation, or physical examination suggests more than a simple viral illness or streptococcal pharyngitis, then laboratory testing is indicated. The tempo and complexity of the workup will depend on the pace of

the illness, diagnostic considerations, and the immune status of the host. If findings are focal or if the history, epidemiologic setting, or physical examination suggests certain diagnoses, the laboratory examination can be focused. If fever is undifferentiated, the diagnostic nets must be cast farther, and certain guidelines are indicated, as follows.

CLINICAL PATHOLOGY The workup should include a complete blood count; a differential count should be performed manually or with an instrument sensitive to the identification of eosinophils, juvenile or band forms, toxic granulations, and Dohle bodies, the last three of which are suggestive of bacterial infection. Neutropenia may be present with some viral infections, particularly parvovirus B19 infection; drug reactions; systemic lupus erythematosus; typhoid; brucellosis; and infiltrative diseases of the bone marrow, including lymphoma, leukemia, tuberculosis, and histoplasmosis. Lymphocytosis may occur with typhoid, brucellosis, tuberculosis, and viral disease. Atypical lymphocytes are documented in many viral diseases, including infection with Epstein-Barr virus, cytomegalovirus, or HIV; dengue; rubella; varicella; measles; and viral hepatitis. This abnormality also occurs in serum sickness and toxoplasmosis. Monocytosis is a feature of typhoid, tuberculosis, brucellosis, and lymphoma. Eosinophilia may be associated with hypersensitivity drug reactions, Hodgkin's disease, adrenal insufficiency, and certain metazoan infections. If the febrile illness appears to be severe or is prolonged, the smear should be examined carefully for malarial or babesial pathogens (where appropriate) as well as for classic morphologic features, and the erythrocyte sedimentation rate should be determined. Urinalysis, with examination of urinary sediment, is indicated. It is axiomatic that any abnormal fluid accumulation (pleural, peritoneal, joint), even if previously sampled, merits reexamination in the presence of undiagnosed fever. Joint fluids should be examined for bacteria as well as crystals. Bone marrow biopsy (not simple aspiration) for histopathologic studies (as well as culture) is indicated when marrow infiltration by pathogens or tumor cells is possible. Stool should be inspected for occult blood; an inspection for fecal leukocytes, ova, or parasites also may be indicated.

CHEMISTRY Electrolyte, glucose, blood urea nitrogen, and creatinine levels should be measured. Liver function tests are usually indicated if efforts to identify the cause of fever do not point to the involvement of another organ. Additional assessments (e.g., measurement of creatinine phosphokinase or amylase) can be added as the workup progresses.

MICROBIOLOGY Smears and cultures of specimens from the throat, urethra, anus, cervix, and vagina should be assessed when there are no localizing findings or when findings suggest the involvement of the pelvis or the gastrointestinal tract. If respiratory tract infection is suspected, sputum evaluation (Gram's staining, staining for acid-fast bacilli, culture) is indicated. Cultures of blood, abnormal fluid collections, and urine are indicated when fever is thought to reflect more than uncomplicated viral illness. Cerebrospinal fluid should be examined and cultured if meningismus, severe headache, or a change in mental status is noted.

RADIOLOGY A chest x-ray is usually part of the evaluation for any significant febrile illness.

Outcome of Diagnostic Efforts In most cases of fever, either the patient recovers

spontaneously or the history, physical examination, and initial screening laboratory studies lead to a diagnosis. When fever continues for 2 to 3 weeks, during which time repeat physical examinations and laboratory tests are unrevealing, the patient is provisionally diagnosed as having fever of unknown origin (Chap. 125).

TREATMENT

The Decision to Treat Fever Most fevers are associated with self-limited infections, most commonly of viral origin. In these cases, the general cause of the fever is easily identified. The routine use of antipyretics given automatically as "standing," "routine," or "prn" orders to treat low-grade fevers in adult patients on hospital wards is entirely unacceptable. This practice masks not only fever but also other important clinical indicators of a patient's course. The assumption underlying any decision to reduce fever with antipyretics is that there is no diagnostic benefit to be gained by allowing the fever to persist. However, there may be such a diagnostic benefit. For example, the daily highs and lows of normal temperature are exaggerated in most fevers, but the usual times of peak and trough temperatures may be reversed in typhoid fever and disseminated tuberculosis. Temperature-pulse dissociation (relative bradycardia) occurs in typhoid fever, brucellosis, leptospirosis, some drug-induced fevers, and factitious fever. In newborns, the elderly, patients with chronic renal failure, and patients taking glucocorticoids, fever may not be present despite infection, or core temperature may be hypothermic. Hypothermia is observed in patients with septic shock.

Some febrile diseases have characteristic patterns. With *relapsing* fevers, febrile episodes are separated by intervals of normal temperature; when paroxysms occur on the first and third days, the fever is called *tertian*. *Plasmodium vivax* causes tertian fevers. *Quartan* fevers are associated with paroxysms on the first and fourth days and are seen with *P. malariae*. Other relapsing fevers are related to *Borrelia* infections and rat-bite fever, which are both associated with days of fever followed by a several-day afebrile period and then a relapse of days of fever. Pel-Ebstein fever, with fevers lasting 3 to 10 days followed by afebrile periods of 3 to 10 days, is classic for Hodgkin's disease and other lymphomas. Another characteristic fever is that of cyclic neutropenia, in which fevers occur every 21 days and accompany the neutropenia. There is no periodicity of fever in patients with familial Mediterranean fever.

Mechanisms of Antipyretic Agents The synthesis of PGE_2 depends on the constitutively expressed enzyme cyclooxygenase. The substrate for cyclooxygenase is arachidonic acid released from the cell membrane, and this release is the rate-limiting step in the synthesis of PGE_2. Inhibitors of cyclooxygenase are potent antipyretics. The antipyretic potency of various drugs is directly correlated with the inhibition of brain cyclooxygenase. Acetaminophen is a poor cyclooxygenase inhibitor in peripheral tissue and is without noteworthy anti-inflammatory activity; in the brain, however, acetaminophen is oxidized by the p450 cytochrome system, and the oxidized form inhibits cyclooxygenase activity.

Oral aspirin and acetaminophen are equally effective in reducing fever in humans. Nonsteroidal anti-inflammatory agents (NSAIDs) such as indomethacin and ibuprofen are also excellent antipyretics. Chronic high-dose therapy with antipyretics such as aspirin or the NSAIDs used in arthritis does not reduce normal core body temperature.

Thus, PGE_2 appears to play no role in normal thermoregulation.

As effective antipyretics, glucocorticoids act at two levels. First, similar to the cyclooxygenase inhibitors, glucocorticoids reduce PGE_2 synthesis by inhibiting the activity of phospholipase A_2, which is needed to release arachidonic acid from the cell membrane. Second, glucocorticoids block the transcription of the mRNA for the pyrogenic cytokines.

Drugs that interfere with vasoconstriction (phenothiazines, for example) can act as antipyretics, as can drugs that block muscle contractions. However, these agents are not true antipyretics since they can also reduce core temperature independently of hypothalamic control.

Indications and Regimens for the Treatment of Fever The objectives in treating fever are first to reduce the elevated hypothalamic set point and second to facilitate heat loss. There is no evidence that fever itself facilitates the recovery from infection or acts as an adjuvant to the immune system. In fact, peripheral PGE_2 production is a potent immunosuppressant. Hence, treating fever and its symptoms does no harm and does not slow the resolution of common viral and bacterial infections. Reducing fever with antipyretics also reduces systemic symptoms of headache, myalgias, and arthralgias.

Oral aspirin and NSAIDs effectively reduce fever but can adversely affect platelets and the gastrointestinal tract. Therefore, acetaminophen is preferred to all of these agents as an antipyretic. In children, acetaminophen must be used because aspirin increases the risk of Reye's syndrome. If the patient cannot take oral antipyretics, parenteral preparations of NSAIDs and rectal suppository preparations of various antipyretics can be used.

Treatment of fever in some patient groups is recommended. Fever increases the demand for oxygen (i.e., for every increase of 1°C over 37°C, there is a 13% increase in oxygen consumption) and can aggravate preexisting cardiac, cerebrovascular, or pulmonary insufficiency. Elevated temperature can induce mental changes in patients with organic brain disease. Children with a history of febrile or nonfebrile seizure should be aggressively treated to reduce fever, although it is unclear what triggers the febrile seizure and there is no correlation between absolute temperature elevation and onset of a febrile seizure in susceptible children.

In hyperpyrexia, the use of cooling blankets facilitates the reduction of temperature; however, cooling blankets should not be used without oral antipyretics. In hyperpyretic patients with central nervous system disease or trauma, reducing core temperature mitigates the ill effects of high temperature on the brain.

Treating Hyperthermia A high core temperature in a patient with an appropriate history (e.g., environmental heat exposure or treatment with anticholinergic or neuroleptic drugs, tricyclic antidepressants, succinylcholine, or halothane) along with appropriate clinical findings (dry skin, hallucinations, delirium, pupil dilation, muscle rigidity, and/or elevated levels of creatine phosphokinase) suggests hyperthermia. The attempt to lower the already normal hypothalamic set point is of little use. Physical cooling with sponging, fans, cooling blankets, and even ice baths should be initiated immediately in conjunction

with the administration of intravenous fluids and appropriate pharmacologic agents (see below). If insufficient cooling is achieved by external means, internal cooling can be achieved by gastric or peritoneal lavage with iced saline. In extreme circumstances, hemodialysis or even cardiopulmonary bypass with cooling of blood may be performed.

Malignant hyperthermia should be treated immediately with cessation of anesthesia and intravenous administration of dantrolene sodium. The recommended dose of dantrolene is 1 to 2.5 mg/kg of body weight given intravenously every 6 h for at least 24 to 48 h -- until oral dantrolene can be administered, if needed. Procainamide should also be administered to patients with malignant hyperthermia because of the likelihood of ventricular fibrillation in this syndrome. Dantrolene at similar doses is indicated in the neuroleptic malignant syndrome and in drug-induced hyperthermia and may even be useful in the hyperthermia of thyrotoxicosis. The neuroleptic malignant syndrome may also be treated with bromocriptine, levodopa, amantadine, or nifedipine or by induction of muscle paralysis with curare and pancuronium. Tricyclic antidepressant overdose may be treated with physostigmine.

(Bibliography omitted in Palm version)

Back to Table of Contents

18. FEVER AND RASH - *Elaine T. Kaye, Kenneth M. Kaye*

The acutely ill patient with fever and rash (Fig. 18-CD1) often presents a diagnostic challenge for physicians. The distinctive appearance of an eruption in concert with a clinical syndrome may facilitate a prompt diagnosis and the institution of life-saving therapy or critical infection-control interventions.

Approach to the Patient

A thorough history of patients with fever and rash includes the following relevant information: immune status, medications taken within the previous month, specific travel history, immunization status, exposure to domestic pets and other animals, history of animal or arthropod bites, existence of cardiac abnormalities, presence of prosthetic material, recent exposure to ill individuals, and exposure to sexually transmitted diseases. The history should also include the site of onset of the rash and its direction and rate of spread.

A thorough physical examination entails close attention to the rash, with an assessment and precise definition of its salient features. First, it is critical to determine the *type* of lesions that make up the eruption. *Macules* are flat lesions defined by an area of changed color (i.e., a blanchable erythema). *Papules* are raised, solid lesions <5 mm in diameter; *plaques* are lesions >5 mm in diameter with a flat, plateau-like surface; and *nodules* are lesions >5 mm in diameter with a more rounded configuration. *Wheals* (urticaria, hives) are papules or plaques that are pale pink and may appear annular (ringlike) as they enlarge; classic (nonvasculitic) wheals are transient, lasting only 24 to 48 h in any defined area. *Vesicles* (<5 mm) and *bullae* (>5 mm) are circumscribed, elevated lesions containing fluid. *Pustules* are raised lesions containing purulent exudate; vesicular processes such as varicella or herpes simplex may evolve to pustules. *Nonpalpable purpura* is a flat lesion that is due to bleeding into the skin; if <3 mm in diameter, the purpuric lesions are termed *petechiae*; if >3 mm, they are termed *ecchymoses*. *Palpable purpura* is a raised lesion that is due to inflammation of the vessel wall (vasculitis) with subsequent hemorrhage. An *ulcer* is a defect in the skin extending at least into the upper layer of the dermis, and an *eschar* (tache noire) is a necrotic lesion covered with a black crust.

Other pertinent features of rashes include their *configuration* (i.e., annular or target), the *arrangement* of their lesions, and their *distribution* (i.e., central or peripheral).*For further discussion, see Chaps. 55 and 57.*

CLASSIFICATION OF RASH

This chapter reviews rashes that reflect systemic disease but does not include localized skin eruptions (i.e., cellulitis, impetigo) that may also be associated with fever (Chap. 128). Rashes are classified herein on the basis of the morphology and distribution of lesions. For practical purposes, this classification system is based on the most typical disease presentations. However, morphology may vary as rashes evolve, and the presentation of diseases with rashes is subject to many variations (Chap. 57). For instance, the classic petechial rash of Rocky Mountain spotted fever (RMSF) may initially consist of blanchable erythematous macules distributed peripherally; at times,

the rash associated with RMSF may not be predominantly acral, or a rash may not develop at all.

Diseases with fever and rash may be classified by type of eruption: centrally distributed maculopapular, peripheral, confluent desquamative erythematous, vesiculobullous, urticarial, nodular, purpuric, ulcerated, or eschars (Table 18-1). For a more detailed discussion of each disease associated with a rash, the reader is referred to the chapter dealing with that specific disease. (Reference chapters and color plates are cited in the text and listed in Table 18-1.)

Centrally Distributed Maculopapular Eruptions Centrally distributed rashes, in which lesions are primarily truncal, are the most common type of eruption. The rash of *measles* (rubeola) starts at the hairline 2 to 3 days into the illness and moves down the body, sparing the palms and soles (Chap. 194). It begins as discrete erythematous lesions, which become confluent as the rash spreads. Koplik's spots (1- to 2-mm white or bluish lesions with an erythematous halo on the buccal mucosa) are pathognomonic for measles and are generally seen during the first 2 days of symptoms. They should not be confused with Fordyce's spots (ectopic sebaceous glands), which have no erythematous halos and are found in the mouth of healthy individuals. Koplik's spots may briefly overlap with the measles exanthem.

German measles (rubella) also spreads from the hairline downward; unlike that of measles, however, the rash of rubella tends to clear from originally affected areas as it migrates and may be pruritic (Chap. 195). Forchheimer spots (palatal petechiae) may develop but are nonspecific since they also develop in mononucleosis (Chap. 184) and scarlet fever (Chap. 140). Postauricular and suboccipital adenopathy and arthritis are common among adults with German measles. Exposure of pregnant women to ill individuals should be avoided, as rubella causes severe congenital abnormalities. Numerous strains of enteroviruses (Chap. 193), primarily echoviruses and coxsackieviruses, cause nonspecific syndromes of fever and eruptions that may mimic rubella or measles. Patients with infectious mononucleosis caused by Epstein-Barr virus or with primary infection caused by HIV (Chap. 309) may exhibit pharyngitis, lymphadenopathy, and a nonspecific maculopapular exanthem.

The rash of *erythema infectiosum* (fifth disease), which is caused by human parvovirus B19, primarily affects children 3 to 12 years old; it develops after fever has resolved as a bright blanchable erythema on the cheeks ("slapped cheeks") with perioral pallor (Chap. 187). A more diffuse rash (often pruritic) appears the next day on the trunk and extremities and then rapidly develops into a lacy reticular eruption that may wax and wane (especially with temperature change) over 3 weeks. Adults with fifth disease often have arthritis, and fetal hydrops can develop in association with this condition in pregnant women.

Exanthem subitum (roseola,Fig. 18-CD2) is most common among children under 3 years of age (Chap. 185). As in erythema infectiosum, the rash usually appears after fever has subsided. It consists of 2- to 3-mm rose-pink macules and papules that rarely coalesce, occur initially on the trunk and sometimes on the extremities (sparing the face), and fade within 2 days.

Though drug reactions have many manifestations, including urticaria, exanthematous *drug-induced eruptions* (Chap. 59) are most common and are often difficult to distinguish from viral exanthems. Eruptions elicited by drugs are usually more intensely erythematous and pruritic than viral exanthems, but this distinction is not reliable. A history of new medications and an absence of prostration may help to distinguish a drug-related rash from an eruption of another etiology. Rashes may persist for up to 2 weeks after administration of the offending agent is discontinued. Certain populations are more prone than others to drug rashes. Of HIV-infected patients, 50 to 60% develop a rash in response to sulfa drugs; 50 to 100% of patients with mononucleosis due to Epstein-Barr virus develop a rash when given ampicillin.

Rickettsial illnesses (Chap. 177) should be considered in the evaluation of individuals with centrally distributed maculopapular eruptions. The usual setting for *epidemic typhus* is a site of war or natural disaster in which people are exposed to body lice. A diagnosis of recrudescent typhus should be considered in European immigrants to the United States. However, an indigenous form of typhus, presumably transmitted by flying squirrels, has been reported in the southeastern United States. *Endemic typhus* or *leptospirosis* (the latter caused by a spirochete;Chap. 174) may be seen in urban environments where rodents proliferate. Outside the United States, other rickettsial diseases cause a spotted-fever syndrome and should be considered in residents of or travelers to endemic areas. Similarly, *typhoid fever*, a nonrickettsial disease caused by *Salmonella typhi* (Chap. 156), is usually acquired during travel outside the United States.

Some centrally distributed maculopapular eruptions have distinctive features. Erythema chronicum migrans (ECM), the rash of Lyme disease (Chap. 176), typically manifests as singular or multiple annular plaques. Untreated ECM lesions usually fade within a month but may persist for more than a year. *Erythema marginatum*, the rash of acute rheumatic fever (Chap. 235), has a distinctive pattern of enlarging and shifting transient annular lesions.

Collagen vascular diseases may cause fever and rash. Patients with *systemic lupus erythematosus* (Chap. 311) typically develop a sharply defined, erythematous eruption in a butterfly distribution on the cheeks (malar rash) as well as many other skin manifestations. *Still's disease* (Chap. 326) manifests as an evanescent salmon-colored rash on the trunk and proximal extremities that coincides with fever spikes.

Peripheral Eruptions These rashes are alike in that they are most prominent peripherally or begin in peripheral (acral) areas before spreading centripetally. Early diagnosis and therapy are critical inRMSF(Chap. 177) because of its grave prognosis if untreated. Lesions evolve from macular to petechial, start on the wrists and ankles, spread centripetally, and appear on the palms and soles only later in the disease. The rash of *secondary syphilis* (Chap. 172), which may be diffuse but is prominent on the palms and soles, should be considered in the differential diagnosis of pityriasis rosea, especially in sexually active patients. *Atypical measles* (Chap. 194) is seen in individuals contracting measles who received the killed measles vaccine between 1963 and 1967 in the United States and who were not subsequently protected with the live vaccine. *Hand-foot-and-mouth disease* (Chap. 193) is distinguished by tender vesicles distributed peripherally and in the mouth; outbreaks commonly occur within families. The classic

target lesions of *erythema multiforme* appear symmetrically on the elbows, knees, palms, and soles. In relatively severe cases, these lesions may spread diffusely and involve mucosal surfaces. Lesions may develop on the hands and feet in *endocarditis* (Chap. 126).

Confluent Desquamative Erythemas These eruptions consist of diffuse erythema frequently followed by desquamation. The eruptions caused by group A *Streptococcus* or *Staphylococcus aureus* are toxin mediated. Certain disease features may provide diagnostic clues. *Scarlet fever* (Chap. 140) usually follows pharyngitis; patients have a facial flush, a "strawberry" tongue, and accentuated petechiae in body folds (Pastia's lines). *Kawasaki disease* (Chaps. 57 and317) presents in the pediatric population as fissuring of the lips, a strawberry tongue, conjunctivitis, adenopathy, and sometimes cardiac abnormalities. *Streptococcal toxic shock syndrome* (Chap. 140,Fig. 18-CD3) manifests with hypotension, multiorgan failure, and often a severe group A streptococcal infection (e.g., necrotizing fasciitis, Fig. 18-CD4). *Staphylococcal toxic shock syndrome* (Chap. 139) also presents with hypotension and multiorgan failure, but usually only *S. aureus* colonization -- not a severe *S. aureus* infection -- is documented. *Staphylococcal scalded-skin syndrome* (Chap. 139) is seen primarily in children and in immunocompromised adults. Generalized erythema is often evident during the prodrome of fever and malaise; profound tenderness of the skin is distinctive. In the exfoliative stage, the skin can be induced to form bullae with light lateral pressure (Nikolsky's sign). In a mild form, a scarlatiniform eruption mimics scarlet fever, but the patient does not exhibit a strawberry tongue or circumoral pallor. In contrast to the staphylococcal scalded-skin syndrome, in which the cleavage plane is superficial in the epidermis, *toxic epidermal necrolysis* (Chap. 59) (Fig. 18-CD5) involves sloughing of the entire epidermis, resulting in severe disease. *Exfoliative erythroderma syndrome* (Chaps. 56 and59) is a serious reaction associated with systemic toxicity that is often due to eczema, psoriasis, mycosis fungoides, or a severe drug reaction.

Vesiculobullous Eruptions *Varicella* (Chap. 183) is highly contagious, often occurring in winter or spring. At a given time within a given region of the body, varicella lesions are in different stages of development. In immunocompromised hosts, varicella vesicles may lack the characteristic erythematous base or may appear hemorrhagic. *Rickettsialpox* (Chap. 177) is often documented in urban settings and is characterized by vesicles. It can be distinguished from varicella by an eschar at the site of the mouse-mite bite and the papule/plaque base of each vesicle. Disseminated *Vibrio vulnificus* infection (Chap. 159) or *ecthyma gangrenosum* due to *Pseudomonas aeruginosa* (Chap. 155) should be considered in immunosuppressed individuals with sepsis and hemorrhagic bullae.

Urticarial Eruptions Individuals with classic urticaria ("hives") usually have a hypersensitivity reaction without associated fever. In the presence of fever, urticarial eruptions are usually due to *urticarial vasculitis* (Chap. 317). Unlike individual lesions of classic urticaria, which last up to 48 h, these lesions may last up to 5 days. Etiologies include serum sickness (often induced by drugs such as penicillins, sulfas, salicylates, or barbiturates), connective-tissue disease (e.g., systemic lupus erythematosus or Sjogren's syndrome), and infection (e.g., with hepatitis B virus, coxsackievirus A9, or parasites). Malignancy may be associated with fever and chronic urticaria (Chap. 57).

Nodular Eruptions In immunocompromised hosts, nodular lesions often represent disseminated infection. Patients with disseminated *candidiasis* (often due to *Candida tropicalis*) may have a triad of fever, myalgias, and eruptive nodules (Chap. 205). Disseminated *cryptococcosis* lesions (Chap. 204) may resemble molluscum contagiosum. Necrosis of nodules should raise the suspicion of *aspergillosis* (Chap. 206) or *mucormycosis* (Chap. 207). *Erythema nodosum* presents with exquisitely tender nodules on the lower extremities. *Sweet's syndrome* (Chap. 57) should be considered in individuals with multiple nodules and plaques, often so edematous that they give the appearance of vesicles or bullae. Sweet's syndrome may affect either healthy individuals or persons with lymphoproliferative disease.

Purpuric Eruptions *Acute meningococcemia* (Chap. 146) classically presents in children as a petechial eruption, but initial lesions may appear as blanchable macules or urticaria.RMSFshould be considered in the differential diagnosis of acute meningococcemia. *Echovirus 9 infection* (Chap. 193) may mimic acute meningococcemia; patients should be treated as if they have bacterial sepsis since prompt differentiation of these conditions may be impossible. Large ecchymotic areas of *purpura fulminans* (Chaps. 124 and146) reflect severe underlying disseminated intravascular coagulation, which may be due to infectious or noninfectious causes. The lesions of *chronic meningococcemia* (Chap. 146) may have a variety of morphologies, including petechial. Purpuric nodules may develop on the legs and resemble erythema nodosum but lack its exquisite tenderness. Lesions of *disseminated gonococcemia* (Chap. 147) are distinctive, sparse, countable hemorrhagic pustules, usually located near joints. The lesions of chronic meningococcemia and those of gonococcemia may be indistinguishable in terms of appearance and distribution. *Viral hemorrhagic fever* (Chaps. 198 and199) should be considered in patients with an appropriate travel history and a petechial rash. *Thrombotic thrombocytopenic purpura* (Chaps. 57,108, and116) is a noninfectious cause of fever and petechiae. *Cutaneous small-vessel vasculitis* (*leukocytoclastic vasculitis*) typically manifests as palpable purpura and has a wide variety of causes (Chap. 57).

Eruptions with Ulcers or Eschars The presence of an ulcer or eschar in the setting of a more widespread eruption can provide an important diagnostic clue. For example, the presence of an eschar may suggest the diagnosis of scrub typhus or rickettsialpox in the appropriate setting. In other illnesses (e.g., anthrax), an ulcer or eschar may be the only skin manifestation.

(Bibliography omitted in Palm version)

Back to Table of Contents

19. APPROACH TO THE ACUTELY ILL INFECTED FEBRILE PATIENT - *Tamar F. Barlam*, *Dennis L. Kasper*

The physician treating the acutely ill febrile patient must be able to recognize infections that require emergent attention. If such infections are not adequately evaluated and treated at initial presentation, the opportunity to alter an adverse outcome may be lost. In this chapter, the clinical presentations of and approach to patients with relatively common infectious disease emergencies are discussed. These infectious processes are discussed in detail in other chapters.**Noninfectious causes of fever are not covered in this chapter; information on the approach to fever of unknown origin, including that eventually shown to be of noninfectious etiology, is presented in Chap. 125.*

GENERAL CONSIDERATIONS

APPEARANCE

A physician must have a consistent approach to acutely ill patients. Even before the history is elicited and a physical examination performed, an immediate assessment of the patient's general appearance yields valuable information. The perceptive physician's subjective sense that a patient is septic or toxic often proves accurate. Visible agitation or anxiety in a febrile patient can be a harbinger of critical illness.

HISTORY

Presenting symptoms are frequently nonspecific. In addition to a general description of symptoms, it is important to obtain a sense of disease progression. Detailed questions should be asked about the onset and duration of symptoms and about changes in severity or rate of progression over time. Host factors and comorbid conditions may enhance the risk of infection with certain organisms or of a more fulminant course than is usually seen. Lack of splenic function, alcoholism with significant liver disease, intravenous drug use, HIV infection, diabetes, malignancy, and chemotherapy all predispose to specific infections and frequently to increased severity. The patient should be questioned about factors that might help identify a nidus for invasive infection, such as recent upper respiratory tract infections, influenza, or varicella; prior trauma; disruption of cutaneous barriers due to lacerations, burns, surgery, or decubiti; and the presence of foreign bodies, such as nasal packing after rhinoplasty, barrier contraceptives, tampons, arteriovenous fistulas, or prosthetic joints. Travel, contact with pets or other animals, or activities that might result in tick exposure can lead to diagnoses that would not otherwise be considered. Recent dietary intake, medication use, social contact with ill individuals, vaccination history, and menstrual history may be relevant. A review of systems should focus on any neurologic signs or sensorium alterations, rashes or skin lesions, and focal pain or tenderness and should also include a general review of respiratory, gastrointestinal, or genitourinary symptoms. It is especially important to determine the duration and progression of these symptoms in order to gain an appreciation of the pace and urgency of the process.

PHYSICAL EXAMINATION

A complete physical examination should be performed, with special attention to some

areas that are sometimes given short shrift in routine examinations. Assessment of the patient's general appearance and vital signs, skin and soft tissue examination, and the neurologic evaluation are of particular importance.

The patient may appear either anxious and agitated or lethargic and apathetic. Fever is usually present, although the elderly and compromised hosts, such as those who are uremic or cirrhotic and patients who are taking glucocorticoids or nonsteroidal anti-inflammatory agents, may be afebrile despite serious underlying infection. Measurement of blood pressure, heart rate, and respiratory rate helps determine the degree of hemodynamic and metabolic compromise. The patient's airway must be evaluated to rule out the risk of obstruction from an invasive oropharyngeal infection.

The etiologic diagnosis may become evident in the context of a thorough skin examination. Petechial rashes are typically seen with meningococcemia or Rocky Mountain spotted fever (RMSF); erythroderma is usual with toxic shock syndrome (TSS) and drug fever. The soft tissue and muscle examination is critical. Areas of erythema or duskiness, edema, and tenderness may indicate underlying necrotizing fasciitis, myositis, or myonecrosis. The neurologic examination must include a careful assessment of mental status for signs of early encephalopathy. Evidence of nuchal rigidity or focal neurologic findings should be sought. Focal findings, depressed mental status, or papilledema should be evaluated by brain imaging prior to lumbar puncture, which, in this setting, could initiate herniation.

SPECIFIC PRESENTATIONS

For most infections, there is time for careful evaluation, diagnostic testing, and consultation with other physicians. However, the infections considered below according to common clinical presentation can have rapidly catastrophic outcomes, and their immediate recognition can be life-saving. Recommended therapeutic regimens are presented in Table 19-1.

SEPSIS WITHOUT AN OBVIOUS FOCUS OF PRIMARY INFECTION

These patients initially have a brief prodrome of nonspecific symptoms and signs that progresses quickly to hemodynamic instability with hypotension, tachycardia, tachypnea, or respiratory distress. A patient may display altered mental status. Disseminated intravascular coagulation (DIC) with clinical evidence of a hemorrhagic diathesis is a poor prognostic sign.

Septic Shock (See also Chap. 124) Patients with bacteremia leading to septic shock may have a primary site of infection (e.g., pneumonia, pyelonephritis, or cholangitis) that is not evident initially. Elderly patients with comorbid conditions, hosts compromised by malignancy and neutropenia, or patients who have recently undergone a surgical procedure or hospitalization are at increased risk for an adverse outcome. Gram-negative bacteremia with organisms such as *Pseudomonas aeruginosa*, *Aeromonas hydrophila*, or *Escherichia coli* and gram-positive infection with organisms such as *Staphylococcus aureus* or group A streptococci can present as intractable hypotension and multiorgan failure. Treatment can usually be initiated empirically on the basis of the presentation (Table 124-3).

Overwhelming Infection in Asplenic Patients (See also Chap. 124) Patients without splenic function are at risk for overwhelming bacterial sepsis. Asplenic patients succumb to sepsis at 600 times the rate of the general population; 50 to 70% of cases occur within the first 2 years after splenectomy, with a mortality rate of up to 80%. However, in the asplenic individual, an increased risk of overwhelming sepsis continues throughout life. In asplenia, encapsulated bacteria cause the majority of infections, and adults are at lower risk than children because they are more likely to have antibody to these organisms. *Streptococcus pneumoniae* infection is most common, but the risk of infection with *Haemophilus influenzae* or *Neisseria meningitidis* is also high. Severe clinical manifestations of infections due to *E. coli*, *S. aureus*, group B streptococci, *P. aeruginosa*, *Capnocytophaga*, *Babesia*, and *Plasmodium* have been described.

Babesiosis (See also Chap. 214) A history of recent travel to endemic areas should raise the possibility of infection with *Babesia*. Between 1 and 4 weeks after a tick bite, the patient experiences chills, fatigue, anorexia, myalgia, arthralgia, nausea, and headache; ecchymosis and/or petechiae are occasionally seen. The tick that most commonly transmits *Babesia*, *Ixodes scapularis*, also transmits *Borrelia burgdorferi* (the agent of Lyme disease) and *Ehrlichia*, and co-infection can occur, resulting in more severe disease. Infection with the European species *Babesia divergens* is more frequently fulminant than that due to the U.S. species *B. microti*, causing a febrile syndrome with hemolysis, jaundice, hemoglobinemia, and renal failure and a mortality rate of>50%. Severe babesiosis is especially common in asplenic hosts but does occur in hosts with normal splenic function.

Other Sepsis Syndromes Tularemia (Chap. 161) is seen throughout the United States, but primarily in Arkansas, Oklahoma, and Missouri, in association with wild rabbit, tick, and tabanid fly contact. The uncommon typhoidal form can be associated with gram-negative septic shock and a mortality rate of>30%. In the United States, plague (Chap. 162) is found primarily in New Mexico, Arizona, and Colorado after contact with ground squirrels, prairie dogs, or chipmunks. The septic form is particularly rare and is associated with shock, multiorgan failure, and a 30% mortality rate. These rare infections should be considered in the appropriate epidemiologic setting.

SEPSIS WITH SKIN MANIFESTATIONS (See also Chap. 18)

Maculopapular rashes may reflect early meningococcal or rickettsial disease but are usually associated with nonemergent infections. Exanthems are usually viral.

Petechiae Petechial rashes caused by viruses are seldom associated with hypotension or a toxic appearance, although severe measles can be an exception. In other settings, petechial rashes require more urgent attention.

Meningococcemia (See also Chap. 146) Almost three-quarters of patients with bacteremic *N. meningitidis* infection have a rash. Meningococcemia most often affects young children (i.e., those 6 months to 5 years old, often in daycare). However, sporadic cases and outbreaks occur in schools (grade school through college) and army barracks. Between 10 and 20% of all cases have a fulminant course, with shock, DIC, and multiorgan failure. Of these patients, 50 to 60% die, and survivors often require

extensive debridement or amputation of gangrenous extremities. Patients may exhibit fever, headache, nausea, vomiting, myalgias, change in mental status, and meningismus. However, the rapidly progressive form of disease is not usually associated with meningitis. The rash is initially pink, blanching, and maculopapular, appearing on the trunk and extremities, but then becomes hemorrhagic, forming petechiae. Petechiae are first seen at the ankles, wrists, axillae, mucosal surfaces, and palpebral and bulbar conjunctiva, with subsequent spread to the lower extremities and trunk. A cluster of petechiae may be seen at pressure points, e.g., where a blood pressure cuff has been inflated. In rapidly progressive meningococcemia, the petechial rash quickly becomes purpuric (Plate IID-44) and patients develop DIC. Hypotension with petechiae for <12 h is associated with significant mortality. The mortality rate can exceed 90% in patients without meningitis who have rash, hypotension, and a normal or low white blood cell count and erythrocyte sedimentation rate. A better prognosis has been reported in cases where antibiotics are given before admission by the primary care provider. This observation suggests that early initiation of treatment may be life-saving.

Rocky Mountain spotted fever (See also Chap. 177) RMSFoccurs throughout the United States. A history of tick bite is common; however, if such a history is lacking, a history of travel or outdoor activity (e.g., camping in tick-infested areas) can be ascertained. RMSF is caused by *Rickettsia rickettsii*. For the first 3 days, headache, fever, malaise, myalgias, nausea, vomiting, and anorexia are present. By day 3, half of patients have skin findings. Blanching macules develop initially on the wrists and ankles and then spread over the legs and trunk. The lesions become hemorrhagic and are frequently petechial. The rash spreads to palms and soles later in the course (Plate IID-45). The centripetal spread is a classic feature of RMSF. However, 10 to 15% of patients with RMSF never develop a rash. The patient can be hypotensive and develop noncardiogenic pulmonary edema, confusion, lethargy, and encephalitis progressing to coma. The cerebrospinal fluid (CSF) contains 10 to 100 cells/uL, usually with a predominance of mononuclear cells. The CSF glucose level is often normal; the protein concentration may be slightly elevated. Renal and hepatic injury and bleeding secondary to vascular damage are noted. Untreated infection has a mortality rate of 30%.

Purpura Fulminans (See alsoChaps. 124 and146) This is the cutaneous manifestation ofDIC and presents as large ecchymotic areas and hemorrhagic bullae. Progression of petechiae to purpura and ecchymoses is associated with congestive heart failure, septic shock, acute renal failure, acidosis, hypoxia, hypotension, and death. Purpura fulminans has primarily been associated with *N. meningitidis* but, in the splenectomized patient, has been described in association with *S. pneumoniae* and *H. influenzae*.

Ecthyma Gangrenosum Septic shock caused by *P. aeruginosa* and *A. hydrophila* can be associated with ecthyma gangrenosum (Plate IID-57C,Fig. 19-CD1): hemorrhagic vesicles surrounded by a rim of erythema with central necrosis and ulceration. These gram-negative bacteremias are most common among patients with neutropenia, extensive burns, and hypogammaglobulinemia.

Other Emergent Infections Associated with Rash *Vibrio vulnificus* and other noncholera *Vibrio* bacteremic infections (Chap. 159) can cause focal skin lesions and overwhelming sepsis in the host with liver disease. After ingestion of contaminated

shellfish, there is a sudden onset of malaise, chills, fever, and hypotension. The patient develops bullous or hemorrhagic skin lesions, usually on the lower extremities, and 75% of patients have leg pain. The mortality rate can be as high as 50%. *Capnocytophaga canimorsus* (Chap. 127) can cause septic shock in asplenic patients. Infection with this fastidious gram-negative rod typically presents after a dog bite as fever, chills, myalgia, vomiting, diarrhea, dyspnea, confusion, and headache. Findings can include an exanthem or erythema multiforme (Plate IIE-67), cyanotic mottling or peripheral cyanosis, petechiae, and ecchymosis. About 30% of patients with this fulminant form die of overwhelming sepsis andDIC, and survivors may require amputation to treat gangrene.

Erythroderma TSS(Chaps. 139 and140) is usually associated with erythroderma (Fig. 18-CD3). The patient presents with fever, malaise, myalgias, nausea, vomiting, diarrhea, and confusion. There is a sunburn-type rash that may be subtle and patchy but is usually diffuse and is found on the face, trunk, and extremities. Erythroderma, which desquamates after 1 to 2 weeks, is more common in *Staphylococcus*-associated than in *Streptococcus*-associated TSS. Hypotension develops rapidly after onset of symptoms, often within hours. Multiorgan failure is seen. Often there is no indication of a primary focal infection. Colonization rather than overt infection of the vagina or a postoperative wound, for example, is typical with staphylococcal TSS, and the mucosal areas appear hyperemic but not infected. Early renal failure may distinguish this syndrome from other septic shock syndromes. Clinical evaluation constitutes the diagnosis because TSS is defined by the clinical criteria of fever, rash, hypotension, and multiorgan involvement. The mortality rate is 5% for menstruation-associated TSS, 10 to 15% for nonmenstrual TSS, and 30 to 70% for streptococcal TSS.

SEPSIS WITH A SOFT TISSUE/MUSCLE PRIMARY FOCUS (See also Chap. 128)

Necrotizing Fasciitis This infection may arise at a site of minimal trauma or postoperative incision and may also be associated with recent varicella, childbirth, or muscle strain. The most common causes of necrotizing fasciitis are group A streptococci alone (Chap. 140) and a mixed facultative and anaerobic flora (Chap. 128). Diabetes mellitus, peripheral vascular disease, and intravenous drug use are associated risk factors. Use of nonsteroidal anti-inflammatory agents adversely affects granulocyte chemotaxis, phagocytosis, and bacterial killing, allowing progression of skin or soft tissue infections. The patient may have bacteremia and hypotension without other organ-system failure. Physical findings are minimal compared to the severity of pain and the degree of fever. The examination is often unremarkable except for soft tissue edema and erythema. The infected area is red, hot, shiny, swollen, and exquisitely tender. In untreated infection, the overlying skin develops blue-gray patches after 36 h, and cutaneous bullae and necrosis develop after 3 to 5 days. Necrotizing fasciitis due to a mixed flora, but not that due to group A streptococci, can be associated with gas production. Without treatment, pain decreases because of thrombosis of the small blood vessels and destruction of the peripheral nerves -- an ominous sign. The mortality rate is >30% overall,>70% in association withTSS, and nearly 100% without surgical intervention. Life-threatening necrotizing fasciitis may also be due to *Clostridium perfringens* (Chap. 145); in this condition, the patient is extremely toxic and the mortality rate is high. Within 48 h, rapid tissue invasion and systemic toxicity associated with hemolysis and death ensue. The distinction between this entity and clostridial

myonecrosis is made by muscle biopsy.

Clostridial Myonecrosis (See alsoChap. 145) Myonecrosis is often associated with trauma or surgery but can be spontaneous. The incubation period is usually 12 to 24 h long, and massive necrotizing gangrene develops within hours of onset. Systemic toxicity, shock, and death can occur within 12 h. The patient's pain and toxic appearance are out of proportion to physical findings. On examination, the patient is febrile, apathetic, tachycardic, and tachypneic and may express a feeling of impending doom. Hypotension and renal failure develop later, and hyperalertness is evident preterminally. The skin over the affected area is bronze-brown, mottled, and edematous. Bullous lesions with serosanguineous drainage and a mousy or sweet odor can be present. Crepitus can occur secondary to gas production in muscle tissue. The mortality rate is>65% with spontaneous myonecrosis, which is often associated with *C. septicum* and underlying malignancy. The mortality rates associated with trunk and limb infection are 63% and 12%, respectively, and any delay in surgical treatment increases the risk of death.

NEUROLOGIC INFECTIONS WITH OR WITHOUT SEPTIC SHOCK

Bacterial Meningitis (See also Chap. 372) Bacterial meningitis is one of the most common infectious emergencies involving the central nervous system. Although hosts with cell-mediated immune deficiency, including transplant recipients, diabetic patients, the elderly, and cancer patients treated with certain chemotherapeutic agents, are at particular risk for *Listeria monocytogenes* meningitis, most cases in adults are due to *S. pneumoniae* (30 to 50%) and *N. meningitidis* (10 to 35%). An early presentation of headache, meningismus, and fever is classic but is seen in only half of patients. The elderly can present without fever or meningeal signs despite lethargy and confusion. Cerebral dysfunction is evidenced by confusion, delirium, and lethargy that can progress to coma. The presentation is fulminant, with sepsis and brain edema, in some cases; papilledema at presentation is unusual and suggests another diagnosis (e.g., an intracranial lesion). Focal signs, including cranial nerve palsies (IV, VI, VII), can be seen in 10 to 20% of cases; 50 to 60% of patients have bacteremia. A poor neurologic outcome is associated with coma at any time during the course or with aCSFglucose level of <0.6 mmol/L (<10 mg/dL). Mortality is associated with coma, respiratory distress, shock, a CSF protein level of>2.5 g/L, a peripheral white blood cell count of<5000/uL, and a serum sodium level of<135 mmol/L.

Suppurative Intracranial Infections (See also Chap. 372) Other rare intracranial lesions that present with sepsis and hemodynamic instability are subdural empyema, septic cavernous sinus thrombosis, and septic superior sagittal sinus thrombosis. Rapid recognition of the toxic patient with central neurologic signs is crucial to improvement of the dismal prognosis of these entities.

Subdural Empyema This infection arises from the paranasal sinus in 60 to 70% of cases. Microaerophilic streptococci and staphylococci are the predominant etiologic organisms. The patient is toxic, with fever, headache, and nuchal rigidity. Of all patients, 75% have focal signs and 6 to 20% die.

Septic Cavernous Sinus Thrombosis This condition follows a facial or sphenoid sinus

infection; 70% of cases are due to staphylococci and the remainder to aerobic or anaerobic streptococci. A unilateral or retroorbital headache progresses to a toxic appearance and fever within days. Three-quarters of patients have unilateral periorbital edema that becomes bilateral and then progresses to ptosis, proptosis, ophthalmoplegia, and papilledema. The mortality rate is as high as 30%.

Septic Thrombosis of the Superior Sagittal Sinus This infection spreads from the ethmoid or maxillary sinuses. Its bacterial causes include *S. pneumoniae*, other streptococci, and staphylococci. The fulminant course is characterized by headache, nausea, vomiting, rapid progression to confusion and coma, nuchal rigidity, and brainstem signs. If the sinus is totally thrombosed, the mortality rate exceeds 80%.

Brain Abscess (See alsoChap. 372) Brain abscess often occurs without systemic signs. Almost half of patients are afebrile, and presentations are more consistent with a space-occupying lesion in the brain; 70% have headache, 50% have focal neurologic signs, and 25% have papilledema. Abscesses can present as single or multiple lesions resulting from contiguous foci or hematogenous infection, such as unrecognized endocarditis. The infection progresses over several days from cerebritis to an abscess with a mature capsule. Abscesses arising hematogenously are especially apt to rupture into the ventricular space, causing a sudden and severe deterioration in clinical status and high mortality. Otherwise, mortality is low but morbidity is high (30 to 55%). Patients presenting with stroke and a parameningeal infectious focus, such as sinusitis or otitis, may have a brain abscess, and physicians must maintain a high level of suspicion. Prognosis worsens in patients with a fulminant course, delayed diagnosis, abscess rupture into the ventricles, multiple abscesses, or abnormal neurologic status at presentation.

Cerebral Malaria (See alsoChap. 214) This entity should be urgently considered if patients who have recently traveled to areas endemic for malaria present with a febrile illness and lethargy or other neurologic signs. Fulminant malaria is caused by *Plasmodium falciparum* and is associated with temperatures of>40°C (>104°F), hypotension, jaundice, adult respiratory distress syndrome, and bleeding. By definition, any patient with a change in mental status or repeated seizure in the setting of fulminant malaria has cerebral malaria. In adults this nonspecific febrile illness progresses to coma over several days; occasionally, coma occurs within hours and death within 24 h. Nuchal rigidity and photophobia are rare. On physical examination, symmetric encephalopathy is typical, and upper motor neuron dysfunction with decorticate and decerebrate posturing can be seen with advanced disease. Unrecognized infection results in a 30% mortality rate.

Spinal Epidural Abscesses (See alsoChap. 368) Patients with spinal epidural abscesses often present with back pain and develop neurologic deficits late in their course. At-risk patients include those with diabetes mellitus; intravenous drug use; recent spinal trauma, surgery, or epidural anesthesia; and other comorbid conditions, such as HIV infection. The thoracic or lumbar spine is the most common location, and staphylococci are the most common etiologic agents; in HIV-infected intravenous drug users, therapy must cover gram-negative rods and methicillin-resistant *S. aureus*. If a patient gives a history of antecedent back pain and has new neurologic symptoms, this diagnosis must immediately be considered. Almost 60% of patients have fever and

almost 90% have back pain. Paresthesia, bowel and bladder dysfunction, radicular pain, and weakness are frequent neurologic complaints, and examination of the patient may reveal abnormal reflexes and motor and sensory deficits. Rapid recognition and treatment, including immediate drainage, can prevent or minimize permanent neurologic sequelae.

FOCAL SYNDROMES WITH A FULMINANT COURSE

Infection at virtually any primary focus (e.g., osteomyelitis, pneumonia, pyelonephritis, or cholangitis) can result in bacteremia and sepsis.TSS has been associated with focal infections such as septic arthritis, peritonitis, sinusitis, and wound infection. Death occurs secondary to septic shock or toxin production with hemodynamic instability and multiorgan failure. Rapid clinical deterioration and death can be associated with destruction of the primary site of infection, as is seen in endocarditis and in necrotizing infections of the oropharynx (in which edema suddenly compromises the airway).

Rhinocerebral Mucormycosis (See also Chap. 207) Patients with diabetes or malignancy are at risk for invasive rhinocerebral mucormycosis. Patients present with low-grade fever, dull sinus pain, diplopia, decreased mental status, decreased ocular motion, chemosis, proptosis, dusky or necrotic nasal turbinates, and necrotic hard-palate lesions that respect the midline. Without rapid recognition and intervention, the process continues an inexorable invasive course with high mortality.

Acute Bacterial Endocarditis (See also Chap. 126) This entity presents with a much more aggressive course than subacute endocarditis. Bacteria such as *S. aureus*, *S. pneumoniae*, *L. monocytogenes*, *Haemophilus* spp., and streptococci of groups A, B, and G attack native valves. Mortality rates range from 10 to 40%. The host may have comorbid conditions such as underlying malignancy, diabetes mellitus, intravenous drug use, or alcoholism. The patient presents with fever, fatigue, and malaise<2 weeks after onset of infection. On physical examination, a changing murmur and congestive heart failure may be noted. Hemorrhagic macules on palms or soles (*Janeway lesions*,Fig. 19-CD2) sometimes develop. Petechiae, Roth's spots (Fig. 19-CD3), splinter hemorrhages (Fig. 19-CD4), and splenomegaly are unusual. Rapid valvular destruction, particularly of the aortic valve, results in pulmonary edema and hypotension. Myocardial abscesses can form, eroding through the septum or into the conduction system and causing life-threatening arrhythmias or high-degree conduction block. Large friable vegetations can result in major arterial emboli, metastatic infection, or tissue infarction. Emboli can lead to stroke, change in mental status, visual disturbances, aphasia, ataxia, headache, meningismus, brain abscess, cerebritis, spinal cord infarct with paraplegia, arthralgia, osteomyelitis, splenic abscess, septic arthritis, and hematuria. Rapid intervention is crucial for a successful outcome.

DIAGNOSTIC WORKUP OF THE ACUTELY ILL PATIENT

After a quick clinical assessment, diagnostic material should be obtained rapidly and antibiotic and supportive treatment begun. In the sepsis syndromes, blood (for cultures; baseline complete blood count with differential; measurement of serum electrolytes, blood urea nitrogen, serum creatinine, and serum glucose; and liver function tests) can be obtained at the time an intravenous line is placed and before antibiotics are

administered. For patients with possible acute endocarditis, three sets of blood cultures should be performed. Asplenic patients should have a blood smear examined to confirm the presence of Howell-Jolly bodies (indicating the absence of splenic function) and a buffy coat examined for bacteria; these patients can have>10^6organisms per milliliter of blood (compared to 10^4/mL in patients with an intact spleen). Blood smears from patients with possible cerebral malaria or babesiosis must be examined for the diagnosis and quantitation of parasitemia. Blood smears may also be diagnostic in ehrlichiosis.

Patients with meningitis should have CSFobtained before the initiation of antibiotic therapy. *If focal neurologic signs, abnormal mental status, or papilledema mandates brain imaging before a lumbar puncture, antibiotics should be administered prior to imaging but after blood for cultures has been drawn.* If CSF cultures are negative, laboratory examination of CSF by latex agglutination or immunoprecipitation can be attempted to make an etiologic diagnosis. However, blood cultures will provide the diagnosis in 50 to 70% of cases.

Focal abscesses necessitate immediate computed tomography or magnetic resonance imaging as part of an evaluation for surgical intervention. Other diagnostic procedures, such as cultures of wounds or scraping of skin lesions, should not delay the initiation of treatment for more than minutes. Once emergent evaluation, diagnostic procedures, and (if appropriate) surgical consultation (see below) have been completed, other laboratory tests can be conducted. Appropriate radiography, computed axial tomography, magnetic resonance imaging, urinalysis, erythrocyte sedimentation rate determination, and transthoracic or transesophageal echocardiography may all prove important.

TREATMENT

Table 19-1 lists first-line treatments for the infections considered in this chapter. (For a more detailed discussion of treatment, see specific chapters.) In addition to the initiation of parenteral antibiotic therapy, several of these infections require urgent surgical attention. General surgery for possible necrotizing fasciitis or myonecrosis, neurosurgical evaluation for subdural empyema or spinal epidural abscess, otolaryngologic surgery for possible mucormycosis, and cardiothoracic surgery for critically ill patients with acute endocarditis are as important as the rapid commencement of antibiotic therapy. For infections such as necrotizing fasciitis and clostridial myonecrosis, rapid surgical intervention supercedes other diagnostic or therapeutic maneuvers.

Acutely ill febrile patients require close observation, aggressive supportive measures, and -- in most cases -- admission to intensive care units. Adjunctive treatments, such as intravenous immunoglobulin administration forTSS, can be considered after initial stabilization. The most important task of the physician is to recognize the acute infectious emergency and proceed with appropriate urgency.

(Bibliography omitted in Palm version)

Back to Table of Contents

20. HYPOTHERMIA AND FROSTBITE - *Daniel F. Danzl*

HYPOTHERMIA

Accidental hypothermia occurs when there is an unintentional drop in the body's core temperature below 35°C (95°F). At this temperature, many of the compensatory physiologic mechanisms to conserve heat begin to fail. *Primary accidental hypothermia* is a result of the direct exposure of a previously healthy individual to the cold. The mortality rate is much higher for those patients who develop *secondary hypothermia* as a complication of a serious systemic disorder.

CAUSES

Primary accidental hypothermia is geographically and seasonally pervasive. Although most cases occur in the winter months and in colder climates, it is surprisingly common in warmer regions as well. In the United States, hypothermia accounts for more than 700 deaths each year, half of which occur in people age 65 or older.

Multiple variables make individuals at the extremes of age, the elderly and neonates, particularly vulnerable to hypothermia (Table 20-1). The elderly have diminished thermal proprioception and are more susceptible to immobility, malnutrition, and systemic illnesses that interfere with heat generation or conservation. Dementia, psychiatric illness, and socioeconomic factors often compound these problems by impeding adequate measures to prevent hypothermia. Neonates have high rates of heat loss because of their increased surface-to-mass ratio and their lack of effective shivering and adaptive behavioral responses. In addition, malnutrition can contribute to heat loss because of diminished subcutaneous fat and because of its association with depleted energy stores used for thermogenesis.

Individuals whose occupations or hobbies entail extensive exposure to cold weather are clearly at increased risk for hypothermia. Military history is replete with hypothermic tragedies. Hunters, sailors, skiers, and climbers also are at great risk of exposure, whether it involves injury, changes in weather, or lack of preparedness.

Ethanol causes vasodilatation (which increases heat loss), reduces thermogenesis and gluconeogenesis, and may impair judgment or lead to obtundation. Hypothermia is not an uncommon feature in Wernicke's encephalopathy and may mask its other manifestations. A number of medications are associated with altered thermal regulation. Phenothiazines, barbiturates, benzodiazepines, cyclic antidepressants, and many other medications reduce centrally-mediated vasoconstriction. Up to one-quarter of patients admitted to an intensive care unit because of drug overdose are hypothermic. Anesthetics can block the shivering responses; their effects may be compounded when patients are not covered adequately in the operating or recovery rooms.

Several types of endocrine dysfunction can lead to hypothermia. Hypothyroidism -- particularly when extreme, as in myxedema coma -- reduces the metabolic rate and impairs thermogenesis and behavioral responses. Myxedema is more common in women than in men and may be occult. Adrenal insufficiency and hypopituitarism can also increase susceptibility to hypothermia. Hypoglycemia, most commonly caused by

insulin or oral hypoglycemic drugs, is associated with hypothermia, in part the result of neuroglycopenic effects on hypothalamic function. Increased osmolality and metabolic derangements associated with uremia, diabetic ketoacidosis, and lactic acidosis can lead to altered hypothalamic thermoregulation.

Neurologic injury from trauma, cerebrovascular accident, subarachnoid hemorrhage, or hypothalamic lesions increases susceptibility to hypothermia. Agenesis of the corpus callosum, or Shapiro syndrome, is one cause of episodic hypothermia, characterized by profuse perspiration followed by a rapid fall in temperature. Acute spinal cord injury disrupts the autonomic pathways that lead to shivering and prevents cold-induced reflex vasoconstrictive responses.

Hypothermia associated with sepsis is a poor prognostic sign. Hepatic failure causes decreased glycogen stores and gluconeogenesis, as well as a diminished shivering response. In acute myocardial infarction associated with low cardiac output, hypothermia may be reversed after adequate resuscitation. With extensive burns, psoriasis, erythrodermas, and other skin diseases, increased peripheral blood flow leads to excessive heat loss.

THERMOREGULATION

Heat loss occurs through five mechanisms: radiation (55 to 65% of heat loss), conduction (10 to 15% of heat loss, but much greater in cold water), convection (increase in the wind), respiration, and evaporation (which are affected by the ambient temperature and the relative humidity).

The preoptic anterior hypothalamus normally orchestrates thermoregulation (Chap. 17). The immediate defense of thermoneutrality is via the autonomic nervous system (Chap. 72), whereas delayed control is mediated by the endocrine system. Autonomic nervous system responses include the release of norepinephrine, increased muscle tone, and shivering, leading to thermogenesis and an increase in the basal metabolic rate. Cutaneous cold thermoreception causes direct reflex vasoconstriction to converse heat. Prolonged exposure to cold also stimulates hypothalamic release of thyrotropin releasing hormone; this leads to increased levels of thyroid stimulating hormone (TSH), which stimulates the thyroid gland to produce thyroxine, a hormone that increases metabolic rate.

CLINICAL PRESENTATION

In most cases of hypothermia, the history of exposure to environmental factors, such as prolonged exposure to the outdoors without adequate clothing, makes the diagnosis straightforward. In urban settings, however, the presentation is often more subtle and the clinician may focus on other disease processes, toxin exposures, or psychiatric diagnoses.

After initial stimulation by hypothermia, there is progressive depression of all organ systems. The timing of the appearance of these clinical manifestations varies widely (Table 20-2). Without knowing the core temperature, it can be difficult to interpret other vital signs. For example, a tachycardia disproportionate to the core temperature

suggests secondary hypothermia resulting from hypoglycemia, hypovolemia, or a toxin overdose. Because carbon dioxide production declines progressively, the respiratory rate should be low; persistent hyperventilation suggests a central nervous system (CNS) lesion or one of the organic acidoses. A markedly depressed level of consciousness in a patient with mild hypothermia should raise suspicion of an overdose or CNS dysfunction due to infection or trauma.

Physical examination findings can also be altered by hypothermia. For instance, the assumption that areflexia is solely attributable to hypothermia can obscure and delay the diagnosis of a spinal cord injury. Patients with hypothermia may be confused or combative; these symptoms abate more rapidly with rewarming than with the use of restraints. A classic example of maladaptive behavior in patients with hypothermia is paradoxical undressing, which involves the inappropriate removal of clothing in response to a cold stress. The cold-induced ileus and abdominal rectus spasm can mimic, or mask, the presentation of an acute abdomen (Chap. 14).

When a patient in hypothermic cardiac arrest is first discovered, cardiopulmonary resuscitation is indicated, unless (1) a do-not-resuscitate status is verified, (2) obviously lethal injuries are identified, or (3) the depression of a frozen chest wall is not possible. As the resuscitation proceeds, the prognosis is grave if there is evidence of widespread cell lysis, as reflected by potassium levels exceeding 10 mEq/L. Other findings that may preclude continuing resuscitation include a core temperature<12°C, a pH<6.5, or evidence of intravascular thrombosis with a fibrinogen value <50 mg/dL. The decision to terminate resuscitation before rewarming the patient to 35°C is extremely difficult. There are no validated prognostic indicators for recovery from hypothermia. A history of asphyxia with secondary cooling is the most important negative predictor of survival.

DIAGNOSIS AND STABILIZATION

Hypothermia is confirmed by measuring the core temperature, preferably at two sites. Rectal probes should be placed to a depth of 15 cm and not adjacent to cold feces. A simultaneous esophageal measurement will be falsely high during heated inhalation therapy. The probe should be placed 24 cm below the larynx. The greatest discordance between the readings is usually during the transition phase before effective rewarming. Relying solely on infrared tympanic thermography is not advisable.

After a diagnosis of hypothermia is established, cardiac monitoring should be instituted, along with attempts to limit further heat loss. If the patient is in ventricular fibrillation, one sequence of 3 defibrillation attempts (2 J/kg) should be administered. If unsuccessful, active rewarming should be continued past 30° to 32°C. Supplemental oxygenation is always warranted, since tissue oxygenation is adversely affected by the leftward shift of the oxyhemoglobin dissociation curve. Pulse oximetry may be unreliable in patients with vasoconstriction. If protective airway reflexes are absent, gentle endotracheal intubation should be performed. Adequate pre-oxygenation will prevent ventricular arrhythmias.

Insertion of a gastric tube prevents dilatation secondary to decreased bowel motility. Indwelling bladder catheters facilitate monitoring of cold-induced diuresis. Dehydration is commonly encountered with chronic hypothermia, and most patients benefit from a bolus of crystalloid. Normal saline containing 5% dextrose is preferable to lactated

Ringer's solution, as the liver in hypothermic patients inefficiently metabolizes lactate. The placement of a pulmonary artery catheter, although of potential value, risks perforation of the less compliant pulmonary artery. The use of a central venous catheter should be avoided because of right atrial irritability.

Arterial blood gases should not be corrected for temperature (Chap. 50). This is termed the ectothermic or alpha-stat approach, which maximizes enzymatic function and maintains the normal distribution of charged metabolic intermediates. An uncorrected pH of 7.42 and a P_{CO2} of 40 mmHg reflects appropriate alveolar ventilation and acid-base balance at any core temperature. Acid-base imbalances should be corrected gradually, since the bicarbonate buffering system is inefficient. When the P_{CO2} increases 10 mmHg at 28°C, it doubles the pH decline of 0.08 that is normally induced at 37°C.

The severity of anemia may be underestimated because the hematocrit increases 2% for each 1°C drop in temperature. White blood cell sequestration and bone marrow suppression are common, potentially masking an infection. Although hypokalemia is more common in chronic hypothermia, hyperkalemia also occurs; the expected electrocardiographic changes can be obscured by hypothermia. Patients with renal insufficiency, metabolic acidoses, or rhabdomyolysis are most at risk for electrolyte disturbances.

Coagulopathies are common because cold inhibits the enzymatic reactions required for activation of the intrinsic cascade. In addition, the production of thromboxane B_2 by platelets is temperature-dependent, and platelet function is impaired. The administration of platelets and fresh frozen plasma is, therefore, not effective. The prothrombin or partial thromboplastin times reported by the laboratory appear deceptively normal and contrast with the observed coagulopathy. This contradiction appears because all coagulation tests are routinely performed at 37°C, and the enzymes are thus rewarmed.

REWARMING STRATEGIES

The key initial decision is whether to rewarm the patient passively or actively. *Passive external rewarming* simply involves covering and insulating the patient in a warm environment. With the head covered, the rate of rewarming is usually 0.5° to 2.0°C per hour. This technique is ideal for previously healthy patients who develop acute, mild primary accidental hypothermia. The patient must have sufficient fuel and glycogen to support endogenous thermogenesis.

There are reservations about the application of heat directly to the extremities of patients with chronic severe hypothermia. Extinguishing peripheral vasoconstriction in the dehydrated patient may precipitate core temperature "afterdrop" -- the continual decline in the core temperature after removal of the patient from the cold. This phenomenon results from conductive temperature equilibration and a circulatory convective mechanism. Rewarming frostbitten extremities before stabilization of the core temperature causes a significant core temperature afterdrop. In contrast, truncal heat application may minimize the risk of afterdrop.

Active rewarming is necessary under the following circumstances: core temperature<32°C (poikilothermia), cardiovascular instability, age extremes, CNS

dysfunction, endocrine insufficiency, or any suspicion of secondary hypothermia. *Active external rewarming* is best accomplished with forced-air heating blankets. Other options include radiant heat sources and hot packs. Monitoring a patient with hypothermia in a heated tub is extremely difficult. Electric blankets should be avoided because vasoconstricted skin is easily burned. Widely available *active core rewarming* options include heated inhalation, heated infusion, and lavage (gastric, colonic, mediastinal, thoracic, pleural). The therapeutic options also include hemodialysis, venovenous, and continuous arteriovenous rewarming, in addition to formal cardiopulmonary bypass.

Arteriovenous anastomoses (AVA) rewarming provides exogenous heat by immersion of the hands, forearms, feet, and calves in 44° to 45°C water. Airway rewarming with heated humidified oxygen (40° to 45°C) is a convenient option via mask or endotracheal tube. Although airway rewarming provides less heat than some other forms of active core rewarming, it eliminates respiratory heat loss and adds 1° to 2°C to the overall rewarming rate. Crystalloids should be heated to 40° to 42°C. The quantity of heat provided is significant only during massive volume resuscitations. The most efficient method for heating and delivering fluid or blood is with a countercurrent in-line heat exchanger. Heated irrigation of the gastrointestinal tract or bladder transfers minimal heat because of the limited available surface area. These methods should be reserved for patients in cardiac arrest and then used in combination with all available active rewarming techniques. Closed thoracic lavage is far more efficient in severely hypothermic patients with cardiac arrest. The hemithoraces are irrigated through two large-bore thoracostomy tubes that are inserted into the left or both of the hemithoraces. Thoracostomy tubes should not be placed in the left chest of a spontaneously perfusing patient for purposes of rewarming. Peritoneal lavage with the dialysate at 40° to 45°C efficiently transfers heat when delivered through two catheters with outflow suction. Like peritoneal dialysis, standard hemodialysis is especially useful for patients with electrolyte abnormalities, rhabdomyolysis, or toxin ingestions.

With extracorporeal venovenous rewarming, the blood is removed from a central venous catheter, heated to 40°C, and returned through a second central or peripheral venous catheter. Continuous arteriovenous rewarming involves the use of percutaneously inserted femoral arterial and contralateral femoral venous 8.5 Fr catheters. The blood pressure must be at least 60 mmHg. Heparin-bonded tubing obviates the need for systemic anticoagulation. Full circulatory support with an oxygenator can only be provided through formal cardiopulmonary bypass (CPB). Femoral flow rates of 2 to 3 L/min elevate the core temperature 1° to 2°C every 3 to 5 min. CPB should be considered in nonperfusing patients without documented contraindications to resuscitation. Circulatory support may also be the only effective option in patients with completely frozen extremities, or those with significant tissue destruction coupled with rhabdomyolysis.

There is no evidence that extremely rapid rewarming improves survival in perfusing patients. The best strategy is usually a combination of passive, truncal active, and active core rewarming techniques.

DRUG THERAPY

When a patient is hypothermic, target organs and the cardiovascular system respond

minimally to most medications. Moreover, cumulative doses can cause toxicity during rewarming because of increased binding of drugs to proteins, and impaired metabolism and excretion. As an example, the administration of repeated doses of digoxin or insulin would be ineffective while the patient is hypothermic, and the residual drugs are potentially toxic during rewarming.

Any pharmacologic manipulation of the depressed and vasoconstricted cardiovascular system should generally be avoided. If the hypotension does not respond to crystalloid infusion and rewarming, low-dose dopamine (2 to 5 ug/kg per min) support should be considered. Atrial arrhythmias should initially be monitored without intervention, as the ventricular response will be slow, and most will convert spontaneously during rewarming. When indicated, bretylium tosylate is the class III ventricular antiarrhythmic of choice. During ventricular fibrillation, it should initially be administered at a dose of 10 mg/kg. Bretylium uniquely increases the ventricular arrhythmia threshold at low temperatures, although the wisdom of prophylaxis is unresolved.

Initiating empirical therapy for adrenal insufficiency is usually not warranted unless there is a history suggesting steroid dependence, hypoadrenalism, or a failure to rewarm with standard therapy. However, the administration of parenteral levothyroxine to euthyroid patients with hypothermia is potentially hazardous. Because laboratory results can be delayed and confounded by the presence of the sick euthyroid syndrome (Chap. 330), historical clues or physical findings suggestive of hypothyroidism should be sought. When myxedema is the cause of hypothermia, the relaxation phase of the Achilles reflex is prolonged more than the contraction phase.

Hypothermia obscures most of the symptoms and signs of infection, notably fever and leukocytosis. Shaking rigors from infection may be mistaken for shivering. Except in mild cases, extensive cultures and repeated physical examinations are essential. Unless an infectious source is identified, empirical antibiotic prophylaxis is most warranted in the elderly, neonates, and immunocompromised patients.

Preventive measures should be discussed with high-risk individuals, such as the elderly or people whose work frequently exposes them to extreme cold. The importance of layered clothing and headgear, adequate shelter, increased caloric intake, and the avoidance of ethanol should be emphasized, along with access to rescue services.

FROSTBITE

Peripheral cold injuries include both freezing and nonfreezing injuries to tissue. Frostbite occurs when the tissue temperature drops below 0°C. Ice crystal formation subsequently distorts and destroys the cellular architecture. Once the vascular endothelium is damaged, stasis progresses rapidly to microvascular thrombosis. Tissue freezes quickly when in contact with thermal conductors such as metal or volatile solutions. Other predisposing factors include constrictive clothing or boots, immobility, or vasoconstrictive medications.

Clinically, it is most practical to classify frostbite as superficial or deep. Superficial does not entail tissue loss. Classically, frostbite is retrospectively graded like a burn once the resultant pathology is demarcated over time. First-degree frostbite causes only

anesthesia and erythema. The appearance of superficial vesiculation surrounded by edema and erythema is considered second degree (Plates IIA-18,IIA-19). Hemorrhagic vesicles reflect a serious injury to the microvasculature, and indicate third-degree frostbite. Fourth-degree injuries damage subcuticular, muscular, and osseous tissues.

PATHOPHYSIOLOGY

Peripheral cold injury involves a cascade of events. Endothelial cells are very susceptible to cold injury. In the prefreeze phase, plasma leaks and there is the development of microvascular vasoconstriction. The radiation of heat from underlying tissues initially prevents crystallization. The freeze phase usually begins with extracellular fluid crystallization. Water exits the cell and causes intracellular dehydration, hyperosmolality, and ultimately cellular shrinkage and demise. Damaged tissue releases thromboxane A_2 and prostaglandin F_{2a}, which produce platelet aggregation, leukocyte immobilization, and vasoconstriction.

After the tissue thaws, the second phase of the cascade causes progressive dermal ischemia. The microvasculature begins to collapse, arteriovenous shunting increases tissue pressures, and there is progressive formation of edema. Finally, thrombosis, ischemia, and superficial necrosis appear. The development of mummification and demarcation may take weeks to months.

CLINICAL PRESENTATION

The initial presentation of frostbite can be deceptively benign. The symptoms always include a sensory deficiency affecting light touch, pain, and temperature perception. The acral areas and distal extremities are the most common insensate areas. Some patients complain of a clumsy or "chunk of wood" sensation in the extremity.

Deep frostbitten tissue can appear waxy, mottled, yellow, or violaceous-white. Favorable presenting signs include some warmth or sensation with normal color. The injury is often superficial if the subcutaneous tissue is pliable or if the dermis can be rolled over boney prominences.

The two most common nonfreezing peripheral cold injuries are *chilblain (pernio)* and *immersion (trench) foot*. Chilblain results from neuronal and endothelial damage induced by repetitive exposure to dry cold. Young females, particularly those with a history of Raynaud's phenomenon, are most at risk. Persistent vasospasticity and vasculitis can cause erythema, mild edema, and pruritus. Eventually plaques, blue nodules, and ulcerations develop. These lesions typically involve the dorsa of the hands and feet. In contrast, immersion (trench) foot results from repetitive exposure to wet cold above the freezing point. The feet initially appear cyanotic, cold, and edematous. The subsequent development of bullae is often indistinguishable from frostbite. This vesiculation rapidly progresses to ulceration and liquefaction gangrene. Patients with milder cases complain of hyperhidrosis, cold sensitivity, and painful ambulation for many years.

Various ancillary tests have been used in an attempt to diagnose the severity of peripheral cold injuries. None consistently predicts the extent of injury at presentation. For example, angiography and magnetic resonance imaging can demonstrate the

patency of large vessels but not the microvasculature. Ultrasonography and digital plethysmography are also insensitive. Thermography and technetium scintigraphy help evaluate perfusion several days after rewarming.

TREATMENT

Frozen tissue should be rapidly and completely thawed by immersion in circulating water at 37° to 40°C. Rapid rewarming often produces an initial hyperemia. The early formation of clear distal large blebs is more favorable than smaller proximal dark hemorrhagic blebs. A common error is the premature termination of thawing, since the reestablishment of perfusion is intensely painful. Parenteral narcotics will be necessary with deep frostbite. If cyanosis persists after rewarming, the tissue compartment pressures should be monitored carefully.

Numerous experimental antithrombotic and vasodilatory treatment regimens have been evaluated. There is no conclusive evidence that dextran, heparin, steroids, calcium channel blockers, or hyperbaric oxygen salvage tissue. A treatment protocol for frostbite is summarized in Table 20-3.

Unless infection develops, any decision regarding debridement or amputation should be deferred until there is clear evidence of demarcation, mummification, and sloughing. The most common symptomatic sequelae reflect neuronal injury and the persistently abnormal sympathetic tone, including paresthesias, thermal misperception, and hyperhidrosis. Delayed findings include nail deformities, cutaneous carcinomas, and epiphyseal damage in children.

(Bibliography omitted in Palm version)

Back to Table of Contents

SECTION 3 -NERVOUS SYSTEM DYSFUNCTION

21. FAINTNESS, SYNCOPE, DIZZINESS, AND VERTIGO - *Robert B. Daroff*, *Mark D. Carlson*

Syncope is defined as transient loss of consciousness due to reduced cerebral blood flow. Syncope is associated with postural collapse and spontaneous recovery. It may occur suddenly, without warning, or may be preceded by symptoms of varying duration. These include faintness or lightheadedness, "dizziness" without true vertigo, a feeling of warmth, diaphoresis, nausea, and visual blurring occasionally proceeding to blindness. These presyncopal symptoms may increase in severity until loss of consciousness occurs or may resolve prior to loss of consciousness if the cerebral ischemia is corrected. The differentiation of syncope from seizure is an important, sometimes difficult, diagnostic problem.

Syncope may be benign when it occurs as a result of normal cardiovascular reflex effects on heart rate and vascular tone or malignant when due to a life-threatening arrhythmia. Syncope may occur as a single event or may be recurrent. Recurrent, unexplained syncope, particularly in an individual with structural heart disease, is associated with a high risk for death (40% mortality within 2 years).

SYNCOPE

At the beginning of a syncopal attack, the patient is nearly always in the upright position, either sitting or standing. A cardiac etiology, such as an arrhythmia, is exceptional in this respect. The patient is warned of the impending faint by a sense of "feeling bad," of giddiness, and of movement or swaying of the floor or surrounding objects. The patient becomes confused and may yawn, visual spots and dimming may occur, and the ears may ring. Nausea and vomiting sometimes accompany these symptoms. There is a striking pallor or ashen gray color of the face, and generalized perspiration ensues. In some patients, a gradual onset with presyncopal symptoms may allow time for protection against injury; in others the syncope is sudden and without warning. The onset varies from instantaneous to 10 to 30 s, rarely longer.

The depth and duration of unconsciousness vary. Sometimes the patient remains partly aware of the surroundings, or there may be profound coma. The patient may remain in this state for seconds or minutes. Usually the patient lies motionless with skeletal muscles relaxed, but a few clonic jerks of the limbs and face may occur shortly after consciousness is lost. Sphincter control is usually maintained, in contrast to a seizure. The pulse is feeble or apparently absent, the blood pressure may be low or undetectable, and breathing may be almost imperceptible. Once the patient is in a horizontal position, gravity no longer hinders the flow of blood to the brain. The strength of the pulse may then improve, color begins to return to the face, breathing becomes quicker and deeper, and consciousness is regained. There is usually an immediate recovery of consciousness. Some patients, however, may be keenly aware of physical weakness, and rising too soon may precipitate another faint. In other patients, particularly those with tachyarrhythmias, there may be no residual symptoms following the initial syncope. Headache and drowsiness, which with mental confusion are the usual sequelae of a seizure, do not follow a syncopal attack.

PATHOPHYSIOLOGY

The more common types of faint are reducible to a few simple mechanisms. Syncope results from a sudden impairment of brain metabolism, usually brought about by hypotension with reduction of cerebral blood flow. Several mechanisms subserve circulatory adjustments to the upright posture. Approximately three-fourths of the systemic blood volume is contained in the venous bed, and any interference in venous return may lead to a reduction in cardiac output. Cerebral blood flow may still be maintained, as long as systemic arterial vasoconstriction occurs, but when this adjustment fails, serious hypotension with resultant cerebral underperfusion to less than half of normal results in syncope. Normally, the pooling of blood in the lower parts of the body is prevented by: (1) pressor reflexes that induce constriction of peripheral arterioles and venules, (2) reflex acceleration of the heart by means of aortic and carotid reflexes, and (3) improvement of venous return to the heart by activity of the muscles of the limbs. Placing a normal person on a tilt table to relax the muscles and tilting upright slightly diminishes cardiac output and allows the blood to accumulate in the legs to a slight degree; this may then be followed by a slight transitory fall in systolic arterial pressure and, in patients with defective vasomotor reflexes, may produce faints.

CAUSES OF SYNCOPE

Transiently decreased cerebral blood flow is usually due to one of three general mechanisms: disorders of vascular tone or blood volume, cardiovascular disorders including cardiac arrhythmias, or cerebrovascular disease (Table 21-1). Not infrequently, however, the cause of syncope is multifactorial.

Disorders of Vascular Tone or Blood Volume Disorders of autonomic control of the heart and circulation account for at least half of syncopal episodes. These disorders share common pathophysiologic mechanisms: a cardioinhibitory component (e.g., bradycardia due to increased efferent vagal activity), a vasodepressor component (e.g., inappropriate vasodilatation due to sympathetic withdrawal), or both.

Vasovagal (Vasodepressor, Neurocardiogenic) Syncope This form of syncope is the common faint that may be experienced by normal persons and accounts for approximately half of all episodes of syncope. It is frequently recurrent and commonly precipitated by a hot or crowded environment, alcohol, extreme fatigue, severe pain, hunger, prolonged standing, and emotional or stressful situations. Episodes are often preceded by a presyncopal prodrome lasting seconds to minutes. Vasovagal syncope rarely occurs in the supine position. The individual is usually sitting or standing and experiences weakness, nausea, diaphoresis, lightheadedness, blurred vision, and often a forceful heart beat with tachycardia followed by cardiac slowing prior to loss of consciousness. The individual appears pallid and has decreasing blood pressure prior to syncope. The duration of unconsciousness is rarely longer than a few minutes if the conditions that provoke the episode are reversed. Consciousness is usually regained shortly after assuming a recumbent posture, but unconsciousness may be prolonged if an individual remains upright. Although commonly benign, vasovagal syncope can be associated with prolonged asystole and hypotension, resulting in injury.

Vasovagal syncope occurs in the setting of increased sympathetic activity and venous pooling. Under these conditions, vigorous myocardial contraction of a relatively empty left ventricle activates ventricular mechanoreceptors and vagal afferent nerve fibers, inhibiting sympathetic efferent activity and increasing parasympathetic efferent activity. The resultant vasodilatation and bradycardia induce hypotension and syncope.

The central nervous system (CNS) mechanisms responsible for vasovagal syncope are not clear. Animal studies, not confirmed in humans, suggest that endogenous opiates (endorphins) may play a role. Serotonin (5-hydroxytryptamine) participates in blood pressure regulation and may also be involved in inhibition of sympathetic efferent activity (and arterial vasodilatation) associated with vasovagal syncope.

Although the reflex involving myocardial mechanoreceptors is the mechanism usually accepted as responsible for vasovagal syncope, other reflexes may also be operative. Patients with transplanted (denervated) hearts have experienced cardiovascular responses identical to those present during vasovagal syncope. This should not be possible if the response depends solely on the reflex mechanisms described above unless the transplanted heart has become reinnervated. Moreover, vasovagal syncope often occurs in response to stimuli (fear, emotional stress, or pain) that may not be associated with venous pooling in the lower extremities, which suggests a cognitive or cortical component to the reflex. Thus, a variety of afferent and efferent responses may cause vasovagal syncope.

Postural (Orthostatic) Hypotension This occurs in patients who have a chronic defect in, or variable instability of, vasomotor reflexes. Systemic arterial blood pressure falls on assumption of upright posture due to loss of vasoconstriction reflexes in resistance and capacitance vessels of the lower extremities. Although the syncopal attack differs little from vasodepressor syncope, the effect of posture is critical. Sudden rising from a recumbent position or standing quietly are precipitating circumstances. *Orthostatic hypotension may be the cause of syncope in up to 30% of the elderly; polypharmacy with antihypertensive or antidepressant drugs is often a contributor in these patients.*

Postural syncope may occur in otherwise normal persons with defective postural reflexes. Patients with *idiopathic postural hypotension* may be identified by a characteristic response to upright tilt on a table. Initially, the blood pressure diminishes slightly before stabilizing at a lower level. Shortly thereafter, the compensatory reflexes fail and the systemic arterial pressure falls precipitously. The condition is often familial.

Orthostatic hypotension, often accompanied by disturbances in sweating, impotence, and sphincter difficulties, is also a primary feature of the autonomic nervous system disorders discussed in Chap. 366 and listed in Table 366-1. The most common causes of neurogenic orthostatic hypotension are chronic diseases of the peripheral nervous system that involve postganglionic unmyelinated fibers (e.g., diabetic, nutritional, and amyloid polyneuropathy). Much less common are the multisystem atrophies, which are CNS disorders in which orthostatic hypotension is associated with (1) parkinsonism (Shy-Drager syndrome), (2) progressive cerebellar degeneration, or (3) a more variable parkinsonian and cerebellar syndrome (striatonigral degeneration). Very rarely, an acute postganglionic dysautonomia has been reported that appears to represent a variant of Guillain-Barre syndrome (Chap. 378).

There are several additional causes of postural syncope: (1) After physical deconditioning (such as after prolonged illness with recumbency, especially in elderly individuals with reduced muscle tone) or after prolonged weightlessness, as in space flight; (2) after sympathectomy that has abolished vasopressor reflexes; and (3) in patients receiving antihypertensive or vasodilator drugs and those who are hypovolemic because of diuretics, excessive sweating, diarrhea, vomiting, hemorrhage, or adrenal insufficiency.

Carotid Sinus Hypersensitivity Syncope due to carotid sinus hypersensitivity is precipitated by pressure on the carotid sinus baroreceptors, which are located just cephalad to the bifurcation of the common carotid artery. This typically occurs in the setting of shaving, a tight collar, or turning the head to one side. Carotid sinus hypersensitivity occurs predominantly in men, most of whom are 50 years of age or older. Activation of carotid sinus baroreceptors gives rise to impulses carried via the nerve of Hering, a branch of the glossopharyngeal nerve, to the medulla oblongata. These afferent impulses activate efferent sympathetic nerve fibers to the heart and blood vessels, cardiac vagal efferent nerve fibers, or both. In patients with carotid sinus hypersensitivity, these responses may cause sinus arrest or atrioventricular (AV) block (a cardioinhibitory response), vasodilatation (a vasodepressor response), or both (a mixed response). Although originally described in 1933, the mechanisms responsible for the syndrome are not clear, and validated diagnostic criteria do not exist; some authorities have questioned its very existence.

Situational Syncope A variety of activities, including cough, deglutition, micturition, and defecation, are associated with syncope in susceptible individuals. These syndromes are caused, at least in part, by abnormal autonomic control and may involve a cardioinhibitory response, a vasodepressor response, or both. Cough, micturition, and defecation are associated with maneuvers (such as Valsalva, straining, and coughing) that may contribute to hypotension and syncope by decreasing venous return. Increased intracranial pressure secondary to the increased intrathoracic pressure may also contribute by decreasing cerebral blood flow.

Cough syncope typically occurs in men with chronic bronchitis or chronic obstructive lung disease during or immediately after prolonged coughing fits. Micturition syncope occurs predominantly in middle-aged and older men, particularly those with prostatic hypertrophy and obstruction of the bladder neck; loss of consciousness usually occurs at night during or immediately after voiding. Deglutition and defecation syncope occur in men and women. Deglutition syncope may be associated with esophageal disorders, particularly esophageal spasm. In some individuals, particular foods and carbonated or cold beverages initiate episodes by activating esophageal sensory receptors that trigger reflex sinus bradycardia or AV block. Defecation syncope is probably secondary to a Valsalva maneuver in older individuals with constipation.

Glossopharyngeal Neuralgia Syncope due to glossopharyngeal neuralgia is preceded by pain in the oropharynx, tonsillar fossa, or tongue. Loss of consciousness is usually associated with asystole rather than vasodilatation. The mechanism is thought to involve activation of afferent impulses in the glossopharyngeal nerve which terminate in the nucleus solitarius of the medulla and, via collaterals, activate the dorsal motor

nucleus of the vagus nerve.

Cardiovascular Disorders Cardiac syncope results from a sudden reduction in cardiac output, caused most commonly by a cardiac arrhythmia. In normal individuals, heart rates between 30 and 180 beats per minutes (bpm) do not reduce cerebral blood flow, especially if the person is in the supine position. As the heart rate decreases, ventricular filling time and stroke volume increase to maintain normal cardiac output. At rates below 30 bpm, stroke volume can no longer increase to compensate adequately for the decreased heart rate. At rates greater than approximately 180 bpm, ventricular filling time is inadequate to maintain adequate stroke volume. In either case, cerebral hypoperfusion and syncope may occur. Upright posture, cerebrovascular disease, anemia, and coronary, myocardial, or valvular disease all reduce the tolerance to alterations in rate.

Bradyarrhythmias (Chap. 229) may occur as a result of an abnormality of impulse generation (e.g., sinoatrial arrest) or impulse conduction (e.g.,AVblock). Either may cause syncope if the escape pacemaker rate is insufficient to maintain cardiac output. Syncope due to bradyarrhythmias may occur abruptly, without presyncopal symptoms, and recur several times daily. Patients with *sick sinus syndrome* may have sinus pauses (>3 s), and those with syncope due to high-degree AV block (*Stokes-Adams-Morgagni syndrome*) may have evidence of conduction system disease (e.g., prolonged PR interval, bundle branch block). However, the arrhythmia is often transitory, and the surface electrocardiogram or continuous electrocardiographic monitor (Holter monitor) taken later may not reveal the abnormality. The *bradycardia-tachycardia syndrome* is a common form of sinus node dysfunction in which syncope generally occurs as a result of marked sinus pauses following termination of paroxysmal supraventricular tachycardia. Drugs are a common cause for bradyarrhythmias, particularly in patients with underlying structural heart disease. Digoxin,b-adrenergic receptor antagonists, calcium channel blockers, and many antiarrhythmic drugs may suppress sinoatrial node impulse generation or slow AV nodal conduction.

Syncope due to a *tachyarrhythmia* (Chap. 230) is usually preceded by palpitation or lightheadedness but may occur abruptly with no warning symptoms. *Supraventricular tachyarrhythmias* are unlikely to cause syncope in individuals with structurally normal hearts but may do so if they occur in patients with: (1) heart disease that also compromises cardiac output, (2) cerebrovascular disease, (3) a disorder of vascular tone or blood volume, or (4) a rapid ventricular rate. These tachycardias result most commonly from paroxysmal atrial flutter, atrial fibrillation, or reentry involving theAV node or accessory pathways that bypass part or all of the AV conduction system. Patients with the *Wolff-Parkinson-White syndrome* may experience syncope when a very rapid ventricular rate occurs due to reentry across an accessory AV connection.

In patients with structural heart disease, ventricular tachycardia, sometimes associated with ventricular fibrillation, is a common cause of syncope, particularly in patients with a prior myocardial infarction. Patients with aortic valvular stenosis and hypertrophic obstructive cardiomyopathy are also at risk for ventricular tachycardia. Individuals with abnormalities of ventricular repolarization (prolongation of the QT interval) are at risk to develop polymorphic ventricular tachycardia (*torsade de pointes*). Those with the inherited form of this syndrome often have a family history of sudden death in young

individuals. Genetic markers can identify some patients with familial long-QT syndrome but the clinical utility of these markers remains unproven. Drugs (i.e., certain antiarrhythmics and erythromycin) and electrolyte disorders (i.e., hypokalemia, hypocalcemia, hypomagnesemia) can prolong the QT interval and predispose to torsade de pointes. Antiarrhythmic medications may precipitate ventricular tachycardia, particularly in patients with structural heart disease.

In addition to arrythmias, syncope may also occur with a variety of structural cardiovascular disorders. Episodes are usually precipitated when the cardiac output cannot increase to compensate adequately for peripheral vasodilatation. Peripheral vasodilatation may be appropriate, such as following exercise, or may occur due to inappropriate activation of left ventricular mechanoreceptor reflexes, as occurs in aortic outflow tract obstruction (aortic valvular stenosis or hypertrophic obstructive cardiomyopathy). Obstruction to forward flow is the most common reason that cardiac output cannot increase. Pericardial tamponade is a rare cause of syncope. Syncope occurs in up to 10% of patients with massive pulmonary embolism and may occur with exertion in patients with severe primary pulmonary hypertension. The cause is an inability of the right ventricle to provide appropriate cardiac output in the presence of obstruction or increased pulmonary vascular resistance. Loss of consciousness is usually accompanied by other symptoms such as chest pain and dyspnea. Atrial myxoma, a prosthetic valve thrombus, and, rarely, mitral stenosis may impair left ventricular filling, decrease cardiac output, and cause syncope.

Cerebrovascular Disease Cerebrovascular disease alone rarely causes syncope but may lower the threshold for syncope in patients with other causes. The vertebrobasilar arteries, which supply brainstem centers responsible for maintaining consciousness, are usually involved when cerebrovascular disease causes or contributes to syncope. An exception is the rare patient with tight bilateral carotid stenosis and recurrent syncope, often precipitated by standing or walking. Most patients who experience lightheadedness or syncope due to cerebrovascular disease also have symptoms of focal neurologic ischemia, such as arm or leg weakness, diplopia, ataxia, dysarthria, or sensory disturbances. Basilar artery migraine is a rare disorder that causes syncope in adolescents.

DIFFERENTIAL DIAGNOSIS

Anxiety Attacks and the Hyperventilation Syndrome Anxiety, such as occurs in panic attacks, is frequently interpreted as a feeling of faintness or dizziness resembling presyncope. The symptoms are not accompanied by facial pallor and are not relieved by recumbency. The diagnosis is made on the basis of the associated symptoms such as a feeling of impending doom, air hunger, palpitations, and tingling of the fingers and perioral region. Attacks can often be reproduced by hyperventilation, resulting in hypocapnia, alkalosis, increased cerebrovascular resistance, and decreased cerebral blood flow. The release of epinephrine in anxiety states also contributes to the symptoms.

Seizures A seizure may be heralded by an aura, which is caused by a focal seizure discharge and hence has localizing significance. The aura is usually followed by a rapid return to normal or by a loss of consciousness. Injury from falling is frequent in a seizure

and rare in syncope, because only in seizures are protective reflexes abolished instantaneously. Tonic-convulsive movements are characteristic of seizures and usually do not occur with syncope, although, as stated above, brief tonic-clonic seizure-like activity can accompany fainting episodes. The period of unconsciousness tends to be longer in seizures than in syncope. Urinary incontinence is frequent in seizures and rare in syncope. The return of consciousness is prompt in syncope, slow after a seizure. Mental confusion, headache, and drowsiness are common sequelae of seizures; physical weakness with a clear sensorium characterizes the postsyncopal state. Repeated spells of unconsciousness in a young person at a rate of several per day or month are more suggestive of epilepsy than syncope.

Hypoglycemia Severe hypoglycemia is usually due to a serious disease such as a tumor of the islets of Langerhans; advanced adrenal, pituitary, or hepatic disease; or to excessive administration of insulin.

Acute Hemorrhage Hemorrhage, usually within the gastrointestinal tract, is an occasional cause of syncope. In the absence of pain and hematemesis, the cause of the weakness, faintness, or even unconsciousness may remain obscure until the passage of a black stool.

Hysterical Fainting The attack is usually unattended by an outward display of anxiety. Lack of change in pulse and blood pressure or color of the skin and mucous membranes distinguish it from the vasodepressor faint.

Approach to the Patient

The diagnosis of syncope is often challenging. The cause may only be apparent at the time of the event, leaving few, if any, clues when the patient is seen later by the physician. In dealing with patients who have fainted, the physician should think first of those causes of fainting that constitute a therapeutic emergency. Among them are massive internal hemorrhage or myocardial infarction, which may be painless, and cardiac arrythmias. In elderly persons, a sudden faint, without obvious cause, should arouse the suspicion of complete heart block or a tachyarrhythmia, even though all findings are negative when the patient is seen.

An algorithmic approach to syncope is presented in Fig. 21-1. A careful history is the most important diagnostic tool, both to suggest the correct cause and to exclude other important potential causes (Table 21-1). Although no single element of the history is specific for a particular etiology of syncope, the nature of the events and their time course immediately prior to, during, and after an episode often provide valuable etiologic clues. Loss of consciousness in particular situations, such as during venipuncture, micturition, or in association with volume depletion, suggests an abnormality of vascular tone. The position of the patient at the time of the syncopal episode is very important; syncope in the supine position is unlikely to be vasovagal and suggests an arrhythmia or a seizure. Syncope due to carotid sinus syndrome may occur when the individual is wearing a shirt with a tight collar, turning the head (turning to look while driving in reverse), or manipulating the neck (as in shaving). The patient's medications must be noted, including nonprescription drugs or health store supplements, with particular attention to recent changes.

The physical examination should include evaluation of heart rate and blood pressure in the supine, sitting, and standing positions. In patients with unexplained recurrent syncope, an attempt to reproduce an attack may assist in diagnosis. Anxiety attacks induced by hyperventilation can be reproduced readily by having the patient breathe rapidly and deeply for 2 to 3 min. Cough syncope may be reproduced by inducing the Valsalva maneuver. Carotid sinus massage should generally be avoided, even in patients with suspected carotid sinus hypersensitivity; it is a risky procedure that can cause a transient ischemic attack (TIA) or stroke in susceptible individuals.

Diagnostic Tests The choice of diagnostic tests should be guided by the history and the physical examination. Measurements of serum electrolytes, glucose, and the hematocrit may help to establish the cause of syncope. Cardiac enzymes should be evaluated if myocardial ischemia is suspected. Blood and urine toxicology screens may reveal the presence of alcohol or other drugs. In patients with possible adrenocortical insufficiency, plasma aldosterone and mineralocorticoid levels should be obtained.

Although the surface electrocardiogram is unlikely to provide a definitive diagnosis, it may provide clues to the cause of syncope *and should be performed in almost all patients*. The presence of conduction abnormalities (PR prolongation and bundle branch block) suggests a bradyarrhythmia, whereas pathologic Q waves or prolongation of the QT interval suggests a ventricular tachyarrhythmia. Inpatients should undergo continuous electrocardiographic monitoring; outpatients should wear a Holter monitor for 24 to 48 h. Whenever possible, symptoms should be correlated with the occurrence of arrhythmias. Continuous electrocardiographic monitoring may establish the cause of syncope in as many as 15% of patients. Cardiac event monitors may be useful in patients with infrequent symptoms, particularly in patients with presyncope. The presence of a late potential on a signal-averaged electrocardiogram is associated with increased risk for ventricular tacharrhythmias in patients with a prior myocardial infarction. Low-voltage (visually inapparent) T wave alternans is also associated with development of sustained ventricular arrhythmias.

Invasive cardiac electrophysiologic testing provides diagnostic and prognostic information regarding sinus node function,AVconduction, and supraventricular and ventricular arrhythmias. Abnormal findings include prolongation of the sinus node recovery time, prolongation of the histoventricle (HV) interval, induction of a supraventricular arrhythmia associated with hypotension, or induction of a supraventricular arrhythmia. Prolongation of the sinus node recovery time (>1500 ms) is a specific finding (85 to 100%) for diagnosis of sinus node dysfunction but has a low sensitivity; continuous electrocardiographic monitoring is usually more effective for diagnosing this abnormality. Prolongation of the HV interval and conduction block below the His bundle indicate that His-Purkinje disease may be responsible for syncope. Although an HV interval>100 ms is abnormal, this finding is not common in patients with syncope, and some patients with shorter intervals are also at risk for AV block. Programmed stimulation for ventricular arrhythmias is most useful in patients who have experienced a myocardial infarction; the sensitivity and specificity of this technique is lower in patients with normal hearts or those with heart disease other than coronary artery disease.

Upright tilt table testing is indicated for recurrent syncope, a single syncopal episode that caused injury, or a single syncopal event in a "high-risk" setting (pilot, commercial vehicle driver, etc.), whether or not there is a history of preexisting heart disease or prior vasovagal episodes. In susceptible patients, upright tilt at an angle between 60 and 80° for 30 to 60 min induces a vasovagal episode. The protocol can be shortened if upright tilt is combined with intravenous administration of drugs that cause venous pooling or increase adrenergic stimulation (isoproterenol, nitroglycerin, edrophonium, or adenosine). The sensitivity and specificity of tilt table testing is difficult to ascertain because of the lack of validated criteria. Moreover, the reflexes responsible for vasovagal syncope can be elicited in most, if not all, individuals given the appropriate stimulus. The reported accuracy of the test ranges from 30 to 80%, depending on the population studied and the techniques used. Whereas the reproducibility of a negative test is 85 to 100%, the reproducibility of a positive tilt table test is only between 62 and 88%.

A variety of other tests may be useful to determine the presence of structural heart disease that may cause syncope. The echocardiogram with Doppler examination detects valvular, myocardial, and pericardial abnormalities. The echocardiogram is the "gold standard" for the diagnosis of hypertrophic cardiomyopathy and atrial myxoma. Cardiac cine magnetic resonance (MR) imaging provides an alternative noninvasive modality that may be useful for patients in whom diagnostic-quality echocardiographic images cannot be obtained. This test is also indicated for patients suspected of having arrhythmogenic right ventricular dysplasia or right ventricular outflow tract ventricular tachycardia. Both are associated with right ventricular structural abnormalities that are better visualized on MR imaging than by echocardiogram. Exercise testing may detect ischemia or exercise-induced arrhythmias. In some patients, cardiac catheterization may be necessary to diagnose the presence or severity of coronary artery disease or valvular abnormalities. Ultrafast computed tomographic scan, ventilation-perfusion scan, or pulmonary angiography are indicated in patients in whom syncope may be due to pulmonary embolus.

In possible cases of cerebrovascular syncope, a variety of neuroimaging tests may be indicated, including Doppler ultrasound studies of the carotid and vertebrobasilar systems,MRimaging, MR angiography, and x-ray angiography of the cerebral vasculature (Chaps. 358 and361). Electroencephalography is indicated if seizures are suspected.

TREATMENT

The treatment of syncope is directed toward the underlying cause. This discussion will focus on the treatment of disorders of autonomic control.**Arrythmias are discussed in Chaps. 229 and 230, valvular heart diseases in Chap. 236, and cerebrovascular disorders in Chap. 361.*

Certain precautions should be taken regardless of the cause of syncope. At the first sign of symptoms, patients should make every effort to avoid injury should they lose consciousness. Patients with frequent episodes, or those who have experienced syncope without warning symptoms should avoid situations in which sudden loss of consciousness might result in injury (e.g., climbing ladders, swimming alone, operating

heavy machinery, driving). Patients should lower their head to the extent possible, and preferably should lie down. Lowering the head by bending at the waist should be avoided because it may further compromise venous return to the heart. When appropriate, family members or other close contacts should be educated as to the problem. This will ensure appropriate therapy and may prevent delivery of inappropriate therapy (chest compressions associated with cardiopulmonary resuscitation) that may inflict trauma.

Patients who have lost consciousness should be placed in a position that maximizes cerebral blood flow, offers protection from trauma, and secures the airway. Whenever possible, the patient should be placed supine with the head turned to the side to prevent aspiration and the tongue from blocking the airway. Assessment of the pulse and direct cardiac auscultation may assist in determining if the episode is associated with a bradyarrythmia or tachyarrhythmia. Clothing that fits tightly around the neck or waist should be loosened. Peripheral stimulation, such as by sprinkling cold water on the face, may be helpful. Patients should not be given anything by mouth or be permitted to rise until the sense of physical weakness has passed.

Patients with vasovagal syncope should be instructed to avoid situations or stimuli that have caused them to lose consciousness. Episodes associated with intravascular volume depletion may be prevented by salt and fluid loading prior to provocative events.b-Adrenoceptor antagonists, the most widely used agents, mitigate the increase in myocardial contractility that stimulates left ventricular mechanoreceptors and also block central serotonin receptors. Disopyramide, a vagolytic with negative inotropic properties, and another vagolytic, transdermal scopolamine, are used to treat vasovagal syncope. Paroxetine, a serotonin reuptake inhibitor used for depression, appears to be an effective treatment, as are theophylline and ephedrine. Midodrine, ana agonist, has been a first-line agent for some patients. Permanent cardiac pacing is effective for patients with frequent episodes of vasovagal syncope and is indicated for those with prolonged asystole associated with vasovagal episodes.

Patients with orthostatic hypotension should be instructed to rise slowly and systematically (supine to seated, seated to standing) from the bed or a chair. Movement of the legs prior to rising facilitates venous return from the lower extremities. Whenever possible, medications that aggravate the problem (vasodilators, diuretics, etc.) should be discontinued. Elevation of the head of the bed [20 to 30 cm (8 to 12 in.)] and use of elastic stockings may help.

Therapeutic modalities include devices that prevent lower limb blood pooling, such as an antigravity or g suit or elastic stockings; salt loading; and a variety of pharmacologic agents including sympathomimetic amines, monamine oxidase inhibitors, beta blockers, and levodopa.**The treatment of orthostatic hypotension secondary to central or peripheral disorders of the autonomic nervous system is discussed in Chap. 366.*

Glossopharyngeal neuralgia is treated with carbamazepine, which is effective for the syncope as well as for the pain. Patients with carotid sinus syndrome should be instructed to avoid clothing and situations that stimulate carotid sinus baroreceptors. Patients should turn their entire body, rather than just their head, to look to one side. Those with intractable syncope due to the cardioinhibitory response to carotid sinus

stimulation should undergo permanent pacemaker implantation.

Patients with syncope should be hospitalized when the episode may have resulted from a life-threatening abnormality or if recurrence with significant injury seems likely. These individuals should be admitted to a bed with continuous electrocardiographic monitoring. Patients who are known to have a normal heart and for whom the history strongly suggests vasovagal or situational syncope may be treated as outpatients if the episodes are neither frequent nor severe.

DIZZINESS AND VERTIGO

Dizziness is a common and often vexing symptom. Patients use the term to encompass a variety of sensations, including those that seem semantically appropriate (e.g., lightheadedness, faintness, spinning, giddiness, etc.) and those that are misleadingly inappropriate, such as mental confusion, blurred vision, headache, or tingling. Moreover, some individuals with gait disorders complain of dizziness despite the absence of vertigo or other abnormal cephalic sensations. The causes include peripheral neuropathy, myelopathy, spasticity, parkinsonian rigidity, and cerebellar ataxia. In this context, the term *dizziness* is being used to describe disturbed mobility. There may be mild associated lightheadedness, particularly with impaired sensation from the feet or poor vision; this is known as *multiple-sensory-defect dizziness* and occurs in elderly individuals who complain of dizziness only during ambulation. Decreased position sense (secondary to neuropathy or myelopathy) and poor vision (from cataracts or retinal degeneration) create an overreliance on the aging vestibular apparatus. A less precise but sometimes comforting designation to patients is *benign dysequilibrium of aging*. Thus, a careful history is necessary to determine exactly what a patient who states, "Doctor, I'm dizzy," is experiencing. After eliminating the misleading symptoms or gait disturbance, "dizziness" usually means either *faintness* (presyncope) or *vertigo* (an illusory or hallucinatory sense of movement of the body or environment, most often a feeling of spinning). Operationally, dizziness is classified into three categories: (1) faintness, (2) vertigo, and (3) miscellaneous head sensations.

FAINTNESS

Prior to an actual faint (syncope), there are often prodromal presyncopal symptoms (faintness) reflecting ischemia to a degree insufficient to impair consciousness (see above).

VERTIGO

Vertigo is usually due to a disturbance in the vestibular system. The end organs of this system, situated in the bony labyrinths of the inner ears, consist of the three semicircular canals and the otolithic apparatus (utricle and saccule) on each side. The canals transduce angular acceleration, while the otoliths transduce linear acceleration and static gravitational forces, the latter providing a sense of head position in space. The neural output of the end organs is conveyed to the vestibular nuclei in the brainstem via the eighth cranial nerve. The principal projections from the vestibular nuclei are to the nuclei of cranial nerves III, IV, and VI, the spinal cord, the cerebral cortex, and the cerebellum. The vestibuloocular reflex (VOR) serves to maintain visual

stability during head movement and depends on direct projections from the vestibular nuclei to the sixth cranial nerve (abducens) nuclei in the pons and, via the medial longitudinal fasciculus, to the third (oculomotor) and fourth (trochlear) cranial nerve nuclei in the midbrain. These connections account for the nystagmus (to-and-fro oscillation of the eyes) that is an almost invariable accompaniment of vestibular dysfunction. The vestibular nerves and nuclei project to areas of the cerebellum (primarily the flocculus and nodulus) that modulate the VOR. The vestibulospinal pathways assist in the maintenance of postural stability. Projections to the cerebral cortex, via the thalamus, provide conscious awareness of head position and movement.

The vestibular system is one of three sensory systems subserving spatial orientation and posture; the other two are the visual system (retina to occipital cortex) and the somatosensory system that conveys peripheral information from skin, joint, and muscle receptors. The three stabilizing systems overlap sufficiently to compensate (partially or completely) for each other's deficiencies. Vertigo may represent either physiologic stimulation or pathologic dysfunction in any of the three systems.

Physiologic Vertigo This occurs when (1) the brain is confronted with a mismatch among the three stabilizing sensory systems; (2) the vestibular system is subjected to unfamiliar head movements to which it has never adapted, such as in seasickness; or (3) unusual head/neck positions, such as the extreme extension when painting a ceiling. Intersensory mismatch explains carsickness, height vertigo, and the visual vertigo most commonly experienced during motion picture chase scenes; in the latter, the visual sensation of environmental movement is unaccompanied by concomitant vestibular and somatosensory movement cues. *Space sickness*, a frequent transient effect of active head movement in the weightless zero-gravity environment, is another example of physiologic vertigo.

Pathologic Vertigo This results from lesions of the visual, somatosensory, or vestibular systems. Visual vertigo is caused by new or incorrect spectacles or by the sudden onset of an extraocular muscle paresis with diplopia; in either instance,CNScompensation rapidly counteracts the vertigo. Somatosensory vertigo, rare in isolation, is usually due to a peripheral neuropathy that reduces the sensory input necessary for central compensation when there is dysfunction of the vestibular or visual systems.

The most common cause of pathologic vertigo is vestibular dysfunction. The vertigo is frequently accompanied by nausea, jerk nystagmus, postural unsteadiness, and gait ataxia. Since vertigo increases with rapid head movements, patients tend to hold their heads still.

Labyrinthine Dysfunction This causes severe rotational or linear vertigo. When rotational, the hallucination of movement, whether of environment or self, is directed away from the side of the lesion. The fast phases of nystagmus beat away from the lesion side, and the tendency to fall is toward the side of the lesion.

When the head is straight and immobile, the vestibular end organs generate a tonic resting firing frequency that is equal from the two sides. With any rotational acceleration, the anatomic positions of the semicircular canals on each side necessitate an increased firing rate from one and a commensurate decrease from the other. This change in

neural activity is ultimately projected to the cerebral cortex, where it is summed with inputs from the visual and somatosensory systems to produce the appropriate conscious sense of rotational movement. After cessation of movement, the firing frequencies of the two end organs reverse; the side with the initially increased rate decreases, and the other side increases. A sense of rotation in the opposite direction is experienced; since there is no actual head movement, this hallucinatory sensation is *physiologic postrotational vertigo*.

Any disease state that changes the firing frequency of an end organ, producing unequal neural input to the brainstem and ultimately the cerebral cortex, causes vertigo. The symptom can be conceptualized as the cortex inappropriately interpreting the abnormal neural input from the brainstem as indicating actual head rotation. Transient abnormalities produce short-lived symptoms. With a fixed unilateral deficit, central compensatory mechanisms ultimately diminish the vertigo. Since compensation depends on the plasticity of connections between the vestibular nuclei and the cerebellum, patients with brainstem or cerebellar disease have diminished adaptive capacity, and symptoms may persist indefinitely. Compensation is always inadequate for severe fixed bilateral lesions despite normal cerebellar connections: these patients are permanently symptomatic.

Acute unilateral labyrinthine dysfunction is caused by infection, trauma, and ischemia. Often, no specific etiology is uncovered, and the nonspecific terms *acute labyrinthitis*, *acute peripheral vestibulopathy*, or *vestibular neuritis* are used to describe the event. The attacks are brief and leave the patient for some days with a mild positional vertigo. Infection with herpes simplex virus type 1 has been implicated. It is impossible to predict whether a patient recovering from the first bout of vertigo will have recurrent episodes.

Acute bilateral labyrinthine dysfunction is usually the result of toxins such as drugs or alcohol. The most common offending drugs are the aminoglycoside antibiotics which damage the fine hair cells of the vestibular end organs and may cause a permanent disorder of equilibrium.

Recurrent unilateral labyrinthine dysfunction, in association with signs and symptoms of cochlear disease (progressive hearing loss and tinnitus), is usually due to Meniere's disease (Chap. 29). When auditory manifestations are absent, the term *vestibular neuronitis* denotes recurrent monosymptomatic vertigo.TIAs of the posterior cerebral circulation (vertebrobasilar insufficiency) very infrequently cause recurrent vertigo without concomitant motor, sensory, visual, cranial nerve, or cerebellar signs.

Positional vertigo is precipitated by a recumbent head position, either to the right or to the left. Benign paroxysmal positional (or positioning) vertigo (BPPV) of the posterior semicircular canal is particularly common. Although the condition may be due to head trauma, usually no precipitating factors are identified. It generally abates spontaneously after weeks or months. The vertigo and accompanying nystagmus have a distinct pattern of latency, fatigability, and habituation that differs from the less common central positional vertigo (Table 21-2) due to lesions in and around the fourth ventricle. Moreover, the pattern of nystagmus in posterior canal BPPV is distinctive. The lower eye displays a large-amplitude torsional nystagmus, and the upper eye has a lesser degree of torsion combined with upbeating nystagmus. If the eyes are directed to the

upper ear, the vertical nystagmus in the upper eye increases in amplitude.

Vertigo of vestibular nerve origin may occur with diseases that involve the nerve in the petrous bone or the cerebellopontine angle. Except that it is less severe and less frequently paroxysmal, it has many of the characteristics of labyrinthine vertigo. The adjacent auditory division of the eighth cranial nerve also may be affected, which explains the frequent association of vertigo with tinnitus and deafness. The function of the eighth cranial nerve may be disturbed by tumors of the lateral recess (especially schwannomas), less frequently by meningeal inflammation in this region and, rarely, by an abnormal vessel that compresses the nerve.

Schwannomas involving the eighth cranial nerve (*acoustic neuroma*) grow slowly and produce such a gradual reduction of labyrinthine output that central compensatory mechanisms can prevent or minimize the vertigo; auditory symptoms of hearing loss and tinnitus are the most common manifestations. While lesions of the brainstem or cerebellum can cause acute vertigo, associated signs and symptoms usually permit distinction from a labyrinthine etiology (Table 21-3). However, labyrinthine ischemia, presumably due to occlusion of the labyrinthine branch of the internal auditory artery, may be the sole manifestation of vertebrobasilar insufficiency; patients with this syndrome present with the abrupt onset of severe vertigo, nausea and vomiting without tinnitus or hearing loss. Occasionally, an acute lesion of the vestibulocerebellum may present with monosymptomatic vertigo indistinguishable from a labyrinthopathy.

Vestibular epilepsy, vertigo secondary to temporal lobe epileptic activity, is rare and almost always intermixed with other epileptic manifestations.

Psychogenic vertigo, usually a concomitant of panic attacks or agoraphobia (fear of large open spaces, crowds, or leaving the safety of home), should be suspected in patients so "incapacitated" by their symptoms that they adopt a prolonged housebound status. Most patients with organic vertigo attempt to function despite their discomfort. Organic vertigo is accompanied by nystagmus; a psychogenic etiology is almost certain when nystagmus is absent during a vertiginous episode.

Miscellaneous Head Sensations This designation is used, primarily for purposes of initial classification, to describe dizziness that is neither faintness nor vertigo. Cephalic ischemia or vestibular dysfunction may be of such low intensity that the usual symptomatology is not clearly identified. For example, a small decrease in blood pressure or a slight vestibular imbalance may cause sensations different from distinct faintness or vertigo but that may be identified properly during provocative testing techniques. Other causes of dizziness in this category are hyperventilation syndrome, hypoglycemia, and the somatic symptoms of a clinical depression; these patients should have normal neurologic examinations and vestibular function tests.

Approach to the Patient

The most important diagnostic tool is a careful history focused on the meaning of "dizziness" to the patient. Is it faintness? Is there a sensation of spinning? If either of these is affirmed and the neurologic examination is normal, appropriate investigations for the multiple etiologies of cephalic ischemia or vestibular dysfunction are undertaken.

When the meaning of "dizziness" is uncertain, provocative tests may be helpful. These office procedures simulate either cephalic ischemia or vestibular dysfunction. Cephalic ischemia is obvious if the dizziness is duplicated during maneuvers that produce orthostatic hypotension. Further provocation involves the Valsalva maneuver, which decreases cerebral blood flow and should reproduce ischemic symptoms.

The simplest provocative test for vestibular dysfunction is rapid rotation and abrupt cessation of movement in a swivel chair. This always induces vertigo that the patients can compare with their symptomatic dizziness. The intense induced vertigo may be unlike the spontaneous symptoms, but shortly thereafter, when the vertigo has all but subsided, a lightheadedness supervenes that may be identified as "my dizziness." When this occurs, the dizzy patient, originally classified as suffering from "miscellaneous head sensations," is now properly diagnosed as having mild vertigo secondary to a vestibulopathy.

Patients with symptoms of positional vertigo should be appropriately tested (Table 21-2); positional testing is more sensitive with special spectacles that preclude visual fixation (Frenzel lenses).

A final provocative test, requiring the use of Frenzel lenses, is vigorous head shaking in the horizontal plane for about 10 s. If nystagmus develops after the shaking stops, even in the absence of vertigo, vestibular dysfunction is demonstrated. The maneuver can then be repeated in the vertical plane. If the provocative tests establish the dizziness as a vestibular symptom, an evaluation of vestibular vertigo is undertaken.

Evaluation of Patients with Pathologic Vestibular Vertigo The evaluation depends on whether a central etiology is suspected (Table 21-3). If so, MR imaging of the head is mandatory. Such an examination is rarely helpful in cases of recurrent monosymptomatic vertigo with a normal neurologic examination. Typical BPPV requires no investigation after the diagnosis is made (Table 21-2).

Vestibular function tests serve to (1) demonstrate an abnormality when the distinction between organic and psychogenic is uncertain, (2) establish the side of the abnormality, and (3) distinguish between peripheral and central etiologies. The standard test is electronystagmography (calorics), where warm and cold water (or air) are applied, in a prescribed fashion, to the tympanic membranes, and the slow-phase velocities of the resultant nystagmus from the right and left ears are compared. A velocity decrease from one side indicates hypofunction ("canal paresis"). An inability to induce nystagmus with ice water denotes a "dead labyrinth." Some institutions have the capability of quantitatively determining various aspects of the vestibuloocular reflex using computer-driven rotational chairs and precise oculographic recording of the eye movements.

Hyperventilation is the cause of dizziness in many anxious individuals; tingling of the hands and face may be absent. Forced hyperventilation for 1 min is indicated for patients with enigmatic dizziness and normal neurologic examinations. Similarly, depressive symptoms (which patients usually insist are "secondary" to the dizziness) must alert the examiner to a clinical depression as the *cause*, rather than the effect, of

the dizziness.

CNSdisease can produce dizzy sensations of all types. Consequently, a neurologic examination is always required even if the history or provocative tests suggest a cardiac, peripheral vestibular, or psychogenic etiology. Any abnormality on the neurologic examination should prompt appropriate neurodiagnostic studies.

TREATMENT

Treatment of acute vertigo consists of bed rest and vestibular suppressant drugs such as antihistaminics (meclizine, dimenhydrinate, promethazine), or a tranquilizer with GABA-ergic effects (diazepam). If the vertigo persists beyond a few days, most authorities advise ambulation in an attempt to induce central compensatory mechanisms, despite the short-term discomfort to the patient. Chronic vertigo of labyrinthine origin may be treated with a systematized vestibular rehabilitation program to facilitate central compensation (see alsoTable 21-4).

BPPVis often self-limited but, when persistent, responds dramatically to specific repositioning exercise programs designed to empty particulate debris from the posterior semicircular canal. One of these exercises, the Epley procedure, is graphically demonstrated, in four languages, on a website for use in both physician's offices and self-treatment (http://www.charite.de/ch/neuro/vertigo.html).

Prophylactic measures to prevent recurrent vertigo are variably effective. Antihistamines are commonly utilized. Meniere's disease may respond to a diuretic or, more effectively, to a very low salt diet (1 g/day).

There are a variety of inner ear surgical procedures for refractory Meniere's disease, but these are only rarely necessary.

(Bibliography omitted in Palm version)

Back to Table of Contents

22. WEAKNESS, MYALGIAS, DISORDERS OF MOVEMENT, AND IMBALANCE - *Richard K. Olney*, *Michael J. Aminoff*

Normal motor function requires integrated muscle activity with appropriate modulation by neuronal activity in the cerebral cortex, basal ganglia, cerebellum, and spinal cord. Symptoms and signs of motor system dysfunction may include weakness, fatigue, myalgias, spasms, cramps, dyskinetic movement, ataxia, imbalance, or disorders in the initiation or planning of movement.

WEAKNESS

Weakness is a reduction in normal power of one or more muscles. Patients may use the term differently; thus one or more specific examples of weakness should be elicited during the history. Increased fatigability or limitation in function due to pain is often confused with weakness by patients. *Increased fatigability* is the inability to sustain the performance of an activity that should be normal for a person of the same age, gender, and size.

Weakness is described commonly by severity and distribution. Paralysis and the suffix "-plegia" indicate weakness that is so severe that it is complete or nearly complete. "Paresis" refers to weakness that is mild or moderate. The prefix "hemi-" refers to one half of the body, "para-" to both legs, and "quadri-" to all four limbs.

Tone is the resistance of a muscle to passive stretch. Central nervous system (CNS) abnormalities that cause weakness generally produce *spasticity*, an increase in tone due to upper motor neuron disease. Spasticity is velocity-dependent, has a sudden release after reaching a maximum (the "clasp-knife" phenomenon), and predominantly affects antigravity muscles (i.e., upper limb flexors and lower limb extensors). Spasticity is distinct from rigidity and paratonia, two other types of increased tone. *Rigidity* is increased tone that is present throughout the range of motion (a "lead pipe" or "plastic" stiffness) and affects flexors and extensors equally. In some patients, rigidity has a cogwheel quality that is enhanced by voluntary movement of the contralateral limb (reinforcement). Rigidity occurs with certain extrapyramidal disorders. *Paratonia*, also referred to as *gegenhalten*, is increased tone that varies irregularly in a manner that may seem related to the degree of relaxation, is present throughout the range of motion, and affects flexors and extensors equally. Paratonia usually results from disease of the frontal lobes. Weakness with decreased tone (flaccidity) or normal tone occurs with disorders of the *motor unit*, that is, a single lower motor neuron and all of the muscle fibers it innervates.

Three basic patterns of weakness can usually be recognized based on the signs summarized in Table 22-1. One results from upper motor neuron pathology, and the other two from disorders of the motor unit (lower motor neuron and myopathic weakness). Fasciculations and early atrophy help to distinguish lower motor neuron (neurogenic) weakness from myopathic weakness. A *fasciculation* is a visible or palpable twitch within a single muscle due to the spontaneous discharge of one motor unit. Neurogenic weakness also produces more prominent hypotonia and greater depression of tendon reflexes than myopathic weakness.

PATHOGENESIS

Upper Motor Neuron Weakness This pattern of weakness results from disorders that affect the upper motor neurons or their axons in the cerebral cortex, subcortical white matter, internal capsule, brainstem, or spinal cord (Fig. 22-1). Both the pyramidal and bulbospinal pathways contribute to normal strength, tone, coordination, and gait. Upper motor neuron lesions produce weakness through decreased activation of the lower motor neurons. In general, distal muscle groups are affected more severely than proximal ones, and axial movements are spared unless the lesion is severe and bilateral. With corticobulbar involvement, weakness is usually observed only in the lower face and tongue; extraocular, upper facial, pharyngeal, and jaw muscles are almost always spared. With bilateral corticobulbar lesions, *pseudobulbar palsy* often develops, in which dysarthria, dysphagia, dysphonia, and emotional lability accompany bilateral facial weakness. Spasticity accompanies upper motor neuron weakness but may not be present in the acute phase.

Upper motor neuron lesions also affect the ability to perform rapid repetitive movements. Such movements are slow and coarse, but normal rhythmicity is maintained. Finger-nose-finger and heel-knee-shin are performed slowly but adequately.

Lower Motor Neuron Weakness This pattern results from disorders of cell bodies of lower motor neurons in the brainstem motor nuclei and the anterior horn of the spinal cord, or from dysfunction of the axons of these neurons as they pass to skeletal muscle (Fig. 22-2).

Lower motor weakness is produced by a decrease in the number of motor units that can be activated, through a loss of the a motor neurons or disruption of their connections to muscle. With a decreased number of motor units, fewer muscle fibers are activated with full effort and maximum power is reduced. Loss of g motor neurons does not cause weakness but decreases tension on the muscle spindles. Muscle tone and tendon reflexes depend on g motor neurons, muscle spindles, spindle afferent fibers, and the a motor neurons. A tap on a tendon stretches muscle spindles and activates the primary spindle afferent fibers. These monosynaptically stimulate thea motor neurons in the spinal cord, producing a brief muscle contraction, which is the familiar tendon reflex.

When a motor unit becomes diseased, especially in anterior horn cell diseases, it may spontaneously discharge, producing a fasciculation. These isolated small twitches may be seen or felt clinically or recorded by electromyography (EMG) (Chap. 357). When a motor neurons or their axons degenerate, the denervated muscle fibers spontaneously discharge in a manner that cannot be seen or felt but can be recorded with EMG. These small single muscle fiber discharges are called *fibrillation potentials*. If significant lower motor neuron weakness is present, recruitment of motor units is delayed or reduced, with fewer than normal activated at a given discharge frequency. This contrasts with upper motor neuron weakness, in which a normal number of motor units are activated at a given frequency but in which the maximum discharge frequency is decreased.

Myopathic Weakness This pattern of weakness is produced by disorders within the motor unit that affect the muscle fibers or the neuromuscular junctions.

Two types of muscle fibers exist. Type I muscle fibers are rich in mitochondria and oxidative enzymes, produce relatively low force, but have low energy demands that can be supplied by ongoing aerobic metabolism. They produce sustained postural and nonforceful movements. Type II muscle fibers are rich in glycolytic enzymes, can produce relatively high force, but have high energy demands that cannot be supplied for long by ongoing aerobic metabolism. Thus, these units can be activated maximally for only brief periods of time to produce high-force movements.

For graded voluntary movements, type I muscle fibers are activated earlier in recruitment. For each muscle fiber, if the nerve terminal releases a normal number of acetylcholine molecules presynaptically and a sufficient number of postsynaptic acetylcholine receptors are opened, the end plate reaches threshold and thereby generates an action potential that spreads across the muscle fiber membrane and into the transverse tubular system. This electrical excitation activates intracellular events that produce an energy-dependent contraction of the muscle fiber (excitation-contraction coupling).

Myopathic weakness is produced by a decrease in the number or contractile force of muscle fibers activated within the motor unit. With muscular dystrophies, inflammatory myopathies, or myopathies with muscle fiber necrosis, decreased numbers of muscle fibers survive within many motor units. As demonstrated withEMG, the size of each motor unit action potential is decreased so that motor units must be recruited more rapidly than normal to produce the power necessary for a certain movement. Neuromuscular junction diseases, such as myasthenia gravis, produce weakness in a similar manner, although the loss of muscle fibers within the motor unit is functional rather than actual. Furthermore, the number of muscle fibers activated can vary over time, depending on the state of rest of the neuromuscular junctions. Thus, fatigable weakness is suggestive of myasthenia gravis or another neuromuscular junction disease. Some myopathies produce weakness through loss of contractile force of muscle fibers or through relatively selective involvement of the type II muscle fibers. These may not affect the size of individual motor unit action potentials observed with EMG and are detected by a discrepancy between the electrical activity and force of a muscle.

Integrated Movements Most purposeful movements require the integrated coordination of many muscle groups. Consider a simple movement, such as grasping a ball. The primary movement is a flexion of the thumb and fingers of one hand, with opposition of the thumb and little finger. This requires the contraction of several muscles, including flexor digitorum superficialis, flexor digitorum profundus, flexor pollicis longus, flexor pollicis brevis, opponens pollicis, and opponens digiti minimi. These prime movers for this action are called *agonists*. In order for the grasping to be smooth and forceful, the thumb and finger extensors need to relax at the same rate as the flexors contract. The muscles that act in a directly opposing manner to the agonists are *antagonists*. A secondary action of the thumb and finger flexors is to flex the wrist; because wrist flexion tends to weaken finger flexion if both occur, activation of wrist extensors assists the grasping movement. Muscles that produce such complementary movements are *synergists*. Finally, the arm needs to be held in a stable position as the grasp occurs, so that the ball is not knocked away before it is secured. Muscles that stabilize the arm

position are *fixators*.

The coordination of activity by agonists, antagonists, synergists, and fixators is regulated by a three-level hierarchy of motor control. The lowest level of control is mediated through segmental reflexes in the spinal cord. These reflexes facilitate agonists and reciprocally inhibit the antagonists. Spinal segments also control rhythmic patterns of movement that involve more than a single pair of agonists and antagonists. For example, the lumbosacral spinal cord contains the basic programming for cyclical stepping movements that involve the synergistic activation of different muscle groups over time. The intermediate level of control is mediated through the descending bulbospinal pathways, which integrate visual, proprioceptive, and vestibular feedback into the execution of an action. For example, the locomotor center in the midbrain is required to modify the cyclical stepping movements in order that balance be maintained and forward movement occur. The highest level of control is mediated by the cerebral cortex. Superimposition of this highest level of control is necessary for activities such as walking to be goal-directed. Precise movements that are learned and improved through practice are also initiated and controlled by the motor cortex. Although only the agonists are directly activated, during the course of a complex sequence of actions such as playing the piano, the sequential activation of different groups of agonists for each note or chord is a part of the learned motor program. Further, the execution of these actions also involves input from the basal ganglia and cerebellar hemispheres to facilitate agonists, synergists, and fixators and to inhibit undesired antagonists.

Apraxia is a disorder of planning and initiating a skilled or learned movement (Chap. 25). Unilateral apraxia of the right hand may be due to a lesion of the left frontal lobe (especially anterior or inferior), the left temporoparietal region (especially the supramarginal gyrus), or their connections. Left body apraxia is produced by lesions of these regions in the right hemisphere or by lesions in the corpus callosum that disconnect the right temporoparietal or frontal regions from those on the left. Bilateral apraxia is often due to bilateral frontal lobe lesions or diffuse bilateral hemispheric disease.

Approach to the Patient

The mode of onset, distribution, and associated features of weakness should be carefully defined. When there is a discrepancy between the history and physical findings, it is usually because the patient complains of weakness, whereas symptoms are actually due to other causes, such as incoordination or pain limiting effort. Power may be examined in a variety of ways. The patient is asked to push or pull in a specified direction against resistance, and the strength in each muscle group is graded from 0 to 5 by the scale developed by the Medical Research Council (Table 22-2). A second method is indirect testing through observation of task performance such as holding the arms outstretched. This is especially useful in detecting mild, asymmetric upper motor neuron weakness through the observation of a downward drift with pronation of the forearm on one side. A third method is functional testing, which involves quantitation of activities. Common tests include counting the number of times a person can perform a deep-knee bend or step on a stool or chair, or timing the length of time the arms can be held abducted to 90 degrees. When performed serially, functional tests provide useful estimates of changes in the patient's status over time.

Other elements of the motor examination include appraisal of muscular bulk, inspection for fasciculations, and assessment of tone. Fasciculations are most easily determined by observing relaxed limbs that are illuminated from behind, but they can also be palpated as irregular low-amplitude twitches within the muscle. Tone is assessed by passive movement of each limb at its various joints and at several different speeds. In the clinical context of weakness, tone may be spastic or flaccid. The presence of cogwheel rigidity, lead-pipe rigidity, or paratonia suggests a disorder of integrated movements, rather than true weakness.

Hemiparesis Hemiparesis results from an upper motor neuron lesion above the midcervical spinal cord; most lesions that produce hemiparesis are located above the foramen magnum. The presence of language disorders, cortical sensory disturbances, cognitive abnormalities, disorders of visual-spatial integration, apraxia, or seizures indicates a cortical lesion. Homonymous visual field defects reflect either a cortical or a subcortical hemispheric lesion. A "pure motor" hemiparesis of the face, arm, and/or leg is due to a small, discrete lesion in the posterior limb of the internal capsule, cerebral peduncle, or upper pons. Some brainstem lesions produce the classic findings of ipsilateral cranial nerve signs and contralateral hemiparesis. These "crossed paralyses" are discussed further inChap. 361. The absence of cranial nerve signs or facial weakness suggests that a hemiparesis is due to a lesion in the high cervical spinal cord, especially if associated with ipsilateral loss of proprioception and contralateral loss of pain and temperature sense (the Brown-Sequard syndrome). However, most spinal cord lesions produce quadriparesis or paraparesis.

Acute or episodic hemiparesis usually has a vascular pathogenesis, either ischemia or a primary hemorrhage (Chap. 361). Less commonly, hemorrhage may occur into brain tumors (Chap. 370) or from rupture of normal vessels due to trauma (Chap. 369); the trauma may be trivial in patients who are anticoagulated or elderly. Less likely possibilities include a focal inflammatory lesion from multiple sclerosis (Chap. 371), abscess, or sarcoidosis (Chap. 318). Evaluation begins immediately with a computed tomography (CT) scan of the brain (Fig. 22-3). If CT is normal and an ischemic stroke is unlikely, magnetic resonance imaging (MRI) of the brain or cervical spine may be indicated.

Subacute hemiparesis that evolves over days or weeks has a long differential diagnosis. A common cause is subdural hematoma; this readily treatable condition must always be considered, especially in elderly or anticoagulated patients, even in the absence of a history of trauma (Chap. 369). Infectious possibilities include cerebral bacterial abscess (Chap. 372), fungal granuloma or meningitis (Chap. 374), and parasitic infection. Weakness from malignant primary and metastatic neoplasms may evolve over days to weeks (Chap. 370). AIDS (Chap. 309) may present with subacute hemiparesis due to toxoplasmosis or primaryCNSlymphoma. Noninfectious inflammatory processes, such as multiple sclerosis (Chap. 371) or, less commonly, sarcoidosis, are further considerations. If the brainMRI is normal and if cortical and hemispheric signs are not present, MRI of the cervical spine may be required.

Chronic hemiparesis that evolves over months is usually due to a neoplasm (Chap. 370), an unruptured arteriovenous malformation (Chap. 361), a chronic subdural

hematoma (Chap. 369), or a degenerative disease (Chaps. 363 to 366). The initial diagnostic test is often an MRIof the brain, especially if the clinical findings suggest brainstem pathology. If MRI of the brain is normal, the possibility of a foramen magnum or high cervical spinal cord lesion should be considered.

Paraparesis An intraspinal lesion at or below the upper thoracic spinal cord level is most commonly responsible. A sensory level over the trunk identifies the approximate level of the cord lesion. Paraparesis can also result from lesions at other locations that disturb upper motor neurons (especially parasagittal lesions and hydrocephalus) and lower motor neurons (anterior horn cell disorders, cauda equina syndromes, and occasionally peripheral neuropathies).

Acute or episodic paraparesis due to spinal cord disease may be difficult to distinguish from disorders affecting lower motor neurons or cerebral hemispheres. Recurrent episodes of paraparesis are often due to multiple sclerosis or to vascular malformations of the spinal cord. With acute spinal cord disease, the upper motor neuron deficit is usually associated with incontinence and a sensory disturbance of the lower limbs that extends rostrally to a level on the trunk; tone is typically flaccid, and tendon reflexes absent. In such cases, the diagnostic approach starts with an imaging study of the spinal cord (Fig. 22-3). Compressive lesions (particularly epidural tumor, abscess, or hematoma), spinal cord infarction (proprioception is usually spared), an arteriovenous fistula or other vascular anomaly, and transverse myelitis, among other causes may be responsible (Chap. 368). Diseases of the cerebral hemispheres that produce acute paraparesis include anterior cerebral artery ischemia (shoulder shrug also affected), superior sagittal sinus or cortical venous thrombosis, and acute hydrocephalus. If upper motor neuron signs are associated with drowsiness, confusion, seizures, or other hemispheric signs but not a sensory level over the trunk, the diagnostic approach starts with anMRI of the brain. Paraparesis is part of the cauda equina syndrome, which may result from trauma to the low back, a midline disk herniation, or intraspinal tumor; although sphincters are affected, hip flexion is often spared, as is sensation over the anterolateral thighs. Rarely, paraparesis is caused by a rapidly evolving peripheral neuropathy such as Guillain-Barre syndrome or by a myopathy. In such cases, electrophysiologic studies are diagnostically helpful and refocus the subsequent evaluation (Chaps. 378 and381).

Subacute or chronic paraparesis with spasticity is caused by upper motor neuron disease. When paraparesis evolves over weeks or months with lower limb sensory loss and sphincter involvement, possible spinal cord disorders include multiple sclerosis, intraparenchymal tumor, chronic spinal cord compression from degenerative disease of the spine, subacute combined degeneration due to vitamin B_{12}deficiency, viral infections (especially human T cell leukemia/lymphoma virus I), and hereditary or other degenerative diseases. Primary progressive multiple sclerosis usually presents in the fourth or fifth decade as progressive paraparesis (Chap. 371). Gliomas of the spinal cord typically produce a progressive myelopathy that is painful (Chap. 370). The clinical approach begins with anMRI of the spinal cord. If the imaging study is normal and spasticity is present, MRI of the brain may be indicated. If hemispheric signs are present, parasagittal meningioma or chronic hydrocephalus is likely and MRI of the brain is the initial test. Progression over months to years is typical of degenerative disorders such as primary lateral sclerosis (Chap. 365) and hereditary disorders such as

familial spastic paraparesis and adrenomyeloneuropathy (Chap. 368). In the rare situations when a chronic paraparesis is due to a lower motor neuron or myopathic etiology, the localization is usually suspected on clinical grounds by the absence of spasticity and confirmed byEMG and nerve conduction tests.

Quadriparesis or Generalized Weakness Generalized weakness may be due to disorders of the central nervous system or of the motor unit. Although the terms *quadriparesis* and *generalized weakness* are often used interchangeably, quadriparesis is more often chosen when an upper motor neuron cause is suspected and generalized weakness when a disease of the motor unit is likely. Weakness fromCNSdisorders is usually associated with changes in consciousness or cognition, with increased muscle tone and muscle stretch reflexes, and with alterations of sensation. Most neuromuscular causes of intermittent weakness are associated with normal mental function, diminished muscle tone, and hypoactive muscle stretch reflexes. Exceptions are some causes of acute quadriparesis due to upper motor neuron disorders in which transient hypotonia is present. The major causes of intermittent weakness are listed inTable 22-3. A patient with generalized fatigability without objective weakness may have the *chronic fatigue syndrome* (Chap. 384).

Acute Quadriparesis Acute quadriparesis with onset over minutes may result from disorders of upper motor neurons (e.g., anoxia, hypotension, brainstem or cervical cord ischemia, trauma, and systemic metabolic abnormalities) or muscle (electrolyte disturbances, certain inborn errors of muscle energy metabolism, toxins, or periodic paralyses). Onset over hours to weeks may, in addition to the above, be due to lower motor neuron disorders. Guillain-Barre syndrome (Chap. 378) is the most common lower motor neuron weakness that progresses over days to several weeks; the finding of an elevated protein level in the cerebrospinal fluid is helpful but may be absent early in the course. If stupor or coma is present, the evaluation begins with aCT scan of the brain. If upper motor neuron signs are present but the patient is alert, the initial test is usually anMRI of the cervical cord. If weakness is lower motor neuron, myopathic, or uncertain in origin, the clinical approach starts with blood studies for muscle enzymes and electrolytes and anEMG and nerve conduction study.

Subacute or Chronic Quadriparesis When quadriparesis due to upper motor neuron disease develops over weeks, months, or years, the distinction between disorders of the cerebral hemispheres, brainstem, and cervical spinal cord is usually possible by clinical criteria alone. The diagnostic approach begins with anMRI of the clinically suspected site of pathology. Lower motor neuron disease usually presents with weakness that is most profound distally, whereas myopathic weakness is typically proximal; the evaluation then begins withEMG and nerve conduction studies.

Monoparesis This is usually due to lower motor neuron disease, with or without associated sensory involvement. Upper motor neuron weakness occasionally presents with a monoparesis of distal and nonantigravity muscles. Myopathic weakness is rarely limited to one limb.

Acute Monoparesis Distinguishing between upper and lower motor neuron disorders may be difficult clinically because tone and reflexes are frequently decreased in both at presentation. If the weakness is predominantly in distal and nonantigravity muscles and

not associated with sensory impairment or pain, focal cortical ischemia is likely (Chap. 361); in this setting, diagnostic possibilities are similar to those for acute hemiparesis. Sensory loss and pain usually accompany acute lower motor neuron weakness. The distribution of weakness is commonly localized to a single nerve root or peripheral nerve within one limb but occasionally reflects involvement of the brachial or lumbosacral plexus. If lower motor neuron weakness is suspected, or if the pattern of weakness is uncertain, the clinical approach begins with anEMG and nerve conduction study.

Subacute or Chronic Monoparesis Weakness with atrophy of one limb that develops over weeks or months is almost always lower motor neuron in origin. If the weakness is associated with numbness, a peripheral nerve or spinal root origin is likely; uncommonly, the brachial or lumbosacral plexus is affected. If numbness is absent, anterior horn cell disease is likely. In either case, an electrodiagnostic study is indicated. If upper rather than lower motor neuron signs are present, a tumor, vascular malformation, or other cortical lesion affecting the precentral gyrus may be responsible. Alternatively, if the leg is affected, a small thoracic cord lesion, often a tumor or multiple sclerosis, may be present. In these situations, the approach begins with an imaging study of the suspicious area.

Distal Weakness Involvement of two or four limbs distally suggests lower motor neuron or peripheral nerve disease. Acute distal lower limb weakness occurs occasionally from an acute toxic polyneuropathy or cauda equina syndrome. Distal symmetric weakness usually develops over weeks, months, or years and is due to metabolic, toxic, hereditary, degenerative, or inflammatory diseases of peripheral nerves (Chap. 377). With peripheral nerve disease, weakness is usually less severe than numbness. Anterior horn cell disease may begin distally but is typically asymmetric and is not associated with numbness (Chap. 365). Rarely, myopathies also present with distal weakness (Chap. 381). The first step in evaluation is an electrodiagnostic study (Fig. 22-3).

Proximal Weakness Proximal weakness of two or four limbs suggests a disorder of muscle or, less commonly, neuromuscular junction or anterior horn cell. Myopathy often produces symmetric weakness of the pelvic or shoulder girdle muscles (Chap. 381). Diseases of the neuromuscular junction (such as myasthenia gravis) may present with symmetric proximal weakness (Chap. 380), often associated with ptosis, diplopia, or bulbar weakness and fluctuating in severity during the day. Extreme fatigability present in some cases of myasthenia gravis may even suggest episodic weakness, but strength rarely returns fully to normal. The proximal weakness of anterior horn cell disease is most often asymmetric, but may be symmetric if familial (Chap. 365). Numbness does not occur with any of these diseases. The evaluation usually begins with determination of the serum creatine kinase level and electrophysiologic studies.

Weakness in a Restricted Distribution In some patients, weakness does not fit any of the above patterns. Examples include weakness limited to the extraocular, hemifacial, bulbar, or respiratory muscles. If unilateral, restricted weakness is usually due to lower motor neuron or peripheral nerve disease, such as in a facial palsy (Chap. 367) or an isolated superior oblique muscle paresis (Chap. 28). Relatively symmetric weakness of extraocular or bulbar muscles is usually due to a myopathy (Chap. 381) or neuromuscular junction disorder (Chap. 380). Bilateral facial palsy with areflexia

suggests Guillain-Barre syndrome (Chap. 378). Worsening of relatively symmetric weakness with fatigue is characteristic of neuromuscular junction disorders (Chap. 380). Asymmetric bulbar weakness is usually due to motor neuron disease. Weakness limited to respiratory muscles is uncommon and is usually due to motor neuron disease, myasthenia gravis, or polymyositis/dermatomyositis (Chap. 382).

MYALGIAS, SPASMS, AND CRAMPS

Spontaneous or exercise-related discomfort from muscles is usually benign and is rarely caused by a definable neuromuscular disease. However, a number of disorders of the motor system are characteristically painful. Some terms for muscular discomfort or involuntary contractions, such as myalgias, spasms, and cramps, are often used interchangeably by patients but have a more specific meaning to physicians. Other terms, such as aching, heaviness, and stiffness, are less specific. *Myalgias* are pains that are felt in muscle; the term does not imply an involuntary contraction. *Spasms* and *cramps* refer to episodes of involuntary contraction of one or more muscles. Cramps are usually painful, whereas spasms are not necessarily uncomfortable.

MYALGIAS

Proximal or generalized weakness associated with myalgias is usually due to an inflammatory, metabolic, endocrine, or toxic myopathy (Chap. 381). Spontaneous myalgias not accompanied by objective weakness are often without a clear cause unless associated with a well-defined systemic illness. Myalgias are a common manifestation of fever or infection, especially influenza. Muscle pains and stiffness with elevated serum creatine kinase concentration is common in hypothyroidism, even in patients without objective weakness. *Polymyalgia rheumatica* (Chap. 317) is characterized by diffuse myalgias and joint stiffness that predominantly affect the pelvic and shoulder girdles in a patient over 50 years of age who has anorexia, mild weight loss, and low-grade fever. Limitation of activity from the myalgias and joint stiffness also leads to disuse atrophy and may give the impression of weakness. However,EMG, serum creatine kinase levels, and muscle biopsy are normal. The erythrocyte sedimentation rate is elevated in most patients, and features of giant-cell arteritis are present in 25%. Diffuse myalgias are common in many rheumatologic diseases, in which the diagnosis and treatment are based on other symptoms and signs. Myalgias are occasionally present in dermatomyositis/polymyositis, but most patients have weakness without significant pain. *Fibromyalgia* (fibrositis, fibromyositis) is associated with pain and tenderness of muscle and adjacent connective tissue (Chap. 325). Fatigue, insomnia, and depression are often present, but objective weakness, elevation of serum creatine kinase level, or elevation of the erythrocyte sedimentation rate does not occur. The diagnosis is dependent upon identifying characteristic focal "trigger points."

Focal Myalgias Focal muscle pain is often traumatic. Rupture of muscle tendons such as the biceps or gastrocnemius muscle may produce visible muscle shortening. Many such tears resolve without surgery but leave an abnormal appearance to the muscle belly. Nontraumatic focal muscle pain is often related to adjacent nonmuscular disorders (e.g., unilateral gastrocnemius pain due to deep venous thrombosis). Rarely, focal muscle pain may be caused by ischemic infarction or bacterial myositis, if acute, or by

neoplasm, parasitic infection, sarcoidosis, or other inflammation or infection, if subacute or chronic.

Exertional Myalgias Myalgias following unaccustomed, strenuous physical activity occur in normal individuals are often associated with laboratory evidence for muscle damage, such as an elevation of serum creatine kinase, edema of muscles onMRI, necrosis of muscle fibers on biopsy, and rarely myoglobinuria. Similar symptoms and laboratory abnormalities characterize certain metabolic disorders of muscle, such as carnitine palmitoyl transferase and glycolytic pathway enzyme deficiencies. The association of objective weakness during an episode of myalgias suggests a metabolic muscle disease. The development of an acute contracture (the inability to relax a muscle due to energy depletion) with the myalgias suggests a metabolic muscle disease with a glycolytic enzyme deficiency (Chap. 383). Exertional myalgias with muscle fiber necrosis also occur in muscular dystrophy with partial deficiency of dystrophin, and certain mitochondrial cytopathies (Chap. 383). Exertional myalgias with elevated creatine kinase concentration but without weakness also occur in hypothyroidism, and when confined to the legs may be due to vascular or neurogenic intermittent claudication. Most patients with exertional myalgias and no weakness do not have a definable abnormality.

SPASMS AND CRAMPS

Involuntary contraction of muscle may occur with disorders of theCNS, lower motor neuron, or muscle. Contractions that originate within the CNS and are associated with upper motor neuron signs are usually referred to as spasms and generally affect the flexors or extensors of one or more limbs. Those that originate within the CNS and are not associated with upper motor neuron signs include movement disorders discussed below, as well as the rare stiff-person syndrome and tetanus. Muscle rigidity from active muscle contraction can occur in the malignant hyperthermia syndrome, usually associated with general anesthesia. In the neuroleptic malignant syndrome, muscle rigidity arises from CNS overactivity and is present in muscle. Involuntary contractions that originate in the lower motor neurons are usually cramps, occasionally tetany, or rarely neuromyotonia. Spasms that originate in muscle or muscle membrane are usually a delayed relaxation after voluntary contraction, either myotonia or rarely a contracture. These conditions may be difficult to distinguish clinically but are often well characterized byEMGstudies.

Stiff-Person Syndrome This rare syndrome is characterized by slowly progressive muscle stiffness and superimposed spasms. The stiffness commonly begins in the low back and spreads over months up the spine and into the limbs but not into the jaw. The gait becomes stiff, and there is hyperlordosis of the lumbar spine. Spasms are often produced by startle. Emotional stress tends to worsen the stiffness as well as the frequency and severity of spasms. The spontaneous motor activity disappears during sleep. The syndrome is often associated with diabetes mellitus and can be paraneoplastic, accompanying Hodgkin's lymphoma, small cell cancer of the lung, and breast cancer. Most patients have a serum antibody against glutamic acid decarboxylase, an enzyme responsible for synthesis of the inhibitory neurotransmitter g-aminobutyric acid (GABA). Stiffness results from loss of descending brainstem or segmental spinal inhibitory influences on the lower motor neurons.EMGstudies reveal

continuous motor unit activity that is similar to voluntary effort with preservation of the silent period to muscle stretch. Stiffness and spasms typically respond partially to treatment with baclofen or benzodiazepines.

Tetanus This rare hyperexcitable state results from exposure to tetanus toxin in patients infected with *Clostridium tetani* (Chap. 143). Painful spasms typically begin with jaw closure (trismus) and soon become generalized.EMGstudies reveal continuous motor unit activity that is similar to voluntary effort except for loss of the silent period to muscle stretch.

Cramps These are the most common type of involuntary muscle contraction. Cramps are a painful contraction of a single muscle that produces a palpable knot within the muscle for seconds to minutes and is relieved by passive stretch of the muscle or spontaneously.EMGstudies reveal motor unit activity that has too high a discharge frequency to be voluntary. If cramps are associated with weakness, the weakness is almost always lower motor neuron in origin. When strength is normal, no definable condition is usually found, although dehydration, hypothyroidism (Video 330-1), or uremia is occasionally present. If prominent, membrane stabilizing drugs, such as carbamazepine, may provide symptomatic benefit.

Tetany Tetany is characterized by contraction of distal muscles of the hands (carpal spasm with extension of interphalangeal joints and adduction and flexion of the metacarpophalangeal joints) and feet (pedal spasm) and is associated with tingling around the mouth and distally in the limbs. Tetany with carpopedal spasms is a common manifestation of hypocalcemia or respiratory alkalosis (even from hyperventilation).EMGstudies reveal single or more often grouped motor unit discharges at low discharge frequency.

Neuromyotonia (Isaac's Syndrome) Neuromyotonia is characterized by muscle stiffness at rest that persists during sleep and by delayed relaxation after voluntary effort. Distal limb muscles are usually affected most severely, but all skeletal muscle may be involved. Gait may be stiff, and close inspection of the muscle reveals undulation of the overlying skin due to continuous muscle fiber contractions (myokymia). The continuous muscle fiber activity generates heat, and excessive sweating is common.EMGstudies commonly reveal myokymic discharges, especially in familial cases. Rarely, EMGs record high-frequency neuromyotonic discharges. Autoantibodies against voltage-gated potassium channels have been demonstrated in some cases, and plasma exchange may be effective.

Myotonia This is a nonpainful delay in the relaxation of muscle after voluntary activity. Delay in opening the hand after a forceful grip (grip myotonia) is common. These disorders are usually familial and worsen in cold weather.EMGdemonstrates a waxing and waning discharge of individual muscle fibers.

Contracture A painful inability to relax a muscle after voluntary activity due to energy depletion characterizes certain metabolic disorders with failure of energy production, such as myophosphorylase deficiency (McArdle's disease).EMGstudies reveal electrical silence.

MOVEMENT DISORDERS

Movement disorders are neurologic syndromes in which abnormal movements (or *dyskinesias*) occur due to a disturbance of fluency and speed of voluntary movement or the presence of unintended extra movements. Because they are so distinct from the pyramidal disorders that cause upper motor neuron weakness, movement disorders are often referred to as *extrapyramidal diseases*. *Hyperkinetic movement disorders* are those in which an excessive amount of spontaneous motor activity is seen or in which abnormal involuntary movements occur. *Hypokinetic movement disorders* are characterized by *akinesia* or *bradykinesia*, in which purposeful motor activity is absent or reduced. This is often described as "poverty of movement."

PATHOGENESIS

Movement disorders result from disease of the basal ganglia, paired subcortical gray matter structures consisting of the caudate and the putamen (which together are called the striatum), the internal and external segments of the globus pallidus, the subthalamic nucleus, and the substantia nigra. The major interconnections and neurotransmitters involved in basal ganglia circuits are illustrated in Fig. 22-4*A*. An understanding of this circuitry can explain, in part, the perturbation that occurs in both the hypo- and hyperkinetic disorders.

Parkinson's disease (Video 361-1) (Chap. 363), the prototypic hypokinetic movement disorder, results from a loss of dopaminergic neurons in the substantia nigra pars compacta. This leads to less excitation of striatal neurons that express the D_1 type of dopamine receptors and less inhibition of D_2 striatal neurons, both contributing to reduced facilitation of cortically initiated movement (Fig. 22-4*B*). The resting tremor of Parkinson's disease is less readily explained by this model but may result from effects on cholinergic interneurons in the striatum. *Huntington's disease* (Chap. 362), a hyperkinetic movement disorder, may be explained by selective loss of D_2 striatal neurons, resulting in disinhibition of cortically initiated movements without normal feedback control. The pathogenesis of hemiballismus is similar -- a direct lesion of the glutamatergic neurons in the subthalamic nucleus (usually from a stroke) leads to disinhibition of thalamocortical projections.

Approach to the Patient

An algorithm for the interpretation of abnormal movements is illustrated in Fig. 22-5. The initial step is to determine if the movement disorder is due to an excess or a poverty of movement (i.e., a hyperkinetic or a hypokinetic movement disorder).

Hyperkinetic Movement Disorders Abnormal involuntary movements are divided into those that are rhythmical and those that are irregular. Those that are rhythmical are termed *tremors*, with the uncommon exception of *palatal and segmental myoclonus*. Tremors are divided into three types: rest, postural, and intention tremor. A *rest tremor* is maximal at rest and becomes less prominent with activity. It is characteristic of parkinsonism, a hypokinetic movement disorder, and is therefore commonly associated with bradykinesia and cogwheel rigidity. A rest tremor that develops acutely is usually due to toxins [such as exposure to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine

(MPTP)] or dopamine blocking drugs (such as phenothiazines). If insidious in onset, the diagnostic approach is the same as for Parkinson's disease (Chap. 363). A *postural tremor* is maximal while limb posture is actively maintained against gravity; it is lessened by rest and is not markedly enhanced during voluntary movement toward a target. A postural tremor that develops acutely is usually due to toxic or metabolic factors (for example, hyperthyroidism) or stress. The insidious onset of a postural tremor suggests a benign or familial essential tremor (Chap. 363). An *intention tremor* is most prominent during voluntary movement toward a target and is not present during postural maintenance or at rest. It is a sign of cerebellar disease (Chap. 364). *Asterixis*, which may superficially resemble a tremor, is an intermittent inhibition of muscle contraction that occurs with metabolic encephalopathy (Chap. 376). This leads, for example, to a momentary and repetitive partial flexion of the wrists during attempted sustained wrist extension.

Involuntary movements that are irregular are characterized further by their speed and site of occurrence and by whether they can be suppressed voluntarily. The slowest are athetosis and dystonia. *Athetosis* is a slow, writhing, sinuous movement that occurs nearly continuously in distal muscles. *Dystonia* is a slowly varying but nearly continuous deviation of posture about one or more joints; it may occur in a proximal or distal limb or in axial structures. Dystonia is a more sustained deviation of posture than athetosis, although these two phenomena overlap considerably. The further evaluation of athetosis and dystonia are discussed inChap. 363.

Among the rapid irregular movements, *tics* are controlled with voluntary effort, while the others are not. Tics often occur repetitively in a single location but are sometimes multifocal (Chap. 363).

Chorea, hemiballismus, and myoclonus are rapid, irregular jerks that cannot be consciously suppressed. *Hemiballismus* is the most distinctive among them. It is manifest as a sudden and often violent flinging movement of a proximal limb, usually an arm (Chap. 363). Hemiballismus usually develops acutely due to infarction of the contralateral subthalamic nucleus but occasionally develops subacutely or chronically due to other lesions of this nucleus.

Chorea is a rapid, jerky, irregular movement that tends to occur in the distal limbs or face but may also occur in proximal limb and axial structures. Acute or subacute onset is usually toxic due to excess levodopa or dopamine-agonist therapy or, less often, neuroleptics, birth control pills, pregnancy (chorea gravidarum), hyperthyroidism, or the antiphospholipid syndrome. In children, it may be associated with rheumatic fever and, in such cases, is referred to as *Sydenham's chorea*. The gradual onset of chorea is typical of degenerative neurologic diseases, such as Huntington's chorea (Chap. 362).

Myoclonus is a rapid, brief, irregular movement that is usually multifocal. Myoclonus can occur spontaneously at rest, in response to sensory stimuli, or with voluntary movements. It is a symptom that occurs in a wide variety of metabolic and neurologic disorders. Posthypoxic intention myoclonus is a special myoclonic syndrome that occurs as a sequel to transient cerebral anoxia. Myoclonus may result from lipid storage disease, encephalitis, Creutzfeldt-Jakob disease, or metabolic encephalopathies due to respiratory failure, chronic renal failure, hepatic failure, or electrolyte imbalance.

Myoclonus is also a feature of certain types of epilepsy, as discussed in Chap. 360. *Palatal and segmental myoclonus* are uncommon rhythmic forms of myoclonus that may resemble tremor; they are caused by structural disease of the brainstem or spinal cord at the level of the abnormal movement.

Hypokinetic Movement Disorders These syndromes are manifest as bradykinesia, with a masked, expressionless facial appearance, loss of associated limb movements during walking, and rigid en bloc turning. If bradykinesia is associated only with a rest tremor, cogwheel rigidity, or impairment of postural reflexes (especially with a tendency to fall backwards), Parkinson's disease is likely (Chap. 363). If cognitive, language, upper motor neuron, sensory, or autonomic signs are also present, a *multisystem degenerative neurologic disease* is present.**These disorders are discussed in Chaps. 363, 364, and 366.*

IMBALANCE AND DISORDERS OF GAIT

Imbalance is the impaired ability to maintain the intended orientation of the body in space. It is generally manifest as difficulty in maintaining an upright posture while standing or walking; a severe imbalance may also affect the ability to maintain posture while seated. Patients with imbalance commonly complain of a feeling of unsteadiness or dysequilibrium. Whereas imbalance and unsteadiness are synonymous, *dysequilibrium* implies the additional component of impaired spatial orientation even while lying down. Patients with dysequilibrium commonly also experience *vertigo*, defined as an hallucination of rotatory movement.

PATHOGENESIS

Imbalance and Limb Ataxia Imbalance results from disorders of the spinal cord (spinocerebellar) or vestibular sensory input, the integration of these inputs in the brainstem or midline cerebellum, or the motor output to the spinal neurons that control axial and proximal muscles. Limb ataxia results from disorders of the spinocerebellar and corticopontocerebellar inputs, the integration of these inputs in the intermediate and lateral cerebellum, or the output to the spinal neurons (via the red nucleus and rubrospinal tract) or to the cortex. These pathways ensure adequate speed, fluency, and integration of limb movements. The lateral cerebellar hemispheres coordinate a complex feedback circuit that modulates cortically initiated limb movement.

Sensory ataxia is caused by lesions that affect the peripheral sensory fibers, dorsal root ganglia cells, posterior columns of the spinal cord, lemniscal system in the brainstem, thalamus, or parietal cortex; relevant anatomy is discussed in Chap. 23. Impairment of the proprioceptive sensory feedback to the cerebellum, basal ganglia, and cortex produces sensory ataxia. Sensory ataxia results in imbalance and disturbs the fluency and integration of movements that can be partially alleviated by visual feedback.

Disorders of Gait Walking is one of the most complicated motor activities. Essentially all structures discussed in this chapter participate in normal walking. Cyclical stepping movements produced by the lumbosacral spinal cord centers are modified by cortical, basal ganglionic, brainstem, and cerebellar influences based on proprioceptive, vestibular, and visual feedback.

Approach to the Patient

Examination of coordination, balance, and gait is typically performed at the same time. The finger-nose-finger and the heel-knee-shin maneuvers are observed for signs of incoordination in general and dysmetria in particular. *Dysmetria* consists of irregular errors in the amplitude and force of limb movements. This is accentuated near the target or point of intention and hence termed *intention tremor*. The patient is also asked to maintain the arms outstretched against a resistance that is suddenly removed; excessive *rebound* indicates cerebellar dysfunction. The ability of the patient to rapidly and repetitively tap the hands and feet is assessed for speed and rhythmicity. Errors in rhythm (irregular rate, velocity, or force) indicate *dysdiadochokinesia*. Slow, coarse, but rhythmical movements indicate upper motor neuron disorders. The patient is asked to demonstrate how to comb the hair or brush the teeth to assess the ability to initiate and execute a simple sequence of activity. Balance is examined by having the patient stand stationary with the feet together. If this position can be maintained, the eyes are closed for 5 to 10 s. Accentuation of sway or actual loss of balance is assessed. If balance is momentarily lost, several trials may be necessary to determine if the loss is consistently in the same direction. Walking along an uncrowded space, such as a hallway, is observed. Symmetry of arm swing and various phases of the gait cycle are observed. Walking is then performed for several steps on the heels, on the toes, and in tandem.

Imbalance An algorithm for interpretation of imbalance is presented inFig. 22-6.

Cerebellar ataxia results from disorders of the cerebellum or of its afferent inputs or efferent projections. Abnormalities of the midline cerebellar vermis or the flocculonodular lobe produce truncal ataxia which is usually revealed during the process of rising from a chair, assuming the upright stance with the feet together, or performing some other activity while standing. Once a desired position is reached, imbalance may be surprisingly mild. As walking begins, the imbalance recurs. Patients usually learn to lessen the imbalance by walking with the legs widely separated. The imbalance is usually not lateralized and may be accompanied by symmetric nystagmus.

Abnormalities of the intermediate and lateral portions of the cerebellum typically produce impaired limb movements rather than truncal ataxia. If involvement is asymmetric, lateralized imbalance is common and usually associated with asymmetric nystagmus. Clinical signs of cerebellar limb ataxia include dysmetria, intention tremor, dysdiadochokinesia, and abnormal rebound. Muscle tone is often modestly reduced; this contributes to the abnormal rebound due to decreased activation of segmental spinal cord reflexes and also to pendular reflexes, i.e., a tendency for a tendon reflex to produce multiple swings to and fro after a single tap.**For further discussion of cerebellar diseases, see Chap. 364.*

Imbalance with vestibular dysfunction is characterized by a consistent tendency to fall to one side. The patient commonly complains of vertigo rather than imbalance, especially if the onset is acute. Acute vertigo associated with lateralized imbalance but no other neurologic signs is often due to disorders of the semicircular canal (Chap. 21); the presence of other neurologic signs suggests brainstem ischemia (Chap. 361) or multiple sclerosis (Chap. 371). When the vestibular dysfunction is peripheral, positional

nystagmus and vertigo tend to resolve if a provocative position is maintained (extinction) or repeated (habituation). Lateralized imbalance of gradual onset or persisting for more than 2 weeks, accompanied by nystagmus, may result from lesions of the semicircular canal or vestibular nerve, brainstem, or cerebellum.

Imbalance with sensory ataxia is characterized by marked worsening when visual feedback is removed. The patient can often assume the upright stance with feet together cautiously with eyes open. With eye closure, balance is rapidly lost (positive Romberg sign) in various directions at random. Sensory examination reveals impairment of proprioception at the toes and ankles, usually associated with an even more prominent abnormality of vibratory perception. Prompt evaluation for vitamin B_{12}deficiency is important, as this disorder is reversible if recognized early (Chap. 368). Depression or absence of reflexes points to peripheral nerve disorders (Chap. 377). Spasticity with extensor plantar responses suggests posterior column and spinal cord disorders (Chap. 368). Rarely, sensory ataxia produces lateralized imbalance. In these cases, the disorder is usually in the parietal lobe or thalamus (Chap. 23), but may also be due to an asymmetric sensory neuropathy (Chap. 377) or posterior column disease (Chap. 368).

Sensory limb ataxia is similar to cerebellar limb ataxia but is markedly worse when the eyes are closed. Examination also reveals abnormal proprioception and vibratory perception. The approach focuses on localizing the proprioceptive impairment to the peripheral nerves (Chap. 377), the posterior columns of the spinal cord (Chap. 368), or rarely the parietal lobe.

Other forms of imbalance occur, but the fundamental problem is usually a primary disorder of strength, extrapyramidal function, or cortical initiation of movement.

Abnormal Gait Each of the disorders discussed in this chapter produces a characteristic gait disturbance. If the neurologic examination is normal except for an abnormal gait, diagnosis may be difficult even for the experienced clinician.

Hemiparetic gait characterizes spastic hemiparesis. In its most severe form, an abnormal posture of the limbs is produced by spasticity. The arm is adducted and internally rotated, with flexion of the elbow, wrist, and fingers and with extension of the hip, knee, and ankle. Forward swing of the spastic leg during walking requires abduction and circumduction at the hip, often with contralateral tilt of the trunk to prevent the toes catching on the floor as the leg is advanced. In its mildest form, the affected arm is held in a normal position, but swings less than the normal arm. The affected leg is flexed less than the normal leg during its forward swing and is more externally rotated. A hemiparetic gait is a common residual sign of a stroke (Chap. 361).

Paraparetic gait (Video 361-3) is a walking pattern in which both legs are moved in a slow, stiff manner with circumduction, similar to the leg movement in a hemiparetic gait. In many patients, the legs tend to cross with each forward swing ("scissoring"). A paraparetic gait is a common sign of spinal cord disease (Chap. 368) and also occurs in cerebral palsy.

Steppage gait is produced by weakness of ankle dorsiflexion. Because of the partial or

complete foot drop, the leg must be lifted higher than usual to avoid catching the toe on the floor during the forward swing of the leg. If unilateral, steppage gait is usually due to L5 radiculopathy, sciatic neuropathy, or peroneal neuropathy (Chap. 377). If bilateral, it is the common result of a distal polyneuropathy or lumbosacral polyradiculopathy (Chap. 377).

Waddling gait results from proximal lower limb weakness, most often from myopathy (Chap. 381) but occasionally from neuromuscular junction disease (Chap. 380) or a proximal symmetric spinal muscular atrophy (Chap. 365). With weakness of hip flexion, the trunk is tilted away from the leg that is being moved to lift the hip and provide extra distance between the foot and the floor, and the pelvis is rotated forward to assist with forward motion of the leg. Because pelvic girdle weakness is customarily bilateral, the pelvic lift and rotation alternates from side to side, giving the waddling appearance to the gait.

Parkinsonian gait (Video 361-1) is characterized by a forward stoop, with modest flexion at the hips and knees. The arms are flexed at the elbows and adducted at the shoulders, often with a 4- to 6-Hz resting pronation-supination tremor but little other movement, even during walking. Walking is initiated slowly by leaning forward and maintained with short rapid steps, during which the feet shuffle along the floor. The pace tends to accelerate (festination) as the upper body gradually leans further ahead of the feet, whether movement is forward (propulsion) or backward (retropulsion). The postural instability leads to falls (Chap. 363).

Apraxic gait (Video 361-4) results from bilateral frontal lobe disease with impaired ability to plan and execute sequential movements. This gait superficially resembles that of parkinsonism, in that the posture is stooped and any steps taken are short and shuffling. However, initiation and maintenance of walking are impaired in a different manner. Each movement that is required for walking can usually be performed, if tested in isolation while sitting or lying. However, when asked to step forward while standing, a long pause often occurs before any attempt is made to flex at the hip and advance, as if the patient is "glued to the ground." Once walking is initiated, it is not maintained, even in an abnormal festinating manner. Rather, after one or several steps are taken, walking is stopped for several seconds or longer. The process is then repeated. Dementia and incontinence may coexist.

Choreoathetotic gait is characterized by an intermittent, irregular movement that disrupts the smooth flow of a normal gait. Flexion or extension movements at the hip are common and unpredictable but readily observed as a pelvic lurch (Chap. 363).

Cerebellar ataxic gait (Video 361-3) is a broad-based gait disorder in which the speed and length of stride varies irregularly from step to step. With midline cerebellar disease, as in alcoholics, posture is erect but the feet are separated; lower limb ataxia is commonly present as well. Assumption of a particular stance or a change in position may cause instability, yet balance can usually be maintained well with the eyes open or closed. Walking may be rapid, but cadence is irregular. Although patients commonly lack confidence in the stability of their walking, only minimal support is often required for reassurance. With disease of the cerebellar hemispheres, limb ataxia and nystagmus are commonly present as well (Chap. 364).

Sensory ataxic gait may resemble a cerebellar gait, with its broad-based stance and difficulty with change in position. However, although balance may be maintained with the eyes open, loss of visual input through eye closure results in rapid loss of balance with a fall (positive Romberg sign), unless the physician assists the patient.

Vestibular gait is one in which the patient consistently tends to fall to one side, whether walking or standing. Cranial nerve examination demonstrates an obviously asymmetric nystagmus. The possibilities of unilateral sensory ataxia and hemiparesis are excluded by the findings of normal proprioception and strength (Chap. 21).

Astasia-abasia is a typical hysterical gait disorder. Although the patient usually has normal coordination of leg movements in bed or while sitting, the patient is unable to stand or walk without assistance. If distracted, stationary balance is sometimes maintained and several steps are taken normally, followed by a dramatic demonstration of imbalance with a lunge toward the examiner's arms or a nearby bed.

(Bibliography omitted in Palm version)

Back to Table of Contents

23. NUMBNESS, TINGLING, AND SENSORY LOSS - *Arthur K. Asbury*

NORMAL SENSATION

Normal somatic sensation reflects a continuous day and night monitoring process that occupies considerable moment-to-moment nervous system capacity. Little of this activity reaches consciousness under ordinary conditions. In contrast, disordered sensation, particularly if experienced as painful, is alarming and dominates the sufferer's attention. Abnormalities of sensation, especially if painful, tend to make those suffering seek medical help. The physician must be able to recognize abnormal sensations by how they are described, know their type and likely site of origin, and understand their implications.**For a consideration of pain, see Chap. 12.*

Positive and Negative Phenomena Abnormal sensory phenomena may be divided into two categories, positive and negative. The prototypical positive phenomenon is tingling (pins-and-needles), and the principal negative phenomenon is numbness. In addition to tingling, positive sensory phenomena include other altered sensations that are often described as pricking, bandlike, lightning-like shooting feelings (lancinations), aching, knifelike, twisting, drawing, pulling, tightening, burning, searing, electrical, or raw feelings. These descriptors are frequently the actual words used by patients. Such sensations may or may not be experienced as painful.

Positive phenomena usually result from trains of impulses generated at a site or sites of lowered threshold or heightened excitability along a sensory pathway, either peripheral or central. The nature and severity of an abnormal sensation depend on the number, rate, timing, and distribution of ectopic impulses and the type and function of nervous tissue in which they arise. Because positive phenomena represent excessive activity in sensory pathways, they are not necessarily associated with any sensory deficit (loss) upon examination.

Negative phenomena represent loss of sensory function and are characterized by diminished or absent feeling, often experienced as numbness. In contrast to positive phenomena, negative phenomena are accompanied by abnormal findings on sensory examination. In disorders affecting peripheral sensation, it is estimated that at least half the afferent axons innervating a given site are lost or functionless before sensory deficit can be demonstrated by clinical examination. This estimate probably varies according to how rapidly sensory nerve fibers have lost function. If the rate of loss is slow and chronic, lack of cutaneous feeling may be unnoticed by the patient and difficult to demonstrate on examination, even though few sensory fibers are functioning. Rapidly evolving sensory abnormality usually evokes both positive and negative phenomena and is readily recognized by patients. Subclinical degrees of sensory dysfunction not demonstrable on clinical sensory examination may be revealed by sensory nerve conduction studies or somatosensory cerebral evoked potentials (Chap. 357). Sensory symptoms may be either positive or negative, but sensory signs on examination are always a measure of negative phenomena.

Terminology Words used to characterize sensory disturbance are descriptive and have been arrived at mainly by convention. Paresthesia and dysesthesia are general terms used to denote sensory symptoms (positive phenomena) and are usually stated in the

plural form. *Paresthesias* usually refer to tingling or pins-and-needles sensations but may also include a wide variety of other abnormal sensations, excepting pain. Sometimes "paresthesias" carry the implication that the abnormal sensations are perceived without an apparent stimulus. *Dysesthesia* is a more general term used to subsume all types of abnormal sensations, even painful ones, whether a stimulus is evident or not.

While dysesthesias and paresthesias refer to sensations described by patients, another set of terms refers to sensory abnormalities found on examination. These include *hypesthesia* or *hypoesthesia* (reduction of cutaneous sensation to a specific type of testing such as pressure, light touch, and warm or cold stimuli); *anesthesia* (complete absence of skin sensation to the same stimuli plus pinprick); and *hypalgesia* (referring to reduced pain perception, i.e., nociception, such as the pricking quality elicited by a pin). *Hyperesthesia* means pain in response to touch. Similarly, *allodynia* describes the situation in which a nonpainful stimulus, once perceived, is experienced as painful, even excruciating. An example is elicitation of a painful sensation by application of a vibrating tuning fork. *Hyperalgesia* denotes severe pain in response to a mildly noxious stimulus, and *hyperpathia*, a broad term, encompasses all the phenomena described by hyperesthesia, allodynia, and hyperalgesia. With hyperpathia, the threshold for a sensory stimulus is increased and the perception is delayed but once felt, is unduly painful.

Disorders of deep sensation, arising from muscle spindles, tendons, and joints, affect proprioception (position sense). Manifestations include imbalance (particularly with eyes closed or in the dark), clumsiness of precision movements, and unsteadiness of gait, which are referred to collectively as *sensory ataxia* (Chap. 22). Other findings on examination usually, but not invariably, include reduced or absent joint position and vibratory sensibility and absent deep tendon reflexes in the affected limbs. Romberg's sign is positive, which means that the patient sways or topples when asked to stand with feet close together and eyes closed. In severe states of deafferentation involving deep sensation, the patient cannot walk or stand unaided or even sit unsupported. Continuous, sometimes wormlike involuntary movements, called *pseudoathetosis*, of the outstretched hands and fingers occur, particularly with eyes closed. Such patients are severely disabled.

Anatomy of Sensation Cutaneous afferent innervation is conveyed by a rich variety of receptors, both naked nerve endings (nociceptors and thermoreceptors) and encapsulated terminals (mechanoreceptors). Each type of receptor has its own set of sensitivities to specific stimuli, size and distinctness of receptive fields, and adaptational qualities. Much of the knowledge about these receptors has come from the development of techniques to study single intact nerve fibers intraneurally in awake unanesthetized human subjects. It is possible not only to record from single nerve fibers, large or small, but also to stimulate single fibers in isolation. A single impulse, whether elicited by a natural stimulus or evoked by electrical microstimulation, in a large myelinated afferent fiber may be both perceived and localized.

Afferent fibers of all sizes in peripheral nerve trunks traverse the dorsal roots and enter the dorsal horn of the spinal cord (Fig. 23-1). From there the smaller fibers take a different route to the parietal cortex than the larger fibers. The polysynaptic projections

of the smaller fibers (unmyelinated and small myelinated), which subserve mainly nociception, temperature sensibility, and touch, cross and ascend in the opposite anterior and lateral columns of the spinal cord, through the brainstem, to the ventral posterolateral (VPL) nucleus of the thalamus, and ultimately project to the postcentral gyrus of the parietal cortex (Chap. 12). This is referred to as the *spinothalamic pathway*, or *anterolateral system*. The larger fibers, which subserve tactile and position sense and kinesthesia, project rostrally in the posterior column on the same side of the spinal cord and make their first synapse in the gracile or cuneate nuclei of the lower medulla. The second-order neuron decussates and ascends in the medial lemniscus located medially in the medulla and in the tegmentum of the pons and midbrain and synapses in the VPL. The third-order neuron projects to parietal cortex; this large fiber system is referred to as the *posterior column-medial lemniscal pathway* (lemniscal, for short). Note that although the lemniscal and the anterolateral pathways both project up the spinal cord to the thalamus, it is the (crossed) anterolateral pathway that is referred to as the *spinothalamic tract*, by convention.

Although the fiber types and functions that make up the spinothalamic and lemniscal systems are relatively well known, it has been found that many other fibers, particularly those associated with touch, pressure, and position sense, ascend in a diffusely distributed pattern both ipsilaterally and contralaterally in the anterolateral quadrants of the spinal cord. This explains why an individual with a complete lesion of the posterior columns of the spinal cord may have little sensory deficit on examination.

EXAMINATION OF SENSATION

The main tasks of the sensory examination are tests of primary sensation. By convention these include the sense of pain, touch, vibration, joint position, and thermal sensation, both hot and cold (Table 23-1). Detailed descriptions of how to perform the various tests of the sensory examination can be found in standard texts (see "Bibliography").

Some general principles pertain. First, the examiner must depend on subjective patient response, particularly when using cutaneous stimuli (pin, touch, vibration, warm or cold). This factor may complicate the interpretation of the sensory examination. Second, with complaints of numbness, patients should be asked to outline on themselves the borders of numb areas. Third, some patients are only partially examinable. In a stuporous patient, sensory examination is reduced to observing the briskness of withdrawal in response to a pinch or other noxious stimulus. Comparison of response on one side of the body to the other is essential. In the alert but uncooperative patient, cutaneous sensation may be unexaminable. However, it is usually possible to get some idea of proprioceptive function by noting the patient's best performance of movements requiring balance and precision. Fourth, sensory examination of a patient who has no neurologic complaints should be abbreviated and may consist of pin, touch, and vibration testing in the hands and feet plus evaluation of stance and gait, including the Romberg maneuver. Evaluation of stance and gait also tests the integrity of motor and cerebellar systems.

Primary Sensation (See Table 23-1) The sense of pain is usually tested with a pin, asking the patient to focus on the pricking or unpleasant quality of the stimulus and not just the pressure or touch sensation elicited. Areas of hypalgesia should be mapped by

proceeding radially from the most hypalgesic site (Figs. 23-2 and23-3).

Temperature sensation, to both hot and cold, is probably best tested with water flasks filled with water of the desired temperature, using a thermometer to verify the temperature. This is impractical in most settings. An alternative way to test cold sensation is to touch a metal object, such as a tuning fork at room temperature, to the skin. For testing warm temperatures, the tuning fork or other metal object may be held under warm water of the desired temperature and then used. Both cold and warm should be tested because different receptors respond to each.

Touch is usually tested with a wisp of cotton or a fine camelhair brush. In general, it is better to avoid testing touch on hairy skin because of the profusion of sensory endings that surround each hair follicle.

Joint position testing is a measure of proprioception, one of the most important functions of the sensory system. With the patient keeping eyes closed, joint position is tested in the great toe and in the fingers. If errors are made in recognizing the direction of passive movements of the toe or the finger, more proximal joints should be tested. A test of proximal joint position sense, primarily at the shoulder, is performed by asking the patient to bring the two index fingers together with the arms extended and the eyes closed. Normal individuals should be able to do this quite accurately, with errors of a centimeter or less.

The sense of vibration is tested with a tuning fork, preferably a large one that vibrates at 128 Hz. Vibration is usually tested at bony prominences, beginning distally at the malleoli of the ankles, and at the knuckles. If abnormalities are found, more proximal sites can be examined. Vibratory thesholds at the same site in the patient and the examiner can be compared for control purposes.

Quantitative Sensory Testing Effective sensory testing devices have been developed over the past two decades. Quantitative sensory testing is particularly useful for serial evaluation of cutaneous sensation in clinical trials. Threshold testing for touch and vibratory and thermal sensation is the most widely used application.

Cortical Sensation Cortical sensory testing includes two-point discrimination, touch localization, and bilateral simultaneous stimulation and tests for graphesthesia and stereognosis, to name the most commonly used methods. Abnormalities of these sensory tests, in the presence of normal primary sensation in an alert cooperative patient, signify a lesion of the parietal cortex or thalamocortical projections to the parietal lobe. If primary sensation is altered, these cortical discriminative functions will usually be abnormal, too. Comparisons should always be made between analogous sites on the two sides of the body because the deficit with a specific parietal lesion is likely to be hemilateral. Side-to-side comparisons hold true for all cortical sensory testing.

Two-point discrimination is tested by special calipers, the points of which may be set from 2 mm to several centimeters apart and then applied simultaneously to the site to be tested. The pulp of the fingertips is a common site to test; a normal individual can distinguish about 3-mm separation of points there.

Touch localization is usually carried out by light pressure with the examiner's fingertip, asking the patient, whose eyes are closed, to identify the site of touch. It is usual to ask the patient to touch the same site with a fingertip.

Bilateral simultaneous stimulation at analogous sites (e.g., the dorsa of both hands) can be carried out to determine whether the perception of touch is extinguished consistently on one side or the other. The phenomenon is referred to as *extinction* on bilateral simultaneous stimulation.

Graphesthesia means the capacity to recognize with eyes closed letters or numbers drawn by the examiner's fingertip on the palm of the hand. Once again, the comparison of one side with the other is of prime importance. Inability to recognize numbers or letters is termed *agraphesthesia*.

Stereognosis refers to the ability to identify common objects by palpation, recognizing their shape, texture, and size. Common standard objects are the best test objects, such as a marble, a paper clip, or coins. Patients with normal stereognosis should be able to distinguish a dime from a penny and a nickel from a quarter without looking. Patients should only be allowed to feel the object with one hand at a time. If they are unable to identify it in one hand, it should be placed in the other for comparison. Individuals unable to identify common objects and coins in one hand who can do so in the other are said to have *astereognosis* of the abnormal hand.

LOCALIZATION OF SENSORY ABNORMALITIES

Sensory symptoms and signs can result from lesions at almost any level of the nervous system, including parietal cortex, deep white matter, thalamus, brainstem, spinal cord, spinal root, peripheral nerve, and sensory receptor. Noting the distribution and nature of sensory symptoms and signs is the most important way to localize their source. The extent, configuration, symmetry, quality, and severity are the key observations.

Dysesthesias without sensory findings by examination can be difficult to interpret. To illustrate, tingling dysesthesias in an acral distribution (hands and feet) can have more than one interpretation. Distal dysesthesias can be systemic in origin, e.g., secondary to hyperventilation, or can be induced by a medication, such as the diuretic acetazolamide. Distal dysesthesias can also be an early event in an evolving polyneuropathy or can herald a myelopathy, such as with vitamin B_{12}deficiency. Sometimes distal dysesthesias have no definable basis. In contrast, dysesthesias that correspond to a particular peripheral nerve territory denote a lesion of that nerve trunk. For instance, dysesthesias restricted to the fifth digit and the adjacent one-half of the fourth finger on one hand reliably point to disorder of the ulnar nerve, most commonly at the elbow.

Nerve and Root In focal nerve trunk lesions severe enough to cause a deficit, sensory abnormalities are readily mapped and generally have discrete boundaries (Figs. 23-2 and23-3). Root lesions, referred to as radicular, are frequently accompanied by deep, aching pain along the course of the related nerve trunk. With compression of a fifth lumbar (L5) or first sacral (S1) root, as may occur with a ruptured intervertebral disc, sciatica is a frequent manifestation. With a lesion affecting a single root, sensory deficit

in the distribution of that root is often minimal or not demonstrable at all. This is because adjacent root territories overlap extensively.

Polyneuropathies are generally graded, distal, and symmetric in distribution of deficit (Chap. 377). Dyesthesias begin in the toes and ascend symmetrically, followed by numbness. When dyesthesias reach the knees, they have usually also appeared in the fingertips. The process appears to be nerve length-dependent, and the deficit is often described as "stocking-glove" in type. Although most polyneuropathies are pansensory and affect all modalities of sensation, selective sensory dysfunction according to nerve fiber size may occur. In polyneuropathies that affect small nerve fibers selectively, the hallmark is burning, painful dysesthesias with reduced pinprick and thermal sensation but with sparing of proprioception, motor function, and even deep tendon jerks. Touch is variably involved, but when spared, the sensory pattern is referred to as *sensory dissociation*. Sensory dissociation patterns can be seen with spinal cord lesions (see below) as well as with small fiber neuropathies. In contrast to small fiber polyneuropathies, large fiber polyneuropathies are characterized by position sense deficit, imbalance, absent tendon jerks, and variable motor dysfunction but preservation of most cutaneous sensation. Dyesthesias, if present at all, tend to be tingling or bandlike.

Spinal Cord (SeeChap. 368) If the spinal cord is transected, all sensation is lost below the level of transection. Bladder and bowel function are also lost, as is motor function. Hemisection of the spinal cord produces the Brown-Sequard syndrome, which involves absent pain and temperature sensation on the opposite side below the lesion, and loss of proprioceptive sensation and loss of motor power on the same side below the lesion (seeFigs. 23-1 and368-1). Dissociated sensory deficit patterns (see above) are also a sign of spinothalamic tract involvement in the spinal cord, especially if the deficit is unilateral and has an upper level on the torso. Bilateral spinothalamic tract involvement occurs with lesions affecting the center of the spinal cord, such as happens with expansion of the central canal in syringomyelia. Sensory dissociation is characteristic of syringomyelia.

Brainstem Harlequin patterns of sensory disturbance, in which one side of the face and the opposite side of the body are affected, localize to the lateral medulla. Here a small lesion may damage both the ipsilateral descending trigeminal tract and ascending spinothalamic fibers subserving the opposite arm, leg, and hemitorso (see "Lateral medullary syndrome" in Fig. 361-7). In the tegmentum of the pons and midbrain, where the lemniscal and spinothalamic tracts merge, a lesion here causes pansensory loss on the contralateral body.

Thalamus Hemisensory disturbance with tingling numbness from head to foot is often thalamic in origin but can also be anterior parietal. If abrupt in onset, the lesion is likely to be due to a small stroke (lacunar infarction), particularly if localized to the thalamus. Occasionally, with lesions affecting theVPL or adjacent white matter, a syndrome of thalamic pain, also called *Dejerine-Roussy syndrome*, may ensue. This persistent unrelenting hemipainful state is often described in dramatic terms such as "like the flesh is being torn from my limbs" or "as though that side is bathed in acid" (Chap. 12).

Cortex With lesions of the parietal lobe, either of the cortex or of subjacent white matter,

the most prominent symptoms are contralateral hemineglect, hemi-inattention, and a tendency not to use the affected hand and arm. Tests of primary sensation may be normal or altered. Anterior parietal infarction may present as a pseudothalamic syndrome with crossed hemilateral loss of primary sensation. Dysesthesias or a sense of numbness may also occur, and rarely a painful state.

Focal Sensory Seizures These are generally due to lesions in or near the postcentral gyrus. Symptoms of focal sensory seizures are usually combinations of numbness and tingling, but frequently additional more complex sensations are present, such as a rushing feeling, a sense of warmth, a sense of movement without visible motion, or other unpleasant dysesthesias. Duration of seizures is variable; they may be transient, lasting only seconds, or they may persist for hours. Focal motor features (clonic jerking) may supervene, and seizures can become generalized with loss of consciousness. Likely sites of symptoms are unilaterally in the lips, face, digits, or foot, and symptoms may spread as in a Jacksonian march. On occasion, symptoms may occur in a symmetric bilateral fashion, for instance, in both hands; this results from involvement of the second sensory area (unilaterally) located in the rolandic area at and just above the Sylvian fissure.

(Bibliography omitted in Palm version)

Back to Table of Contents

24. ACUTE CONFUSIONAL STATES AND COMA - *Allan H. Ropper*

Confusional states and coma are among the most common problems in general medicine. They account for a substantial portion of admissions to emergency wards and are a frequent cause of distress on all hospital services. Because clouding of consciousness and a diminished level of consciousness frequently coexist and result from many of the same diseases, they are presented together here, but from a medical perspective they have different clinical characteristics and physiologic explanations.

The basis of consciousness has long been a topic of great interest to psychologists and philosophers and is the subject of a vast literature. Physicians have been mainly concerned with impairments in the level of consciousness (coma, stupor, drowsiness) and with alterations of consciousness, meaning an inability to think coherently, i.e., with accustomed clarity and speed. The latter, broadly termed *confusion* relates to lessened awareness, perception, apperception (the interpretation of perceptions), thinking, expression in language and action, and all forms of intellection that are dependent on the continuous integration of mental processes. Normal awareness provides a background to our inner mental life, which flows from infancy to death like a "stream of thought," to use William James's metaphor. Self-awareness requires that a person experience these thoughts and be able to reflect and operate upon them.

Almost all instances of diminished alertness can be traced to widespread abnormalities of the cerebral hemispheres or to reduced activity of a special thalamocortical alerting system termed the *reticular activating system* (RAS). The proper functioning of this system, its ascending projections to the cortex, the cortex itself, and corticothalamic connections are required to maintain alertness and coherence of thought.

THE CONFUSIONAL STATE

Confusion is a mental and behavioral state of reduced comprehension, coherence, and capacity to reason. Inattention, as defined by the inability to sustain uninterrupted thought and actions, and disorientation are its earliest outward signs. As the state of confusion worsens, there are more global mental failings, including impairments of memory, perception, comprehension, problem solving, language, praxis, visuospatial function, and various aspects of emotional behavior that are each attributable to particular regions of the brain. In other instances an apparent confusional state may arise from an isolated deficit in mental function such as an impairment of language (*aphasia*), loss of memory (*amnesia*), or lack of appreciation of spatial relations of self and the external environment (*agnosia*), but the attributes of the problem are then quite different (Chap. 25). Confusion is also a feature of dementia, in which case the chronicity of the process, as in the instance of Alzheimer's disease, distinguishes it from an acute encephalopathy (Chap. 26).

The confused patient is usually subdued, not inclined to speak, and is physically inactive. A state of confusion that is accompanied by agitation, hallucinations, tremor, and illusions (misperceptions of environmental sight, sound, or touch) is termed *delirium*, as typified by delirium tremens from alcohol or drug withdrawal. In psychiatric circles, delirium often refers, albeit imprecisely, to all acute states of confusion with clouding of consciousness and incoherence of thought.

Approach to the Patient

Confusion and delirium always signify a disorder of the nervous system. They may be the major manifestation of a head injury; a seizure; drug toxicity (or drug withdrawal); a metabolic disorder resulting from hepatic, renal, pulmonary or cardiac failure; a systemic infection; meningitis or encephalitis; or a chronic dementing disease.

The search for these manifold causes begins with a careful history emphasizing the patient's condition before the onset of confusion. The clinical examination should focus on signs of diminished attentiveness, disorientation, and drowsiness and on the presence of localizing neurologic signs. From the clinical data the clinician is directed to the appropriate laboratory tests discussed further on. Often, even after all diagnostic tests are completed, one may still not know the cause of a confusional state. The proper approach is to observe the patient in the hospital for a number of days under stable conditions. New clues may appear or an obscure confusion perhaps related to a medication, may clear up, while other causes such as renal or hepatic failure may worsen and lead to coma.

Orientation and memory are tested by asking the patient in a forthright manner the date, inclusive of month, day, year, and day of week; the precise place; and some items of generally acknowledged and universally known information (the names of the President and Vice President, a recent national catastrophe, the state capital). Further probing may be necessary to reveal a defect -- why is the patient in the hospital; what is his or her address, zip code, telephone number, social security number? Problems of increasing complexity may be pursued, but they usually provide little additional information. Attention and coherence of thought can be gauged by the clarity of and speed of responses while the history is being given but are examined more explicitly by having the patient repeat strings of numbers (most adults easily retain seven digits forward and four backward), spell a word such as "world" backwards, and perform serial calculations -- tests of serial subtraction of 3 from 30 or 7 from 100 are useful. It is the inability to sustain coherent mental activity in performing tasks such as these that exposes the most subtle confusional states.

Other salient neurologic findings are the level of alertness, which fluctuates if there is drowsiness; indications of focal damage of the cerebrum such as hemiparesis, hemianopia, and aphasia; or adventitious movements of myoclonus or partial convulsions. The language of the confused patient may be disorganized and rambling, even to the extent of incorporating paraphasic words. These features, along with impaired comprehension that is due mainly to inattention, may be mistaken for aphasia.

One of the most specific signs of a metabolic encephalopathy is asterixis, which is an arrhythmic flapping tremor that is typically elicited by asking the patient to hold the arms outstretched with the wrists and hands fully extended. After a few seconds, there is a large jerking lapse in the posture of the hand and then a rapid return to the original position. The same movements can be appreciated in any tonically held posture, even of the tongue, and in extreme form the movements may intrude on voluntary limb motion. Bilateral asterixis always signifies a metabolic encephalopathy, e.g., from hepatic failure, hypercapnia, or from drug ingestion, especially with anticonvulsant

medications. Myoclonic jerking and tremor in an awake patient are typical of uremic encephalopathy or the use of antipsychotic drugs such as lithium, phenothiazines, or butyrophenones; myoclonus with coma may also signify anoxic cerebral damage.

Confusion in the postoperative period is common but at times so subtle as to escape attention. Cardiac and orthopedic procedures are particularly likely to produce disorientation or delirium in susceptible patients. Often a careful history will reveal that a mild but compensated dementia existed prior to the operation. Medications, particularly those with anticholinergic activity (including meperidine), inadvertent withdrawal from sleeping pills or alcohol, fever, and any of the endogenous metabolic derangements listed above may be responsible, or a stroke may have occurred.

Frequently confusion cannot be attributed to any single factor and it clears in several days. In many cases, particularly in the elderly, transient confusion and drowsiness arise with a febrile infection of the urinary tract, lungs, blood, or peritoneum. The term *septic encephalopathy* is currently used to describe this association, but the mechanism by which infection or inflammation leads to cerebral dysfunction is unknown. Fever can also alter brain function in a way that makes preexisting focal signs worse.

Distinguishing dementia from an acute confusional state is a great problem, especially in the elderly, since the two may coexist if a fever, other acute medical problem, or a poorly tolerated medication supervenes in a mildly demented patient, producing a so-called beclouded dementia. The memory loss of dementia brings about a confusional state that varies little in severity from hour to hour and day to day. Poor mental performance is derived mainly from incomplete recollection, inadequate access to names and ideas, and the inability to retain new information, thus affecting orientation and factual knowledge. In contrast to the acute confusional states, attention, alertness, and coherence are preserved until the most advanced stages. Eventually dementia produces a chronic confusion with breakdown of all types of mental performance, and the distinction from an acute encephalopathy depends mainly on the longstanding nature of the condition.

Treatment of the confusional state requires that all unnecessary medication be stopped, metabolic alterations be rectified, and infection be treated. Skilled nursing and a quiet room with a window are important. Careful explanations should be given at regular intervals to the family. In the elderly, regular reorientation and active measures to avoid risk factors (sleep deprivation, immobility, and vision and hearing impairments) reduce the number and severity of episodes of delirium in hospitalized patients.

COMA AND RELATED DISORDERS OF CONSCIOUSNESS

The unnatural situation of reduced alertness and responsiveness represents a continuum that in severest form is called *coma*, a deep sleeplike state from which the patient cannot be aroused. *Stupor* defines lesser degrees of unarousability in which the patient can be awakened only by vigorous stimuli, accompanied by motor behavior that leads to avoidance of uncomfortable or aggravating stimuli. *Drowsiness*, which is familiar to all persons, simulates light sleep and is characterized by easy arousal and the persistence of alertness for brief periods. Drowsiness and stupor are usually attended by some degree of confusion. In clinical practice these terms should be

supplemented by a narrative description of the level of arousal and of the type of responses evoked by various stimuli precisely as observed at the bedside. Such an account is preferable to ambiguous terms such as semicoma or obtundation, the definitions of which differ between physicians.

Several other neurologic conditions render patients apparently unresponsive and simulate coma, and certain other subsyndromes of coma must be considered separately because of their special significance. Among the latter, the *vegetative state* signifies an awake but unresponsive state. Most of these patients were earlier comatose and after a period of days or weeks emerge to an unresponsive state in which their eyelids are open, giving the appearance of wakefulness. Yawning, grunting, swallowing, as well as limb and head movements persist, but there are few, if any, meaningful responses to the external and internal environment -- in essence, an "awake coma." Although respiratory and autonomic functions are retained, the term "vegetative" is nonetheless unfortunate as it is subject to misinterpretation by lay persons. Always there are accompanying signs that indicate extensive damage in both cerebral hemispheres, e.g., decerebrate or decorticate limb posturing and absent responses to visual stimuli (see below). Cardiac arrest and head injuries are the most common causes of the vegetative state (Chaps. 369 and376). The prognosis for regaining mental faculties once the vegetative state has supervened for several months is almost nil hence the term *persistent vegetative state*. Most instances of dramatic recovery, when investigated carefully, are found to yield to the usual rules for prognosis, but it must be acknowledged that rare instances of awakening to a condition of dementia and paralysis have been documented.

Certain other clinical states are prone to be misinterpreted as stupor or coma. *Akinetic mutism* refers to a partially or fully awake patient who is able to form impressions and think but remains immobile and mute, particularly when unstimulated. The condition may result from damage in the regions of the medial thalamic nuclei, the frontal lobes (particularly situated deeply or on the orbitofrontal surfaces), or from hydrocephalus. The term *abulia* is used to describe a mental and physical slowness and lack of impulse to activity that is in essence a mild form of akinetic mutism, with the same anatomic origins. *Catatonia* is a curious hypomobile and mute syndrome associated with a major psychosis. In the typical form patients appear awake with eyes open but make no voluntary or responsive movements, although they blink spontaneously, swallow, and may not appear distressed. As often, the eyes are half-open as if the patient is in a fog or light sleep. There are signs that indicate voluntary attempts to appear less than fully responsive, though it may take some ingenuity on the part of the examiner to demonstrate these. Eyelid elevation is actively resisted, blinking occurs in response to a visual threat, and the eyes move concomitantly with head rotation, all signs belying a brain lesion. It is characteristic but not invariable for the limbs to retain the posture, no matter how bizarre, in which they have been placed by the examiner ("waxy flexibility," or catalepsy.) Upon recovery, such patients have some memory of events that occurred during their catatonic stupor. The appearance is superficially similar to akinetic mutism, but clinical evidence of brain damage is lacking.

The *locked-in state* describes a pseudocoma in which an awake patient has no means of producing speech or volitional limb, face, and pharyngeal movements in order to indicate that he or she is awake, but vertical eye movements and lid elevation remain

unimpaired, thus allowing the patient to signal. Such individuals have written entire treatises using Morse code. Infarction or hemorrhage of the ventral pons, which transects all descending corticospinal and corticobulbar pathways, is the usual cause. A similar awake but deefferented state occurs as a result of total paralysis of the musculature in severe cases of Guillain-Barre syndrome (Chap. 378), critical illness neuropathy (Chap. 376), and pharmacologic neuromuscular blockade.

THE ANATOMY AND PHYSIOLOGY OF UNCONSCIOUSNESS

To the extent that all complex waking behaviors require the widespread participation of the cerebral cortex, consciousness cannot exist without the activity of these structures. A loosely grouped aggregation of neurons located in the upper brainstem and medial thalamus, theRAS, maintains the cerebral cortex in a state of wakeful consciousness. It follows that the principal causes of coma are (1) lesions that damage a substantial portion of the RAS; (2) destruction of large portions of both cerebral hemispheres; and (3) suppression of thalamocerebral function by drugs, toxins, or by internal metabolic derangements such as hypoglycemia, anoxia, azotemia, or hepatic failure.

The classic animal experiments of Moruzzi and Magoun, published in 1949, and subsequent human clinicopathologic observations have established that the regions of the reticular formation that are critical to the maintenance of wakefulness extend from the caudal midbrain to the lower thalamus. A most important practical consideration derives from the anatomic proximity of theRAS to structures that are concerned with pupillary function and eye movements. Pupillary enlargement and loss of vertical and adduction movements of the globes suggest that upper brainstem damage may be the source of coma. Although circumscribed lesions confined to one or both cerebral hemispheres do not affect the brainstem RAS, a large mass on one side of the brain may cause coma by secondarily compressing the upper brainstem and consequently producing abnormalities of the pupils and eye movements (see discussion of transtentorial herniation below). This type of indirect effect is most typical of cerebral hemorrhages and of rapidly expanding tumors within a cerebral hemisphere. In all cases the degree of diminished alertness also relates to the rapidity of evolution and the extent of compression of the RAS.

The neurons of theRAS are thought to project rostrally to the cortex primarily via thalamic relay nuclei that in turn exert a tonic influence on the activity of the entire cerebral cortex. The behavioral arousal effected by somesthetic, auditory, and visual stimuli depends upon the rich reciprocal innervation that the RAS receives from these sensory systems. The relays between the RAS and the thalamic and cortical areas utilize a variety of neurotransmittors. Of these, the effect of arousal on acetylcholine and on the biogenic amines has been studied more extensively. Cholinergic fibers connect the midbrain to other areas of the upper brainstem, thalamus, and cortex. Serotonin and norepinephrine also subserve important functions in regulation of the sleep-wake cycle (Chap. 27). Their roles in arousal and coma have not been clearly established, although the alerting effects of amphetamines are likely to be mediated by catecholamine release.

Coma Due to Cerebral Mass Lesions and Herniations The cranial cavity is separated into compartments by infoldings of the dura -- the two cerebral hemispheres are

separated by the falx, and the anterior and posterior fossae by the tentorium. *Herniation* refers to displacement of brain tissue away from a mass and into a compartment that it normally does not occupy. Many of the signs associated with coma, and indeed coma itself, can be attributed to these tissue shifts. Herniation can be *transfalcial* (displacement of the cingulate gyrus under the falx and across the midline), *transtentorial* (displacement of the medial temporal lobe into the tentorial opening), and *foraminal* (downward forcing of the cerebellar tonsils into the foramen magnum; Fig. 24-1).

Uncal transtentorial herniation refers to impaction of the anterior medial temporal gyrus (the uncus) into the anterior portion of the tentorial opening. The displaced tissue compresses the third nerve as it traverses the subarachnoid space and results in enlargement of the ipsilateral pupil (putatively because the fibers subserving parasympathetic pupillary function are located peripherally in the nerve). The coma that follows may be due to lateral compression of the midbrain against the opposite tentorial edge by the displaced parahippocampal gyrus (Fig. 24-2). In some cases the lateral displacement causes compression of the opposite cerebral peduncle, producing a Babinski response and hemiparesis contralateral to the original hemiparesis (the Kernohan-Woltman sign). In addition to compressing the upper brainstem, tissue shifts, including herniations, may compress major blood vessels, particularly the anterior and posterior cerebral arteries as they pass over the tentorial reflections, thus producing brain infarctions. The distortions may also entrap portions of the ventricular system, resulting in regional hydrocephalus.

Central transtentorial herniation denotes a symmetric downward movement of the upper thalamic region through the tentorial opening. Miotic pupils and drowsiness are the heralding signs. Both temporal and central herniations are thought to cause progressive compression of the brainstem from above: first the midbrain, then the pons, and finally the medulla. The result is a sequential appearance of neurologic signs that corresponds to the affected level.

A direct relationship between the various configurations of transtentorial herniations and coma is, at best, tenuous. The orderly progression of signs from midbrain to medulla is often bypassed in catastrophic lesions where all brainstem functions are lost almost simultaneously. It is also clear that displacement of deep brain structures by a mass in any direction, with or without herniation, compresses the region of theRAS and results in coma. Furthermore, drowsiness and stupor typically occur with moderate lateral shifts at the level of the diencephalon (thalami) well before transtentorial or other herniations are evident. Lateral shift is easily quantified on axial images of computed tomography (CT) and magnetic resonance imaging (MRI) scans (Fig. 24-2). In cases of *acutely appearing masses*, a fairly consistent and simple relationship exists between the degree of horizontal displacement of midline structures and the level consciousness. Specifically, horizontal displacement of the pineal calcification of 3 to 5 mm is generally associated with drowsiness, 6 to 8 mm with stupor, and>9 mm with coma. At the same time, intrusion of the medial temporal lobe into the tentorial opening may be apparent as an obliteration of the cisterns that surround the upper brainstem.

Coma and Confusional States Due to Metabolic Disorders A large variety of systemic metabolic abnormalities cause coma by interrupting the delivery of energy

substrates (hypoxia, ischemia, hypoglycemia) or by altering neuronal excitability (drug and alcohol intoxication, anesthesia, and epilepsy). The same metabolic abnormalities that produce coma may in milder form induce widespread cortical dysfunction and an acute confusional state. Thus, in metabolic encephalopathies, clouded consciousness and coma are a continuum. Neuropathologic changes in the various metabolic failures are variable -- very evident in hypoxia-ischemia, manifest as astrocytic changes in hepatic coma, and negligible in renal and other metabolic encephalopathies.

Cerebral neurons are fully dependent on cerebral blood flow (CBF) and the related delivery of oxygen and glucose. CBF approximates 75 mL per 100 g/min in gray matter and 30 mL per 100 g/min in white matter (mean = 55 mL per 100 g/min); oxygen consumption is 3.5 mL per 100 g/min, and glucose utilization is 5 mg per 100 g/min. Brain stores of glucose provide energy for approximately 2 min after blood flow is interrupted, and oxygen stores last 8 to 10 s after the cessation of blood flow. Simultaneous hypoxia and ischemia exhaust glucose more rapidly. The electroencephalogram (EEG) rhythm in these circumstances becomes diffusely slowed, typical of metabolic encephalopathies, and as conditions of substrate delivery worsen, eventually all recordable brain electrical activity ceases. In almost all instances of metabolic encephalopathy, the global metabolic activity of the brain is reduced in proportion to the degree of unconsciousness.

Conditions such as hyponatremia, hyperosmolarity, hypercapnia, hypercalcemia, and hepatic and renal failure are associated with a variety of alterations in neurons and astrocytes. It should be stated at the outset that the reversible effects of these conditions on the brain are not understood, but they may in different circumstances impair energy supplies, change ion fluxes across neuronal membranes, and cause neurotransmitter abnormalities. For example, the high brain ammonia concentration that is associated with hepatic coma interferes with cerebral energy metabolism and with the Na^+, K^+-ATPase pump, increases the number and size of astrocytes, alters nerve cell function, and causes increased concentrations of potentially toxic products of ammonia metabolism; it may also result in abnormalities of neurotransmitters, including possible "false" neurotransmitters that may be active at receptor sites. Apart from hyperammonemia, which of these mechanisms is of critical importance is not clear. The mechanism of the encephalopathy of renal failure is also not known. Unlike ammonia, urea itself does not produce central nervous system (CNS) toxicity. A multifactorial causation has been proposed, including increased permeability of the blood-brain barrier to toxic substances such as organic acids and an increase in brain calcium or cerebrospinal fluid (CSF) phosphate content. Likewise, the basis of confusion and drowsiness that commonly accompanies the septic state has not been clarified.

Coma and seizures are a common accompaniment of any large shifts in sodium and water balance. These changes in osmolarity may be the result of a number of systemic medical disorders including diabetic ketoacidosis, the nonketotic hyperosmolar state, and hyponatremia from any cause (e.g., water intoxication, excessive secretion of antidiuretic hormone or atrial natriuretic peptides). The volume of brain water correlates with the level of consciousness in these states, but other factors also play a role. Sodium levels below 125 mmol/L induce confusion, and below 115 mmol/L are associated with coma and convulsions. In hyperosmolar coma the serum osmolarity generally exceeds 350 mosmol/L. *As in most other metabolic encephalopathies, the*

severity of neurologic change depends to a large degree on the rapidity with which the serum changes occur. Hypercapnia depresses the level of consciousness in proportion to the rise in CO_2tension in the blood and depends very much on the rapidity of change. The pathophysiology of other metabolic encephalopathies such as hypercalcemia, hypothyroidism, vitamin B_{12}deficiency, and hypothermia are incompletely understood but must also reflect derangements ofCNSbiochemistry and membrane function.

Epileptic Coma Although all metabolic derangements in some way alter neuronal electrophysiologic function, epilepsy is the only primary excitatory disturbance of brain electrical activity that is encountered in clinical practice. Continuous, generalized electrical discharges of the cortex (*seizures*) are associated with coma even in the absence of epileptic motor activity (*convulsions*). The self-limited coma that follows seizures, termed the *postictal state*, may be due to exhaustion of energy reserves or effects of locally toxic molecules that are the byproduct of seizures. The postictal state produces a pattern of continuous, generalized slowing of the backgroundEEGactivity similar to that of other metabolic encephalopathies.

Pharmacologic Coma This class of encephalopathy is in large measure reversible and leaves no residual damage providing hypoxia does not supervene. Many drugs and toxins are capable of depressing nervous system function. Some produce coma by affecting both the brainstem nuclei, including theRAS, and the cerebral cortex. The combination of cortical and brainstem signs, which occurs in certain drug overdoses, may lead to an incorrect diagnosis of structural brainstem disease.

Approach to the Patient

The diagnosis and management of coma depend on knowledge of its main causes (see "Differential Diagnosis," below) and on interpretation of salient clinical signs, notably brainstem reflexes and motor function. Acute respiratory and cardiovascular problems should be attended to prior to neurologic assessment. A complete medical evaluation, except for the vital signs, funduscopy, and examination for nuchal rigidity, may be deferred until the neurologic evaluation has established the severity and nature of coma.

History In many cases, the cause of coma is immediately evident (e.g., trauma, cardiac arrest, or known drug ingestion). In the remainder, historic information about the onset of coma is often sparse, but certain historic points are especially useful: (1) the circumstances and rapidity with which neurologic symptoms developed; (2) the details of any immediately preceding medical and neurologic symptoms (confusion, weakness, headache, fever, seizures, dizziness, double vision, or vomiting); (3) the use of medications, illicit drugs, or alcohol; and (4) chronic liver, kidney, lung, heart, or other medical disease. Direct interrogation or telephone calls to family and observers on the scene are an important part of the initial evaluation. Ambulance technicians often provide the most useful information in an enigmatic case.

General Physical Examination The temperature, pulse, respiratory rate and pattern, and blood pressure should be measured quickly as the evaluation is getting under way. Fever suggests a systemic infection, bacterial meningitis, or encephalitis; only rarely is it attributable to a brain lesion that has disturbed temperature-regulating centers. A slight elevation in temperature may follow vigorous convulsions. High body temperature, 42 to

44°C, associated with dry skin should arouse the suspicion of heat stroke or anticholinergic drug intoxication. Hypothermia is observed with bodily exposure to lowered environmental temperature; alcoholic, barbiturate, sedative, or phenothiazine intoxication; hypoglycemia; peripheral circulatory failure; or hypothyroidism. Hypothermia itself causes coma only when the temperature is<31°C. Tachypnea may indicate acidosis or pneumonia. Aberrant respiratory patterns that may reflect brainstem disorders are discussed below. Marked hypertension, a sign of hypertensive encephalopathy or a rapid rise in intracranial pressure, may occur acutely after head injury. Hypotension is characteristic of coma from alcohol or barbiturate intoxication, internal hemorrhage, myocardial infarction, sepsis, profound hypothyroidism, or Addisonian crisis. The funduscopic examination is invaluable in detecting subarachnoid hemorrhage (subhyaloid hemorrhages), hypertensive encephalopathy (exudates, hemorrhages, vessel-crossing changes, papilledema), and increased intracranial pressure (papilledema). Generalized cutaneous petechiae suggest thrombotic thrombocytopenic purpura, meningococcemia, or a bleeding diathesis from which an intracerebral hemorrhage arises.

Neurologic Assessment The patient should be observed first without examiner intervention. Patients who toss about, reach up toward the face, cross their legs, yawn, swallow, cough, or moan are close to being awake. Lack of restless movements on one side or an outturned leg at rest suggests a hemiplegia. Intermittent twitching movements of a foot, finger, or facial muscle may be the only sign of seizures. Multifocal myoclonus almost always indicates a metabolic disorder, particularly azotemia, anoxia, or drug ingestion (lithium and haloperidol are particularly prone to cause this sign), or the rarer conditions of spongiform encephalopathy and Hashitmoto disease. In a drowsy and confused patient bilateral asterixis is a certain sign of metabolic encephalopathy or drug ingestion.

The terms *decorticate rigidity* and *decerebrate rigidity*, or "posturing," describe stereotyped arm and leg movements occurring spontaneously or elicited by sensory stimulation. Flexion of the elbows and wrists and supination of the arm (decortication) suggests severe bilateral damage rostral to the midbrain, whereas extension of the elbows and wrists with pronation (decerebration) indicates damage to motor tracts in the midbrain or caudal diencephalon. The less frequent combination of arm extension with leg flexion or flaccid legs is associated with lesions in the pons. These concepts have been adapted from animal work and cannot be applied with the same precision to coma in humans. In fact, acute and widespread cerebral disorders of any type, regardless of location, frequently cause limb extension, and almost all such extensor posturing becomes predominantly flexor as time passes. Thus, posturing alone cannot be utilized for precise anatomic localization. Posturing may also be unilateral and may coexist with purposeful limb movements, usually reflecting incomplete damage to the motor system.

Level of Arousal and Elicited Movements If the patient is not aroused by a conversational volume of voice, a sequence of increasingly intense stimuli is used to determine the patient's threshold of arousal and the optimal motor response of each limb. It should be recognized that the results of this testing may vary from minute to minute and that serial examinations are most useful. Tickling the nostrils with a cotton wisp is a moderate stimulus to arousal -- all but deeply stuporous and comatose patients will move the head away and rouse to some degree. Using the hand to remove

an offending stimulus such as this one represents an even lesser degree of unresponsiveness.

Responses to noxious stimuli should be appraised critically. Stereotyped posturing indicates severe dysfunction of the corticospinal system. Abduction-avoidance movement of a limb is usually purposeful and denotes an intact corticospinal system extending from the contralateral cortex to the ipsilateral spinal cord. Pressure on the knuckles or bony prominences and pinprick are humane forms of noxious stimulus; pinching the skin causes unsightly ecchymoses and is generally not necessary but may be useful in eliciting abduction withdrawal movements of the limbs. Conversely, consistent (obligatory) adduction and flexion of stimulated limbs may be reflexive in origin and implies damage to the corticospinal system. Brief clonus or twitching may occur at the end of extensor posturing movements and should not be mistaken for convulsions.

Brainstem Reflexes Assessment of brainstem damage is essential to the localization of the lesion in coma (Fig. 24-3). The brainstem reflexes that are conveniently assessed are pupillary responses to light, spontaneous and elicited eye movements, corneal responses, and the respiratory pattern. As a rule, when these brainstem activities are preserved, particularly the pupil reactions and eye movements, coma must necessarily be ascribed to bilateral hemispheral disease. The converse, however, is not always true as a mass in the hemispheres may be the proximate cause of coma but nonetheless produce brainstem signs.

PUPILS Pupillary reactions are examined with a bright, diffuse light (not an ophthalmoscope); if the response is absent, this should be confirmed by observation through a magnifying lens. Reaction to light is often difficult to appreciate in pupils<2 mm in diameter, and bright room lighting mutes pupillary reactivity. Normally reactive and round pupils of midsize (2.5 to 5 mm) essentially exclude midbrain damage, either primary or secondary to compression. One unreactive and enlarged pupil (>6 mm) or one that is poorly reactive signifies a compression or stretching of the third nerve from the effects of a mass above. Enlargement of the pupil contralateral to a mass may occur but is infrequent. It may be found in cases of subdural hematoma or brain hemorrhage, possibly as a result of compression of the midbrain or third nerve against the opposite tentorial margin. An oval and slightly eccentric pupil is a transitional sign that accompanies early midbrain-third nerve compression. The most extreme pupillary sign, bilaterally dilated and unreactive pupils, indicates severe midbrain damage, usually from compression by a mass or from ingestion of drugs with anticholinergic activity. The use of mydriatic eye drops, by a previous examiner or self-administered by the patient, and direct ocular trauma are among the causes of misleading pupillary enlargement.

Unilateral miosis in coma has been attributed to dysfunction of sympathetic efferents originating in the posterior hypothalamus and descending in the tegmentum of the brainstem to the cervical cord. Reactive and bilaterally small (1 to 2.5 mm) but not pinpoint pupils are seen in metabolic encephalopathies or in deep bilateral hemispheral lesions such as hydrocephalus or thalamic hemorrhage. Very small but reactive pupils (<1 mm) characterize narcotic or barbiturate overdoses but also occur with extensive pontine hemorrhage. The response to naloxone and the presence of reflex eye movements (see below) distinguish these. The unilaterally small pupil of the Horner

syndrome is detected by failure of the pupil to enlarge in the dark. It is an occasional finding with a large cerebral hemorrhage that affects the thalamus.

OCULAR MOVEMENTS Eye movements are the second sign of importance in determining if the brainstem has been damaged. Abnormalities, implicate both midbrain and pontine functions, thus permitting the analysis of a large portion of the brainstem. The eyes are first observed by elevating the lids and noting the resting position and spontaneous movements of the globes. Lid tone, tested by lifting the eyelids and noting their resistance to opening and the speed of closure, is reduced progressively as coma deepens. Horizontal divergence of the eyes at rest is normal in drowsiness. As coma deepens, the ocular axes may become parallel again. An abducted eye indicates a medial rectus paresis due to third nerve dysfunction and has the same significance as pupillary enlargement. An adducted eye indicates lateral rectus paresis due to a sixth nerve lesion and, when bilateral, is often a sign of increased intracranial pressure. With few exceptions, vertical separation of the ocular axes (one eye lower than the other, i.e. skew deviation) results from pontine or cerebellar lesions but may also be a manifestation of a partial third nerve palsy.

Spontaneous eye movements in coma often take the form of conjugate horizontal roving. This finding alone exonerates the midbrain and pons and has the same meaning as normal reflex eye movements (see below). Cyclic vertical downward movements are seen in some circumstances. "Ocular bobbing" describes a brisk downward and slow upward movement of the eyes associated with loss of horizontal eye movements and is diagnostic of bilateral pontine damage, characteristically from thrombosis of the basilar artery. "Ocular dipping" is a slower, arrhythmic downward movement followed by a faster upward movement in patients with normal reflex horizontal gaze; it usually indicates diffuse cortical anoxic damage. The eyes may turn down and inward as a result of thalamic and upper midbrain lesions, typically with thalamic hemorrhage or dilatation of the third ventricle from hydrocephalus. Conjugate horizontal ocular deviation to one extreme at rest indicates damage to the pons on the side of the gaze paresis or a lesion in the frontal lobe on the opposite side. This phenomenon may be summarized by the following maxim: *The eyes look toward a hemispheral lesion and away from a brainstem lesion.* On rare occasions, the eyes may turn paradoxically away from the side of a deep hemispheral lesion ("wrong-way eyes"). Many other complex and interesting eye movements are known but do not have the same salience in coma as the ones already mentioned.

Oculocephalic reflexes are automatic movements of the eyes elicited by moving the head from side to side or vertically. As the activity of the hemispheres is subdued from whatever cause, eye movements are evoked in the direction opposite to the head movement (Fig. 24-3). These movements, called somewhat inappropriately "doll's eyes" (which more accurately refers to the reflex elevation of the eyelids with flexion of the neck) are suppressed by visual fixation, which requires the patient to be awake. Induced adduction of the globes tends to be less complete than abduction, hence subtle abnormalities in the doll's-eye manuever should be interpreted with caution. Oculocephalic reflexes are generated by brainstem mechanisms originating in the labyrinths and in cervical proprioceptors and require the undiminished activity of the third nerve nucleus in the midbrain, the contralateral sixth nerve nucleus in the pons, and the medial longitudinal fasciculus (MLF) that runs virtually the length of the

brainstem and links the two. Preservation of reflex eye movements (particularly adduction) therefore informs the examiner that coma is probably not due to an upper brainstem lesion and by implication that the origin of unconsciousness lies in the cerebral diencephalic structures. However, the opposite -- the absence of eye movements -- may signify either damage within the brainstem or profound metabolic depression of all neuronal function including the brainstem nuclei. Metabolic causes of depressed neuronal function include overdoses of phenytoin, tricyclic antidepressants, barbiturates, alcohol, phenothiazines, diazepam, and neuromuscular blocking agents. The presence of normal pupillary size and light reaction will distinguish most drug-induced comas from structural brainstem damage.

Thermal, or "caloric," stimulation of the vestibular apparatus (oculovestibular response) provides a more intense stimulus that may be used to confirm the absence of the oculocephalic reflex but gives fundamentally the same information. The test is performed by irrigating the external auditory canal with cool water in order to induce convection currents in the labyrinths. After a brief latency, the result is tonic deviation of both eyes (lasting 30 to 120 s) to the side of cool-water irrigation. The integrity of the third and sixth nerve complexes and brainstem pathways from the labyrinths to the midbrain are thereby confimed, thus excluding a brainstem lesion as the cause of coma. If the cerebral hemispheres are functioning, as in catatonic or hysterical pseudocoma, an obligate rapid corrective nystagmus is generated away from the side of tonic deviation. (The acronym "COWS" has been used to remind generations of medical students of the direction of compensatory nystagmus -- "cold water opposite, warm water same"). The absence of this nystagmus despite conjugate deviation of the globes signifies that the cerebral hemispheres are damaged or profoundly suppressed.

By touching the cornea with a wisp of cotton, a response consisting of brief bilateral lid closure is normally observed. Although the corneal reflexes are rarely useful alone, they may corroborate eye-movement abnormalities because they also depend on the integrity of pontine pathways. The response is lost if the reflex connections between the fifth (afferent) and both seventh (efferent) cranial nerves within the pons are damaged.CNSdepressant drugs diminish or eliminate the corneal responses soon after reflex eye movements are paralyzed but before the pupils become unreactive to light. The corneal (and pharyngeal) response may be lost for a time on the side of an acute hemiplegia.

RESPIRATION Respiratory patterns have received much attention in coma diagnosis but are of less localizing value in comparison to other brainstem signs. Shallow, slow, but regular breathing suggests metabolic or drug depression. Cheyne-Stokes respiration in its classic cyclic form, ending with a brief apneic period, signifies bihemispheral damage or metabolic suppression and commonly accompanies light coma. Rapid, deep (Kussmaul) breathing usually implies metabolic acidosis but may also occur with pontomesencephalic lesions and, of course, severe pneumonia. Agonal gasps reflect bilateral lower brainstem damage and are well known as the terminal respiratory pattern of severe brain damage. A number of other cyclic breathing variations are of lesser significance for localization.

LABORATORY STUDIES AND IMAGING

The following studies are most useful in the diagnosis of confusional states and coma: chemical-toxicologic analysis of blood and urine, cranialCT orMRI,EEG, andCSFexamination. Arterial blood-gas analysis is helpful in patients with lung disease and acid-base disorders. Chemical blood determinations are obtained routinely to disclose metabolic, toxic, or drug-induced encephalopathies. The metabolic aberrations commonly encountered in clinical practice require measurements of electrolytes, glucose, calcium, osmolarity, and renal (blood urea nitrogen) and hepatic (NH_3) function. Toxicologic analysis is necessary in any case of coma where the diagnosis is not immediately clear. However, the presence of exogenous drugs or toxins, especially alcohol, does not exclude the possibility that other factors, particularly head trauma, are also contributing to the clinical state. An ethanol level of 43 mmol/L (200 mg/dL) in nonhabituated patients generally causes confusion and impaired mental activity and of>65 mmol/L (300 mg/dL) is associated with stupor. The development of tolerance may allow the chronic alcoholic to remain awake at levels >87 mmol/L (400 mg/dL).

The increased availability ofCT andMRI has focused attention on causes of coma that are radiologically detectable (e.g., hemorrhages, tumors, or hydrocephalus). Resorting primarily to this approach, although at times expedient, is imprudent because most cases of coma (and confusion) are metabolic or toxic in origin. The notion that a normal CT scan excludes anatomic lesions as the cause of coma is also erroneous. Bilateral hemisphere infarction, small brainstem lesions, encephalitis, meningitis, mechanical shearing of axons as a result of closed head trauma, absent cerebral perfusion associated with brain death, sagittal sinus thrombosis, and subdural hematomas that are isodense to adjacent brain are some of the lesions that may not be visible. Nevertheless, if the source of coma remains unknown, a scan should be obtained.

The EEG is useful in metabolic or drug-induced confusional states but is rarely diagnostic, with the important exceptions of coma due to clinically unrecognized seizures, to herpesvirus encephalitis and Creutzfeldt-Jakob disease. The amount of background slowing of the EEG is a useful reflection of the severity of any diffuse encephalopathy. Predominant high-voltage slowing (d or triphasic waves) in the frontal regions is typical of metabolic coma, as from hepatic failure, and widespread fast (b) activity implicates sedative drugs (diazepines, barbiturates). A pattern of "a coma," defined by widespread, variable 8- to 12-Hz activity, superficially resembles the normal a rhythm of waking but is unresponsive to environmental stimuli. It results from pontine or diffuse cortical damage and has a poor prognosis. Most importantly, EEG recordings reveal coma that is due to persistent epileptic discharges that are not clinically manifested as convulsions. Normal aactivity on the EEG may also alert the clinician to the locked-in syndrome or to hysteria or catatonia.

Lumbar puncture is used more judiciously than in prior decades in cases of coma or confusion because neuroimaging scans effectively exclude intracerebral hemorrhage and most cases of subarachnoid hemorrhages. However, examination of theCSF is indispensable in the diagnosis of meningitis and encephalitis and in instances of suspected subarachnoid hemorrhage in which the scan is normal. Lumbar puncture should therefore not be deferred if meningitis is a possibility. Xanthochromia, indicating preexisting blood in the CSF, is documented by spinning the CSF in a large tube and comparing the supernatant to water. Measurement of the opening pressure within the subarachnoid space is of further help in interpreting abnormalities of the cell count and

protein content of the CSF.

DIFFERENTIAL DIAGNOSIS OF COMA (Table 24-1)

In most instances confusion and coma are part of an obvious medical problem such as overt drug ingestion, hypoxia, stroke, trauma, or liver or kidney failure. Attention is then appropriately focused on the primary illness. Some general rules are helpful. Illnesses that cause sudden onset of coma are due to drug ingestion or to cerebral hemorrhage, trauma, cardiac arrest, epilepsy, or basilar artery embolism. Coma that appears subacutely is usually related to a preceding medical or neurologic problem, including the secondary brain swelling that surrounds a preexisting lesion such as a tumor or cerebral infarction.

The structural causes of coma can also be conceptualized in three broad categories: those without focal or lateralizing neurologic signs (e.g., metabolic encephalopathies); meningitis syndromes, characterized by stiff neck and an excess of cells in the spinal fluid (e.g., bacterial meningitis, subarachnoid hemorrhage); and those with prominent focal signs (e.g., stroke, cerebral hemorrhage). These are elaborated in Table 24-1.

Cerebrovascular diseases cause the greatest difficulty in coma diagnosis. These are described in more detail in Chap. 361 but may be summarized as follows: (1) basal ganglia and thalamic hemorrhage (acute but not instantaneous onset, vomiting, headache, hemiplegia, and characteristic eye signs); (2) pontine hemorrhage (sudden onset, pinpoint pupils, loss of reflex eye movements and corneal responses, ocular bobbing, posturing, hyperventilation, and excessive sweating); (3) cerebellar hemorrhage (occipital headache, vomiting, gaze paresis, and inability to stand); (4) basilar artery thrombosis (neurologic prodrome or warning spells, diplopia, dysarthria, vomiting, eye movement and corneal response abnormalities, and asymmetric limb paresis); and (5) subarachnoid hemorrhage (precipitous coma after headache and vomiting). The most common stroke, infarction in the territory of the middle cerebral artery, does not cause coma acutely but the surrounding edema may expand and act as a mass in a limited number of patients with large infarcts. The syndrome of acute hydrocephalus may accompany many intracranial diseases, particularly subarachnoid hemorrhage. Acute symmetric enlargement of both lateral ventricles causes headache and sometimes vomiting that may progress quickly to coma, with extensor posturing of the limbs, bilateral Babinski signs, small nonreactive pupils, and impaired vertical oculocephalic movements in the vertical direction.

If the history and examination do not suggest a large cerebral lesion or meningitic syndrome or a metabolic or drug cause, then information obtained fromCT orMRI may be needed as outlined in Table 24-1. As mentioned earlier, the majority of medical causes of coma can be established without a neuroimaging study.

BRAIN DEATH

This is a state in which there has been cessation of cerebral blood flow; as a result, global ischemia of the brain occurs while respiration is maintained by artificial means and the heart continues to function. It is the only type of brain damage that is unequivocally recognized as death. Many roughly equivalent criteria have been

advanced for the diagnosis of brain death, and it is essential to adhere to those endorsed as standards by the local medical community. Ideal criteria are simple, can be conducted at the bedside, and allow no chance of diagnostic error. They contain three essential elements: (1) widespread cortical destruction shown by deep coma -- unresponsiveness to all forms of stimulation; (2) global brainstem damage demonstrated by absent pupillary light reaction and the loss of oculovestibular and corneal reflexes; and (3) lower brainstem destruction indicated by complete apnea. The pulse rate is also invariant and unresponsive to atropine. Most patients have diabetes insipidus, but in some it develops only hours or days after the clinical signs of brain death. The pupils are often enlarged and may be mid-sized but should not be constricted. The absence of deep tendon reflexes is not required because the spinal cord may remain functional.

The proof that apnea is due to irreversible medullary damage requires that the P_{CO2} be high enough to stimulate respiration during a test of spontaneous breathing (apnea test). This can be done safely in most patients by the use, prior to removing the ventilator, of diffusion oxygenation. This is accomplished by preoxygenation with 100% oxygen and then sustained during the test by a tracheal cannula connected to an oxygen supply. CO_2tension increases approximately 0.3 to 0.4 kPa/min (2 to 3 mmHg/min) during apnea. At the end of the period of observation, typically several minutes in duration, arterial P_{CO2}should be at least >6.6 to 8.0 kPa (50 to 60 mmHg) for the test to be valid.

The possibility of profound drug-induced or hypothermic depression of the nervous system should be excluded, and some period of observation, usually 6 to 24 h, is desirable during which this state is shown to be sustained. It is particularly advisable to delay clinical testing for up to 24 h if a cardiac arrest has caused brain death or if the inciting disease is not known. An isoelectricEEG may be used as a confirmatory test for total cerebral damage but is not absolutely necessary. Radionuclide brain scanning, cerebral angiography, or transcranial Doppler measurements may also be used to demonstrate the absence of cerebral blood flow, but with the exception of the latter, they are cumbersome and have not been correlated extensively with pathology.

There is no compelling reason to demonstrate brain death except when organ transplantation is involved. Although it is largely accepted in western society that the respirator can be disconnected from a brain-dead patient, problems frequently arise because of inadequate explanation and preparation of the family by the physician. Moreover, there is no proscription in reasonable medical practice to removing such support from patients who are not brain dead but whose condition is nonetheless hopeless and are likely to live for only a brief time.

TREATMENT

The immediate goal in acute coma is the prevention of further nervous system damage. Hypotension, hypoglycemia, hypercalcemia, hypoxia, hypercapnia, and hyperthermia should be corrected rapidly and assiduously. An oropharyngeal airway is adequate to keep the pharynx open in drowsy patients who are breathing normally. Tracheal intubation is indicated if there is apnea, upper airway obstruction, hypoventilation, or emesis, or if the patient is liable to aspirate because of coma. Mechanical ventilation is

required if there is hypoventilation or if there is an intracranial mass and a need to induce hypocapnia in order to lower intracranial pressure (ICP) as described below. Intravenous access is established and naloxone and dextrose are administered if narcotic overdose or hypoglycemia are even remote possibilities, and thiamine is given with glucose in order to avoid eliciting Wernicke disease in malnourished patients. In cases of suspected basilar thrombosis with brainstem ischemia, intravenous heparin or a thrombolytic agent is often utilized, keeping in mind that cerebellar and pontine hemorrhages resemble basilar artery occlusion. Physostigmine, when used by experienced physicians and with careful monitoring, may awaken patients with anticholinergic-type drug overdose, but many physicians believe that this is justified only to treat cardiac arrhythmias resulting from these overdoses. The use of benzodiazepine antagonists offers some prospect of improvement after overdoses of soporific drugs and has transient benefit in hepatic encephalopathy. Intravenous administration of hypotonic solutions should be monitored carefully in any serious acute brain illness because of the potential for exacerbating brain swelling. Cervical spine injuries must not be overlooked, particularly prior to attempting intubation or the evaluation of oculocephalic responses. Headache accompanied by fever and meningismus indicates an urgent need for examination of theCSF to diagnose meningitis, and it is worth reemphasizing that lumbar puncture should not be delayed while awaiting aCT scan. If the lumbar puncture in a case of suspected meningitis is delayed for any reason, an antibiotic such as a third-generation cephalosporin should be administered as soon as possible, preferably after obtaining blood cultures.

Enlargement of one pupil usually indicates secondary midbrain or third nerve compression by a hemispherical mass and requires thatICP be reduced (Chap. 376). Surgical evacuation of the mass may be appropriate in some cases (e.g., subdural and epidural hematoma.). Medical management to reduce ICP begins with the infusion of normal saline (safe because it is slightly hyperosmolar to serum). Therapeutic hyperventilation may be used to reduce ICP by inducing an arterial P_{CO2} of 3.7 to 4.2 kPa (28 to 32 mmHg), but its effects are brief. Hyperosmolar therapy with mannitol or an equivalent agent is the mainstay of ICP reduction. It is used simultaneously with hyperventilation in critical cases. A ventricular puncture is necessary to decompress hydrocephalus if medical measures fail to improve alertness. The routine use of high-dose barbiturates and other neuronal-sparing agents soon after cardiac arrest or head trauma has not been shown in clinical studies to be beneficial, and glucocorticoids, although often still used, have no proven value except in cases of brain tumor with edema.

PROGNOSIS

The prediction of the outcome of coma must be considered in reference to long-term care and medical resources. One hopes to avoid the emotionally painful, hopeless outcomes associated with patients who are left severely disabled or vegetative. Several general rules pertain. The uniformly pessimistic outcome of the persistent vegetative state has already been mentioned. Children and young adults may have ominous early clinical findings such as abnormal brainstem reflexes and yet recover, so that temporization in offering a prognosis in this group of patients is wise. Metabolic comas have a far better prognosis than traumatic comas. All schemes for prognosis in adults should be taken as approximate indicators, and medical judgments must be tempered

by factors such as age, underlying systemic disease, and general medical condition. In an attempt to collect prognostic information from large numbers of patients with head injury, the Glasgow Coma Scale was devised; empirically it has predictive value in cases of brain trauma (Chap. 369). For anoxic and metabolic coma, clinical signs such as the pupillary and motor responses after 1 day, 3 days, and 1 week have been shown to have predictive value (Chap. 376). The absence of the cortical waves of the somatosensory evoked potentials has also proved a strong indicator of poor outcome in coma from any cause.

(Bibliography omitted in Palm version)

Back to Table of Contents

25. APHASIAS AND OTHER FOCAL CEREBRAL DISORDERS - *M.-Marsel Mesulam*

The cerebral cortex of the human brain contains approximately 20 billion neurons spread over an area of 2 m^2. The *primary sensory* areas provide an obligatory portal for the entry of sensory information into cortical circuitry, whereas the *primary motor* areas provide a final common pathway for coordinating complex motor acts. The primary sensory and motor areas constitute <10% of the cerebral cortex. The rest is subsumed by unimodal, heteromodal, paralimbic, and limbic areas, collectively known as the *association cortex* (Fig. 25-1). The association cortex mediates the integrative processes that subserve cognition, emotion, and comportment. A systematic testing of these mental functions is necessary for the effective clinical assessment of the association cortex and its diseases.

According to current thinking, there are no centers for "hearing words," "perceiving space," or "storing memories." Cognitive and behavioral functions (domains) are coordinated by intersecting *large-scale neural networks* that contain interconnected cortical and subcortical components. The network approach to higher cerebral function has at least four implications of clinical relevance: (1) a single domain such as language or memory can be disrupted by damage to any one of several areas, as long as these areas belong to the same network; (2) damage confined to a single area can give rise to multiple deficits, involving the functions of all networks that intersect in that region; (3) damage to a network component may give rise to minimal or transient deficits if other parts of the network undergo compensatory reorganization; and (4) individual anatomic sites within a network display a relative (but not absolute) specialization for different behavioral aspects of the relevant function. Five anatomically defined large-scale networks are most relevant to clinical practice: a perisylvian network for language; a parietofrontal network for spatial cognition; an occipitotemporal network for face and object recognition; a limbic network for retentive memory; and a prefrontal network for attention and comportment.

THE LEFT PERISYLVIAN NETWORK FOR LANGUAGE: APHASIAS AND RELATED CONDITIONS

Language allows the communication and reshaping of thoughts and experiences by linking them to arbitrary symbols known as words. The neural substrate of language is composed of a distributed network centered in the perisylvian region of the *left* hemisphere. The posterior pole of this network is known as *Wernicke's area* and includes the posterior third of the superior temporal gyrus and a surrounding rim of the inferior parietal lobule. An essential function of Wernicke's area is to transform sensory inputs into their neural word representations so that these can establish the distributed associations that give the word its meaning. The anterior pole of the language network, known as *Broca's area*, includes the posterior part of the inferior frontal gyrus and a surrounding rim of prefrontal heteromodal cortex. An essential function of this area is to transform neural word representations into their articulatory sequences so that the words can be uttered in the form of spoken language. The sequencing function of Broca's area also appears to involve the ordering of words into sentences that contain a meaning-appropriate *syntax* (grammar). Wernicke's and Broca's areas are interconnected with each other and with additional perisylvian, temporal, prefrontal, and posterior parietal regions, making up a neural network subserving the various aspects of

language function. Damage to any one of these components or to their interconnections can give rise to language disturbances (*aphasia*). Aphasia should be diagnosed only when there are deficits in the formal aspects of language such as naming, word choice, comprehension, spelling, and syntax. Dysarthria and mutism do not, by themselves, lead to a diagnosis of aphasia. The language network shows a left hemisphere dominance pattern in the vast majority of the population. In approximately 90% of right handers and 60% of left handers, aphasia occurs only after lesions of the left hemisphere. In some individuals no hemispheric dominance for language can be discerned, and in some others (including a small minority of right handers) there is a right hemisphere dominance for language. A language disturbance occurring after a right hemisphere lesion in a right hander is called *crossed aphasia*.

Clinical Examination The clinical examination of language should include the assessment of naming, spontaneous speech, comprehension, repetition, reading, and writing. A deficit of naming (*anomia*) is the single most common finding in aphasic patients. When asked to name common objects (pencil or wristwatch) or their parts (eraser, lead, stem, band), the patient may fail to come up with the appropriate word, may provide a circumlocutious description of the object ("the thing for writing"), or may come up with the wrong word (*paraphasia*). If the patient offers an incorrect but legitimate word ("pen" for "pencil"), the naming error is known as a *semantic paraphasia*; if the word approximates the correct answer but is phonetically inaccurate ("plentil" for "pencil"), it is known as a *phonemic paraphasia*. Asking the patient to name body parts, geometric shapes, and component parts of objects (lapel of coat, cap of pen) can elicit mild forms of anomia in patients who can otherwise name common objects. In most anomias, the patient cannot retrieve the appropriate name when shown an object but can point to the appropriate object when the name is provided by the examiner. This is known as a one-way (or retrieval-based) naming deficit. A two-way naming deficit exists if the patient can neither provide nor recognize the correct name, indicating the presence of a language comprehension impairment. *Spontaneous speech* is described as "fluent" if it maintains appropriate output volume, phrase length, and melody or as "nonfluent" if it is sparse, halting, and average phrase length is below four words. The examiner should also note if the speech is paraphasic or circumlocutious; if it shows a relative paucity of substantive nouns and action verbs versus function words (prepositions, conjunctions); and if word order, tenses, suffixes, prefixes, plurals, and possessives are appropriate. *Comprehension* can be tested by assessing the patient's ability to follow conversation, by asking yes-no questions ("Can a dog fly?", "Does it snow in summer?") or asking the patient to point to appropriate objects ("Where is the source of illumination in this room?"). Statements with embedded clauses or passive voice construction ("If a tiger is eaten by a lion, which animal stays alive?") help to assess the ability to comprehend complex syntactic structure. Commands to close or open the eyes, stand up, sit down, or roll over should not be used to assess overall comprehension since appropriate responses aimed at such axial movements can be preserved in patients who otherwise have profound comprehension deficits.

Repetition is assessed by asking the patient to repeat single words, short sentences, or strings of words such as "No ifs, ands, or buts." The testing of repetition with tongue-twisters such as "hippopotamus" or "Irish constabulary" provides a better assessment of dysarthria and pallilalia than aphasia. Aphasic patients may have little difficulty with tongue-twisters but have a particularly hard time repeating a string of

function words. It is important to make sure that the number of words does not exceed the patient's attention span. Otherwise, the failure of repetition becomes a reflection of the narrowed attention span rather than an indication of an aphasic deficit. *Reading* should be assessed for deficits in reading aloud as well as comprehension. *Writing* is assessed for spelling errors, word order, and grammar. *Alexia* describes an inability to either read aloud or comprehend single words and simple sentences; *agraphia* (or dysgraphia) is used to describe an acquired deficit in the spelling or grammar of written language.

The correspondence between individual deficits of language function and lesion location does not display a rigid one-to-one relationship and should be conceptualized within the context of the distributed network model. Nonetheless, the classification of aphasic patients into specific clinical syndromes helps to determine the most likely anatomic distribution of the underlying neurologic disease and has implications for etiology and prognosis (Table 25-1). Aphasic syndromes can be divided into "central" syndromes, which result from damage to the two epicenters of the language network (Broca's and Wernicke's areas), and "disconnection" syndromes, which arise from lesions that interrupt the functional connectivity of these centers with each other and with the other components of the language network. The syndromes outlined below are idealizations; pure syndromes occur rarely.

Wernicke's Aphasia Comprehension is impaired for spoken and written language. Language output is fluent but is highly paraphasic and circumlocutious. The tendency for paraphasic errors may be so pronounced that it leads to strings of neologisms, which form the basis of what is known as "jargon aphasia." Speech contains large numbers of function words (e.g., prepositions, conjunctions) but few substantive nouns or verbs that refer to specific actions. The output is therefore voluminous but uninformative. For example, a patient attempts to describe how his wife accidentally threw away something important, perhaps his dentures: "We don't need it anymore, she says. And with it when that was downstairs was my teethtick...a...den...dentith...my dentist. And they happened to be in that bag...see? How could this have happened? How could a thing like this happen...So she says we won't need it anymore...I didn't think we'd use it. And now if I have any problems anybody coming a month from now, four months from now, or six months from now, I have a new dentist. Where my two...two little pieces of dentist that I use...that I...all gone. If she throws the whole thing away...visit some friends of hers and she can't throw them away."

Gestures and pantomime do not improve communication. The patient does not seem to realize that his or her language is incomprehensible and may appear angry and impatient when the examiner fails to decipher the meaning of a severely paraphasic statement. In some patients this type of aphasia can be associated with severe agitation and paranoid behaviors. One area of comprehension that may be preserved is the ability to follow commands aimed at axial musculature. The dissociation between the failure to understand simple questions ("What is your name") in a patient who rapidly closes his or her eyes, sits up, or rolls over when asked to do so is characteristic of Wernicke's aphasia and helps to differentiate it from deafness, psychiatric disease, or malingering. Patients with Wernicke's aphasia cannot express their thoughts in meaning-appropriate words and cannot decode the meaning of words in any modality of input. This aphasia therefore has expressive as well as receptive components.

Repetition, naming, reading, and writing are also impaired.

The lesion site most commonly associated with Wernicke's aphasia is the posterior portion of the language network and tends to involve at least parts of Wernicke's area. An embolus to the inferior division of the middle cerebral artery, and to the posterior temporal or angular branches in particular, is the most common etiology (Chap. 361). Intracerebral hemorrhage, severe head trauma, or neoplasm are other causes. A coexisting right hemi- or superior quadrantanopia is common, and mild right nasolabial flattening may be found, but otherwise the examination is often unrevealing. The paraphasic, neologistic speech in an agitated patient with an otherwise unremarkable neurologic examination may lead to the suspicion of a primary psychiatric disorder such as schizophrenia or mania, but the other components characteristic of acquired aphasia and the absence of prior psychiatric disease usually settle the issue. Some patients with Wernicke's aphasia due to intracerebral hemorrhage or head trauma may improve as the hemorrhage or the injury heals. In most other patients, prognosis for recovery is guarded.

Broca's Aphasia Speech is nonfluent, labored, interrupted by many word-finding pauses, and usually dysarthric. It is impoverished in function words but enriched in meaning-appropriate nouns and verbs. Abnormal word order and the inappropriate deployment of *bound morphemes* (word endings used to denote tenses, possessives, or plurals) lead to a characteristic agrammatism. Speech is telegraphic and pithy but quite informative. In the following passage, a patient with Broca's aphasia describes his medical history: "I see...the dotor, dotor sent me...Bosson. Go to hospital. Dotor...kept me beside. Two, tee days, doctor send me home."

Output may be reduced to a grunt or single word ("yes" or "no"), which is emitted with different intonations in an attempt to express approval or disapproval. In addition to fluency, naming and repetition are also impaired. Comprehension of spoken language is intact, except for syntactically difficult sentences with passive voice structure or embedded clauses. Reading comprehension is also preserved, with the occasional exception of a specific inability to read small grammatical words such as conjunctions and pronouns. The last two features indicate that Broca's aphasia is not just an "expressive" or "motor" disorder and that it may also involve a comprehension deficit for function words and syntax. Patients with Broca's aphasia can be tearful, easily frustrated, and profoundly depressed. Insight into their condition is preserved, in contrast to Wernicke's aphasia. Even when spontaneous speech is severely dysarthric, the patient may be able to display a relatively normal articulation of words when singing. This dissociation has been used to develop specific therapeutic approaches (melodic intonation therapy) for Broca's aphasia. Additional neurologic deficits usually include right facial weakness, hemiparesis or hemiplegia, and a buccofacial apraxia characterized by an inability to carry out motor commands involving oropharyngeal and facial musculature (e.g., patients are unable to demonstrate how to blow out a match or suck through a straw). Visual fields are intact. The cause is most often infarction of Broca's area (the inferior frontal convolution;Fig. 25-1) and surrounding anterior perisylvian and insular cortex, due to occlusion of the superior division of the middle cerebral artery (Chap. 361). Mass lesions including tumor, intracerebral hemorrhage, or abscess may also be responsible. Small lesions confined to the posterior part of Broca's area may lead to a nonaphasic and often reversible deficit of speech articulation, usually

accompanied by mild right facial weakness. When the cause of Broca's aphasia is stroke, recovery of language function generally peaks within 2 to 6 months, after which time further progress is limited.

Global Aphasia Speech output is nonfluent, and comprehension of spoken language is severely impaired. Naming, repetition, reading, and writing are also impaired. This syndrome represents the combined dysfunction of Broca's and Wernicke's areas and usually results from strokes that involve the entire middle cerebral artery distribution in the left hemisphere. Most patients are initially mute or say a few words, such as "hi" or "yes." Related signs include right hemiplegia, hemisensory loss, and homonymous hemianopia. Occasionally, a patient with a lesion in Wernicke's area will present with a global aphasia that soon resolves into Wernicke's aphasia.

Conduction Aphasia Speech output is fluent but paraphasic, comprehension of spoken language is intact, and repetition is severely impaired. Naming and writing are also impaired. Reading aloud is impaired, but reading comprehension is preserved. The lesion sites spare Broca's and Wernicke's areas but may induce a functional disconnection between the two so that neural word representations formed in Wernicke's area and adjacent regions cannot be conveyed to Broca's area for assembly into corresponding articulatory patterns. Occasionally, a Wernicke's area lesion gives rise to a transient Wernicke's aphasia that rapidly resolves into a conduction aphasia. The paraphasic output in conduction aphasia interferes with the ability to express meaning, but this deficit is not nearly as severe as the one displayed by patients with Wernicke's aphasia. Associated neurologic signs in conduction aphasia vary according to the primary lesion site.

Nonfluent Transcortical Aphasia (Transcortical Motor Aphasia) The features are similar to Broca's aphasia, but repetition is intact and agrammatism may be less pronounced. The neurologic examination may be otherwise intact, but a right hemiparesis can also exist. The lesion site disconnects the intact language network from prefrontal areas of the brain and usually involves the anterior watershed zone between anterior and middle cerebral artery territories or the supplementary motor cortex in the territory of the anterior cerebral artery.

Fluent Transcortical Aphasia (Transcortical Sensory Aphasia) Clinical features are similar to those of Wernicke's aphasia, but repetition is intact. The lesion site disconnects the intact core of the language network from other temporoparietal association areas. Associated neurologic findings may include hemianopia. Cerebrovascular lesions (e.g., infarctions in the posterior watershed zone) or neoplasms that involve the temporoparietal cortex posterior to Wernicke's area are the most common causes.

Isolation Aphasia This rare syndrome represents a combination of the two transcortical aphasias. Comprehension is severely impaired, and there is no purposeful speech output. The patient may parrot fragments of heard conversations (*echolalia*), indicating that the neural mechanisms for repetition are at least partially intact. This condition represents the pathologic function of the language network when it is isolated from other regions of the brain. Broca's and Wernicke's areas tend to be spared, but there is damage in surrounding frontal, parietal, and temporal cortex. Lesions are patchy and

can be associated with anoxia, carbon monoxide poisoning, or complete watershed zone infarctions.

Anomic Aphasia This form of aphasia may be considered the "minimal dysfunction" syndrome of the language network. Articulation, comprehension, and repetition are intact, but confrontation naming, word finding, and spelling are impaired. Speech is enriched in function words but impoverished in substantive nouns and verbs denoting specific actions. Language output is fluent but paraphasic, circumlocutious, and uninformative. The lesion sites can be anywhere within the left hemisphere language network, including the middle and inferior temporal gyri. *Anomic aphasia is the single most common language disturbance seen in head trauma, metabolic encephalopathy, and Alzheimer's disease*. The language impairment of Alzheimer's disease almost always leads to fluent aphasias (e.g., anomic, Wernicke's, conduction, or fluent transcortical aphasia). The insidious onset and relentless progression of nonfluent language disturbances (Broca's or nonfluent transcortical aphasia) can be seen in *primary progressive aphasia*, a degenerative syndrome most commonly associated with focal nonspecific neuronal loss or Pick's disease.

Pure Word Deafness This is not a true aphasic syndrome because the language deficit is modality-specific. The most common lesions are either bilateral or left-sided in the superior temporal gyrus. The net effect of the underlying lesion is to interrupt the flow of information from the unimodal auditory association cortex to Wernicke's area. Patients have no difficulty understanding written language and can express themselves well in spoken or written language. They have no difficulty interpreting and reacting to environmental sounds since primary auditory cortex and subcortical auditory relays are intact. Since auditory information cannot be conveyed to the language network, however, it cannot be decoded into neural word representations and the patient reacts to speech as if it were in an alien tongue that cannot be deciphered. Patients cannot repeat spoken language but have no difficulty naming objects. In time, patients with pure word deafness teach themselves lip reading and may appear to have improved. There may be no additional neurologic findings, but agitated paranoid reactions are frequent in the acute stages. Cerebrovascular lesions are the most frequent cause.

Pure Alexia Without Agraphia This is the visual equivalent of pure word deafness. The lesions (usually a combination of damage to the left occipital cortex and to a posterior sector of the corpus callosum -- the splenium) interrupt the flow of visual input into the language network. There is usually a right hemianopia, but the core language network remains unaffected. The patient can understand and produce spoken language, name objects in the left visual hemifield, repeat, and write. However, the patient acts as if illiterate when asked to read even the simplest sentence because the visual information from the written words (presented to the intact left visual hemifield) cannot reach the language network. Objects in the left hemifield may be named accurately because they activate nonvisual associations in the right hemisphere, which, in turn, can access the language network through transcallosal pathways anterior to the splenium. Patients with this syndrome may also lose the ability to name colors, although they can match colors. This is known as a *color anomia*. The most common etiology of pure alexia is a vascular lesion in the territory of the posterior cerebral artery or an infiltrating neoplasm in the left occipital cortex that involves the optic radiations as well as the crossing fibers of the splenium. Since the posterior cerebral artery also supplies medial temporal components

of the limbic system, the patient with pure alexia may also experience an amnesia, but this is usually transient because the limbic lesion is unilateral.

Aphemia There is an acute onset of severely impaired fluency (often mutism), which cannot be accounted by corticobulbar, cerebellar, or extrapyramidal dysfunction. Recovery is the rule and involves an intermediate stage of hoarse whispering. Writing, reading, and comprehension are intact, so this is not a true aphasic syndrome. Partial lesions of Broca's area or subcortical lesions that undercut its connections with other parts of the brain may be present. Occasionally, the lesion site is on the medial aspects of the frontal lobes and may involve the supplementary motor cortex of the left hemisphere.

Apraxia This generic term designates a complex motor deficit that cannot be attributed to pyramidal, extrapyramidal, cerebellar, or sensory dysfunction and that does not arise from the patient's failure to understand the nature of the task. The form that is most frequently encountered in clinical practice is known as *ideomotor apraxia*. Commands to perform a specific motor act ("cough," "blow out a match") or to pantomime the use of a common tool (a comb, hammer, straw, or toothbrush) in the absence of the real object cannot be followed. The patient's ability to comprehend the command is ascertained by demonstrating multiple movements and establishing that the correct one can be recognized. Some patients with this type of apraxia can imitate the appropriate movement (when it is demonstrated by the examiner) and show no impairment when handed the real object, indicating that the sensorimotor mechanisms necessary for the movement are intact. Some forms of ideomotor apraxia represent a disconnection of the language network from pyramidal motor systems: commands to execute complex movements are understood but cannot be conveyed to the appropriate motor areas, even though the relevant motor mechanisms are intact. *Buccofacial apraxia* involves apraxic deficits in movements of the face and mouth. *Limb apraxia* encompasses apraxic deficits in movements of the arms and legs. Ideomotor apraxia is almost always caused by lesions in the left hemisphere and is commonly associated with aphasic syndromes, especially Broca's aphasia and conduction aphasia. Its presence cannot be ascertained in patients with language comprehension deficits. The ability to follow commands aimed at axial musculature ("close the eyes," "stand up") is subserved by different pathways and may be intact in otherwise severely aphasic and apraxic patients. Patients with lesions of the anterior corpus callosum can display a special type of ideomotor apraxia confined to the left side of the body. Since the handling of real objects is not impaired, ideomotor apraxia, by itself, causes no limitation of daily living activities.

Ideational apraxia refers to a deficit in the execution of a goal-directed sequence of movements in patients who have no difficulty executing the individual components of the sequence. For example, when asked to pick up a pen and write, the sequence of uncapping the pen, placing the cap at the opposite end, turning the point towards the writing surface, and writing may be disrupted, and the patient may be seen trying to write with the wrong end of the pen or even with the removed cap. These motor sequencing problems are usually seen in the context of confusional states and dementias rather than focal lesions associated with aphasic conditions. *Limb-kinetic apraxia* involves a clumsiness in the actual use of tools that cannot be attributed to sensory, pyramidal, extrapyramidal, or cerebellar dysfunction. This condition can

emerge in the context of focal premotor cortex lesions or *corticobasal ganglionic degeneration*.

Gerstmann's Syndrome The combination of *acalculia* (impairment of simple arithmetic), *dysgraphia* (impaired writing), *finger anomia* (an inability to name individual fingers such as the index or thumb), and *right-left confusion* (an inability to tell whether a hand, foot, or arm of the patient or examiner is on the right or left side of the body) is known as Gerstmann's syndrome. In making this diagnosis it is important to establish that the finger and left-right naming deficits are not part of a more generalized anomia and that the patient is not otherwise aphasic. When Gerstmann's syndrome is seen in isolation, it is commonly associated with damage to the inferior parietal lobule (especially the angular gyrus) in the left hemisphere.

Aprosodia Variations of melodic stress and intonation influence the meaning and impact of spoken language. For example, the two statements "He *is* clever." and "He is clever?" contain an identical word choice and syntax but convey vastly different messages because of differences in the intonation and stress with which the statements are uttered. This aspect of language is known as *prosody*. Damage to perisylvian areas in the right hemisphere can interfere with speech prosody and can lead to syndromes of aprosodia. Ross has pointed out that damage to right hemisphere regions corresponding to Wernicke's area yields a greater impairment in the decoding of speech prosody, whereas damage to right hemisphere regions corresponding to Broca's area yields a greater impairment in the ability to introduce meaning-appropriate prosody into spoken language. The latter deficit is the most common type of aprosodia identified in clinical practice -- the patient produces grammatically correct language with accurate word choice but the statements are uttered in a monotone that interferes with the ability to convey the intended stress and affect. Patients with this type of aprosodia give the mistaken impression of being depressed or indifferent.

Subcortical Aphasias Damage to subcortical components of the language network (e.g., the striatum and thalamus of the left hemisphere) can also lead to aphasia. The resulting syndromes contain combinations of deficits in the various aspects of language but rarely fit the specific patterns described in Table 25-1. An anomic aphasia accompanied by dysarthria or a fluent aphasia with hemiparesis should raise the suspicion of a subcortical lesion site.

THE PARIETOFRONTAL NETWORK FOR SPATIAL ORIENTATION: NEGLECT AND RELATED CONDITIONS

Hemispatial Neglect Adaptive orientation to significant events within the extrapersonal space is subserved by a large-scale network containing three major cortical components. The *cingulate cortex* provides access to a limbic-motivational mapping of the extrapersonal space, the *posterior parietal cortex* to a sensorimotor representation of salient extrapersonal events, and the *frontal eye fields* to motor strategies for attentional behaviors (Fig. 25-2). Subcortical components of this network include the striatum and the thalamus. Contralesional hemispatial neglect represents one outcome of damage to any of the cortical or subcortical components of this network. *The traditional view that hemispatial neglect always denotes a parietal lobe lesion is inaccurate*. In keeping with this anatomic organization, the clinical manifestations of

neglect display three behavioral components: sensory events (or their mental representations) within the neglected hemispace have a lesser impact on overall awareness; there is a paucity of exploratory and orienting acts directed toward the neglected hemispace; and the patient behaves as if the neglected hemispace was motivationally devalued.

According to one model of spatial cognition, the right hemisphere directs attention within the *entire* extrapersonal space, whereas the left hemisphere directs attention mostly within the contralateral right hemispace. Consequently, unilateral left hemisphere lesions do not give rise to much contralesional neglect since the ipsilateral attentional mechanisms of the right hemisphere can compensate for the loss of the *contralaterally* directed attentional functions of the left hemisphere. Unilateral right hemisphere lesions, however, give rise to severe contralesional left hemispatial neglect because the unaffected left hemisphere does not contain ipsilateral attentional mechanisms. This model is consistent with clinical experience, which shows that contralesional neglect is more common, severe, and lasting after damage to the right hemisphere than after damage to the left hemisphere. Severe neglect for the right hemispace is rare, even in left handers with left hemisphere lesions.

Patients with severe neglect may fail to dress, shave, or groom the left side of the body; may fail to eat food placed on the left side of the tray; and may fail to read the left half of sentences. When the examiner draws a large circle [12 to 16 cm (5 to 6 in.) in diameter] and asks the patient to place the numbers 1 to 12 as if the circle represented the face of a clock, there is a tendency to crowd the numbers on the right side and leave the left side empty. When asked to copy a simple line drawing, the patient fails to copy detail on the left; and when asked to write, there is a tendency to leave an unusually wide margin on the left.

Two bedside tests that are useful in assessing neglect are *simultaneous bilateral stimulation* and *visual target cancellation*. In the former, the examiner provides either unilateral or simultaneous bilateral stimulation in the visual, auditory, and tactile modalities. Following right hemisphere injury, patients who have no difficulty detecting unilateral stimuli on either side experience the bilaterally presented stimulus as coming only from the right. This phenomenon is known as *extinction* and is a manifestation of the sensory-representational aspect of hemispatial neglect. In the target detection task, targets (e.g., A's) are interspersed with foils (e.g., other letters of the alphabet) on a 21.5´ 28.0 cm (8.5 ´11 in.) sheet of paper and the patient is asked to circle all the targets. A failure to detect targets on the left is a manifestation of the exploratory deficit in hemispatial neglect. Hemianopia, by itself, does not interfere with performance in this task since the patient is free to turn the head and eyes to the left. The normal tendency in target detection tasks is to start from the left upper quadrant and move systematically in horizontal or vertical sweeps. Some patients show a tendency to start the process from the right and proceed in a haphazard fashion. This represents a subtle manifestation of left neglect, even if the patient eventually manages to detect all the appropriate targets. Some patients with neglect may also deny the existence of hemiparesis and may even deny ownership of the paralyzed limb, a condition known as *anosognosia*.

Cerebrovascular lesions and neoplasms in the right hemisphere are the most common

causes of hemispatial neglect. Depending on the site of the lesion, the patient with neglect may also have hemiparesis, hemihypesthesia, and hemianopia on the left, but these are not invariant findings. The majority of patients display considerable improvement of hemispatial neglect, usually within the first several weeks.

Balint's Syndrome, Simultanagnosia, Dressing Apraxia, and Construction Apraxia

Bilateral involvement of the network for spatial attention, especially its parietal components, leads to a state of severe spatial disorientation known as *Balint's syndrome*. Balint's syndrome involves deficits in the orderly visuomotor scanning of the environment (*oculomotor apraxia*) and in accurate manual reaching toward visual targets (*optic ataxia*). The third and most dramatic component of Balint's syndrome is known as *simultanagnosia* and reflects an inability to integrate visual information in the center of gaze with more peripheral information. The patient gets stuck on the detail that falls in the center of gaze without attempting to scan the visual environment for additional information. The patient with simultanagnosia "misses the forest for the trees." Complex visual scenes cannot be grasped in their entirety, leading to severe limitations in the visual identification of objects and scenes. For example, a patient who is shown a table lamp and asked to name the object may look at its circular base and call it an ash tray. Some patients with simultanagnosia report that objects they look at may suddenly vanish, probably indicating an inability to look back at the original point of gaze after brief saccadic displacements. Movement and distracting stimuli greatly exacerbate the difficulties of visual perception. Simultanagnosia can sometimes occur without the other two components of Balint's syndrome.

A modification of the letter cancellation task described above can be used for the bedside diagnosis of simultanagnosia. In this modification, some of the targets (e.g., A's) are made to be much larger than the others [7.5 to 10 cm vs. 2.5 cm (3 to 4 in. vs. 1 in.) in height], and all targets are embedded among foils. Patients with simultanagnosia display a counterintuitive but characteristic tendency to miss the larger targets (Fig. 25-3). This occurs because the information needed for the identification of the larger targets cannot be confined to the immediate line of gaze and requires the integration of visual information across a more extensive field of view. The greater difficulty in the detection of the larger targets also indicates that poor acuity is not responsible for the impairment of visual function and that the problem is central rather than peripheral. Balint's syndrome results from bilateral dorsal parietal lesions; common settings include watershed infarction between the middle and posterior cerebral artery territories, hypoglycemia, sagittal sinus thrombosis, or atypical forms of Alzheimer's disease. In patients with Balint's syndrome due to stroke, bilateral visual field defects (usually inferior quadrantanopias) are common.

Another manifestation of bilateral (or right sided) dorsal parietal lobe lesions is *dressing apraxia*. The patient with this condition is unable to align the body axis with the axis of the garment and can be seen struggling as he or she holds a coat from its bottom or extends his or her arm into a fold of the garment rather than into its sleeve. Lesions that involve the posterior parietal cortex also lead to severe difficulties in copying simple line drawings. This is known as a *construction apraxia* and is much more severe if the lesion is in the right hemisphere. In some patients with right hemisphere lesions, the drawing difficulties are confined to the left side of the figure and represent a manifestation of hemispatial neglect; in others, there is a more universal deficit in reproducing contours

and three-dimensional perspective. Dressing apraxia and construction apraxia represent special instances of a more general disturbance in spatial orientation.

THE OCCIPITOTEMPORAL NETWORK FOR FACE AND OBJECT RECOGNITION: PROSOPAGNOSIA AND OBJECT AGNOSIA

Perceptual information about faces and objects is initially encoded in primary (striate) visual cortex and adjacent (upstream) peristriate visual association areas. This information is subsequently relayed first to the downstream visual association areas of occipitotemporal cortex and then to other heteromodal and paralimbic areas of the cerebral cortex. Bilateral lesions in the fusiform and lingual gyri of occipitotemporal cortex disrupt this process and interfere with the ability of otherwise-intact perceptual information to activate the distributed multimodal associations that lead to the recognition of faces and objects. The resultant face and object recognition deficits are known as *prosopagnosia* and *visual object agnosia*.

The patient with prosopagnosia cannot recognize familiar faces, including, sometimes, the reflection of his or her own face in the mirror. This is not a perceptual deficit since prosopagnosic patients can easily tell if two faces are identical or not. Furthermore, a prosopagnosic patient who cannot recognize a familiar face by visual inspection alone can use auditory cues to reach appropriate recognition if allowed to listen to the person's voice. The deficit in prosopagnosia is therefore modality-specific and reflects the existence of a lesion that prevents the activation of otherwise intact multimodal templates by relevant visual input. Damasio has pointed out that the deficit in prosopagnosia is not limited to the recognition of faces but that it can also extend to the recognition of individual members of larger generic object groups. For example, prosopagnosic patients characteristically have no difficulty with the generic identification of a face as a face or of a car as a car, but they cannot recognize the identity of an individual face or the make of an individual car. This reflects a visual recognition deficit for proprietary features that characterize individual members of an object class. When recognition problems become more generalized and extend to the generic identification of common objects, the condition is known as visual object agnosia. In contrast to prosopagnosic patients, those with object agnosia cannot recognize a face as a face or a car as a car. It is important to distinguish visual object agnosia from anomia. The patient with anomia cannot name the object but can describe its use. In contrast, the patient with visual agnosia is unable either to name a visually presented object or to describe its use. The characteristic lesions in prosopagnosia and visual object agnosia consist of bilateral infarctions in the territory of the posterior cerebral arteries. Associated deficits can include visual field defects (especially superior quadrantanopias) or a centrally based color blindness known as achromatopsia. Rarely, the responsible lesion is unilateral. In such cases, prosopagnosia is associated with lesions in the right hemisphere and object agnosia with lesions in the left.

THE LIMBIC NETWORK FOR MEMORY: AMNESIAS

Limbic and paralimbic areas (such as the hippocampus, amygdala, and entorhinal cortex), the anterior and medial nuclei of the thalamus, the medial and basal parts of the striatum, and the hypothalamus collectively constitute a distributed network known as the *limbic system*. The behavioral affiliations of this network include the coordination of

emotion, motivation, autonomic tone, and endocrine function. An additional area of specialization for the limbic network, and the one which is of most relevance to clinical practice, is that of declarative (conscious) memory for recent episodes and experiences. A disturbance in this function is known as *amnestic state*. In the absence of deficits in motivation, attention, language, or visuospatial function, the clinical diagnosis of a persistent global amnestic state is always associated with bilateral damage to the limbic network, usually within the hippocampo-entorhinal complex or the thalamus.

The memory disturbance in the amnestic state is multimodal and includes retrograde and anterograde components. The *retrograde amnesia* involves an inability to recall experiences that occurred before the onset of the amnestic state. Relatively recent events are more vulnerable to retrograde amnesia than more remote events. A patient who comes to the emergency room complaining that he cannot remember his identity but who can remember the events of the previous day is almost certainly not suffering from a neurologic cause of memory disturbance. The second and most important component of the amnestic state is the *anterograde amnesia*, which indicates an inability to store, retain, and recall new knowledge. Patients with amnestic states cannot remember what they ate a few minutes ago or the details of an important event they may have experienced a few hours ago. In the acute stages, there may also be a tendency to fill in memory gaps with inaccurate, fabricated, and often implausible information. This is known as *confabulation*. Patients with the amnestic syndrome forget that they forget and tend to deny the existence of a memory problem when questioned.

The patient with an amnestic state is almost always disoriented, especially to time. Accurate temporal orientation and accurate knowledge of current news rule out a major amnestic state. Memory can be tested with a list of four to five words read aloud by the examiner up to five times or until the patient can immediately repeat the entire list without intervening delay. In the next phase of testing, the patient is allowed to concentrate on the words and to rehearse them internally for 1 min before being asked to recall them. Accurate performance in this phase indicates that the patient is motivated and sufficiently attentive to hold the words on-line for at least 1 min. The final phase of the testing involves a retention period of 5 to 10 min, during which the patient is engaged in other tasks. Adequate recall at the end of this interval requires off-line storage, retention, and retrieval. Amnestic patients fail this phase of the task and may even forget that they were given a list of words to remember. Accurate recognition of the words by multiple choice in a patient who cannot recall them indicates a less severe memory disturbance that affects mostly the retrieval stage of memory.

Many neurologic diseases can give rise to an amnestic state. These include tumors (of the sphenoid wing, posterior corpus callosum, thalamus, or medial temporal lobe), infarctions (in the territories of the anterior or posterior cerebral arteries), head trauma, herpes simplex encephalitis, Wernicke-Korsakoff encephalopathy, paraneoplastic limbic encephalitis, and degenerative dementias such as Alzheimer's or Pick's disease. The one common denominator of all these diseases is that they lead to the bilateral lesions within one or more components in the limbic network, most commonly the hippocampus, entorhinal cortex, the mammillary bodies of the hypothalamus, and the limbic thalamus. Occasionally, unilateral left-sided lesions can give rise to an amnestic state, but the memory disorder tends to be transient. Depending on the nature and distribution of the underlying neurologic disease, the patient may also have visual field deficits, eye

movement limitations, or cerebellar findings. In many patients, such as those with *transient global amnesia* (Chap. 26), there are no associated neurologic findings; this sometimes leads incorrectly to the diagnosis of a psychiatric disorder.

Although the limbic network is the site of damage for amnestic states, it is almost certainly not the storage site for memories. Memories are stored in widely distributed form throughout the association cortex. The role attributed to the limbic network is to bind these distributed fragments into coherent events and experiences that can sustain conscious recall. Damage to the limbic network does not necessarily destroy memories but interferes with their conscious (declarative) recall in coherent form. The individual fragments of information remain preserved despite the limbic lesions and can sustain what is known as *implicit memory*. For example, patients with amnestic states can acquire new motor or perceptual skills, even though they may have no conscious knowledge of the experiences that led to the acquisition of these skills.

THE PREFRONTAL NETWORK FOR ATTENTION AND COMPORTMENT

Approximately one-third of all the cerebral cortex in the human brain is located in the frontal lobes. The frontal lobes can be subdivided into motor-premotor, dorsolateral prefrontal, medial prefrontal, and orbitofrontal components. The terms *frontal lobe syndrome* and *prefrontal cortex* refer only to the last three of these four components. These are the parts of the cerebral cortex that show the greatest phylogenetic expansion in primates and especially in humans. The dorsolateral prefrontal, medial prefrontal, and orbitofrontal areas, and the subcortical structures with which they are interconnected (i.e., the head of the caudate and the dorsomedial nucleus of the thalamus), collectively make up a large-scale network that coordinates exceedingly complex aspects of human cognition and comportment.

The prefrontal network plays an important role in behaviors that require an integration of thought with emotion and motivation. There is no simple formula for summarizing the diverse functional affiliations of the prefrontal network. Its integrity appears important for the simultaneous awareness of context, options, consequences, relevance, and emotional impact so as to allow the formulation of adaptive inferences, decisions, and actions. Damage to this part of the brain impairs mental flexibility, reasoning, hypothesis formation, abstract thinking, foresight, judgment, the on-line (attentive) holding of information, and the ability to inhibit inappropriate responses. Behaviors impaired by prefrontal cortex lesions, especially those related to the manipulation of mental content, are often referred to as "executive functions."

Even very large bilateral prefrontal lesions may leave all sensory, motor, and basic cognitive functions intact while leading to isolated but dramatic alterations of personality and comportment. The most common clinical manifestations of damage to the prefrontal network take the form of two relatively distinct syndromes. In the *frontal abulic syndrome*, the patient shows a loss of initiative, creativity, and curiosity and displays a pervasive emotional blandness and apathy. In the *frontal disinhibition syndrome*, the patient becomes socially disinhibited and shows severe impairments of judgment, insight, and foresight. The dissociation between intact intellectual function and a total lack of even rudimentary common sense is striking. Despite the preservation of all essential memory functions, the patient cannot learn from experience and continues to

display inappropriate behaviors without appearing to feel emotional pain, guilt, or regret when such behaviors repeatedly lead to disastrous consequences. The impairments may emerge only in real-life situations when behavior is under minimal external control and may not be apparent within the structured environment of the medical office. Testing judgment by asking patients what they would do if they detected a fire in a theater or found a stamped and addressed envelope on the road is not very informative since patients who answer these questions wisely in the office may still act very foolishly in the more complex real-life setting. The physician must therefore be prepared to make a diagnosis of frontal lobe disease on the basis of historic information alone even when the office examination of mental state may be quite intact.

The abulic syndrome tends to be associated with damage to the dorsolateral prefrontal cortex, and the disinhibition syndrome with the medial prefrontal or orbitofrontal cortex. These syndromes tend to arise almost exclusively after bilateral lesions, most frequently in the setting of head trauma, stroke, ruptured aneurysms, hydrocephalus, tumors (including metastases, glioblastoma, and falx or olfactory groove meningiomas), or focal degenerative diseases. Unilateral lesions confined to the prefrontal cortex may remain silent until the pathology spreads to the other side. The emergence of developmentally primitive reflexes such as grasping, rooting, and sucking are seen primarily in patients with large structural lesions that extend into the premotor components of the frontal lobes or in the context of metabolic encephalopathies. The vast majority of patients with prefrontal lesions and frontal lobe behavioral syndromes do not display these reflexes.

Damage to the frontal lobe disrupts a variety of attention-related functions including working memory (the transient on-line holding of information), concentration span, verbal fluency, the scanning and retrieval of stored information, the inhibition of immediate but inappropriate responses, and mental flexibility. The capacity for focusing on a trend of thought and the ability to voluntarily shift the focus of attention from one thought or stimulus to another can become impaired. Digit span (which should be seven forward and five reverse) is decreased; the recitation of the months of the year in reverse order (which should take less than 15 s) is slowed; and the number of words starting with a, f, or s that can be generated in 1 min (normally 12 or more per letter) is diminished even in nonaphasic patients. Characteristically, there is a progressive slowing of performance as the task proceeds; e.g., the patient asked to count backwards by 3s may say "100, 97, 94...91,...88," etc., and may not complete the task. In go-no go tasks (where the instruction is to raise the finger upon hearing one tap but to keep it still upon hearing two taps), the patient shows a characteristic inability to keep still in response to the "no go" stimulus; mental flexibility (tested by the ability to shift from one criterion to another in sorting or matching tasks) is impoverished; distractibility by irrelevant stimuli is increased; and there is a pronounced tendency for impersistence and perseveration.

These attentional deficits disrupt the orderly registration and retrieval of new information and lead to*secondary* memory deficits. Such memory deficits can be differentiated from the *primary* memory impairments of the amnestic state by showing that they improve when the attentional load of the task is decreased. Working memory (also known as immediate memory) is an attentional function based on the temporary on-line holding of information. It is closely associated with the integrity of the prefrontal network and the ascending reticular activating system. Retentive memory, on the other hand, depends

on the stable (off-line) storage of information and is associated with the integrity of the limbic network. The distinction of the underlying neural mechanisms is illustrated by the observation that severely amnestic patients who cannot remember events that occurred a few minutes ago may have intact if not superior working memory capacity as shown in tests of digit span.

Lesions in the caudate nucleus or in the dorsomedial nucleus of the thalamus (subcortical components of the prefrontal network) can also produce a frontal lobe syndrome. This is one reason why the mental state changes associated with degenerative basal ganglia diseases, such as Parkinson's or Huntington's disease, may take the form of a frontal lobe syndrome. Because of its widespread connections with other regions of association cortex, one essential computational role of the prefrontal network is to function as an integrator, or "orchestrator," for other networks. Bilateral multifocal lesions of the cerebral hemispheres, none of which are individually large enough to cause specific cognitive deficits such as aphasia or neglect, can collectively interfere with the connectivity and integrating function of prefrontal cortex. A frontal lobe syndrome is the single most common behavioral profile associated with a variety of bilateral multifocal brain diseases including metabolic encephalopathy, multiple sclerosis, vitamin B_{12}deficiency, and others. In fact, the vast majority of patients with the clinical diagnosis of a frontal lobe syndrome tend to have lesions that do not involve prefrontal cortex but involve either the subcortical components of the prefrontal network or its connections with other parts of the brain. In order to avoid making a diagnosis of "frontal lobe syndrome" in a patient with no evidence of frontal cortex disease, it is advisable to use the diagnostic term *frontal network syndrome*, with the understanding that the responsible lesions can lie anywhere within this distributed network.

The patient with frontal lobe disease raises potential dilemmas in differential diagnosis: the abulia and blandness may be misinterpreted as depression, and the disinhibition as mania or acting-out. Appropriate intervention may be delayed while a treatable tumor keeps expanding. An informed approach to frontal lobe disease and its comportmental manifestations may help to avoid such errors.

CARING FOR THE PATIENT WITH DEFICITS OF HIGHER CEREBRAL FUNCTION

Some of the deficits described in this chapter are so complex that they may bewilder not only the patient and family but also the physician. It is imperative to carry out a systematic clinical evaluation in order to characterize the nature of the deficits and explain them in lay terms to the patient and family. Such an explanation can allay at least some of the anxieties, address the mistaken impression that the deficit (e.g., social disinhibition or inability to recognize family members) is psychologically motivated, and lead to practical suggestions for daily living activities. The consultation of a skilled neuropsychologist may aid in the formulation of diagnosis and management. Patients with simultanagnosia, for example, may benefit from the counterintuitive instruction to stand back when they cannot find an item so that a greater search area falls within the immediate field of gaze. In some patients, the history may be more important than the bedside examination. For example, patients with frontal lobe disease can be extremely irritable and abusive to spouses and yet display all the appropriate social graces during the visit to the medical office.

Reactive depression is common in patients with higher cerebral dysfunction and should be treated. These patients may be sensitive to the usual doses of antidepressants or anxiolytics and deserve a careful titration of dosage. Brain damage may cause a dissociation between feeling states and their expression, so that a patient who may superficially appear jocular could still be suffering from an underlying depression that deserves to be treated. In many cases, agitation may be controlled with reassurance. In other cases, treatment with benzodiazepines or sedating antidepressants may become necessary. The use of neuroleptics for the control of agitation should be reserved for refractory cases since extrapyramidal side effects are frequent in patients with coexisting brain damage.

Spontaneous improvement of cognitive deficits due to acute neurologic lesions is common. It is most rapid in the first few weeks but may continue for up to 2 years, especially in young individuals with single brain lesions. The mechanisms for this recovery are incompletely understood. Some of the initial deficits appear to arise from remote dysfunction (diaschisis) in parts of the brain that are interconnected with the site of initial injury. Improvement in these patients may reflect, at least in part, a normalization of the remote dysfunction. Other mechanisms may involve functional reorganization in surviving neurons adjacent to the injury or the compensatory use of homologous structures, e.g., the right superior temporal gyrus with recovery from Wernicke's aphasia. In some patients with large lesions involving Broca's and Wernicke's areas, only Wernicke's area may show contralateral compensatory reorganization (or bilateral functionality), giving rise to a situation where a lesion that should have caused a global aphasia becomes associated with a residual Broca aphasia. Prognosis for recovery from aphasia is best when Wernicke's area is spared. Cognitive rehabilitation procedures have been used in the treatment of higher cortical deficits. There are few controlled studies, but some do show a benefit of rehabilitation in the recovery from hemispatial neglect and aphasia. Some types of deficits may be more prone to recovery than others. For example, patients with nonfluent aphasias are more likely to benefit from speech therapy than patients with fluent aphasias and comprehension deficits. In general, lesions that lead to a denial of illness (e.g., anosognosia) are associated with cognitive deficits that are more resistant to rehabilitation. The recovery of higher cortical dysfunction is rarely complete. Periodic neuropsychological assessment is necessary for quantifying the pace of the improvement and for generating specific recommendations for cognitive rehabilitation, modifications in the home environment, and the timetable for returning to school or work.

(Bibliography omitted in Palm version)

Back to Table of Contents

26. MEMORY LOSS AND DEMENTIA - *Thomas D. Bird*

DEFINITION

Dementia is a serious and common problem that affects more than 4 million Americans and costs society more than $50 billion annually. Ten percent of persons over age 70 and 20 to 40% of individuals over age 85 have clinically identifiable memory loss. Dementia is a syndrome with many causes. A simple definition of dementia is a deterioration in cognitive abilities that impairs the previously successful performance of activities of daily living. Memory is the most common and most important cognitive ability that is lost. Other mental faculties may also be affected such as attention, judgment, comprehension, orientation, learning, calculation, problem solving, mood, and behavior. Agitation or withdrawal, hallucinations, delusions, insomnia, and loss of inhibitions are also common. Individuals with mental retardation and psychosis may become demented if a decline in intellectual function occurs. Many common forms of dementia are progressive, but some dementing illnesses are static and unchanging. Dementia is a chronic condition, whereas delirium is an acute confusional state associated with a change in level of consciousness (ranging from lethargy to agitation).

Memory is a complex function of the brain that has fascinated philosophers and scientists for centuries. Memory is currently viewed as a mental process that uses several storage buffers of differing capacity and duration (Table 26-1). Sensory memory lasts for about 250 ms in the visual mode (iconic memory) and 1 to 2 s in the auditory mode (echoic memory). Immediate (short-term or primary) memory has a duration of about half a minute and a limited capacity of approximately 5 to 10 items. Immediate memory is highly vulnerable to distraction, requiring attention and vigilance to maintain the content. It is often tested at the bedside by asking the patient to recall several digits forward and backward. Recent, or secondary, memory has been called both "short-term" and "long-term." It has a duration of minutes to weeks and exhibits a larger storage capacity than immediate memory. On entering this buffer, information undergoes a process of consolidation of variable duration. Recent memory is commonly tested in the clinical setting by asking a patient to recall three words after 3 to 5 min. Remote, or long-term, memory stores information lasting weeks to a lifetime and contains most of our personal experiences and knowledge. Some information appears to be stored accurately for an indefinite time, whereas other items fade or become distorted. Memory function includes registration (encoding or acquisition), retention (storage or consolidation), stabilization, and retrieval (decoding or recall). Registration and retrieval are conscious processes. Animal experiments have shown that long-term memory requires new protein synthesis, and the stabilization process probably involves physical changes at neuronal synapses.

Several additional classifications of memory are sometimes used by psychologists, particularly in reference to the content or use of the memory stores. Reference memory refers to a filing system that contains recent and remote information gained from previous experience. Working memory refers to an active process that is being updated continually by current experience. Episodic memory contains information about events occurring in a specific place and time. Semantic memory contains unchanging facts, principles, associations, and rules (for example, state capitals and the number of days in a week). Declarative (explicit) memory refers to facts about the world and past personal

events that must be consciously retrieved to be remembered. Procedural (implicit) memory, in contrast, is involved in learning and retaining a skill or procedure such as how to ride a bicycle, get dressed, or drive a car. Abilities stored in procedural memory become automatic and do not require conscious implementation.

Finally, the term *executive function* refers to mental activity involved in planning, initiating, and regulating behavior. It is considered the central organizing function of the brain that results in systematic, goal-directed activity. Executive functions are active in nonroutine situations where reflex or automatic behavior is not adequate. The anatomic and physiologic substrates of executive function are presumed to involve the frontal lobes (Chap. 25). Deficits in executive function occur frequently in patients with dementia.

FUNCTIONAL ANATOMY AND PATHOGENESIS

Dementia results from disorders of cerebral neuronal circuits and is a result of the total quantity of neuronal loss combined with the specific location of such loss (Chap. 25). The anatomic basis of memory was initially clarified from study of the alcohol/thiamine deficiency syndrome of Korsakoff and the consequences of temporal lobe surgery performed for the treatment of epilepsy. In Korsakoff's syndrome lesions in the hypothalamus, mammillary bodies, and dorsomedial nuclei of the thalamus showed that these areas were important for learning, recall, and recognition. Unilateral temporal lobe surgery for epilepsy produced mild to moderate amnesia for either verbal or nonverbal material. Bilateral medial temporal lobe excision involving the hippocampal formation, the parahippocampal gyrus, and part of the amygdala produced a severe anterograde learning disorder, i.e., an inability to store new memories, often with retained ability to recall old ones. The components of the medial temporal lobe memory system include the hippocampus and adjacent cortex, including the entorhinal, perirhinal, and parahippocampal regions (Fig. 26-1). This includes a circular pathway of neurons from the entorhinal cortex to the dentate gyrus, CA3 and CA1 neurons of the hippocampus to the subiculum, and back to the entorhinal cortex; this pathway is heavily damaged in Alzheimer's disease (AD). This system is fast, has limited capacity, and performs a crucial function at the time of learning and establishing declarative memory. Its role continues after learning during a lengthy period of reorganization and consolidation whereby memory stored in neocortex eventually becomes independent of the medial temporal lobe memory system. This process, by which the burden of long-term (permanent) memory storage is gradually assumed by neocortex, assures that the medial temporal lobe system is always available for the acquisition of new information. Recent functional brain imaging studies indicate that learning and memory involve many of the same regions of the cortex that process sensory information and control motor output. The forms of perceptual and motor learning that can occur without conscious recollections are mediated in part by contractions and expansions of representations in the sensory and motor cortex. One study, for example, has shown that the cortical representation of the fingers of the left hand of musical string players is larger than that in control individuals, suggesting that the representation of different parts of the body in the primary somatosensory cortex of humans depends on use and changes to conform to the current needs and experiences of the individual. Discrete cortical regions exist in which object knowledge (such as words related to color, animals, tools, or action) is organized as a distributed system in which the attributes of an object are stored close to

the regions of the cortex that mediate perception of those attributes (Chap. 25). That is, brain regions active during object identification are partly dependent on the intrinsic properties of the object. Procedural (implicit) memory appears to involve centers outside the hippocampus such as amygdala, cerebellum, and sensory cortex. Different frontal regions are activated for different kinds of memory storage. Functional magnetic resonance imaging (MRI) studies show that the magnitude of focal activation in left prefrontal-temporal regions or right prefrontal-bilateral parahippocampal regions predicts how well verbal or visual stimuli, respectively, will be remembered.

Biochemically, the cholinergic system plays an important role in memory. Anticholinergic agents such as atropine and scopolamine interfere with memory. Choline acetyl transferase (the enzyme catalyzing the formation of acetylcholine) and nicotinic cholinergic receptors are known to be deficient in the cortex of patients withAD. The brains of patients with AD show severe neuronal loss in the nucleus basalis of Meynert, a major source of cholinergic input to the cerebral cortex. These findings form the basis for the use of cholinesterase inhibitors in the treatment of AD, with benefit presumably arising from increased available levels of acetylcholine. Behavior and mood are modulated by noradrenergic, serotonergic, and dopaminergic pathways; and norepinephrine has been shown to be reduced in the brainstem locus coeruleus in patients with AD. Neurotrophins are also postulated to play a role in memory in part by preserving cholinergic neurons.

Long-term potentiation (LTP), which refers to a long-lasting enhancement of synaptic transmission resulting from repetitive stimulation of excitatory synapses, is presumed to be involved in memory acquisition and storage. LTP occurs in the hippocampus and is mediated by *N*-methyl-D-aspartate (NMDA) receptors as well as cyclic AMP-responsive element binding protein (CREB). Gene knockout mouse models have been useful in the definition of secondary messenger systems that play a role in hippocampal LTP. For example, disruption of either calcium/calmodulin-dependent protein kinase or cytoplasmic tyrosine kinase (fyn) results in deficient hippocampal LTP and impaired spatial learning. In contrast, mice in which a neuronal glycoprotein thy-1 has been inactivated show regionally selective impairment of LTP but intact spatial learning, suggesting that LTP in the entorhinal projection to the dentate gyrus of the hippocampus (Fig. 26-1) may not be necessary for some forms of spatial learning. Disruption of hippocampal levels of CREB impairs long-term memory in rats.

Most diseases causing dementia do not have highly restricted regions of pathology. Disorders such asADappear to eventually represent relatively diffuse neuronal deterioration throughout the cerebral cortex, whereas multi-infarct dementia associated with recurrent strokes causes more focal damage in a random patchwork of cortical regions. Diffuse white matter damage may disrupt intracerebral connections and cause dementia syndromes such as those associated with leukodystrophies, multiple sclerosis, and Binswanger's disease. Subcortical structures such as the caudate, putamen, thalamus, and substantia nigra also modulate cognition and behavior in ways that are not yet well understood. Some investigators distinguish between cortical and subcortical types of dementia. A cortical dementia such as AD primarily presents as memory loss and is often associated with aphasia or other disturbance of language. Patients with subcortical dementia such as Huntington's disease (HD) are less likely to have memory and language problems and more likely to have difficulties with attention,

judgment, awareness, and behavior. Both the clinical and anatomic characteristics of the cortical and subcortical dementias show considerable overlap, and the conditions are often not distinct.

Lesions of some relatively specific cortical-subcortical pathways may have significant effects on behavior (Chap. 25). The dorsolateral prefrontal cortex has connections with the dorsolateral caudate, globus pallidus, and thalamus. Lesions of these pathways result in poor organization and planning, perseveration, and decreased cognitive flexibility with impaired judgment. The lateral orbital frontal cortex connects with the ventromedial caudate, globus pallidus, and thalamus. Lesions of these connections cause irritability, impulsiveness, and distractibility. The anterior cingulate cortex connects with the nucleus accumbens, globus pallidus, and thalamus. Interruption of these connections produces apathy and poverty of speech or even akinetic mutism.

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade over age 50 and is associated most often with the microscopic changes ofAD at autopsy. Slow accumulation of mutations in neuronal mitochondria is also hypothesized to contribute to the increasing prevalence of dementia with age. Yet many centenarians have intact memory function and no evidence of clinically significant dementia. Whether dementia is an inevitable consequence of normal human aging remains controversial.

DIFFERENTIAL DIAGNOSIS

The many causes of dementia are listed in Table 26-2. The frequency of each condition depends on the age group under study, the country of origin, and perhaps racial or ethnic variations.AD is the most common cause of dementia in western countries, affecting more than half of demented patients. Vascular disease is the second most common cause of dementia in the United States, affecting 10 to 20%; but it is more common than AD in some Asian countries. Dementia associated with chronic alcoholism and Parkinson's disease (PD) represent the next two most common categories. Chronic intoxications including those resulting from prescription drugs are an important, potentially treatable cause of dementia. Other disorders listed in the table are uncommon but important because many are reversible. The classification of dementing illnesses into two broad groups of reversible and irreversible disorders is a useful approach to the differential diagnosis of dementia.

Subtle cumulative memory loss is a natural part of aging. This frustrating experience, often the source of jokes and humor, is referred to as *benign forgetfulness of the elderly*. Benign means that it is not so progressive or serious that it impairs reasonably successful and productive daily functioning, although the distinction between benign and more significant memory loss can be difficult to make. A proportion of persons with benign memory loss progress to frank dementia, usually caused byAD. It remains unclear why some individuals show progression and others do not. It was once assumed that a cumulative loss of hippocampal neurons with normal aging might underlie this forgetfulness, but recent quantitative neuronal counts indicate that this "natural" neuronal loss may not occur.

Alzheimer's disease is a slowly progressive dementing illness associated with diffuse

cortical atrophy and specific neuropathologic hallmarks of amyloid plaques and neurofibrillary tangles. Although quite common in the elderly, it remains a diagnosis of exclusion to be confirmed definitively only at autopsy. The clinical diagnosis of ADestablished by experienced neurologists proves to be correct at autopsy approximately 85 to 90% of the time.*This condition is described in greater detail in Chap. 362.

Two major types of vascular dementia can be identified (Chap. 362). The first, often called multi-infarct dementia, results from an accumulation of discrete cerebral strokes that produce disabling deficits of memory, behavior, and other cognitive abilities. Such patients usually give a history of sudden, separate stroke episodes with stepwise deterioration. On examination, focal neurologic deficits such as hemiparesis, unilateral Babinski reflex, aphasia, or visual field defect are common. Brain imaging shows multiple areas of stroke, which may have been ischemic or hemorrhagic. A second, more subtle and insidious type of vascular dementia, Binswanger's disease, is a dementing illness associated with diffuse, subcortical white matter damage often occurring in patients with chronic hypertension and/or severe atherosclerosis. The white matter changes are dramatically visualized byMRI and have also been called leukoareosis. The pathogenesis of Binswanger's disease is unknown. BecauseAD and vascular dementia are common, occasional patients may have both conditions.

An inherited form of vascular dementia is CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy). It is caused by a mutation in the notch 3 gene and produces characteristic dense bodies in the media of arterioles in brain and skin. Affected persons have diffuse white matter deficits on brainMRIassociated with migraine and recurrent stroke without hypertension.

Frontotemporal dementia (FTD) may represent 10 to 20% of persons with presenile dementia (onset before age 65). Initial symptoms are behavioral, such as disinhibition, apathy, or agitation with relatively intact memory. Brain imaging studies show focal lobar atrophy of the frontal and/or temporal lobes. Pick's disease is a form of FTD. Some cases are familial and associated with neuronal neurofibrillary tangle formation and mutations in the tau gene.

Dementia commonly accompanies chronic alcoholism (Chap. 387). This situation may be a result of associated malnutrition, especially of B vitamins and particularly thiamine. However, other as yet poorly defined aspects of chronic alcohol ingestion may also produce cerebral damage and atrophy. A rare idiopathic syndrome of dementia and seizures with degeneration of the corpus callosum has been reported primarily in male Italian drinkers of red wine (Marchiafava-Bignami disease).

Thiamine (vitamin B_1) deficiency causes Wernicke's encephalopathy. The clinical presentation is a malnourished individual (frequently but not necessarily alcoholic) with confusion, ataxia, and diplopia from ophthalmoplegia (Charcot's triad). Thiamine deficiency damages the thalamus, mammillary bodies, midline cerebellum, periaquaductal gray matter of the midbrain, and peripheral nerves. Damage to medial thalamic regions correlates most closely with memory loss. Prompt administration of parenteral thiamine (100 mg intravenously for 3 days followed by daily oral dosage) may reverse the disease if given in the first few days of symptom onset. However, prolonged

untreated thiamine deficiency can result in an irreversible dementia/amnestic syndrome (Korsakoff's psychosis) or even death.

In Korsakoff's syndrome, the patient is unable to recall new information despite normal immediate memory, attention span, and level of consciousness. Memory for new events is seriously impaired, whereas memory of knowledge prior to the illness is relatively intact. Patients are easily confused, disoriented, and incapable of recalling new information for more than a brief interval. Superficially, they may be conversant, entertaining, able to perform simple tasks, and follow immediate commands. Confabulation is common, although not always present, and may result in obviously erroneous statements and elaborations. There is no specific treatment because the previous thiamine deficiency has produced irreversible damage to the medial thalamic nuclei and mammillary bodies. Mammillary body atrophy may be visible on high-resolutionMRI.

Vitamin B_{12}deficiency, as can occur in pernicious anemia, causes a macrocytic anemia and may also damage the nervous system (Chaps. 107 and368). Neurologically it most commonly produces a spinal cord syndrome (myelopathy) affecting the posterior columns (loss of position and vibratory sense) and the lateral corticospinal tracts (hyperactive tendon reflexes and Babinski responses); it also damages peripheral nerves, resulting in sensory loss with depressed tendon reflexes. Damage to cerebral myelinated fibers may also cause dementia. The mechanism of neurologic damage is unclear but may be related to a deficiency of *S*-adenosylmethionine (required for methylation of myelin phospholipids) due to reduced methionine synthase activity or accumulation of methylmalonate and propionate, providing abnormal substrates for fatty acid synthesis in myelin. The neurologic signs of vitamin B_{12}deficiency are usually associated with macrocytic anemia, but on occasion may occur in its absence. Treatment with parenteral vitamin B_{12}(1000 ug intramuscularly daily for a week, weekly for a month, and monthly for life for pernicious anemia) stops progression of the disease if instituted promptly, but reversal of advanced nervous system damage will not occur.

Deficiency of nicotinic acid (pellagra) is associated with sun-exposed skin rash, glossitis, and angular stomatitis (Chap. 75). Severe dietary deficiency of nicotinic acid along with other B vitamins such as pyridoxine may result in spastic paraparesis, peripheral neuropathy, fatigue, irritability, and dementia. This syndrome has been seen in prisoner-of-war and concentration camps. Low serum folate levels appear to be a rough index of malnutrition, but isolated folate deficiency has not been proven to be a specific cause of dementia.

Approximately 20% of patients withPD(Chap. 363) eventually develop dementia. Treatment with L-dopa neither accelerates nor prevents this process. Some PD patients with dementia have cytoplasmic neuronal inclusions (Lewy bodies) orADchanges in the cerebral cortex, but others have no specific identifiable cortical pathology. Progressive supranuclear palsy is a dementing illness associated with parkinsonian features of rigidity, bradykinesia, and postural instability. Resting tremor is often absent, there is a vertical gaze palsy, and patients are resistant to treatment with L-dopa.

Infections of the central nervous system (CNS) usually cause delirium and other acute neurologic syndromes (Chap. 24). However, some chronic CNS infections such as

tuberculosis or cryptococcosis may produce a dementing illness (Chap. 374). Between 20 and 30% of patients in the advanced stages of infection with HIV become demented (Chap. 309). Cardinal features include psychomotor retardation, apathy, and impaired memory. This condition may result from secondary opportunistic infections but can also be caused by direct involvement of CNS neurons with HIV, where there is a multinucleated giant cell encephalitis and diffuse pallor of white matter. The neuronal toxicity may be mediated by cytokines or the direct neurotoxic effect of the gp 120 envelope glycoprotein. In the absence of CNS opportunistic infection, elevated b_2-microglobulin in cerebrospinal fluid (CSF) is a useful marker for HIV dementia. Herpes simplex encephalitis (Chap. 373) has a predilection for the inferior temporal lobes and may present as subacute confusion and disorientation but more often as an acute syndrome rather than a chronic dementia. Computed tomography (CT), MRI, and electroencephalogram (EEG) may all demonstrate the temporal lobe location of the lesions. CSF usually shows increased protein and a lymphocytic pleocytosis. CNS syphilis (Chap. 172) was a common cause of dementia in the preantibiotic era; it is uncommon now but can still be encountered in individuals with multiple sex partners. Characteristic CSF changes consist of pleocytosis, increased protein, and a positive Venereal Disease Research Laboratory (VDRL) test.

Prion disorders such as Creutzfeldt-Jakob disease (CJD) (Chap. 375) are rare conditions (approximately 1 per million population) that commonly produce dementia. CJD is typically a rapidly progressive disease associated with dementia, rigidity, and myoclonus, causing death in less than 1 to 2 years. These clinical characteristics may also rarely be seen in AD, and the differential diagnosis usually depends on the slower progression of AD and the markedly abnormal periodic EEG discharges seen in CJD. Ataxia or cortical blindness may also accompany CJD. The transmissible agent, or prion, consists principally of an abnormal isoform of a host-encoded protein, the prion protein, which has undergone a physical conformational change and accumulates in affected brains. Bovine spongiform encephalopathy in the United Kingdom is thought to have resulted from cattle feed containing sheep tissues contaminated with infectious prions. Immunoassay for a 14-3-3 brain protein in CSF may be a useful marker for transmissible spongiform encephalopathies in patients with dementia.

Primary and metastatic neoplasms of the CNS (Chap. 370) usually produce focal neurologic findings and seizures rather than dementia. However, if tumor growth begins in the frontal or temporal lobes, the initial manifestations may be memory loss or behavioral changes. A rare paraneoplastic syndrome of dementia associated with occult carcinoma (usually small cell lung cancer) has been termed *limbic encephalitis* (Chap. 101). In this syndrome, confusion, agitation, seizures, poor memory, and frank dementia may occur in association with sensory neuropathy. The CSF often shows an increase in cells and protein. There is neuronal loss and perivascular lymphocytic infiltration in the hippocampus, amygdala, and cingulate and frontal cortex. Circulating antineuronal nuclear antibodies may be present. There is no specific treatment.

The syndrome of normal-pressure hydrocephalus (Chap. 362) is frequently discussed but difficult to diagnose. Clinically, a triad of memory loss, gait disturbance, and bladder incontinence is typical. The gait abnormality is often the initial symptom, and the dementia is usually mild. On imaging studies, the lateral ventricles are enlarged but there is minimal or no cortical atrophy. Lumbar puncture shows a normal or slightly

elevated opening pressure with normalCSF. The condition may be idiopathic or the result of previous meningitis or subarachnoid blood from a ruptured aneurysm or head trauma. The pathogenetic mechanism is presumably a block of normal CSF flow over the convexity and delayed absorption into the venous system, with resulting stretch and distortion of white matter tracts within the corona radiata. Some individuals improve with ventricular shunting but many do not. The condition is difficult to distinguish fromAD(Chap. 362).

A nonconvulsive seizure disorder may underlie a syndrome of confusion, clouding of consciousness, and garbled speech. Psychiatric disease is often suspected, but an EEGdemonstrates the seizure discharges. If recurrent or persistent, the condition may be termed *complex partial status epilepticus*. The cognitive disturbance often responds to anticonvulsant therapy. The etiology may be previous small strokes or head trauma; some cases are idiopathic.

It is important to recognize systemic diseases that indirectly affect the brain and produce chronic confusion or dementia. Such conditions include dysthyroid states (especially hypothyroidism), vasculitis, and hepatic, renal, or pulmonary disease. Hepatic encephalopathy may begin with irritability and confusion and slowly progress to agitation, lethargy, and coma (Chap. 376).

Isolated angiitis of theCNS (CNS granulomatous angiitis) (Chaps. 317 and361) occasionally causes a chronic encephalopathy associated with confusion, disorientation, and clouding of consciousness. Headache is common, and strokes and cranial neuropathies may occur. Brain imaging studies may be normal or nonspecifically abnormal. Studies ofCSFreveal a mild pleocytosis or elevation in the protein level in half of the cases. Cerebral angiography often shows multifocal stenosis and narrowing of vessels. A few patients have only small-vessel disease that is not revealed on angiography. The angiographic appearance is not specific and may be mimicked by atherosclerosis, infection, or other causes of vascular disease. Brain or meningeal biopsy demonstrates abnormal arteries with endothelial cell proliferation and infiltrates of mononuclear cells. Autoantibodies and immune complexes are not present, and a cell-mediated process appears most likely. The prognosis is poor, but some patients respond to glucocorticoids or chemotherapy.

Chronic metal intoxications may also produce a dementing syndrome. The key to diagnosis is the elicitation of a history of exposure at work, home, or even as a consequence of a medical procedure such as dialysis. Lead poisoning has highly variable neurologic manifestations. Fatigue, depression, and confusion may be associated with episodic abdominal pain and peripheral neuropathy. Gray lead lines may appear in the gums. There is usually an associated anemia with basophilic stippling of red cells. The clinical presentation can resemble that of acute intermittent porphyria, including elevated levels of urine porphyrins as a result of the inhibition of d-aminolevulinic acid dehydratase. Chronic lead poisoning from inadequately fired glazed pottery has been reported. The treatment is chelation therapy with agents such as ethylene diaminetetraacetic acid (EDTA). Chronic mercury poisoning may produce dementia, peripheral neuropathy, ataxia, and a fine tremulousness that may progress to a cerebellar intention tremor or choreoathetosis. The confusion and memory loss of chronic arsenic intoxication is also associated with nausea, weight loss, peripheral

neuropathy, pigmentation and scaling of the skin, and transverse white lines of the fingernails (Mee's lines). Treatment is chelation therapy with dimercaprol (BAL). Aluminum poisoning has been best documented with the dialysis dementia syndrome in which water used during renal dialysis was contaminated with excessive amounts of aluminum. This resulted in a progressive encephalopathy associated with confusion, memory loss, agitation, and, later, lethargy and stupor. Speech arrest and myoclonic jerking was common and associated with severe and generalizedEEGchanges. The condition was often fatal. There were no specific pathologic findings, but elevated brain aluminum content was documented. The condition has been eliminated by use of deionized water for dialysis. Although aluminum injected into experimental animals may produce neurofibrillary tangles, patients with dialysis dementia had neither tangles nor amyloid plaques, and there has been no direct association of aluminum poisoning withAD.

Recurrent head trauma in professional boxers may lead to dementia, sometimes called the "punch drunk" syndrome or *dementia pugilistica*. The symptoms can be progressive and may begin late in a boxer's career or even long after retirement. The severity of the syndrome correlates with the length of the boxing career and the total number of bouts. Early in the condition there occurs a personality change associated with social instability and sometimes paranoia and delusions. Later, memory loss progresses to full dementia, often associated with parkinsonian signs and ataxia or intention tremor. At autopsy, the cerebral cortex may show changes similar toAD, although neurofibrillary tangles are usually more predominant than amyloid plaques (which are usually diffuse rather than neuritic). There may also be loss of neurons in the substantia nigra. Chronic subdural hematoma is also occasionally associated with dementia, often in the context of underlying cortical atrophy from conditions such as AD orHD. In these latter cases, evacuation of the subdural hematoma does not alter the underlying degenerative process.

Head injury (Chap. 369) may also be associated with temporary amnesia. The memory disturbance may include events that occurred both before the injury (retrograde amnesia) and during the postinjury period (posttraumatic or anterograde amnesia). Retrograde amnesia after severe head injury may extend back for hours or weeks before the injury; remote memory is usually intact. As patients recover, the extent of retrograde amnesia shrinks and may disappear. Often, retrograde amnesia causes permanent inability to recall the few minutes before the head injury, implying disruption of the immediate memory system and failure to register long-term memory. The length of posttraumatic amnesia generally corresponds to the length of the postconcussive confusional state, but posttraumatic amnesia may persist even in the presence of normal immediate memory and digit span. The duration of posttraumatic amnesia indicates the severity of head injury; the ability to learn new material is often the last cognitive deficit to recover. There are reports of recovery from retrograde amnesia occurring months or years after the initial brain insult; the recovery is sometimes stimulated by hypnosis, amobarbital interview, or electrical stimulation. One theory of such recovery envisions a resetting of distorted patterns of neuronal matrices subserving memory.

Transient global amnesia (TGA) is characterized by sudden onset of complete anterograde loss of memory and learning abilities, usually occurring in persons over age

50. Onset of memory loss may occur in the context of an emotional stimulus or physical exertion. During the attack the individual is alert and communicative, general cognition seems intact, and there are no other neurologic signs or symptoms. The patient may seem confused and repeatedly ask about present events. The ability to form new memories returns after a period of hours, and the individual returns to normal but has no recall for the period of the attack. Frequently no cause can be determined, but cerebrovascular disease, epilepsy (7% in one study), migraine, or cardiac arrhythmia sometimes may be implicated. A Mayo Clinic review of 277 patients with TGA found a past history of migraine in 14% and cerebrovascular disease in 11%, but these conditions were not temporally related to the TGA episodes. About one-fourth of the patients had recurrent attacks, but they were not at increased risk for subsequent stroke. Rare instances of permanent memory loss after sudden onset have been reported.

Psychogenic amnesia for personally important memories is common, although whether this amnesia results from deliberate avoidance of unpleasant memories or from unconscious repression may be impossible to establish. The event-specific amnesia is particularly common after violent crimes such as homicide of a close relative or friend or sexual abuse. It also may occur with severe drug or alcohol intoxication and sometimes with schizophrenia. More prolonged psychogenic amnesia occurs in fugue states that also commonly follow severe emotional stress. The patient with a fugue state suffers from a sudden loss of personal identity and may be found wandering far from home. In contrast to organic amnesia, fugue states are associated with amnesia for personal identity and events closely associated with the personal past. At the same time, memory for other recent events and the ability to learn and use new information are preserved. The episodes usually last hours or days and occasionally weeks or months while the patient takes on a new identity. On recovery, there is a residual amnesic gap for the period of the fugue.

Psychiatric diseases may mimic dementia. Severely depressed individuals may appear demented, a phenomenon called *pseudodementia*. Unlike cortical dementias, memory and language are usually intact when carefully tested in depressed persons. The patients may feel confused and are unable to accomplish routine tasks. Vegetative symptoms are common, such as insomnia, lack of energy, poor appetite, and concern with bowel function. The psychosocial milieu may suggest prominent reasons for depression. The patients respond to antidepressant treatment. Schizophrenia is usually not difficult to distinguish from dementia, but occasionally the distinction can be problematic. (Kraepelin's original term for schizophrenia was *dementia praecox*.) Schizophrenia usually has a much earlier age of onset (second and third decades) than most dementing illnesses. It is associated with intact memory, and the delusions and hallucinations of schizophrenia are usually more complex and bizarre than those of dementia. Some individuals with chronic schizophrenia develop an unexplained progressive dementia late in life that is not related toAD. Memory loss may also be part of a conversion reaction. In this situation, patients commonly complain bitterly of memory loss, but careful cognitive testing either does not confirm the deficits or demonstrates inconsistent or unusual patterns of cognitive problems. The patients' behavior and "wrong" answers to questions often indicate that they both understand the question and know the answer.

Clouding of cognition by chronic drug or medication use, often prescribed by physicians, is an important cause of dementia. Sedatives, tranquilizers, and analgesics used to treat insomnia, pain, anxiety, or agitation may cause confusion, memory loss, and lethargy, especially in the elderly. Discontinuation of such medication often improves mentation.

Approach to the Patient

The approach to the patient with dementia should always keep two major questions in the forefront: What is the most accurate diagnosis, and is there a treatable or reversible condition? A broad overview of this approach is shown inTable 26-3.

History The history should concentrate on the onset, duration, and tempo of the memory loss. Acute or subacute confusion may represent delirium and suggests intoxication, infection, or metabolic derangement. An elderly person with slowly progressive memory loss over several years is likely to haveAD. Initial symptoms often are difficulty with managing money, driving, shopping, following instructions, or finding one's way around town. A change in personality with disinhibition and intact memory may suggestFTD. A history of sudden stroke with an irregular stepwise progression suggests multi-infarct dementia. Stroke is also commonly associated with a history of hypertension, atrial fibrillation, peripheral vascular disease, and diabetes. Rapid progression with rigidity and myoclonus suggestsCJD. Seizures may indicate stroke or neoplasm. Trouble in walking may suggestPD or normal-pressure hydrocephalus, especially the latter when associated with bladder incontinence. A history of multiple sex partners or intravenous drug use may indicateCNSinfection, especially with HIV. A history of recurrent head trauma could indicate chronic subdural hematoma, dementia pugilistica, or normal-pressure hydrocephalus. Alcoholism may suggest malnutrition and thiamine deficiency. A remote history of gastric surgery resulting in loss of intrinsic factor might indicate vitamin B_{12}deficiency. Certain occupations such as working in a battery or chemical factory might indicate heavy metal intoxication. Careful review of medication intake, especially of sedatives and tranquilizers, may raise the issue of chronic drug intoxication. A positive family history of dementia would be elicited inHD, familial AD, and inherited FTD. The recent death of a loved one, insomnia, or poor appetite suggest depression.

Physical Examination A careful examination is essential to document the dementia, look for other signs of nervous system involvement, and search for clues suggesting other systemic disease. Cognitive function should be assessed in terms of orientation, recent and remote memory, and calculation. Many of the simple, commonly used bedside tests of cognitive function (such as serial 7s, digits forward and backward) are most useful when they are performed normally; this makes the diagnosis of dementia unlikely. Mistakes on these simple tests are more difficult to interpret and are of less diagnostic importance. Drawing a clock and the trail-making test are frequently used tests of immediate memory and visual-spatial abilities. The mini-mental status exam (MMSE) is an easily administered 30-points test of cognitive function (Table 26-4). It is used to quickly indicate a dementing process, provide a rough assessment of its severity, and follow progression of the illness. The MMSE is influenced by culture and education and is less useful in the early and late stages of dementia. Language function should be tested by the ability to read, write, comprehend, and name objects. Resting tremor, cogwheel rigidity, bradykinesia, and festinating gait indicate a parkinsonian

syndrome. Gait ataxia or apraxia (inability to initiate and coordinate steps in a sequential fashion) suggests normal-pressure hydrocephalus. Confusion, sixth cranial nerve paresis, and ataxia suggests thiamine deficiency. Myoclonic jerks are present in CJD but also occur in AD. Hemiparesis or other focal neurologic deficits may occur in multi-infarct dementia or brain tumor. Bilateral hyperactive tendon reflexes, Babinski responses, and loss of vibration and position sensation suggest a myelopathy, such as occurs in vitamin B_{12} deficiency. Stocking-glove sensory loss and diminished tendon reflexes suggest a peripheral neuropathy, which could indicate underlying diabetes, vitamin deficiency, or heavy metal intoxication. Dry cool skin, hair loss, and bradycardia suggest hypothyroidism. Confusion associated with repetitive stereotyped movements may indicate ongoing seizure activity. Hearing impairment or visual loss may produce confusion and disorientation misinterpreted as dementia. Such sensory deficits are common in the elderly.

Laboratory Tests The use of multiple laboratory tests in the evaluation of dementia is controversial. The physician does not want to miss a treatable cause, yet no single treatable cause stands out as common; thus a screen must employ multiple different tests, each of which has a low yield. Therefore, cost/benefit ratios are difficult to assess, and many laboratory screening algorithms for dementia discourage multiple tests. Nevertheless, even a test with only a 1 to 2% positive rate is probably worth undertaking if the alternative is missing a reversible or treatable cause of dementia. Table 26-3 lists most screening tests for dementia. Neuroimaging studies (CT and MRI) are especially controversial because of their cost. However, they are clearly of value to identify primary and secondary neoplasms, locate areas of infarction, or suggest normal-pressure hydrocephalus or diffuse white matter disease. They also lend support to the diagnosis of AD, especially if there is hippocampal atrophy in addition to diffuse cortical atrophy, and focal lobar atrophy may suggest FTD. However, attempts to relate cognition to neuroimaging measures of atrophy and white matter changes have shown only modest correlations. A diagnosis of AD is reached primarily by exclusion of other causes of dementia. (The indications for apolipoprotein E testing for AD are discussed in Chap. 362.) Serum levels of vitamin B_{12} and TSH, complete blood count, electrolyte measurements, and a VDRL test are reasonable routine screening measures because they detect treatable conditions. Lumbar puncture need not be done routinely in the evaluation of dementia but is indicated if CNS infection is a serious consideration, for example, in patients with delirium, fever, or nuchal rigidity. CSF levels of tau protein are increased and those of Abamyloid are decreased in some patients with AD; however, the clinical usefulness of these changes is not yet clear. Formal psychometric testing is not necessary in every patient with dementia but can be used to document the severity of dementia, suggest psychogenic causes, and provide a semiquantitative method for following the disease course. EEG is rarely helpful except to suggest CJD (repetitive bursts of diffuse high voltage sharp waves) or an underlying nonconvulsive seizure disorder (epileptiform discharges). Brain biopsy (including meninges) is not commonly advised except to diagnose vasculitis, potentially treatable neoplasms, unusual infections (such as sarcoid), or in young persons where the diagnosis is in doubt. Angiography is not likely to be of use except when multiple strokes or cerebral vasculitis is a possible cause of the dementia.

TREATMENT

The two major goals of management are, first, to treat any correctable cause of the dementia and, second, to provide comfort and support to the patient and caregivers. Treatment of underlying causes might include thyroid replacement for hypothyroidism; vitamin therapy for thiamine and B_{12} deficiency; antibiotics for opportunistic infections; ventricular shunting for normal-pressure hydrocephalus; and appropriate surgical, radiation, and/or chemotherapy for CNS neoplasms. Removal of sedating or cognition-impairing drugs and medications is often beneficial. If the patient is depressed rather than demented (pseudodementia), the depression should be vigorously treated. Patients with degenerative diseases such as AD and HD may also be depressed, and that portion of their condition may respond to antidepressant therapy. Antidepressants should be used with caution in demented patients because they may produce delirium. Antidepressants that have a low incidence of cognitive side effects, such as selective serotonin reuptake inhibitors, and tricyclic antidepressants with low anticholinergic activity such as desipramine and nortriptyline, are advisable. Anticonvulsants are used to control seizures. Agitation, hallucinations, delusions, and confusion are difficult to treat. These behavioral problems represent major causes for nursing home placement and institutionalization. Drugs such as phenothiazines, resperidone, haloperidol, and benzodiazepines may ameliorate the behavior problems but have untoward side effects such as sedation, rigidity, and dyskinesias. Medications that may calm agitation and insomnia without worsening dementia include low-dose haloperidol (0.5 to 2 mg), trazodone, buspirone, and propranolol. Olanzapine is increasingly used for patients with hallucinations. When patients do not respond, it is usually a mistake to advance to higher doses or to use anticholinergics or sedatives (such as barbiturates or benzodiazepines).

Cholinesterase inhibitors are being used to treat AD, and other drugs, such as estrogen, anti-inflammatory agents, and vitamin E are being investigated for the treatment or prevention of AD. These approaches are reviewed in Chap. 362.

A proactive approach has been shown to reduce the occurrence of delirium in hospitalized patients. This scheme includes frequent orientation, cognitive activities, sleep enhancement measures, vision and hearing aids, and correction of dehydration.

Nondrug behavior therapy has an important place in the management of dementia. The primary goal is to make the life of the patient with dementia comfortable, uncomplicated, and safe. Preparing lists, schedules, calendars, and labels can be helpful. It is also useful to stress familiar routines, short-term tasks, brief walks, and simple physical exercises. For many patients with dementia, the memory for facts is worse than that for routine activities, and they still may be able to take part in remembered physical activities such as walking, bowling, dancing, and golf. Patients with dementia usually object to losing control over familiar tasks such as driving, cooking, and handling finances. Attempts to help or take over may be greeted with complaints, depression, or anger. Hostile responses on the part of the caretaker are useless and sometimes harmful. Explanation, reassurance, distraction, and calm statements are more productive responses in this setting. Eventually, tasks such as finances and driving must be assumed by others, and the patient will conform and adjust. Safety is an important issue that includes not only driving but the environment of the kitchen, bathroom, and sleeping area. These areas need to be monitored, supervised, and made as safe as possible. A move to a retirement home, assisted-living center, or nursing home can

initially increase confusion and agitation. Repeated reassurance, reorientation, and careful introduction to the new personnel will help to smooth the process. Provision of activities that are known to be enjoyable to the patient can be of considerable benefit. Attention should also be paid to frustration and depression in family members and caregivers. Caregiver guilt and burn-out are common. Family members often feel overwhelmed and helpless and may vent their frustrations on the patient, each other, and healthcare providers. Caregivers should be encouraged to take advantage of day-care facilities and respite breaks. Education and counseling about dementia are important. Local and national support groups can be of considerable help, such as the Alzheimer's Disease and Related Disorders Association.

(Bibliography omitted in Palm version)

Back to Table of Contents

27. SLEEP DISORDERS - *Charles A. Czeisler*, *John W. Winkelman*, *Gary S. Richardson*

Disturbed sleep is among the most frequent health complaints physicians encounter. More than one-half of adults in the United States experience at least intermittent sleep disturbances. For most, it is an occasional night of poor sleep and/or daytime sleepiness. However, at least 15 to 20% of adults report chronic sleep disturbance or misalignment of circadian timing, which can lead to serious impairment of daytime functioning. In addition, such problems may contribute to or exacerbate medical or psychiatric conditions. Thirty years ago, many such complaints were treated with hypnotic medications without further diagnostic evaluation. Since then, a distinct class of sleep and arousal disorders has been identified, and the field of sleep disorders medicine is now an established clinical discipline. However, most physicians still only receive, on average, 1 h of education in sleep disorders in their medical school curriculum.

PHYSIOLOGY OF SLEEP AND WAKEFULNESS

Most adults sleep 7 to 8 h per night, although the timing, duration, and internal structure of sleep vary among healthy individuals and as a function of age. At the extremes, infants and the elderly have frequent interruptions of sleep. In the United States, adults of intermediate age tend to have one consolidated sleep episode per day, although in some cultures sleep may be divided into a midafternoon nap and a shortened night sleep. Two principal neurobiologic systems govern the sleep-wake cycle: one that actively generates sleep and sleep-related processes and another that times sleep within the 24-h day. Either intrinsic abnormalities in these systems or extrinsic disturbances (environmental, drug- or illness-related) can lead to sleep or circadian rhythm disorders.

STATES AND STAGES OF SLEEP

States and stages of human sleep are defined on the basis of characteristic patterns in the electroencephalogram (EEG), the electrooculogram (EOG -- a measure of eye-movement activity), and the surface electromyogram (EMG) measured on the chin and neck. The continuous recording of this array of electrophysiologic parameters to define sleep and wakefulness is termed *polysomnography*.

Polysomnographic profiles define two states of sleep: (1) rapid-eye-movement (REM) sleep, and (2) non-rapid-eye-movement (NREM) sleep. NREM sleep is in turn subdivided into four stages, characterized by increasing arousal threshold and slowing of the corticalEEG. REM sleep is characterized by a low-amplitude, mixed-frequency EEG similar to that of NREM stage 1 sleep. TheEOGshows bursts of REM similar to those seen during eyes-open wakefulness. ChinEMGactivity is absent, reflecting the brainstem-mediated muscle atonia that is characteristic of that state.

ORGANIZATION OF HUMAN SLEEP

Normal nocturnal sleep in adults displays a consistent organization from night to night (Fig. 27-1). After sleep onset, sleep usually progresses throughNREMstages 1 to 4

within 45 to 60 min. Slow-wave sleep predominates in the first third of the night and comprises 15 to 25% of total nocturnal sleep time in young adults. The percentage of slow-wave sleep is influenced by several factors, most notably age (see below). Prior sleep deprivation increases the rapidity of sleep onset and both the intensity and amount of slow-wave sleep.

The firstREMsleep episode usually occurs in the second hour of sleep. More rapid onset of REM sleep in a young adult (particularly if less than 30 min) may suggest pathology such as endogenous depression, narcolepsy, circadian rhythm disorders, or drug withdrawal.NREM and REM alternate through the night with an average period of 90 to 110 min (the "ultradian" sleep cycle). Overall, REM sleep constitutes 20 to 25% of total sleep, and NREM stages 1 and 2 are 50 to 60% (increasing in elderly subjects).

Age has a profound impact on sleep state organization (Fig. 27-1). Slow-wave sleep is most intense and prominent during childhood, decreasing sharply at puberty and across the second and third decades of life. After age 30, there is a progressive, almost linear decline in the amount of slow-wave sleep, and the amplitude of deltaEEGactivity comprising slow-wave sleep is reduced. In the otherwise healthy older person, slow-wave sleep may be completely absent, particularly in males.

A different age profile exists forREMsleep. In infancy, REM sleep may comprise 50% of total sleep time, and the percentage is inversely proportional to developmental age. The amount of REM sleep falls off sharply over the first postnatal year as a mature REM-NREMcycle develops. During the rest of life into extreme old age, REM sleep occupies a relatively constant percentage of total sleep time.

NEUROANATOMY OF SLEEP

Lesion studies in animals and neurologic diseases in humans have suggested distinct neuroanatomic sites in the generation of normal sleep and wakefulness. Experimental studies in animals have variously implicated the medullary reticular formation, the thalamus, and the basal forebrain in the generation of sleep, while the brainstem reticular formation, the midbrain, the subthalamus, the thalamus, and the basal forebrain have all been suggested to play a role in the generation of wakefulness orEEGarousal (Chap. 24).

Current hypotheses suggest that the capacity for sleep and wakefulness generation is distributed along an axial "core" of neurons extending from the brainstem rostrally to the basal forebrain. Complex commingling of neuronal groups occurs at many points along this brainstem-forebrain axis. It was recently discovered that a cluster of g-aminobutyric acid (GABA) and galaninergic ventrolateral preoptic (VLPO) neurons, which innervate monoaminergic cell groups in the tuberomammilary nucleus that contribute to the ascending arousal system, are activated during sleep. This has led to the hypothesis that these hypothalamic VLPO neurons may play a key role in sleep regulation.

Moreover, the neuroanatomic correlates ofREMsleep appear to be discretely localized. Specific regions in the pons are associated with the neurophysiologic correlates of REM sleep. Small lesions in the dorsal pons result in the loss of the descending muscle inhibition normally associated with REM sleep; microinjections of the cholinergic agonist

carbachol into the pontine reticular formation appear to produce a state with all of the features of REM sleep. These experimental manipulations are mimicked by pathologic conditions in humans and animals. In narcolepsy, for example, abrupt, complete, or partial paralysis (cataplexy) occurs in response to a variety of stimuli. In dogs with this condition, physostigmine, a central cholinesterase inhibitor, increases the frequency of cataplectic attacks, while atropine decreases their frequency. Conversely, in REM sleep behavior disorder (see below), patients suffer from incomplete motor inhibition during REM sleep, resulting in involuntary, occasionally violent movement during REM sleep.

NEUROCHEMISTRY OF SLEEP

Early experimental studies that focused on the raphe nuclei of the brainstem appeared to implicate serotonin as the primary sleep-promoting neurotransmitter, while catecholamines were considered to be responsible for wakefulness. Subsequent work has demonstrated that the raphe-serotonin system may facilitate sleep but is not necessary for its expression. Extensive pharmacologic studies of sleep and wakefulness suggest roles for other neurotransmitters as well. Cholinergic neurotransmission is known to play a role in REM sleep generation. The alerting influence of caffeine implicates adenosine, whereas the hypnotic effect of benzodiazepines and barbiturates suggests a role for endogenous ligands of the $GABA_A$ receptor complex.

A variety of sleep-promoting substances have been identified, although it is not known whether or not they are involved in the endogenous sleep-wake regulatory process. These include prostaglandin D_2, delta sleep-inducing peptide, muramyl dipeptide, interleukin 1, fatty acid primary amides, and melatonin. The hypnotic effect of these substances is commonly limited to NREM or slow-wave sleep, although peptides that increase REM sleep have also been reported. Many putative "sleep factors," including interleukin 1 and prostaglandin D_2, are immunologically active as well, suggesting a link between immune function and sleep-wake states.

PHYSIOLOGY OF CIRCADIAN RHYTHMICITY

The sleep-wake cycle is the most evident of the many 24-h rhythms in humans. Prominent daily variations also occur in endocrine, thermoregulatory, cardiac, pulmonary, renal, gastrointestinal, and neurobehavioral functions. However, in evaluating a daily variation, it is important to distinguish between those rhythmic components passively evoked by periodic environmental or behavioral changes (e.g., the increase in blood pressure and heart rate upon assumption of the upright posture) and those actively driven by an endogenous oscillatory process (e.g., the circadian variation in plasma cortisol that persists under a variety of environmental and behavioral conditions).

The suprachiasmatic nuclei (SCN) of the hypothalamus act as the central neural pacemaker driving endogenous circadian rhythms in mammals. Bilateral destruction of these nuclei results in a loss of endogenous circadian rhythmicity that can only be restored by transplantation of the same structure from a donor animal. The genetically determined period of this endogenous neural oscillator, which averages ~24.2 h in humans, is normally synchronized to the 24-h period of the environmental light-dark cycle. Entrainment of mammalian circadian rhythms by the light-dark cycle is mediated

via the retinohypothalamic tract, a monosynaptic pathway that links the retina to the SCN. Humans are exquisitely sensitive to the resetting effects of light, even at low intensity.

The timing and internal architecture of sleep are directly coupled to the output of the endogenous pacemaker. Paradoxically, the endogenous circadian rhythms of sleep tendency, sleepiness, and REMsleep propensity all peak near the habitual wake time, just after the nadir of the endogenous circadian temperature cycle, whereas the circadian wake propensity rhythm peaks 1 to 3 h before the habitual bedtime. These rhythms are thus timed to oppose the homeostatic decline of sleep tendency during the habitual sleep episode and the rise of sleep tendency throughout the usual waking day, respectively. Misalignment of the output of the endogenous circadian pacemaker with the desired sleep-wake cycle can, therefore, induce insomnia, as well as decrements of alertness and neurobehavioral performance in night-shift workers and after jet lag.

BEHAVIORAL CORRELATES OF SLEEP STATES AND STAGES

Polysomnographic staging of sleep correlates with behavioral changes during specific states and stages. During the transitional state between wakefulness and sleep (stage 1 sleep), subjects may respond to faint auditory or visual signals without "awakening." Furthermore, memory incorporation is inhibited at the onset ofNREMstage 1 sleep, and individuals aroused from that transitional sleep stage frequently deny having been asleep. Such transitions occur spontaneously after chronic partial sleep deprivation (e.g., 4 to 6 h of sleep per night) and acute total sleep deprivation (e.g., 24 h of wakefulness), notwithstanding attempts to remain continuously awake (see "Shift-Work Sleep Disorder," below).

Awakenings fromREMsleep are associated with recall of vivid dream imagery more than 80% of the time. The reliability of dream recall increases with REM sleep episodes occurring later in the night. Imagery may also be reported afterNREMsleep interruptions, though these typically lack the detail and vividness of REM sleep dreams. The incidence of NREM sleep dream recall can be increased by selective REM sleep deprivation, suggesting that REM sleep and dreaming per se are not inexorably linked.

PHYSIOLOGIC CORRELATES OF SLEEP STATES AND STAGES

All major physiologic systems are influenced by sleep. Changes in cardiovascular function include a decrease in blood pressure and heart rate duringNREM, and particularly during slow-wave sleep. DuringREMsleep, phasic activity (bursts of eye movements) is associated with variability in both blood pressure and heart rate mediated principally by the vagus. Cardiac dysrhythmias may occur selectively during REM sleep. Respiratory function also changes (Chap. 263). In comparison to relaxed wakefulness, respiratory rate becomes more regular during NREM sleep (especially slow-wave sleep) and tonic REM sleep and becomes very irregular during phasic REM sleep. Minute ventilation decreases in NREM sleep out of proportion to the decrease in metabolic rate at sleep onset, resulting in a higher P_{CO_2}.

Endocrine function also varies with sleep. The most prominent changes are apparent in neuroendocrine parameters. Slow-wave sleep is associated with secretion of growth

hormone, while sleep in general is associated with augmented secretion of prolactin. Sleep has a complex effect on the secretion of luteinizing hormone (LH): during puberty, sleep is associated with increased LH secretion, whereas sleep in the mature woman inhibits LH secretion in the early follicular phase of the menstrual cycle. Sleep onset (and probably slow-wave sleep) is associated with inhibition of thyroid-stimulating hormone and of the adrenocorticotropic hormone-cortisol axis, an effect that is superimposed on the circadian rhythms in the two systems.

The pineal hormone melatonin is secreted predominantly at night in both day- and night-active species, reflecting the direct modulation of pineal activity by the circadian pacemaker through a circuitous neural pathway from theSCN to the pineal gland. Melatonin secretion is not dependent upon the occurrence of sleep, persisting in individuals kept awake at night. In addition, exogenous melatonin increases sleepiness and may potentiate sleep when administered to good sleepers attempting to sleep during daylight hours at a time when endogenous melatonin levels are low. However, there is little evidence to support the use of melatonin as a hypnotic in such individuals during nighttime hours, when endogenous melatonin levels are high and sleep is already consolidated. A large-scale, double-blind clinical trial is needed to evaluate the efficacy of melatonin as a sleep-promoting therapeutic for patients with insomnia.

Sleep is also associated with alterations of thermoregulatory function.NREMsleep is associated with an attenuation of thermoregulatory responses to either heat or cold stress, and animal studies of thermosensitive neurons in the hypothalamus document an NREM-sleep-dependent reduction of the thermoregulatory set-point.REMsleep is associated with complete absence of thermoregulatory responsiveness, effectively resulting in functional poikilothermy. However, the potential adverse impact of this failure of thermoregulation is blunted by inhibition of REM sleep by extreme ambient temperatures.

DISORDERS OF SLEEP AND WAKEFULNESS

Approach to the Patient

Patients may seek help from a physician because of one of several symptoms: (1) an acute or chronic inability to sleep adequately at night (insomnia); (2) chronic fatigue, sleepiness, or tiredness during the day; or (3) a behavioral manifestation associated with sleep itself. Complaints of insomnia or excessive daytime sleepiness should be viewed as symptoms (much like fever or pain) of underlying disorders. Knowledge of the differential diagnosis of these presenting complaints is essential to identify the underlying medical disorder. Only then can appropriate treatment, rather than nonspecific approaches (e.g., over-the-counter sleeping aids) be applied. Diagnoses of exclusion, such as primary insomnia, should be made only after other diagnoses have been ruled out.Table 27-1 outlines the diagnostic and therapeutic approach to the patient with a complaint of excessive daytime sleepiness.

A careful history is essential in the evaluation of the patient with a sleep complaint. In particular, the duration, severity, and consistency of the complaint are important, along with the patient's estimate -- in the case of an insomnia complaint -- of the consequences of reported sleep loss on subsequent waking function. Information from a

friend or family member can be an invaluable aid in assessing the symptoms and the severity of the complaint for daytime functioning, as some patients may be unaware of, or will underreport, such potentially embarrassing symptoms as heavy snoring or falling asleep while driving.

Completion by the patient of a day-by-day sleep-work-drug log for at least 2 weeks can help the physician better understand the nature of the complaint. Work times and sleep times (including daytime naps and nocturnal awakenings) as well as drug and alcohol use, including caffeine and hypnotics, should be noted each day. The sleep times should be plotted to facilitate recognition of circadian rhythm sleep disorders such as delayed sleep phase syndrome (see below).

Polysomnography is necessary for the diagnosis of specific disorders such as narcolepsy and sleep apnea and may be of utility in other settings as well. In addition to the three electrophysiologic variables used to define sleep states and stages, the standard clinical polysomnogram includes measures of respiration (respiratory effort, air flow, and oxygen saturation), anterior tibialisEMG, and electrocardiogram. Evaluation of penile tumescence during nocturnal sleep can also help determine whether the cause of erectile dysfunction in a patient is psychogenic or organic (Chap. 51).

INSOMNIA

Insomnia is the complaint of inadequate sleep; it can be classified according to the nature of sleep disruption and the duration of the complaint. The nature of the sleep disruption provides important information about the possible etiology of the insomnia and is also central to the selection of specific and appropriate treatment. Insomnia is subdivided into difficulty falling asleep (*sleep onset insomnia*), frequent or sustained awakenings (*sleep maintenance insomnia*), early morning awakenings (*sleep offset insomnia*), or persistent sleepiness despite sleep of adequate duration (*nonrestorative sleep*). Similarly, the duration of the symptom is an important determinant of the nature of appropriate treatment. An insomnia complaint lasting one to several nights (within a single episode) is termed *transient insomnia*. Transient insomnia is typically the result of situational stress or a change in sleep schedule or environment (e.g., jet lag). *Short-term insomnia* lasts from a few days to 3 weeks. Disruption of this duration is usually associated with more protracted stress, such as recovery from surgery or short-term illness. *Long-term insomnia*, or *chronic insomnia*, lasts for months or years and, in contrast with short-term insomnia, requires a thorough evaluation of underlying causes (see below). Chronic insomnia is often a waxing and waning disorder, with spontaneous or stressor-induced exacerbations.

While an occasional night of poor sleep, typically in the setting of stress or excitement about external events, is both common and without lasting consequences, persistent insomnia can have important adverse consequences in the form of impaired daytime function and increased risk of injury due to accidents. There is also clear evidence of increased risk of the development of major depression with insomnia of at least 1 year's duration. In addition, there is emerging evidence that individuals with chronic insomnia have increased utilization of health care resources, even after controlling for comorbid medical and psychiatric disorders.

Extrinsic Insomnia A number of sleep disorders are the result of extrinsic factors that interfere with sleep. *Transient situational insomnia* can occur after a change in the sleeping environment (e.g., in an unfamiliar hotel or hospital bed) or before or after a significant life event, such as a change of occupation, loss of a loved one, illness, or anxiety over a deadline or examination. Increased sleep latency, frequent awakenings from sleep, and early morning awakening can all occur. Recovery generally occurs rapidly, usually within a few weeks. Treatment is usually symptomatic, with intermittent use of hypnotics and resolution of the underlying stress. *Inadequate sleep hygiene* is characterized by a behavior pattern prior to sleep and/or a bedroom environment that is not conducive to sleep. Noise and/or light in the bedroom can interfere with sleep, as can a bed partner with periodic limb movements during sleep or one who snores loudly. Clocks can heighten the anxiety about the time it has taken to fall asleep. Drugs that act on the central nervous system, large meals, vigorous exercise, or hot showers just before sleep may interfere with sleep onset. Many individuals participate in stressful work-related activities in the evening, producing a state incompatible with sleep onset. In preference to hypnotic medications, patients should be counseled to avoid stressful activities before bed, develop a soporific bedtime ritual, and to prepare and reserve the bedroom environment for sleeping. Consistent, regular rising times should be maintained daily, including weekends.

Psychophysiologic Insomnia Persistent *psychophysiologic insomnia* is a behavioral disorder in which patients are preoccupied with a perceived inability to sleep adequately at night. The sleep disturbance is often triggered by an emotionally stressful event; however, the poor sleep habits and beliefs about sleep acquired during the stressful period persist long after the initial incident. Such patients become hyperaroused by their own persistent efforts to sleep and/or the sleep environment, and the insomnia is a conditioned or learned response. They may be able to fall asleep more easily at unscheduled times (when not trying) or outside the home environment. Polysomnographic recording in patients with psychophysiologic insomnia reveals an objective sleep disturbance, often with an abnormally long sleep latency; frequent nocturnal awakenings; and an increased amount of stage 1 transitional sleep. Rigorous attention should be paid to sleep hygiene and correction of counterproductive, arousing behaviors before bedtime. Behavioral therapies are the treatment modality of choice for psychophysiologic insomnia, with only intermittent use of medications. When patients are awake longer than 20 min, they should read or perform other relaxing activities to distract themselves from insomnia-related anxiety. In addition, bedtime and waketime should be scheduled to restrict time in bed to be equal to their perceived total sleep time. This will generally produce sleep deprivation, greater sleep drive, and, eventually, better sleep. Time in bed can then be gradually expanded.

Medication-, Drug-, or Alcohol-Dependent Insomnia Disturbed sleep can result from ingestion of a wide variety of agents. Caffeine is perhaps the most common pharmacologic cause of insomnia. It produces increased latency to sleep onset, more frequent arousals during sleep, and a reduction in total sleep time for up to 8 to 14 h after ingestion. As few as three to five cups of coffee can significantly disturb sleep in some patients; therefore, a 1- to 2-month trial without caffeine should be attempted in patients with these symptoms. Similarly, alcohol and nicotine can interfere with sleep, despite the fact that many patients use them to relax and promote sleep. Although alcohol can increase drowsiness and shorten sleep latency, even moderate amounts of

alcohol increase awakenings in the second half of the night. In addition, alcohol ingestion prior to sleep is contraindicated in patients with sleep apnea because of the inhibitory effects of alcohol on upper airway muscle tone. Acutely, amphetamines and cocaine suppress bothREMsleep and total sleep time, which return to normal with chronic use. Withdrawal leads to a REM sleep rebound.

A number of prescribed medications can produce insomnia. Antidepressants, sympathomimetics, and glucocorticoids are common causes. In addition, severe rebound insomnia can result from the acute withdrawal of hypnotics, especially following the use of high doses of benzodiazepines with a short half-life. For this reason, hypnotic doses should be low to moderate, the total duration of hypnotic therapy should usually be limited to 2 to 3 weeks, and prolonged drug tapering is encouraged.

Altitude Insomnia Sleep disturbance is a common consequence of exposure to high altitude. Periodic breathing of the Cheyne-Stokes type occurs duringNREMsleep about half the time at high altitude, with restoration of a regular breathing pattern duringREMsleep. Both hypoxia and hypocapnia are thought to be involved in the development of periodic breathing. Frequent awakenings and poor quality sleep characterize altitude insomnia, which is generally worst on the first few nights at high altitude but may persist. Treatment with acetazolamide can decrease time spent in periodic breathing and substantially reduce hypoxia during sleep.

Restless Legs Syndrome (RLS) Patients with this sensory-motor disorder report a creeping or crawling dysesthesia deep within the calves or feet, or sometimes even in the upper extemities, that is associated with an irresistible urge to move the affected limbs. For most patients with RLS, the dysesthesias and restlessness are much worse in the evening or night compared to the daytime and frequently interfere with the ability to fall asleep. The disorder is exacerbated by inactivity and temporarily relieved by movement. In contrast, paresthesia secondary to peripheral neuropathy persists with activity. The severity of this chronic disorder may wax and wane with time and can be exacerbated by sleep deprivation, caffeine, and pregnancy. The prevalence is thought to be 5% of adults. Roughly one-third of patients will have multiple affected family members, possibly with an autosomal dominant pattern. Iron deficiency and renal failure may actually cause RLS, which is then considered secondary RLS. The symptoms of RLS are exquisitely sensitive to dopaminergic drugs (e.g., L-dopa or dopamine agonists). Narcotics, benzodiazepines, and certain anticonvulsants may also be of therapeutic value. Most patients with restless legs also experience periodic limb movement disorder during sleep, although the reverse is not the case.

Periodic Limb Movement Disorder *Periodic limb movement disorder*, previously known as *nocturnal myoclonus*, is the principal objective polysomnographic finding in 17% of patients with insomnia and 11% of those with excessive daytime somnolence (Fig. 27-2). It is often unclear whether it is an incidental finding or the cause of disturbed sleep. Stereotyped, 0.5- to 5.0-s extensions of the great toe and dorsiflexion of the foot recur every 20 to 40 s duringNREMsleep, in episodes lasting from minutes to hours. Most such episodes occur during the first half of the night. The disorder occurs in a wide variety of sleep disorders (including narcolepsy, sleep apnea,REMsleep behavior disorder, and various forms of insomnia) and may be associated with frequent arousals and an increased number of sleep-stage transitions. The incidence increases with age:

44% of people over age 65 without a sleep complaint have >five periodic leg movements per hour of sleep. The pathophysiology is not well understood, though individuals with high spinal transections can exhibit periodic leg movements during sleep, suggesting the existence of a spinal generator. Polysomnography with bilateral surfaceEMGrecording of the anterior tibialis is used to establish the diagnosis. Treatment options include dopaminergic medications or benzodiazepines.

Insomnia Associated with Mental Disorders Approximately 80% of patients with psychiatric disorders describe sleep complaints. There is considerable heterogeneity, however, in the nature of the sleep disturbance both between conditions and among patients with the same condition.

Depression can be associated with sleep onset insomnia, sleep maintenance insomnia, and/or early morning wakefulness. However, hypersomnia occurs in some depressed patients, especially adolescents and those with either bipolar or seasonal (fall/winter) depression (Chap. 385). Indeed, sleep disturbance is an important vegetative sign of depression and may commence before any mood changes are perceived by the patient. Consistent polysomnographic findings in depression include decreasedREMsleep latency, lengthened first REM sleep episode, and shortened firstNREMsleep episode; however, these findings are not specific for depression, and the extent of these changes varies with age and symptomatology. Depressed patients also show decreased slow-wave sleep and reduced sleep continuity.

In *mania* and *hypomania*, sleep latency is increased and total sleep time can be reduced. Patients with *anxiety disorders* tend not to show the changes inREMsleep and slow-wave sleep seen in endogenously depressed patients. Finally, *chronic alcoholics* lack slow-wave sleep, have decreased amounts of REM sleep (as an acute response to alcohol), and have frequent arousals throughout the night. This is associated with impaired daytime alertness. The sleep of chronic alcoholics may remain disturbed for years after discontinuance of alcohol usage. Sleep architecture and physiology are disturbed in *schizophrenia* (with a decreased amount of stage 4 sleep and a lack of augmentation of REM sleep following REM sleep deprivation); chronic schizophrenics often show day-night reversal, sleep fragmentation, and insomnia.

Insomnia Associated with Neurologic Disorders A variety of neurologic diseases result in sleep disruption through both indirect, nonspecific mechanisms (e.g., pain in cervical spondylosis or low back pain) or by impairment of central neural structures involved in the generation and control of sleep itself.

For example, *dementia* from any cause has long been associated with disturbances in the timing of the sleep-wake cycle, often characterized by nocturnal wandering and an exacerbation of symptomatology at night (so-called sundowning).

Epilepsy may rarely present as a sleep complaint (Chap. 360). Often the history is of abnormal behavior, at times with convulsive movements, during sleep, and the differential diagnosis includesREMsleep behavior disorder, sleep apnea syndrome, and periodic movements of sleep (see above). Diagnosis requires nocturnalEEGrecording. Other neurologic diseases associated with abnormal movements, such as *Parkinson's disease*, *hemiballismus*, *Huntington's chorea*, and *Gilles de la Tourette syndrome*, are

also associated with disrupted sleep, presumably through secondary mechanisms. However, the abnormal movements themselves are greatly reduced during sleep. Headache syndromes may show sleep-associated exacerbations (*migraine* or *cluster headache*) (Chap. 15) by unknown mechanisms.

Fatal familial insomnia is a rare hereditary disorder caused by bilateral degeneration of anterior and dorsomedial nuclei of the thalamus. Insomnia is a prominent early symptom. Progressively, the syndrome produces autonomic dysfunction, dysarthria, myoclonus, coma, and death. The pathogenesis is a mutation in the prion protein (Chap. 375).

Insomnia Associated with Other Medical Disorders A number of medical conditions are associated with disruptions of sleep. The association is frequently nonspecific, e.g., that between sleep disruption and chronic pain from rheumatologic disorders. Attention to this association is important in that sleep-associated symptoms are the presenting complaint of many such patients. Treatment of the underlying medical disorder or symptom is the most useful approach to such patients. As noted above, sleep disruption can also result from the appropriate use of drugs such as glucocorticoids.

Among the most prominent associations is that between sleep disruption and *asthma*. In many asthmatics there is a prominent daily variation in airway resistance that results in marked increases in asthmatic symptoms at night, especially during sleep. In addition, treatment of asthma with theophylline-based compounds, adrenergic agonists, or glucocorticoids can independently disrupt sleep. When sleep disruption is a prominent side effect of asthma treatment, inhaled steroids (e.g., beclomethasone) that do not disrupt sleep may provide a useful alternative.

Cardiac ischemia may also be associated with sleep disruption. The ischemia itself may result from increases in sympathetic tone as a result of sleep apnea. Patients may present with complaints of nightmares or vivid, disturbing dreams, with or without awareness of the more classic symptoms of angina or of the sleep disordered breathing. Treatment of the sleep apnea may substantially improve the angina and the nocturnal sleep quality. *Paroxysmal nocturnal dyspnea* can also occur as a consequence of sleep-associated cardiac ischemia that causes pulmonary congestion exacerbated by the recumbent posture.

Chronic obstructive pulmonary disease is also associated with sleep disruption, as is *cystic fibrosis*, *menopause*, *hyperthyroidism*, *gastroesophageal reflux*, *chronic renal failure*, and *liver failure*.

EVALUATION OF DAYTIME SLEEPINESS

Daytime impairment due to sleep loss may be difficult to quantify in the clinical setting for several reasons. First, sleepiness is not necessarily proportional to subjectively assessed sleep deprivation. In obstructive sleep apnea, for example, the repeated brief interruptions of sleep associated with resumption of respiration at the end of apneic episodes result in significant waking impairment, despite the fact that the patient may be unaware of the sleep fragmentation. Second, subjective descriptions of waking impairment vary from patient to patient. Patients may describe themselves as "sleepy,"

"fatigued," or "tired" and may have a clear sense of the meaning of those terms, while others may use the same terms to describe a completely different condition. Third, sleepiness, particularly when profound, may affect judgment in a manner analogous to ethanol, such that subjective awareness of the condition and the consequent cognitive and motor impairment is reduced. Finally, patients may be reluctant to admit that sleepiness is a problem, both because they are generally unaware of what constitutes normal alertness and because sleepiness is generally viewed pejoratively, ascribed more often to a deficit in motivation than to an inadequately addressed physiologic sleep need.

In assessing sleepiness in the clinical setting, specific questioning about the occurrence of sleep episodes during normal waking hours, both intentional and unintentional, can overcome the inconsistencies among subjective characterizations and help to interpret the adverse impact of sleepiness on daytime function. Specific areas to be addressed include the occurrence of inadvertent sleep episodes while driving or in other safety-related settings, sleepiness while at work or school (and the relationship of sleepiness to work and school performance), and the effect of sleepiness on social and family life. Evidence for significant daytime impairment [in association either with the diagnosis of a primary sleep disorder, such as narcolepsy or sleep apnea, or with imposed or self-selected sleep-wake schedules (see "Shift-Work Sleep Disorder," below)] raises the question of the physician's responsibility to notify motor vehicle licensing authorities of the increased risk of sleepiness-related vehicle accidents. As with epilepsy, legal requirements vary from state to state, and existing legal precedents do not provide a consistent interpretation of the balance between the physician's responsibility and the patient's right to privacy. At a minimum, physicians should document discussions with the patient regarding the increased risk of operating a vehicle, as well as a recommendation that driving be suspended until successful treatment or schedule modification can be instituted.

The distinction between fatigue and sleepiness can be useful in the differentiation of patients with complaints of fatigue or tiredness in the setting of disorders such as fibromyalgia, chronic fatigue syndrome (Chap. 384), or endocrine deficiencies such as hypothyroidism or Addison's disease. While patients with these disorders can typically distinguish their daytime symptoms from the sleepiness that occurs with sleep deprivation, substantial overlap can occur. This is particularly true when the primary disorder also results in chronic sleep disruption (e.g., sleep apnea in hypothyroidism) or in abnormal sleep (e.g., fibromyalgia).

While clinical evaluation of the complaint of excessive sleepiness is usually adequate, objective quantification is sometimes necessary for diagnostic purposes or for the evaluation of treatment response. Assessment of daytime functioning as an index of the adequacy of sleep can be made with the multiple sleep latency test (MSLT), which involves repeated measurement of sleep latency (time to onset of sleep) under standardized conditions during a day following quantified nocturnal sleep. The average latency across four to six tests (administered every 2 h across the waking day) is taken as an objective measure of daytime sleep tendency. Disorders of sleep that result in pathologic daytime somnolence can be reliably distinguished with the MSLT. In addition, the multiple measurements of sleep onset may identify direct transitions from wakefulness toREMsleep that are suggestive of specific pathologic conditions (e.g.,

narcolepsy).

NARCOLEPSY

Narcolepsy is both a disorder of the ability to sustain wakefulness voluntarily and a disorder ofREMsleep regulation (Table 27-2). The classic "narcolepsy tetrad" consists of excessive daytime somnolence plus three specific symptoms related to an intrusion of REM sleep characteristics (e.g., muscle atonia, vivid dream imagery) into the transition between wakefulness and sleep: (1) sudden weakness or loss of muscle tone without loss of consciousness, often elicited by emotion (cataplexy); (2) hallucinations at sleep onset (hypnogogic hallucinations) or upon awakening (hypnopompic hallucinations); and (3) muscular paralysis upon awakening (sleep paralysis). The severity of cataplexy varies, as patients may have two to three attacks per day or per decade. The extent and duration of an attack may also vary, from a transient sagging of the jaw lasting a few seconds to rare cases of flaccid paralysis of the entire voluntary musculature for up to 20 to 30 min. Symptoms of narcolepsy typically begin in the second decade, although the onset ranges from ages 5 to 50. Once established, the disease is chronic without remissions. Secondary forms of narcolepsy have been described (e.g., after head trauma).

Narcolepsy affects about 1 in 4000 people in the United States and appears to have a genetic basis. Recently, two independent discoveries have revealed that hypothalamic neurons containing the neuropeptide orexin (hypocretin) may play an important role in the regulation of sleep/wakefulness: (1) a mutation in the orexin (hypocretin) receptor 2 gene has been associated with canine narcolepsy; and (2) orexin "knockout" mice that are genetically unable to produce this neuropeptide exhibit a phenotype, as assessed by behavioral and electrophysiologic criteria, that is similar to human narcolepsy. In addition, modafinil, a drug recently approved by the U.S. Food and Drug Administration (FDA) for the treatment of narcolepsy, activates orexin-containing neurons. However, the inheritance pattern of narcolepsy in humans is more complex than that of the canine model. A high rate of discordance in identical twins indicates that one or more nonheritable factors contributes to its development. First-degree relatives of narcoleptic patients nonetheless have about a 1% incidence of narcolepsy, much higher than the general population but much lower than is seen in the animal models. Of note, nearly all narcoleptics with cataplexy are positive for the human leukocyte antigen DQB*0106 (ordinarly found in 20 to 30% of the general population) (Chap. 306).

Diagnosis Definition of the essential and distinctive features of narcolepsy has continued to evolve, and the diagnostic criteria continue to be a matter of debate. Certainly, objective verification of excessive daytime somnolence, typically withMSLTmean sleep latencies <8 min, is an essential if nonspecific diagnostic feature. Other conditions that cause excessive sleepiness, such as sleep apnea or chronic sleep restriction, must be rigorously excluded. The other objective diagnostic feature of narcolepsy is the presence of REMsleep in at least two of the naps during the MSLT. This excessive REM "pressure" is also manifested by the appearance of REM sleep immediately or within minutes after sleep onset in 50% of narcoleptic patients, a rarity in unaffected individuals maintaining a conventional sleep-wake schedule. The REM-related symptoms of the classic narcolepsy tetrad are variably present. There is increasing evidence that narcoleptics with cataplexy (one-half to two-thirds of patients)

may represent a more homogeneous group than those without this symptom. However, a history of cataplexy can be difficult to establish reliably. Hypnogogic and hypnopompic hallucinations and sleep paralysis are often found in nonnarcoleptic individuals and may be present in only one-half of narcoleptics. Nocturnal sleep disruption is commonly observed in narcolepsy but is also a nonspecific symptom. Similarly, history of "automatic behavior" during wakefulness (a trancelike state during which simple motor behaviors persist) is not specific for narcolepsy and serves principally to corroborate the presence of daytime somnolence.

TREATMENT

The treatment of narcolepsy is symptomatic. Somnolence is treated with stimulants. Methylphenidate has long been considered the drug of choice by most; the usual initial dose is 10 mg bid, increasing as needed to a maximum of 20 mg qid. Pemoline, frequently used as an alternative due to its longer half-life, may be less effective and has recently been associated with fatal hepatic failure in several children. Dextroamphetamine, 10 mg bid, and methamphetamine are also frequently used alternatives. Recently, modafinil, a novel wake-promoting agent, has been approved by theFDA for treatment of the excessive daytime somnolence in narcolepsy; the dose is 200-400 mg/d given as a single dose. It is a long-acting agent that may cause fewer side effects than other medications.

Treatment of theREM-related phenomena cataplexy, hypnogogic hallucinations, and sleep paralysis requires the potent REM sleep suppression produced by antidepressant medications. The tricyclic antidepressants [e.g., protriptyline (10-40 mg/d) and clomipramine (25-50 mg/d)] and the selective serotonin reuptake inhibitors (SSRIs) [e.g., fluoxetine (10-20 mg/d)] are commonly used for this purpose in the United States. Efficacy of the antidepressants is limited largely by anticholinergic side effects (tricyclics) and by sleep disturbance and sexual dysfunction (SSRIs). Adequate nocturnal sleep time and planned daytime naps (when possible) are important preventative measures in narcolepsy.

SLEEP APNEA SYNDROMES

Respiratory dysfunction during sleep is a common, serious cause of excessive daytime somnolence as well as of disturbed nocturnal sleep. An estimated 2 to 5 million people in the United States have a reduction or cessation of breathing for 10 to 150 s, from thirty to several hundred times every night during sleep. These episodes may be due to either an occlusion of the airway (*obstructive sleep apnea*), absence of respiratory effort (*central sleep apnea*), or a combination of these factors (*mixed sleep apnea*) (Fig. 27-2). Failure to recognize and treat these conditions appropriately may lead to: significant, and often disabling, impairment of daytime alertness; increased risk of sleep-related motor vehicle accidents; hypertension and other serious cardiovascular complications; and increased mortality. Sleep apnea is particularly prevalent in overweight men and in the elderly, yet it is estimated to remain undiagnosed in 80 to 90% of affected individuals. This is unfortunate since effective treatments are available.**Readers are referred to Chap. 263 for a comprehensive review of the diagnosis and treatment of patients with these conditions.*

PARASOMNIAS

The term *parasomnia* refers to abnormal behaviors that arise from, or occur during, sleep. A continuum of parasomnias arise fromNREMsleep, from brief confusional arousals to sleepwalking and night terrors. The presenting complaint is usually related to the behavior itself, but the parasomnias can disturb sleep continuity or lead to mild impairments in daytime alertness. Only one parasomnia is known to occur inREMsleep, i.e., REM sleep behavior disorder (RBD; see below).

Sleepwalking (Somnambulism) Patients affected by this disorder carry out automatic motor activities that range from simple to complex. Individuals may leave the bed, walk, urinate inappropriately, eat, or exit from the house while remaining only partially aware. Full arousal may be difficult, and some patients may respond to attempted awakening with agitation or even violence. Sleepwalking arises from stage 3 or 4NREMsleep and is most common in children and adolescents, when these sleep stages are most robust. Episodes are usually isolated but may be recurrent in 1 to 6% of patients. The cause is unknown, though it has a familial basis in roughly one-third of cases.

Sleep Terrors This disorder, also called *pavor nocturnus*, occurs primarily in young children during the first several hours after sleep onset, in stages 3 and 4 ofNREMsleep. The child suddenly screams, exhibiting autonomic arousal with sweating, tachycardia, and hyperventilation. The individual may be difficult to arouse and rarely recalls the episode on awakening in the morning. Recurrent attacks are rare, and treatment is usually by way of reassurance of parents. Both sleep terrors and sleepwalking represent abnormalities of arousal. In contrast, *nightmares* (dream anxiety attacks) occur duringREMsleep and cause full arousal, with intact memory for the unpleasant episode.

REM Sleep Behavior Disorder RBD is a rare condition that is distinct from other parasomnias in that it occurs duringREMsleep. It primarily afflicts men of middle age or older, many of whom have a history of prior neurologic disease. In fact, over one-third of patients will go on to develop Parkinson's disease within 10 to 20 years. Presenting symptoms are of agitated or violent behavior during sleep, reported by a bed partner. In contrast to typical somnambulism, injury to patient or bed partner is not uncommon, and, upon awakening, the patient reports vivid, often unpleasant, dream imagery. The principal differential diagnosis is that of nocturnal seizures, which can be excluded with polysomnography. In RBD, seizure activity is absent on theEEG, and disinhibition of the usual motor atonia is observed in theEMGduring REM sleep, at times associated with complex motor behaviors. The pathogenesis is unclear, but damage to brainstem areas mediating descending motor inhibition during REM sleep may be responsible. In support of this hypothesis are the remarkable similarities between RBD and the sleep of animals with bilateral lesions of the pontine tegmentum in areas controlling REM sleep motor inhibition. Treatment with clonazepam provides sustained improvement in almost all reported cases.

Sleep Bruxism Bruxism is an involuntary, forceful grinding of teeth during sleep that affects 10 to 20% of the population. The patient is usually unaware of the problem. The typical age of onset is 17 to 20 years, and spontaneous remission usually occurs by age 40. Sex distribution appears to be equal. Treatment is dictated by the risk of dental injury. In many cases, the diagnosis is made during dental examination, damage is

minor, and no treatment is indicated. In more severe cases, treatment with a rubber tooth guard is necessary to prevent disfiguring tooth injury. Stress management or, in some cases, biofeedback can be useful when bruxism is a manifestation of psychological stress. There are anecdotal reports of benefit using benzodiazepines.

Sleep Enuresis Bedwetting, like sleepwalking and night terrors, is another parasomnia that occurs during slow-wave sleep in the young. Before age 5 or 6, nocturnal enuresis should probably be considered a normal feature of development. The condition usually improves spontaneously at puberty, has a prevalence in late adolescence of 1 to 3%, and is rare in adulthood. The age threshold for initiation of treatment depends on parental and patient concern about the problem. Persistence of enuresis into adolescence or adulthood may reflect a variety of underlying conditions. In older patients with enuresis a distinction must be made between primary and secondary enuresis, the latter being defined as bedwetting in patients who have been fully continent for 6 to 12 months. Treatment of primary enuresis is reserved for patients of appropriate age (older than 5 or 6 years) and consists of bladder training exercises and behavioral therapy. Urologic abnormalities are more common in primary enuresis and must be assessed by urologic examination. Important causes of secondary enuresis include emotional disturbances, urinary tract infections or malformations, cauda equina lesions, epilepsy, sleep apnea, and certain medications. Symptomatic pharmacotherapy is usually accomplished with intranasal desmopressin, or oral oxybutynin chloride or imipramine.

Miscellaneous Parasomnias Other clinical entities fulfill the definition of a parasomnia in that they occur selectively during sleep and are associated with some degree of sleep disruption. Examples include *jactatio capitis nocturna* (nocturnal headbanging), sleep talking, nocturnal paroxysmal dystonia, and nocturnal leg cramps.

CIRCADIAN RHYTHM SLEEP DISORDERS

A subset of patients presenting with either insomnia or hypersomnia may have a disorder of sleep *timing* rather than sleep *generation*. Disorders of sleep timing can either be organic (i.e., due to an intrinsic defect in the circadian pacemaker or its input from entraining stimuli) or environmental (i.e., due to a disruption of exposure to entraining stimuli from the environment). Regardless of etiology, the symptoms reflect the influence of the underlying circadian pacemaker on sleep-wake function. Thus, effective therapeutic approaches should aim to entrain the oscillator at an appropriate phase.

RAPID TIME-ZONE CHANGE (JET LAG) SYNDROME

More than 60 million people experience transmeridian air travel annually, which is often associated with excessive daytime sleepiness, sleep onset insomnia, and frequent arousals from sleep, particularly in the latter half of the night. Gastrointestinal discomfort is common. The syndrome is transient, typically lasting 2 to 14 d depending on the number of time zones crossed, the direction of travel, and the traveler's age and phase-shifting capacity. Travelers who spend more time outdoors reportedly adapt more quickly than those who remain in hotel rooms, presumably due to bright (outdoor) light exposure.

SHIFT-WORK SLEEP DISORDER

More than 7 million workers in the United States regularly work at night, either on a permanent or rotating schedule. In addition, each week millions of Americans elect to remain awake at night to meet deadlines, drive long distances, or participate in recreational activities, leading to both sleep loss and misalignment of their circadian rhythms with respect to their sleep-wake cycle. Chronic shift workers have higher rates of cardiac, gastrointestinal, and reproductive disorders. Studies of regular night-shift workers indicate that the circadian timing system usually fails to adapt successfully to such inverted schedules. This leads to a misalignment between the desired work-rest schedule and the output of the pacemaker and in disturbed daytime sleep. Consequent sleep deprivation, increased length of time awake prior to work, and misalignment of circadian phase produce decreased alertness and performance, increased reaction time, and increased risk of performance lapses, thereby resulting in greater safety hazards among night workers and other sleep-deprived individuals.

Sleep onset is associated with marked attenuation in perception of both auditory and visual stimuli and lapses of consciousness. The sleepy individual may thus attempt to perform routine and familiar motor tasks during the transition state between wakefulness and sleep (stage 1 sleep) in the absence of adequate sensory input from the environment. Motor vehicle operators are especially vulnerable to sleep-related accidents since the sleep-deprived driver or operator often fails to heed the warning signs of fatigue. Such attempts to override the powerful biologic drive for sleep by the sheer force of will can yield a catastrophic outcome when sleep processes intrude involuntarily upon the waking brain. Such intrusions typically last only seconds but are known on occasion to persist for longer durations. These frequent brief intrusions of stage 1 sleep into behavioral wakefulness are a major component of the impaired psychomotor performance seen with sleepiness. Such intrusions and their associated performance lapses, which are preceded by a markedly increased subjective sense of sleepiness, will inevitably occur if the need for sleep is not satiated. There is a marked increase in the risk of sleep-related, fatal-to-the-driver highway crashes in the early morning and late afternoon hours, coincident with peaks in the daily rhythm of sleep tendency.

Safety programs should promote education about sleep and increase awareness of the hazards associated with night work and should be aimed at minimizing both circadian disruption and sleep deprivation. The work schedule should minimize: (1) exposure to night work, (2) the frequency of shift rotation so that shifts do not rotate more than once every 2 to 3 weeks, (3) the number of consecutive night shifts, and (4) the duration of night shifts. In fact, shift durations of greater than 18 h should be universally recognized as increasing the risk of sleep-related errors and performance lapses. Caffeine is undoubtedly the most widely used wake-promoting drug, but it cannot forestall sleep indefinitely and does not protect users from sleep-related performance lapses. Postural changes, exercise, and strategic placement of nap opportunities can sometimes temporarily reduce the risk of fatigue-related performance lapses. Properly timed exposure to bright light can facilitate rapid adaptation to night-shift work, where feasible. An adequate number of safe highway rest areas, shoulder rumble strips, and strict enforcement and compliance monitoring of hours-of-service policies are needed to

reduce the risk of sleep-related transportation crashes. Such steps can lead to improvements in performance and to reduced accident rates both at work and on the roadways.

DELAYED SLEEP PHASE SYNDROME

Delayed sleep phase syndrome is characterized by: (1) reported sleep onset and wake times intractably later than desired, (2) actual sleep times at nearly the same clock hours daily, and (3) essentially normal all-night polysomnography except for delayed sleep onset. Patients exhibit an abnormally delayed endogenous circadian phase, with the temperature minimum during the constant routine occurring later than normal. This delayed phase could be due to: (1) an abnormally long intrinsic period of the endogenous circadian pacemaker; (2) an abnormally reduced phase-advancing capacity of the pacemaker; or (3) an irregular prior sleep-wake schedule, characterized by frequent nights when the patient chooses to remain awake well past midnight (for social, school, or work reasons). In most cases, it is difficult to distinguish among these factors, since patients with an abnormally long intrinsic period are more likely to "choose" such late-night activities because they are unable to sleep at that time. Patients tend to be young adults. This self-perpetuating condition can persist for years and does not usually respond to attempts to reestablish normal bedtime hours.

Treatment methods involving bright-light phototherapy during the morning hours or melatonin administration in the evening hours show promise in these patients, although the relapse rate among such patients is very high.

ADVANCED SLEEP PHASE SYNDROME

Advanced sleep phase syndrome is the converse of the delayed sleep phase syndrome and tends to occur in the elderly. Patients with this condition report excessive daytime sleepiness during the evening hours, when they have great difficulty remaining awake, even in social settings. The patients awaken from 3 to 5 A.M. each day, often several hours before their desired wake times. Although such patients have not been studied extensively, familial inheritance of this condition has been reported. Some of these patients may benefit from bright-light phototherapy during the evening hours, designed to reset the circadian pacemaker to a later hour.

NON-24-H SLEEP-WAKE DISORDER

This condition can occur when the maximal phase-advancing capacity of the circadian pacemaker is not adequate to accommodate the difference between the 24-h geophysical day and the intrinsic period of the pacemaker in the patient. Alternatively, patients' self-selected exposure to artificial light may drive the circadian pacemaker to a longer than 24-h schedule. Affected patients are not able to maintain a stable phase relationship between the output of the pacemaker and the 24-h day. Such patients typically present with an incremental pattern of successive delays in sleep onsets and wake times, progressing in and out of phase with local time. When the patient's endogenous rhythms are out of phase with the local environment, insomnia coexists with excessive daytime sleepiness. Conversely, when the endogenous rhythms are in phase with the local environment, symptoms remit. The intervals between symptomatic

periods may last several weeks to several months. Blind individuals unable to perceive light are particularly susceptible to this disorder. Melatonin administration has been reported to improve sleep, and in some cases even to induce synchronization of the circadian pacemaker.

MEDICAL IMPLICATIONS OF CIRCADIAN RHYTHMICITY

Understanding the role of circadian rhythmicity in the pathophysiology of illness may lead to improvements in diagnosis and treatment. For example, prominent circadian variations have been reported in the incidence of *acute myocardial infarction*, *sudden cardiac death*, and *stroke*, the leading causes of death in the United States. Platelet aggregability is increased after arising in the early morning hours, coincident with the peak incidence of these cardiovascular events. A better understanding of the possible role of circadian rhythmicity in the acute destabilization of a chronic condition such as atherosclerotic disease could improve the understanding of the pathophysiology.

Diagnostic and therapeutic procedures may also be affected by the time of day at which data are collected. Examples include blood pressure, body temperature, the dexamethasone suppression test, and plasma cortisol levels. The timing of chemotherapy administration has been reported to have an effect on the outcome of treatment. Few physicians realize the extent to which routine measures are affected by the time (or sleep/wake state) when the measurement is made.

In addition, both the toxicity and effectiveness of drugs can vary during the day. For example, more than a fivefold difference has been observed in mortality rates following administration of toxic agents to experimental animals at different times of day. Anesthetic agents are particularly sensitive to time-of-day effects. Finally, the physician must be increasingly aware of the public health risks associated with the ever-increasing demands made by the duty-rest-recreation schedules in our round-the-clock society.

(Bibliography omitted in Palm version)

Back to Table of Contents

SECTION 4 -DISORDERS OF EYES, EARS, NOSE, AND THROAT

28. DISORDERS OF THE EYE - *Jonathan C. Horton*

THE HUMAN VISUAL SYSTEM

The visual system provides a supremely efficient means for the rapid assimilation of information from the environment to aid in the guidance of behavior. The act of seeing begins with the capture of images focused by the cornea and lens upon a light-sensitive membrane in the back of the eye, called the *retina*. The retina is actually part of the brain, banished to the periphery to serve as a transducer for the conversion of patterns of light energy into neuronal signals. Light is absorbed by photopigment in two types of receptors: rods and cones. In the human retina there are 100 million rods and 5 million cones. The rods operate in dim (scotopic) illumination. The cones function under daylight (photopic) conditions. The cone system is specialized for color perception and high spatial resolution. The majority of cones are located within the macula, the portion of the retina serving the central 10°of vision. In the middle of the macula a small pit termed the *fovea*, packed exclusively with cones, provides best visual acuity.

Photoreceptors hyperpolarize in response to light, activating bipolar, amacrine, and horizontal cells in the inner nuclear layer. After processing of photoreceptor responses by this complex retinal circuit, the flow of sensory information ultimately converges upon a final common pathway: the ganglion cells. These cells translate the visual image impinging upon the retina into a continuously varying barrage of action potentials that propagates along the primary optic pathway to visual centers within the brain. There are a million ganglion cells in each retina, and hence a million fibers in each optic nerve.

Ganglion cell axons sweep along the inner surface of the retina in the nerve fiber layer, exit the eye at the optic disc, and travel through the optic nerve, optic chiasm, and optic tract to reach targets in the brain. The majority of fibers synapse upon cells in the lateral geniculate body, a thalamic relay station. Cells in the lateral geniculate body project in turn to the primary visual cortex. This massive afferent retinogeniculocortical sensory pathway provides the neural substrate for visual perception. Although the lateral geniculate body is the main target of the retina, separate classes of ganglion cells project to other subcortical visual nuclei involved in different functions. The pupillary reflex is mediated by input to the pretectal olivary nuclei in the midbrain. These pretectal nuclei send their output to the ipsilateral and contralateral Edinger-Westphal nuclei of the oculomotor nuclear complex. Cells in the Edinger-Westphal nuclei provide parasympathetic innervation to the iris sphincter via an interneuron in the ciliary ganglion. Circadian rhythms are timed by a retinal projection to the suprachiasmatic nucleus. Visual orientation and eye movements are served by retinal input to the superior colliculus. Gaze stabilization and optokinetic reflexes are governed by a group of small retinal targets known collectively as the *brainstem accessory optic system*. Finally, there is a sizeable retinal projection to the pulvinar, a large thalamic visual nucleus of obscure function.

The eyes must be rotated constantly within their orbits to place and maintain targets of visual interest upon the fovea. This activity, called *foveation*, or looking, is governed by an elaborate efferent motor system. Each eye is moved by six extraocular muscles,

supplied by cranial nerves from the oculomotor (III), trochlear (IV), and abducens (VI) nuclei. Activity in these ocular motor nuclei is coordinated by pontine and midbrain mechanisms for smooth pursuit, saccades, and gaze stabilization during head and body movements. Large regions of the frontal and parietooccipital cortex control these brainstem eye movement centers by providing descending supranuclear input.

Visual function can be disturbed in myriad ways. The eyes are mounted in a prominent position on the head, where they are vulnerable to trauma, exposure, and infection. Vision can be damaged by diseases intrinsic to the eye, such as glaucoma, cataract, or retinal detachment. Many neurologic diseases produce ocular symptoms, because extensive areas of the cortex, thalamus, cerebellum, and brainstem are devoted to visual perception or to the execution of eye movements. In genetic disorders, eye manifestations are common and often help the clinician to recognize a rare syndrome. Finally, the eyes are affected frequently by acquired systemic diseases.

The eye is a specialized organ, requiring unique optical instruments for proper examination. The slit lamp and ophthalmoscope proffer a beautiful, magnified view of the transparent anatomy of the eye and afford the only opportunity for direct inspection of blood vessels in a living subject. Some physicians do not acquire sufficient facility with these instruments to care for patients with eye problems. This is regrettable, for although it may be determined that a patient requires referral to an ophthalmologist, the initial evaluation of ocular symptoms lies within the purview of all physicians, and the assessment of visual acuity, pupils, eye movements, visual fields, and the fundi remain part of any general physical examination.

CLINICAL ASSESSMENT OF VISUAL FUNCTION

REFRACTIVE STATE

In approaching the patient with reduced vision, the first step is to decide whether refractive error is responsible. In *emmetropia*, parallel rays from infinity are focused perfectly upon the retina. Sadly, this condition is enjoyed by only a minority of the population. In *myopia*, the globe is too long, and light rays come to a focal point in front of the retina. Near objects can be seen clearly, but distant objects require a diverging lens in front of the eye. In *hyperopia*, the globe is too short, and hence a converging lens is used to supplement the refractive power of the eye. In *astigmatism*, the corneal surface is not perfectly spherical, necessitating a cylindrical corrective lens. In recent years it has become possible to correct refractive error with the excimer laser by performing either LASIK (laser in situ keratomileusis) or PRK (photorefractive keratectomy) to alter the curvature of the cornea.

With the onset of middle age, *presbyopia* develops as the lens within the eye becomes unable to increase its refractive power to accommodate upon near objects. To compensate for presbyopia, the emmetropic patient must use reading glasses. The patient already wearing glasses for distance correction usually switches to bifocals. The only exception is the myopic patient, who may achieve clear vision at near simply by removing glasses containing the distance prescription.

Refractive errors usually develop slowly and remain stable after adolescence, except in

unusual circumstances. For example, the acute onset of diabetes mellitus can produce sudden myopia because of fluid imbibition and swelling of the lens induced by hyperglycemia. Testing vision through a pinhole aperture is a useful way to screen quickly for refractive error. If the visual acuity is better through a pinhole than with the unaided eye, the patient needs a refraction to obtain best corrected visual acuity.

VISUAL ACUITY

The Snellen chart is used to test acuity at a distance of 6 m (20 ft). For convenience, a scale version of the Snellen chart, called the Rosenbaum card, is held at 36 cm (14 in) from the patient (Fig. 28-1). All subjects should be able to read the 6/6 m (20/20 ft) line with each eye using their refractive correction, if any. Patients who need reading glasses because of presbyopia must wear them for accurate testing with the Rosenbaum card. If 6/6 (20/20) acuity is not present in each eye, the deficiency in vision must be explained. For acuity worse than 6/240 (20/800), the ability to count fingers, see hand motions, or perceive a bright light should be recorded. Legal blindness is defined by the Internal Revenue Service as a best corrected acuity of 6/60 (20/200) or less in the better eye, or a binocular visual field subtending 20° or less. For driving the laws vary by state, but most require a corrected acuity of 6/12 (20/40) in at least one eye. Patients with a homonymous hemianopia should not drive.

PUPILS

The pupils should be tested individually in dim light with the patient fixating upon a distant target. If they respond briskly to light, there is no need to check the near response, because isolated loss of constriction (miosis) to accommodation does not occur. For this reason, the ubiquitous abbreviation PERRLA (pupils equal, round, and reactive to light and accommodation) implies a wasted effort with the last step. However, it is important to test the near response if the light response is poor or absent. Light-near dissociation occurs with neurosyphilis (Argyll Robertson pupil), lesions of the dorsal midbrain (obstructive hydrocephalus, pineal region tumors), and after aberrant regeneration (oculomotor nerve palsy, Adie's tonic pupil).

An eye with no light perception has no pupillary response to direct light stimulation. If the retina or optic nerve is only partially injured, the direct pupillary response will be weaker than the consensual pupillary response evoked by shining a light into the other eye. This *relative afferent pupillary defect* (Marcus Gunn pupil) can be elicited with the swinging flashlight test (Fig. 28-2). It is an extremely useful sign in retrobulbar optic neuritis and other optic nerve diseases, where it may be the sole objective evidence for disease.

Subtle inequality in pupil size, up to 0.5 mm, is a fairly common finding in normal persons. The diagnosis of essential or physiologic anisocoria is secure as long as the relative pupil asymmetry remains constant as ambient lighting varies. Anisocoria that increases in dim light indicates a sympathetic paresis of the iris dilator muscle. The triad of miosis with ipsilateral ptosis and anhidrosis constitutes Horner's syndrome, although anhidrosis is an inconstant feature. Brainstem stroke, carotid dissection, or neoplasm impinging upon the sympathetic chain are occasionally identified as the cause of Horner's syndrome, but most of cases are idiopathic.

Anisocoria that increases in bright light suggests a parasympathetic palsy. The first concern is an oculomotor nerve paresis. This possibility is excluded if the eye movements are full and the patient has no ptosis or diplopia. Acute pupillary dilation (mydriasis) can occur from damage to the ciliary ganglion in the orbit. Common mechanisms are infection (herpes zoster, influenza), trauma (blunt, penetrating, surgical), or ischemia (diabetes, temporal arteritis). After denervation of the iris sphincter the pupil does not respond well to light, but the response to near is often relatively intact. When the near stimulus is removed, the pupil redilates very slowly compared with the normal pupil, hence the term *tonic pupil*. In Adie's syndrome, a tonic pupil occurs in conjunction with weak or absent tendon reflexes in the lower extremities. This benign disorder, which occurs predominantly in healthy young women, is assumed to represent a mild dysautonomia. Tonic pupils are also associated with Shy-Drager syndrome, segmental hypohidrosis, diabetes, and amyloidosis. Occasionally, a tonic pupil is discovered incidentally in an otherwise completely normal, asymptomatic individual. The diagnosis is confirmed by placing a drop of dilute (0.125%) pilocarpine into each eye. Denervation hypersensitivity produces pupillary constriction in a tonic pupil, whereas the normal pupil shows no response. Pharmacologic dilation from accidental or deliberate instillation of anticholinergic agents (atropine, scopolamine drops) into the eye can also produce pupillary mydriasis. In this situation, normal strength (1%) pilocarpine causes no constriction.

Both pupils are affected equally by systemic medications. They are small with narcotic use (morphine, heroin) and large with anticholinergics (scopolamine). Parasympathetic agents (pilocarpine, demecarium bromide) used to treat glaucoma produce miosis. In any patient with an unexplained pupillary abnormality, a slit-lamp examination is helpful to exclude surgical trauma to the iris, an occult foreign body, perforating injury, intraocular inflammation, adhesions (synechia), angle-closure glaucoma, and iris sphincter rupture from blunt trauma.

EYE MOVEMENTS AND ALIGNMENT

Eye movements are tested by asking the patient with both eyes open to pursue a small target such as a penlight into the cardinal fields of gaze. Normal ocular versions are smooth, symmetric, full, and maintained in all directions without nystagmus. Saccades, or quick refixation eye movements, are assessed by having the patient look back and forth between two stationary targets. The eyes should move rapidly and accurately in a single jump to their target. Ocular alignment can be judged by holding a penlight directly in front of the patient at about 1 m. If the eyes are straight, the corneal light reflex will be centered in the middle of each pupil. To test eye alignment more precisely, the cover test is useful. The patient is instructed to gaze upon a small fixation target in the distance. One eye is covered suddenly while observing the second eye. If the second eye shifts to fixate upon the target, it was misaligned. If it does not move, the first eye is uncovered and the test is repeated on the second eye. If neither eye moves, the eyes are aligned orthotropically. If the eyes are orthotropic in primary gaze but the patient complains of diplopia, the cover test should be performed with the head tilted or turned in whatever direction elicits the patient's diplopia. With practice the examiner can detect an ocular deviation (heterotropia) as small as 1 to 2° with the cover test. Deviations can be measured by placing prisms in front of the misaligned eye to determine the power

required to neutralize the fixation shift evoked by covering the other eye.

STEREOPSIS

Stereoacuity is determined by presenting targets with retinal disparity separately to each eye using polarized images. The most popular office tests measure a range of thresholds from 800 to 40 seconds of arc. Normal stereoacuity is 40 seconds of arc. If a patient achieves this level of stereoacuity, one is assured that the eyes are aligned orthotropically and that vision is intact in each eye. Random dot stereograms have no monocular depth cues and provide an excellent screening test for strabismus and amblyopia in children.

COLOR VISION

The retina contains three classes of cones, with visual pigments of differing peak spectral sensitivity: red (560 nm), green (530 nm), and blue (430 nm). The red and green cone pigments are encoded on the X chromosome; the blue cone pigment on chromosome 7. Mutations of the blue cone pigment are exceedingly rare. Mutations of the red and green pigments cause congenital X-linked color blindness in 8% of males. Affected individuals are not truly color blind; rather, they differ from normal subjects in how they perceive color and how they combine primary monochromatic lights to match a given color. Anomalous trichromats have three cone types, but a mutation in one cone pigment (usually red or green) causes a shift in peak spectral sensitivity, altering the proportion of primary colors required to achieve a color match. Dichromats have only two cone types and will therefore accept a color match based upon only two primary colors. Anomalous trichromats and dichromats have 6/6 (20/20) visual acuity, but their hue discrimination is impaired. Ishihara color plates can be used to detect red-green color blindness. The test plates contain a hidden number, visible only to subjects with color confusion from red-green color blindness. Because color blindness is almost exclusively X-linked, it is worth screening only male children.

The Ishihara plates are often used to detect acquired defects in color vision, although they are intended as a screening test for congenital color blindness. Acquired defects in color vision frequently result from disease of the macula or optic nerve. For example, patients with a history of optic neuritis often complain of color desaturation long after their visual acuity has returned to normal. Color blindness can also occur from bilateral strokes involving the ventral portion of the occipital lobe (cerebral achromatopsia). Such patients can perceive only shades of gray and may also have difficulty recognizing faces (prosopagnosia). Infarcts of the dominant occipital lobe sometimes give rise to color anomia. Affected patients can discriminate colors, but they cannot name them.

VISUAL FIELDS

Vision can be impaired by damage to the visual system anywhere from the eyes to the occipital lobes. One can localize the site of the lesion with considerable accuracy by mapping the visual field deficit by finger confrontation and then correlating it with the topographic anatomy of the visual pathway (Fig. 28-3). More quantitative data can be obtained by formal perimetric examination of the visual fields. In kinetic perimetry, the patient faces a tangent screen or a hemispheric bowl (Goldmann perimeter) while the

examiner moves a small light target from the periphery towards the center. Such manual techniques have largely been supplanted by computer-driven perimeters (Humphrey, Octopus) that present a target of variable intensity at fixed positions in the visual field (Fig. 28-3*A*). By generating an automated printout of light thresholds, these static perimeters provide a sensitive means of detecting scotomas in the visual field. They are also useful for serial assessment of visual function in chronic diseases such as glaucoma or pseudotumor cerebri.

The crux of visual field analysis is to decide whether a lesion is before, at, or behind the optic chiasm. If a scotoma is confined to one eye, it must be due to a lesion anterior to the chiasm, involving either the optic nerve or retina. Retinal lesions produce scotomas that correspond optically to their location in the fundus. For example, a superior-nasal retinal detachment results in an inferior-temporal field cut. Damage to the macula causes a central scotoma (Fig. 28-3*B*).

Optic nerve disease produces characteristic patterns of visual field loss. Glaucoma selectively destroys axons that enter the superotemporal or inferotemporal poles of the optic disc, resulting in arcuate scotomas shaped like a Turkish scimitar, which emanate from the blind spot and curve around fixation to end flat against the horizontal meridian (Fig. 28-3*C*). This type of field defect mirrors the arrangement of the nerve fiber layer in the temporal retina. The superb acuity of humans is achieved by thrusting aside all retinal elements at the fovea except photoreceptors, to minimize absorption and scattering of light. To avoid passing over the fovea, axons from cells in the temporal retina must follow an indirect course arching around the fovea to reach the optic disc. Arcuate or nerve fiber layer scotomas also occur from optic neuritis, ischemic optic neuropathy, optic disc drusen, and branch retinal artery or vein occlusion.

Damage to the entire upper or lower pole of the optic disc causes an altitudinal field cut that follows the horizontal meridian (Fig. 28-3*D*). This pattern of visual field loss is typical of ischemic optic neuropathy but also occurs from retinal vascular occlusion, advanced glaucoma, and optic neuritis.

About half the fibers in the optic nerve originate from ganglion cells serving the macula. Damage to papillomacular fibers causes a cecocentral scotoma encompassing the blind spot and macula (Fig. 28-3*E*). If the damage is irreversible, pallor eventually appears in the temporal portion of the optic disc. Temporal pallor from a cecocentral scotoma may develop in optic neuritis, nutritional optic neuropathy, toxic optic neuropathy, Leber's hereditary optic neuropathy, and compressive optic neuropathy. It is worth mentioning that the temporal side of the optic disc is slightly more pale than the nasal side in most normal individuals. Therefore, it can sometimes be difficult to decide whether the temporal pallor visible on fundus examination represents a pathologic change. Pallor of the nasal rim of the optic disc is a less equivocal sign of optic atrophy.

At the optic chiasm, fibers from nasal ganglion cells decussate into the contralateral optic tract. Crossed fibers are damaged more by compression than uncrossed fibers. As a result, mass lesions of the sellar region cause a temporal hemianopia in each eye. Tumors anterior to the optic chiasm, such as meningiomas of the tuberculum sella, produce a junctional scotoma characterized by an optic neuropathy in one eye and a superior temporal field cut in the other eye (Fig. 28-3*G*). More symmetric compression

of the optic chiasm by a pituitary adenoma (Fig. 28-4), meningioma, craniopharyngioma, glioma, or aneurysm results in a bitemporal hemianopia (Fig. 28-3*H*). The insidious development of a bitemporal hemianopia often goes unnoticed by the patient and will escape detection by the physician unless each eye is tested separately.

It is difficult to localize a postchiasmal lesion accurately, because injury anywhere in the optic tract, lateral geniculate body, optic radiations, or visual cortex can produce a homonymous hemianopia, i.e., a temporal hemifield defect in the contralateral eye and a matching nasal hemifield defect in the ipsilateral eye (Fig. 28-3*I*). A unilateral postchiasmal lesion leaves the visual acuity in each eye unaffected, although the patient may read the letters on only the left or right half of the eye chart. Lesions of the optic radiations tend to cause poorly matched or incongruous field defects in each eye. Damage to the optic radiations in the temporal lobe (Meyer's loop) produces a superior quandrantic homonymous hemianopia (Fig. 28-3*J*), whereas injury to the optic radiations in the parietal lobe results in an inferior quandrantic homonymous hemianopia (Fig. 28-3*K*). Lesions of the primary visual cortex give rise to dense, congruous hemianopic field defects. Occlusion of the posterior cerebral artery supplying the occipital lobe is a frequent cause of total homonymous hemianopia. Some patients with hemianopia after occipital stroke have macular sparing, because the macular representation at the tip of the occipital lobe is supplied by collaterals from the middle cerebral artery (Fig. 28-3*L*). Destruction of both occipital lobes produces cortical blindness. This condition can be distinguished from bilateral prechiasmal visual loss by noting that the pupil responses and optic fundi remain normal.

RED OR PAINFUL EYE

Corneal Abrasions These are seen best by placing a drop of fluorescein in the eye and looking with the slit lamp using a cobalt-blue light. A penlight with a blue filter will suffice if no slit lamp is available. Damage to the corneal epithelium is revealed by yellow fluorescence of the exposed basement membrane underlying the epithelium. It is important to check for foreign bodies. To search the conjunctival fornices, the lower lid should be pulled down and the upper lid everted. A foreign body can be removed with a moistened cotton-tipped applicator after placing a drop of topical anesthetic, such as proparacaine, in the eye. Alternatively, it may be possible to flush the foreign body from the eye by irrigating copiously with saline or artificial tears. If the corneal epithelium has been abraded, antibiotic ointment and a patch should be applied to the eye. A drop of an intermediate-acting cycloplegic, such as cyclopentolate hydrochloride 1%, helps to reduce pain by relaxing the ciliary body. The eye should be reexamined the next day. Minor abrasions may not require patching and cycloplegia.

Subconjunctival Hemorrhage This results from rupture of small vessels bridging the potential space between the episclera and conjunctiva. Blood dissecting into this space can produce a spectacular red eye, but vision is not affected and the hemorrhage resolves without treatment. Subconjunctival hemorrhage is usually spontaneous but can occur from blunt trauma, eye rubbing, or vigorous coughing. Occasionally it is a clue to an underlying bleeding disorder.

Pinguecula This is a small, raised conjunctival nodule at the temporal or nasal limbus. In adults such lesions are extremely common and have little significance, unless they

become inflamed (pingueculitis). A *pterygium* resembles a pinguecula but has crossed the limbus to encroach upon the corneal surface. Removal is justified when symptoms of irritation or blurring develop, but recurrence is a common problem.

Blepharitis This refers to inflammation of the eyelids. The most common form occurs in association with acne rosacea or seborrheic dermatitis. The eyelid margins are usually colonized heavily by staphylococcus. Upon close inspection, they appear greasy, ulcerated, and crusted with scaling debris that clings to the lashes. Treatment consists of warm compresses, strict eyelid hygiene, and topical antibiotics such as erythromycin. An external *hordeolum* (sty) is caused by staphylococcal infection of the superficial accessory glands of Zeis or Moll located in the eyelid margins. An internal hordeolum occurs after suppurative infection of the oil-secreting meibomian glands within the tarsal plate of the eyelid. Systemic antibiotics, usually tetracyclines, are sometimes necessary for treatment of meibomian gland inflammation (meibomitis) or chronic, severe blepharitis. A *chalazion* is a painless, granulomatous inflammation of a meibomian gland that produces a pealike nodule within the eyelid. It can be incised and drained, or injected with glucocorticoids. Basal cell, squamous cell, or meibomian gland carcinoma should be suspected for any nonhealing, ulcerative lesion of the eyelids.

Dacrocystitis An inflammation of the lacrimal drainage system, this can produce epiphora (tearing) and ocular injection. Gentle pressure over the lacrimal sac evokes pain and reflux of mucus or pus from the tear puncta. Dacrocystitis usually occurs after obstruction of the lacrimal system. It is treated with topical and systemic antibiotics, followed by probing or surgery to reestablish patency. *Entropion* (inversion of the eyelid) or *ectropion* (sagging or eversion of the eyelid) can also lead to epiphora and ocular irritation.

Conjunctivitis This is the most common cause of a red, irritated eye. Pain is minimal, and the visual acuity is reduced only slightly. The most common viral etiology is adenovirus infection. It causes a watery discharge, mild foreign-body sensation, and photophobia. Bacterial infection tends to produce a more mucopurulent exudate. Mild cases of infectious conjunctivitis are usually treated empirically with broad-spectrum topical ocular antibiotics, such as sulfacetamide 10%, polymixin-bacitracin-neomycin, or trimethoprim-polymixin combination. Smears and cultures are usually reserved for severe, resistant, or recurrent cases of conjunctivitis. To prevent contagion, patients should be admonished to wash their hands frequently, not to touch their eyes, and to avoid direct contact with others.

Allergic Conjunctivitis This condition is extremely common and often mistaken for infectious conjunctivitis. Three forms of allergic conjunctivitis are recognized, with closely overlapping manifestations. *Hay fever conjunctivitis* has a seasonal incidence, related to the release of airborne antigens into the air by plants. IgE-mediated activation of mast cells in the conjunctiva causes itching, redness, and edema. *Vernal conjunctivitis* is also seasonal, becoming worse during warm months. It affects exclusively children or adolescents and is more common in boys. The cause is unknown, but airborne antigens are thought to trigger symptoms. Itching, photophobia, epiphora, and mucous discharge are typical. The palpebral conjunctiva may become hypertropic with giant excrescences called cobblestone papillae. Irritation from contact lenses or any chronic foreign body can also induce formation of cobblestone papillae.

Atopic conjunctivitis occurs in subjects with atopic dermatitis or asthma. Symptoms caused by allergic conjunctivitis can be alleviated with cold compresses, topical vasoconstrictors, antihistamines, and mast-cell stabilizers such as cromolyn sodium. Topical glucocorticoid solutions provide dramatic relief of immune-mediated forms of conjunctivitis, but their long-term use is ill-advised because of the complications of glaucoma, cataract, and secondary infection. Topical nonsteroidal anti-inflammatory agents (NSAIDs) such as ketorolac tromethamine are a better alternative.

Keratoconjunctivitis Sicca Also known as dry eye, it produces a burning, foreign-body sensation, injection, and photophobia. In mild cases the eye appears surprisingly normal, but tear production measured by wetting of a filter paper (Schirmer strip) is deficient. A variety of systemic drugs, including antihistaminic, anticholinergic, and psychotropic medications, result in dry eye by reducing lacrimal secretion. Disorders that involve the lacrimal gland directly, such as sarcoidosis or Sjogren's syndrome, also cause dry eye. Patients may develop dry eye after radiation therapy if the treatment field includes the orbits. Problems with ocular drying are also common after lesions affecting cranial nerves V or VII. Corneal anesthesia is particularly dangerous, because the absence of a normal blink reflex exposes the cornea to injury without pain to warn the patient. Dry eye is managed by frequent and liberal application of artificial tears and ocular lubricants. In severe cases the tear puncta can be plugged or cauterized to reduce lacrimal outflow.

Keratitis This is a threat to vision because of the risk of corneal clouding, scarring, and perforation. Worldwide, the two leading causes of blindness from keratitis are trachoma from chlamydial infection and vitamin A deficiency related to malnutrition. In the United States, contact lenses play a major role in corneal infection and ulceration. They should not be worn by anyone with an active eye infection. In evaluating the cornea, it is important to differentiate between a superficial infection (*keratoconjunctivitis*) and a deeper, more serious ulcerative process. The latter is accompanied by greater visual loss, pain, photophobia, redness, and discharge. Slit-lamp examination shows disruption of the corneal epithelium, a cloudy infiltrate or abscess in the stroma, and an inflammatory cellular reaction in the anterior chamber. In severe cases, pus settles at the bottom of the anterior chamber, giving rise to a hypopyon. Immediate empirical antibiotic therapy should be initiated after corneal scrapings are obtained for Gram's stain, Giemsa stain, and cultures. Fortified topical antibiotics are most effective, supplemented with subconjunctival antibiotics as required. The most frequent bacterial pathogens are *Staphylococcus*, *Streptococcus* (particularly *S. pneumoniae*), *Pseudomonas*, Enterobacteriaceae, *Haemophilus*, and *Neisseria*. For *Neisseria*, systemic antibiotics should be given in addition to topical antibiotics to eliminate systemic infection. A fungal etiology should always be considered in the patient with keratitis. Fungal infection is common in warm humid climates, especially after penetration of the cornea by plant or vegetable material.

Herpes Simplex The *herpes viruses* are a major cause of blindness from keratitis. Most adults in the United States have serum antibodies to herpes simplex, indicating prior viral infection (Chap. 182). Primary ocular infection is generally caused by herpes simplex type 1, rather than type 2. It manifests as a unilateral follicular blepharoconjunctivitis, easily confused with adenoviral conjunctivitis unless telltale vesicles appear on the periocular skin or conjunctiva. A dendritic pattern of corneal

epithelial ulceration revealed by fluorescein staining is pathognomonic for herpes infection but is seen in only a minority of primary infections. Recurrent ocular infection arises from reactivation of the latent herpes virus. Viral eruption in the corneal epithelium may result in the characteristic herpes dendrite. Involvement of the corneal stroma produces edema, vascularization, and iridocyclitis. Herpes keratitis is treated with topical antiviral agents, cycloplegics, and oral acyclovir. Topical glucocorticoids are effective in mitigating corneal scarring but must be used with extreme caution because of the danger of corneal melting and perforation. Topical glucocorticoids also carry the risk of prolonging infection and inducing glaucoma.

Herpes Zoster Herpes zoster from reactivation of latent varicella (chickenpox) virus causes a dermatomal pattern of painful vesicular dermatitis. Ocular symptoms can occur after zoster eruption in any branch of the trigeminal nerve but are particularly common when vesicles form on the nose, reflecting nasociliary (V1) nerve involvement (Hutchinson's sign). Herpes zoster ophthalmicus produces corneal dendrites, which can be difficult to distinguish from those seen in herpes simplex. Stromal keratitis, anterior uveitis, raised intraocular pressure, ocular motor nerve palsies, acute retinal necrosis, and postherpetic scarring and neuralgia are other common sequelae. Herpes zoster ophthalmicus is treated with antiviral agents and cycloplegics. In severe cases, glucocorticoids may be added to prevent permanent visual loss from corneal scarring.

Episcleritis This is an inflammation of the episclera, a thin layer of connective tissue between the conjunctiva and sclera. Episcleritis resembles conjunctivitis but is a more localized process and discharge is absent. Most cases of episcleritis are idiopathic, but some occur in the setting of an autoimmune disease. *Scleritis* refers to a deeper, more severe inflammatory process, frequently associated with a connective tissue disease such as rheumatoid arthritis, lupus erythematosus, polyarteritis nodosa, Wegener's granulomatosis, or relapsing polychondritis. The inflammation and thickening of the sclera can be diffuse or nodular. In anterior forms of scleritis, the globe assumes a violet hue and the patient complains of severe ocular tenderness and pain. With posterior scleritis the pain and redness may be less marked, but there is often proptosis, choroidal effusion, reduced motility, and visual loss. Episcleritis and scleritis should be treated withNSAIDs. If these agents fail, topical or even systemic glucocorticoid therapy may be necessary, especially if an underlying autoimmune process is active.

Uveitis Involving the anterior structures of the eye, this is called *iritis* or *iridocyclitis*. The diagnosis requires slit-lamp examination to identify inflammatory cells floating in the aqueous humor or deposited upon the corneal endothelium (keratic precipitates). Anterior uveitis develops in sarcoidosis, ankylosing spondylitis, juvenile rheumatoid arthritis, inflammatory bowel disease, psoriasis, Reiter's syndrome, and Behcet's disease. It is also associated with herpes infections, syphilis, Lyme disease, onchocerciasis, tuberculosis, and leprosy. Although anterior uveitis can occur in conjunction with many diseases, no cause is found to explain the majority of cases. For this reason, laboratory evaluation is usually reserved for patients with recurrent or severe anterior uveitis. Treatment is aimed at reducing inflammation and scarring by judicious use of topical glucocorticoids. Dilation of the pupil reduces pain and prevents the formation of synechiae.

Posterior Uveitis This is diagnosed by observing inflammation of the vitreous, retina, or

choroid on fundus examination. It is more likely than anterior uveitis to be associated with an identifiable systemic disease. Some patients have panuveitis, or inflammation of both the anterior and posterior segments of the eye. Posterior uveitis is a manifestation of autoimmune diseases such as sarcoidosis, Behcet's disease, Vogt-Koyanagi-Harada syndrome, and inflammatory bowel disease (see Plate IV-1). It also accompanies diseases such as toxoplasmosis, onchocerciasis, cysticercosis, coccidioidomycosis, toxocariasis, and histoplasmosis; infections caused by organisms such as *Candida*, *Pneumocystis carinii*, *Cryptococcus*, *Aspergillus*, herpes, and cytomegalovirus (see Plate IV-2); and other diseases such as syphilis, Lyme disease, tuberculosis, cat-scratch disease, Whipple's disease, and brucellosis. In multiple sclerosis, chronic inflammatory changes can develop in the extreme periphery of the retina (pars planitis or intermediate uveitis).

Acute Angle-Closure Glaucoma This is a rare and frequently misdiagnosed cause of a red, painful eye. Susceptible eyes have a shallow anterior chamber, either because the eye has a short axial length (hyperopia) or a lens enlarged by the gradual development of cataract. When the pupil becomes mid-dilated, the peripheral iris blocks aqueous outflow via the anterior chamber angle and the intraocular pressure rises abruptly, producing pain, injection, corneal edema, obscurations, and blurred vision. In some patients, ocular symptoms are overshadowed by nausea, vomiting, or headache, prompting a fruitless workup for abdominal or neurologic disease. The diagnosis is made by measuring the intraocular pressure during an acute attack or by performing gonioscopy to reveal the narrowed chamber angle by means of a specially mirrored contact lens. Acute angle closure is treated with oral or intravenous acetazolamide, topical beta blockers, apraclonidine, and pilocarpine to induce miosis. If these measures fail, a laser can be used to create a hole in the peripheral iris to relieve pupillary block. Many physicians are reluctant to dilate patients routinely for fundus examination because they fear precipitating an angle-closure glaucoma. The risk is actually remote and more than outweighed by the potential benefit to patients of discovering a hidden fundus lesion visible only through a fully dilated pupil. Moreover, a single attack of angle closure after pharmacologic dilation rarely causes any permanent damage to the eye and serves as an inadvertent provocative test to identify patients with narrow angles who would benefit from prophylactic laser iridectomy.

Endophthalmitis This occurs from bacterial, viral, fungal, or parasitic infection of the internal structures of the eye. It is usually acquired by hematogenous seeding from a remote site. Chronically ill, diabetic, or immunosuppressed patients, especially those with a history of indwelling intravenous catheters or positive blood cultures, are at greatest risk for endogenous endophthalmitis. Although most patients have ocular pain and injection, visual loss is sometimes the only symptom. Septic emboli, from a diseased heart valve or a dental abscess, that lodge in the retinal circulation can give rise to endophthalmitis. White-centered retinal hemorrhages (Roth's spots) are considered pathognomonic for subacute bacterial endocarditis, but they also appear in leukemia, diabetes, and many other conditions. Endophthalmitis also occurs as a complication of ocular surgery, occasionally months or even years after the operation. An occult penetrating foreign body or unrecognized trauma to the globe should be considered in any patient with unexplained intraocular infection or inflammation.

TRANSIENT OR SUDDEN VISUAL LOSS

Amaurosis Fugax This term refers to a transient ischemic attack of the retina. Because neural tissue has a high rate of metabolism, interruption of blood flow to the retina for more than a few seconds results in *transient monocular blindness*, a term used interchangeably with amaurosis fugax. Patients describe a rapid fading of vision like a curtain descending, sometimes affecting only a portion of the visual field. Amaurosis fugax usually occurs from an embolus that becomes stuck within a retinal arteriole (see Plate IV-3). If the embolus breaks up or passes, flow is restored and vision returns quickly to normal without permanent damage. With prolonged interruption of blood flow, the inner retina suffers infarction. Ophthalmoscopy reveals zones of whitened, edematous retina following the distribution of branch retinal arterioles. Complete occlusion of the central retinal artery produces arrest of blood flow and a milky retina with a cherry-red fovea (see Plate IV-4). Emboli are composed of either cholesterol (Hollenhorst plaque), calcium, or platelet-fibrin debris. The most common source is an atherosclerotic plaque in the carotid artery or aorta, although emboli can also arise from the heart, especially in patients with diseased valves, atrial fibrillation, or wall motion abnormalities.

In rare instances, amaurosis fugax occurs from low central retinal artery perfusion pressure in a patient with a critical stenosis of the ipsilateral carotid artery and poor collateral flow via the circle of Willis. In this situation, amaurosis fugax develops when there is a dip in systemic blood pressure or a slight worsening of the carotid stenosis. Sometimes there is contralateral motor or sensory loss, indicating concomitant hemispheric cerebral ischemia.

Retinal arterial occlusion also occurs rarely in association with retinal migraine, lupus erythematosus, anticardiolipin antibodies (see Plate IV-4), anticoagulant deficiency states (protein S, protein C, and antithrombin III deficiency), pregnancy, intravenous drug abuse, blood dyscrasias, dysproteinemias, and temporal arteritis.

Amaurosis fugax warns of a patient at high risk for stroke. The carotid arteries should be studied by ultrasound. Endarterectomy for a stenosis of ³60%, even in asymptomatic patients, has been shown to reduce the subsequent rate of ipsilateral stroke (Chap. 361). Therapy with aspirin, warfarin, or other anticoagulants is appropriate in selected patients. If no carotid lesion is found, cardiac ultrasound should be performed. Ambulatory electrocardiographic monitoring may reveal that intermittent atrial fibrillation is giving rise to emboli.

Marked *systemic hypertension* causes sclerosis of retinal arterioles, splinter hemorrhages, focal infarcts of the nerve fiber layer (cotton-wool spots), and leakage of lipid and fluid (hard exudate) into the macula (see Plate IV-5). In hypertensive crisis, sudden visual loss can result from vasospasm of retinal arterioles and consequent retinal ischemia. In addition, acute hypertension may produce visual loss from ischemic swelling of the optic disc. Patients with acute hypertensive retinopathy should be treated by lowering the blood pressure. However, the blood pressure should not be reduced precipitously, because there is a danger of optic disc infarction from sudden hypoperfusion.

Impending *branch* or *central retinal vein occlusion* can produce prolonged visual

obscurations that resemble those described by patients with amaurosis fugax. The veins appear engorged and phlebitic, with numerous retinal hemorrhages (see Plate IV-6). In some patients, venous blood flow recovers spontaneously, while others evolve a frank obstruction with extensive retinal bleeding ("blood and thunder" appearance), infarction, and visual loss. Venous occlusion of the retina is often idiopathic, but hypertension, diabetes, and glaucoma are prominent risk factors. The benefit of treatment with anticoagulants is unproven and carries the risk of hemorrhage into the vitreous. Polycythemia, thrombocythemia, or other factors leading to an underlying hypercoagulable state should be corrected.

Anterior Ischemic Optic Neuropathy (AION) This is caused by insufficient blood flow through the posterior ciliary arteries supplying the optic disc. It produces sudden, painless, monocular visual loss, although patients occasionally report premonitory obscurations. The optic disc appears swollen and surrounded by nerve fiber layer splinter hemorrhages (seePlate IV-7). AION is divided into two forms: arteritic and nonarteritic. The nonarteritic form of AION is most common. No specific cause can be identified, although diabetes and hypertension are frequent risk factors. No treatment is available. About 5% of patients, especially those over age 60, develop the arteritic form of AION in conjunction with giant cell (temporal) arteritis (Chap. 317). It is urgent to recognize arteritic AION so that high doses of glucocorticoids can be instituted immediately to prevent blindness in the second eye. Symptoms of polymyalgia rheumatica may be present, and the sedimentation rate is usually elevated. In a patient with visual loss from suspected arteritic AION, temporal artery biopsy is helpful to confirm the diagnosis, but glucocorticoids should be started without waiting for the biopsy to be completed. The diagnosis of arteritic AION is difficult to sustain in the face of a normal sedimentation rate and a negative temporal artery biopsy, but such cases do occur rarely.

Posterior Ischemic Optic Neuropathy This is an infrequent cause of acute visual loss. It is induced by the combination of severe anemia and hypotension, causing infarction of the retrobulbar optic nerve. Cases have been reported after major blood loss during surgery, exsanguinating trauma, gastrointestinal bleeding, and renal dialysis. The fundus usually appears normal, although optic disc swelling develops if the process extends far enough anteriorly. Vision can be salvaged in some patients by prompt blood transfusion and reversal of hypotension.

Optic Neuritis This is a common inflammatory disease of the optic nerve. In the Optic Neuritis Treatment Trial (ONTT), the mean age of patients was 32 years, 77% were female, 92% had ocular pain (especially with eye movements), and 35% had optic disc swelling. In most patients, the demyelinating event was retrobulbar and the ocular fundus appeared normal on initial examination (seePlate IV-8), although optic disc pallor slowly developed over subsequent months.

Virtually all patients experience a gradual recovery of vision after a single episode of optic neuritis, even without treatment. This rule is so reliable that failure of vision to improve considerably after a first attack of optic neuritis casts doubt upon the original diagnosis. Treatment of optic neuritis is controversial because the favorable prognosis for visual recovery has made it difficult to demonstrate any benefit from glucocorticoids. TheONTTshowed that patients treated with a conventional dose of oral glucocorticoids

(prednisone, 1 mg/kg per day for 14 days) did no better than patients treated with a placebo. A recent Danish trial of oral high-dose methylprednisolone (500 mg daily for 5 days, followed by a 10-day taper) reported a slight response at 1 and 3 weeks but none at 8 weeks. From these studies, it is apparent that oral glucocorticoids have little to offer in the treatment of optic neuritis. According to the ONTT, even high-dose intravenous methylprednisolone (250 mg every 6 h for 3 days) followed by oral prednisone (1 mg/kg per day for 11 days) makes no difference in final acuity (measured 6 months after the attack), although the recovery of visual function occurs more rapidly.

For some patients, optic neuritis remains an isolated event. However, the ONTT showed that the 5-year cumulative probability of developing clinically definite multiple sclerosis following optic neuritis is 30%. Remarkably, intravenous glucocorticoids were associated with a reduced rate of development of multiple sclerosis over a 2-year follow-up period, especially in the subgroup of patients with multiple foci of demyelination on their magnetic resonance (MR) scan. However, by the end of a 3-year follow-up period, patients treated with intravenous glucocorticoids versus placebo showed no difference in the rate of multiple sclerosis. Moreover, intravenous glucocorticoids did not reduce the likelihood of subsequent attacks of optic neuritis. To summarize, the organizers of the ONTT recommend an MR scan in patients with optic neuritis. If two or more foci of demyelination are found or visual loss is severe, they suggest treatment with intravenous glucocorticoids. The potential benefits of intravenous glucocorticoids are: (1) a slightly faster recovery of visual function, and (2) a potential reduction in the risk of subsequent neurologic events that would signify multiple sclerosis. Critics of the ONTT have questioned these recommendations, pointing out that: (1) visual outcome is the same in the long run, (2) evidence indicating a reduced risk of eventual multiple sclerosis with intravenous glucocorticoid treatment is based upon follow-up data in a rather small number of patients, and (3) the protection against multiple sclerosis is transient, and no longer apparent beyond 2 years of follow-up. In cases of unilateral optic neuritis, the decision whether to obtain an MR scan or to treat with intravenous glucocorticoids should be based upon clinical judgment and careful discussion with the patient. In cases of bilateral, simultaneous optic neuritis, the rationale for intravenous glucocorticoids is stronger.

Leber's Hereditary Optic Neuropathy This is a disease of young men, characterized by onset over a few weeks of painless, severe, central visual loss in one eye, followed weeks or months later by the same process in the other eye. Acutely, the optic disc appears mildly plethoric with surface capillary telangiectases, but no vascular leakage on fluorescein angiography. Eventually optic atrophy ensues. There is no treatment. Leber's optic neuropathy is caused by a point mutation at codon 11778 in the mitochondrial gene encoding nicotinamide adenine dinucleotide dehydrogenase (NADH) subunit 4. Subsequently, additional mutations responsible for the disease have been identified, most in mitochondrial genes encoding proteins involved in electron transport. Mitochrondrial mutations causing Leber's neuropathy are inherited from the mother by all her children, but usually only sons develop symptoms. This curious male predilection is a mystery.

Toxic Optic Neuropathy This can result in acute visual loss with bilateral optic disc swelling and central or cecocentral scotomas. Such cases have been reported to result from exposure to ethambutol, methyl alcohol (moonshine), ethylene glycol (antifreeze),

or carbon monoxide. In toxic optic neuropathy, visual loss can also develop gradually and produce optic atrophy without a phase of acute optic disc edema (see Plate IV-9). Many agents have been implicated as a cause of toxic optic neuropathy, but the evidence supporting the association for many is weak. The following is a partial list of potential offending drugs or toxins: disulfiram, ethchlorvynol, chloramphenicol, amiodarone, monoclonal anti-CD3 antibody, ciprofloxacin, digitalis, streptomycin, lead, arsenic, thallium, D-penicillamine, isoniazid, emetine, and sulfonamides. Deficiency states, induced either by starvation, malabsorption, or alcoholism, can lead to insidious visual loss. Thiamine, vitamin B_{12}, and folate levels should be checked in any patient with unexplained, bilateral central scotomas and optic pallor.

Papilledema This connotes bilateral optic disc swelling from raised intracranial pressure (seePlate IV-10). Headache is a frequent, but not invariable, accompaniment. All other forms of optic disc swelling, e.g., from optic neuritis or ischemic optic neuropathy, should be called "optic disc edema." This convention is arbitrary but serves to avoid confusion. Often it is difficult to differentiate papilledema from other forms of optic disc edema by fundus examination alone. Transient visual obscurations are a classic symptom of papilledema. They can occur in only one eye or simultaneously in both eyes. They usually last seconds but can persist for minutes if the papilledema is fulminant. Obscurations follow abrupt shifts in posture or happen spontaneously. When obscurations are prolonged or spontaneous, the papilledema is more threatening. Visual acuity is not affected by papilledema unless the papilledema is severe, long-standing, or accompanied by macular edema and hemorrhage. Visual field testing shows enlarged blind spots and peripheral constriction (Fig. 28-3*F*). With unremitting papilledema, peripheral visual field loss progresses in an insidious fashion while the optic nerve develops atrophy. In this setting, reduction of optic disc swelling is an ominous sign of a dying nerve rather than an encouraging indication of resolving papilledema.

Evaluation of papilledema requires computed tomography (CT) orMRimaging to exclude an intracranial lesion. MR angiography is appropriate in selected cases to search for a dural venous sinus occlusion or an arteriovenous shunt. If neuroradiologic studies are negative, the subarachnoid opening pressure should be measured by lumbar puncture. An elevated pressure, with normal cerebrospinal fluid, points by exclusion to the diagnosis of *pseudotumor cerebri* (idiopathic intracranial hypertension). The majority of patients are young, female, and obese. Treatment with a carbonic anhydrase inhibitor such as acetazolamide lowers intracranial pressure by reducing the production of cerebrospinal fluid. Weight reduction is vital but often unsuccessful. If acetazolamide and weight loss fail, and visual field loss is progressive, lumboperitoneal shunting or optic nerve sheath fenestration should be undertaken without delay to prevent blindness. Occasionally, emergency surgery is required for sudden blindness caused by fulminant papilledema.

Optic Disc Drusen These are refractile deposits within the substance of the optic nerve head (seePlate IV-11). They are unrelated to drusen of the retina, which occur in age-related macular degeneration. Optic disc drusen are most common in people of northern European descent, with an incidence of 0.3 to 0.4%. Their diagnosis is obvious when they are visible as glittering particles upon the surface of the optic disc. However, in many patients they are hidden beneath the surface, producing an elevated optic disc with blurred margins that is easily mistaken for papilledema. It is important to recognize

pseudo-papilledema due to optic disc drusen to avoid an unneccessary evaluation for papilledema. Ultrasound orCTscanning are sensitive for detection of buried optic disc drusen because they contain calcium. In most patients, optic disc drusen are an incidental, innocuous finding, but they can produce visual obscurations. On perimetry they give rise to enlarged blind spots and arcuate scotomas from damage to the optic disc. With increasing age, drusen tend to become more exposed on the disc surface as optic atrophy develops. Hemorrhage, choroidal neovascular membrane, and AIONare more likely to occur in patients with optic disc drusen. No treatment for drusen is available.

Vitreous Degeneration This occurs in all individuals with advancing age, leading to chronic and acute visual symptoms. Opacities develop in the vitreous, casting annoying shadows upon the retina. As the eye moves, these distracting "floaters" move synchronously, with a slight lag caused by inertia of the vitreous gel. Vitreous traction upon the retina causes mechanical stimulation, resulting in perception of flashing lights. This photopsia is brief and confined to one eye, in contrast to the bilateral, prolonged scintillations of cortical migraine. Contraction of the vitreous can result in sudden separation from the retina, heralded by an alarming shower of floaters and photopsia. This process, known as *vitreous detachment*, is a frequent involutional event in the elderly. It is not harmful unless it damages the retina. A careful examination of the dilated fundus is mandatory in any patient complaining of floaters or photopsia to search for peripheral tears or holes. If such a lesion is found, laser application or cryotherapy can forestall a retinal detachment. Occasionally a tear ruptures a retinal blood vessel, causing vitreous hemorrhage and sudden loss of vision. On attempted ophthalmoscopy the fundus is hidden by a dark red haze of blood. Ultrasound is required to examine the interior of the eye for a retinal tear or detachment. If the hemorrhage does not resolve spontaneously, the vitreous can be removed surgically. Vitreous hemorrhage also occurs from the fragile neovascular vessels that proliferate on the surface of the retina in diabetes, sickle cell anemia, and other ischemic ocular diseases.

Retinal Detachment This produces symptoms of floaters, flashing lights, and a scotoma in the peripheral visual field corresponding to the detachment (seePlate IV-12). If the detachment includes the fovea, there is an afferent pupil defect and the visual acuity is reduced. In most eyes, retinal detachment starts with a hole, flap, or tear in the peripheral retina (rhegmatogenous retinal detachment). Patients with peripheral retinal thinning (lattice degeneration) are particularly vulnerable to this process. Once a break has developed in the retina, liquified vitreous is free to enter the subretinal space, separating the retina from the pigment epithelium. The combination of vitreous traction upon the retinal surface and passage of fluid behind the retina leads inexorably to detachment. Patients with a history of myopia, trauma, or prior cataract extraction are at greatest risk for retinal detachment. The diagnosis is confirmed by ophthalmoscopic examination of the dilated eye.

Classic Migraine (See alsoChap. 15) This usually occurs with a visual aura lasting about 20 min. In a typical attack, a small central disturbance in the field of vision marches toward the periphery, leaving a transient scotoma in its wake. The expanding border of migraine scotoma has a scintillating, dancing, or zig-zag edge, resembling the bastions of a fortified city, hence the term "fortification spectra." Patients' descriptions of fortification spectra vary widely and can be confused with amaurosis fugax. Migraine

patterns usually last longer and are perceived in both eyes, whereas amaurosis fugax is briefer and occurs in only one eye. Migraine phenomena also remain visible in the dark or with the eyes closed. Generally they are confined to either the right or left visual hemifield, but sometimes both fields are involved simultaneously. Patients often have a long history of stereotypic attacks. After the visual symptoms recede, headache develops in most patients.

Transient ischemic attacks from *vertebrobasilar insufficiency* result in acute homonymous visual symptoms. Many patients mistakenly describe symptoms in their left or right eye, when in fact they are occurring in the left or right hemifield of both eyes. Interruption of blood supply to the visual cortex causes a sudden fogging or graying of vision, occasionally with flashing lights or other positive phenomena that mimic migraine. Cortical ischemic attacks are briefer in duration than migraine, occur in older patients, and are not followed by headache. There may be associated signs of brainstem ischemia, such as diplopia, vertigo, numbness, weakness, or dysarthria.

Stroke This occurs when interruption of blood supply from the posterior cerebral artery to the visual cortex is prolonged. The only finding on examination is a homonymous visual field defect that stops abruptly at the vertical meridian. Occipital lobe stroke is usually due to thrombotic occlusion of the vertebrobasilar system, embolus, or dissection. Lobar hemorrhage, tumor, abscess, and arteriovenous malformation are other common causes of hemianopic cortical visual loss.

Factitious (Functional, Nonorganic) Visual Loss This is claimed by hysterics or malingerers. The latter comprise the vast majority, seeking sympathy, special treatment, or financial gain by feigning loss of sight. The diagnosis is suspected when the history is atypical, physical findings are lacking or contradictory, inconsistencies emerge on testing, and a secondary motive can be identified. In our litigious society, the fraudulent pursuit of recompense has spawned an epidemic of factitious visual loss.

CHRONIC VISUAL LOSS

Cataract This is a clouding of the lens sufficient to reduce vision. Most cataracts develop slowly as a result of aging, leading to gradual impairment of vision. The formation of cataract occurs more rapidly in patients with a history of ocular trauma, uveitis, or diabetes mellitus. Cataracts are acquired in a variety of genetic diseases, such as myotonic dystrophy, neurofibromatosis type 2, and galactosemia. Radiation therapy and glucocorticoid treatment can induce cataract as a side effect. The cataracts associated with radiation or glucocorticoids have a typical posterior subcapsular location. Cataract can be detected by noting an impaired red reflex when viewing light reflected from the fundus with an ophthalmoscope or by examining the dilated eye using the slit lamp.

The only treatment for cataract is surgical extraction of the opacified lens. Over a million cataract operations are performed each year in the United States. The operation is generally done under local anesthesia on an outpatient basis. Remarkable technical innovations have made it possible to aspirate the cataract while leaving the lens capsule intact (extracapsular cataract extraction), rather than removing the entire lens with its capsule (intracapsular cataract extraction). A plastic or silicone intraocular lens is then

placed within the empty lens capsule in the posterior chamber, substituting for the natural lens, and leading to rapid recovery of sight. More than 95% of patients who undergo cataract extraction can expect an improvement in vision. In many patients, the lens capsule remaining in the eye after cataract extraction eventually turns cloudy, causing a secondary loss of vision. A small opening is made in the lens capsule with a laser to restore clarity.

Glaucoma This is a slowly progressive, insidious optic neuropathy, usually associated with chronic elevation of intraocular pressure. In Americans of African descent it is the leading cause of blindness. The mechanism whereby raised intraocular pressure injures the optic nerve is not understood. Axons entering the inferotemporal and superotemporal aspects of the optic disc are damaged first, producing typical nerve fiber bundle or arcuate scotomas on perimetric testing. As fibers are destroyed, the neural rim of the optic disc shrinks and the physiologic cup within the optic disc enlarges (seePlate IV-13). This process is referred to colloquially as pathologic "cupping." The cup-to-disc diameter is expressed as a ratio, e.g., 0.2/1. The cup-to-disc ratio ranges widely in normal individuals, making it difficult to diagnose glaucoma reliably simply by observing an unusually large or deep optic cup. Careful documentation of serial prospective examinations is helpful. In the patient with physiologic cupping, the large cup remains stable, whereas in the patient with glaucoma it expands relentlessly over the years. Detection of visual field loss on formal perimetry also contributes to the diagnosis of glaucoma. Finally, most patients with glaucoma have raised intraocular pressure. However, a surprising number of patients with typical glaucomatous cupping and visual field loss have intraocular pressures that apparently never exceed the normal limit of 20 mmHg (so-called low-tension glaucoma).

In acute angle-closure glaucoma, the eye is red and painful due to abrupt, severe elevation of intraocular pressure. Such cases account for only a handful of patients with glaucoma. Most patients with glaucoma have open, nonoccludable anterior chamber angles. The cause of raised intraocular pressure in these patients is uncertain. Recent studies have implicated mutations in a gene encoding a glycoprotein expressed in the trabecular meshwork. This structure serves as a filter to drain aqueous from the eye. Because the elevation of intraocular pressure develops gradually and is less marked than in angle-closure glaucoma, there is no pain or ocular injection. The central visual field and foveal acuity are spared until end-stage disease is reached. For these reasons, severe and irreversible damage can occur before either the patient or physician recognizes the diagnosis. Screening of patients for glaucoma by noting the cup-to-disc ratio on ophthalmoscopy and by measuring intraocular pressure (using a Schiotz, Tonopen, air-puff, or Goldmann tonometer) is vital. Glaucoma is treated with topical adrenergic agonists (epinephrine, dipivefrin, apraclonidine, brimonidine), cholinergic agonists (pilocarpine), beta blockers (betaxolol, carteolol, levobunolol, metipranolol, and timolol), and prostaglandin analogues (latanaprost). Occasionally, systemic absorption of beta blocker from eye drops can be sufficient to cause side effects of bradycardia, hypotension, heart block, bronchospasm, impotence, or depression. Topical or oral carbonic anhydrase inhibitors are used to lower intraocular pressure by reducing aqueous production. Laser treatment of the trabecular meshwork in the anterior chamber angle improves aqueous outflow from the eye. If medical or laser treatments fail to halt optic nerve damage from glaucoma, a filter must be constructed surgically (trabeculectomy) to release aqueous from the eye in a controlled fashion.

Macular Degeneration This is a major cause of gradual, painless, bilateral central visual loss in the elderly. The old term, "senile macular degeneration," misinterpreted by many patients as an unflattering reference, has been replaced with "age-related macular degeneration." It occurs in a nonexudative (dry) form and an exudative (wet) form. The nonexudative process begins with the accumulation of extracellular deposits, called drusen, underneath the retinal pigment epithelium. On ophthalmoscopy, they are pleomorphic but generally appear as small discrete yellow lesions clustered in the macula (seePlate IV-14). With time they become larger, more numerous, and confluent. The retinal pigment epithelium becomes focally detached and atrophic, causing visual loss by interfering with photoreceptor function. There is currently no way to prevent the development of age-related macular degeneration. Concoctions of various vitamins (A, C, and E) and minerals (zinc, copper, and selenium) have been marketed, without good evidence that they retard the process of macular degeneration.

Exudative macular degeneration, which develops in only a minority of patients, occurs when neovascular vessels from the choroid grow through defects in Bruch's membrane into the potential space beneath the retinal pigment epithelium. Leakage from these vessels produces elevation of the retina and pigment epithelium, with distortion (metamorphopsia) and blurring of vision. Although onset of these symptoms is usually gradual, bleeding from subretinal choroidal neovascular membranes sometimes causes acute visual loss. The neovascular membranes can be difficult to see on fundus examination because they are beneath the retina. Fluorescein or indocyanine green angiography is extremely useful for their detection. In some patients, prompt laser ablation of choroidal neovascular membranes seen on fluorescein angiography can halt the exudative process. However, the neovascular membranes frequently recur, requiring constant vigilance and repeated photocoagulation.

Major or repeated hemorrhage under the retina from neovascular membranes results in fibrosis, development of a round (disciform) macular scar, and permanent loss of central vision. Surgical attempts to remove subretinal membranes in age-related macular degeneration have not improved vision in most patients. However, outcomes have been more encouraging for patients with choroidal neovascular membranes from ocular histoplasmosis syndrome.

Central Serous Chorioretinopathy This primarily affects males between the ages of 20 and 50. Leakage of serous fluid from the choroid causes small, localized detachment of the retinal pigment epithelium and the neurosensory retina. These detachments produce acute or chronic symptoms of metamorphopsia and blurred vision when the macula is involved. They are difficult to visualize with a direct ophthalmoscope because the detached retina is transparent and only slightly elevated. Diagnosis of central serous chorioretinopathy is made easily by fluorescein angiography, which shows dye streaming into the subretinal space. The cause of central serous chorioretinopathy is unknown. Symptoms may resolve spontaneously if the retina reattaches, but recurrent detachment is common. Laser photocoagulation has benefited some patients with this condition.

Diabetic Retinopathy A rare disease until 1921, when the discovery of insulin resulted in a dramatic improvement in life expectancy for patients with diabetes mellitus, it is now

a leading cause of blindness in the United States. The retinopathy of diabetes takes years to develop but eventually appears in nearly all cases. Regular surveillance of the dilated fundus is crucial for any patient with diabetes. In advanced diabetic retinopathy, the proliferation of neovascular vessels leads to blindness from vitreous hemorrhage, retinal detachment, and glaucoma (seePlate IV-15). These complications can be avoided in most patients by administration of panretinal laser photocoagulation at the appropriate point in the evolution of the disease.**For further discussion of the manifestations and management of diabetic retinopathy, see Chap. 333.*

Retinitis Pigmentosa This is a general term for a disparate group of rod and cone dystrophies characterized by progressive night blindness (nyctalopia), visual field constriction with a ring scotoma, loss of acuity, and an abnormal electroretinogram (ERG). It occurs sporadically or in an autosomal recessive, dominant, or X-linked pattern. Irregular black deposits of clumped pigment in the peripheral retina, called bone spicules because of their vague resemblance to the spicules of cancellous bone, give the disease its name (seePlate IV-16). The name is actually a misnomer because retinitis pigmentosa is not an inflammatory process. Most cases are due to a mutation in the gene for rhodopsin, the rod photopigment, or in the gene for peripherin, a glycoprotein located in photoreceptor outer segments. There is no effective treatment for retinitis pigmentosa. Vitamin A (15,000 IU/day) slightly retards the deterioration of the ERG but has no beneficial effect upon visual acuity or visual fields. Some forms of retinitis pigmentosa occur in association with rare, hereditary systemic diseases (olivopontocerebellar degeneration, Bassen-Kornzweig disease, Kearns-Sayre syndrome, Refsum's disease). Chronic treatment with chloroquine, hydroxychloroquine, and phenothiazines (especially thioridazine) can produce visual loss from a toxic retinopathy that resembles retinitis pigmentosa.

Epiretinal Membrane This is a fibrocellular tissue that grows across the inner surface of the retina, causing metamorphopsia and reduced visual acuity from distortion of the macula. With the ophthalmoscope one can see a crinkled, cellophane-like membrane on the retina. Epiretinal membrane is most common in patients over 50 years of age and is usually unilateral. Most cases are idiopathic, but some occur as a result of hypertensive retinopathy, diabetes, retinal detachment, or trauma. When visual acuity is reduced to the level of about 6/24 (20/80), vitrectomy and surgical peeling of the membrane to relieve macular puckering are recommended. Contraction of an epiretinal membrane sometimes gives rise to a *macular hole*. Most macular holes, however, are caused by local vitreous traction within the fovea. Vision is usually depressed to the level of 6/30 (20/100) or worse. Vitrectomy may improve visual acuity in some patients with macular hole. Fortunately, fewer than 10% of patients with a macular hole develop a hole in their other eye.

Melanoma and Other Tumors Melanoma is the most common primary tumor of the eye (seePlate IV-17). It causes photopsia, an enlarging scotoma, and loss of vision. A small melanoma is often difficult to differentiate from a benign choroidal nevus. Careful serial examinations are required to document a malignant pattern of growth. Treatment of melanoma is controversial. Options include enucleation, local resection, and irradiation. *Metastatic tumors* to the eye outnumber primary tumors of uveal origin. Breast and lung carcinoma have a special propensity to spread to the choroid or iris. Leukemia and lymphoma also commonly invade ocular tissues. Sometimes their only

sign on eye examination is cellular debris in the vitreous, which can masquerade as a chronic posterior uveitis. *Retrobulbar tumor* of the optic nerve (meningioma, glioma) or *chiasmal tumor* (pituitary adenoma, meningioma) produces gradual visual loss with few objective findings, except for optic disc pallor. Rarely, sudden expansion of a pituitary adenoma from infarction and bleeding (*pituitary apoplexy*) causes acute retrobulbar visual loss, with headache, nausea, and ocular motor nerve palsies. In any patient with visual field loss or optic atrophy,CT orMRscanning should be considered if the cause remains unknown after careful review of the history and thorough examination of the eye (Fig. 28-4).

PROPTOSIS

When the globes appear asymmetric, the clinician must first decide which eye is abnormal. Is one eye recessed within the orbit (*enophthalmos*) or is the other eye protuberant (*exophthalmos*, or *proptosis*)? A small globe or a Horner's syndrome can give the appearance of enophthalmos. True enophthalmos occurs commonly after trauma, from atrophy of retrobulbar fat, or fracture of the orbital floor. The position of the eyes within the orbits is measured using a Hertel exophthalmometer, a hand-held instrument that records the position of the anterior corneal surface relative to the lateral orbital rim. If this instrument is not available, relative eye position can be judged by bending the patient's head forward and looking down upon the orbits. A proptosis of only 2 mm in one eye is detectable from this perspective. The development of proptosis implies a space-occupying lesion in the orbit. ACT orMR scan should be obtained in any patient with proptosis, unless the diagnosis of Graves' ophthalmopathy is certain.

Graves' Ophthalmopathy This is the leading cause of proptosis in adults (Chap. 330). The proptosis is often asymmetric and can even appear to be unilateral. Orbital inflammation and engorgement of the extraocular muscles, particularly the medial rectus and the inferior rectus, account for the protrusion of the globe. Corneal exposure, lid retraction, conjunctival injection, restriction of gaze, diplopia, and visual loss from optic nerve compression are cardinal symptoms. Acute Graves' ophthalmopathy should be treated with oral prednisone (60 mg/day) for 1 month, followed by a taper over several months. Chronic manifestations can be managed by topical lubricants, eyelid surgery, eye muscle surgery, or radiation treatment. Optic nerve compression should be relieved promptly with glucocorticoids and orbital decompression to prevent permanent visual loss.

Orbital Pseudotumor This is an idiopathic, inflammatory orbital syndrome, frequently confused with Graves' ophthalmopathy. Symptoms are pain, limited eye movements, proptosis, and congestion. Evaluation for sarcoidosis, Wegener's granulomatosis, and other types of orbital vasculitis or collagen-vascular disease is negative. Imaging often shows swollen eye muscles (orbital myositis) with enlarged tendons. By contrast, in Graves' ophthalmopathy the tendons of the eye muscles are usually spared. The Tolosa-Hunt syndrome may be regarded as an extension of orbital pseudotumor through the superior orbital fissure into the cavernous sinus. The diagnosis of orbital pseudotumor is difficult. Biopsy of the orbit frequently yields nonspecific evidence of fat infiltration by lymphocytes, plasma cells, and eosinophils. A dramatic response to a therapeutic trial of systemic glucocorticoids indirectly provides the best confirmation of the diagnosis.

Orbital Cellulitis This causes pain, lid erythema, proptosis, conjunctival chemosis, restricted motility, decreased acuity, afferent pupillary defect, fever, and leukocytosis. It often arises from a paranasal sinus, especially by contiguous spread of infection from the ethmoid sinus through the thin lamina papyracea of the medial orbit. A history of recent upper respiratory tract infection, chronic sinusitis, thick mucous secretions, or dental disease is significant in any patient with suspected orbital cellulitis. Blood cultures should be obtained, but they are usually negative. Most patients respond to empiric therapy with broad-spectrum intravenous antibiotics. Occasionally, orbital cellulitis follows an overwhelming course, with massive proptosis, blindness, septic cavernous sinus thrombosis, and meningitis. To avert this disaster, orbital cellulitis should be managed aggressively in the early stages, with immediate antibiotic therapy and imaging of the orbits. Prompt surgical drainage of an orbital abscess or paranasal sinusitis is indicated if optic nerve function deteriorates despite antibiotics.

Tumors Tumors of the orbit cause painless, progressive proptosis. The most common primary tumors are hemangioma, lymphangioma, neurofibroma, dermoid cyst, adenoid cystic carcinoma, optic nerve glioma, optic nerve meningioma, and benign mixed tumor of the lacrimal gland. Metastatic tumor to the orbit occurs frequently in breast carcinoma, lung carcinoma, and lymphoma. Diagnosis by fine-needle aspiration followed by urgent radiation therapy can sometimes preserve vision.

Carotid Cavernous Fistulas With anterior drainage through the orbit these produce proptosis, diplopia, glaucoma, and tortuous, red conjunctival vessels. Direct fistulas usually result from trauma. They are easily diagnosed because of the dramatic signs produced by high-flow, high-pressure shunting. Indirect fistulas, or dural arteriovenous malformations, are more likely to occur spontaneously, especially in older women. The signs are more subtle and the diagnosis is frequently missed. The combination of slight proptosis, diplopia, enlarged muscles, and an injected eye is often mistaken for thyroid ophthalmopathy. A bruit heard upon auscultation of the head, or reported by the patient, is a valuable diagnostic clue. Imaging shows an enlarged superior ophthalmic vein in the orbits. Carotid cavernous shunts can be eliminated by intravascular embolization.

PTOSIS

Blepharoptosis This is an abnormal drooping of the eyelid. Unilateral or bilateral ptosis can be congenital, from dysgenesis of the levator palpebrae superioris, or from abnormal insertion of its aponeurosis into the eyelid. Acquired ptosis can develop so gradually that the patient is unaware of the problem. Inspection of old photographs is helpful in dating the onset. A history of prior trauma, eye surgery, contact lens use, diplopia, systemic symptoms (e.g., dysphagia or peripheral muscle weakness), or a family history of ptosis should be sought. Fluctuating ptosis that worsens late in the day is typical of myasthenia gravis. Examination should focus upon evidence for proptosis, eyelid masses or deformities, inflammation, pupil inequality, or limitation of motility. The width of the palpebral fissures is measured in primary gaze to quantitate the degree of ptosis. The ptosis will be underestimated if the patient is compensating by lifting the brow with the frontalis muscle.

Mechanical Ptosis This occurs in many elderly patients from stretching and

redundancy of eyelid skin and subcutaneous fat (dermatochalasis). The extra weight of these sagging tissues causes the lid to droop. Enlargement or deformation of the eyelid from infection, tumor, trauma, or inflammation also results in ptosis on a purely mechanical basis.

Aponeurotic Ptosis This is an acquired dehiscence or stretching of the aponeurotic tendon, which connects the levator muscle to the tarsal plate of the eyelid. It occurs commonly in older patients, presumably from loss of connective tissue elasticity. Aponeurotic ptosis is also a frequent sequela of eyelid swelling from infection or blunt trauma to the orbit, cataract surgery, or hard contact lens usage.

Myogenic Ptosis The causes of *myogenic ptosis* include myasthenia gravis (Chap. 380) and a number of rare myopathies that manifest with ptosis. The term *chronic progressive external ophthalmoplegia* refers to a spectrum of systemic diseases caused by mutations of mitochrondrial DNA. As the name implies, the most prominent findings are symmetric, slowly progressive ptosis and limitation of eye movements. In general, diplopia is a late symptom because all eye movements are reduced equally. In the *Kearns-Sayre* variant, retinal pigmentary changes and abnormalities of cardiac conduction develop. Peripheral muscle biopsy shows characteristic "ragged-red fibers." *Oculopharyngeal dystrophy* is a distinct autosomal dominant disease with onset in middle age, characterized by ptosis, limited eye movements, and trouble swallowing. *Myotonic dystrophy*, another autosomal dominant disorder, causes ptosis, ophthalmoparesis, cataract, and pigmentary retinopathy. Patients have muscle wasting, myotonia, frontal balding, and cardiac abnormalities.

Neurogenic Ptosis This results from a lesion affecting the innervation to either of the two muscles that open the eyelid: Muller's muscle or the levator palpebrae superioris. Examination of the pupil helps to distinguish between these two possibilities. In Horner's syndrome, the eye with ptosis has a smaller pupil and the eye movements are full. In an oculomotor nerve palsy, the eye with the ptosis has a larger, or a normal, pupil. If the pupil is normal but there is limitation of adduction, elevation, and depression, a pupil-sparing oculomotor nerve palsy is likely (see next section). Rarely, a lesion affecting the small, central subnucleus of the oculomotor complex will cause bilateral ptosis with normal eye movements and pupils.

DOUBLE VISION

The first point to clarify is whether diplopia persists in either eye after covering the fellow eye. If it does, the diagnosis is monocular diplopia. The cause is usually intrinsic to the eye and therefore has no dire implications for the patient. Corneal aberrations (e.g., keratoconus, pterygium), uncorrected refractive error, cataract, or foveal traction may give rise to monocular diplopia. Occasionally it is a symptom of malingering or psychiatric disease. Diplopia alleviated by covering one eye is binocular diplopia and is caused by disruption of ocular alignment. Inquiry should be made into the nature of the double vision (purely side-by-side versus partial vertical displacement of images), mode of onset, duration, intermittency, diurnal variation, and associated neurologic or systemic symptoms. If the patient has diplopia while being examined, motility testing should reveal a deficiency corresponding to the patient's symptoms. However, subtle limitation of ocular excursions is often difficult to detect. For example, a patient with a

slight left abducens nerve paresis may appear to have full eye movements, despite a complaint of horizontal diplopia upon looking to the left. In this situation, the cover test provides a more sensitive method for demonstrating the ocular malalignment. It should be conducted in primary gaze, and then with the head turned and tilted in each direction. In the above example, a cover test with the head turned to the right will maximize the fixation shift evoked by the cover test.

Occasionally, a cover test performed in an asymptomatic patient during a routine examination will reveal an ocular deviation. If the eye movements are full and the ocular misalignment is equal in all directions of gaze (concomitant deviation), the diagnosis is strabismus. In this condition, which affects about 1% of the population, fusion is disrupted in infancy or early childhood. To avoid diplopia, vision is suppressed from the nonfixating eye. In some children, this leads to impaired vision (amblyopia, or "lazy" eye) in the deviated eye.

Binocular diplopia occurs from a wide range of processes: infectious, neoplastic, metabolic, degenerative, inflammatory, and vascular. One must decide if the diplopia is neurogenic in origin or due to restriction of globe rotation by local disease in the orbit. Orbital pseudotumor, myositis, infection, tumor, thyroid disease, and muscle entrapment (e.g., from a blowout fracture) cause restrictive diplopia. The diagnosis is confirmed by performing a forced duction test in the office. After applying topical anesthesia, the physician grasps the eye with forceps and pulls it toward the direction of deficient motion. If rotation of the globe is prevented by tethering, a restrictive process is at work. The utility of this test is limited by its unpopularity with patients; in practice, the diagnosis of restriction is made by recognizing other associated signs and symptoms of local orbital disease.

Myasthenia Gravis (See also Chap. 380) This is a major cause of diplopia. The diplopia is often intermittent, variable, and not confined to any single ocular motor nerve distribution. The pupils are always normal. Fluctuating ptosis may be present. Many patients have a purely ocular form of the disease, with no evidence of systemic muscular weakness. The diagnosis can be confirmed by an intravenous edrophonium injection or by an assay for antiacetylcholine receptor antibodies. Negative results from these tests do not exclude the diagnosis. *Botulism* from food or wound poisoning can mimic ocular myasthenia.

After restrictive orbital disease and myasthenia gravis are excluded, a lesion of a cranial nerve supplying innervation to the extraocular muscles is the most likely cause of binocular diplopia.

Oculomotor Nerve (Video 28-1) The third cranial nerve innervates the medial, inferior, and superior recti; inferior oblique; levator palpebrae superioris; and the iris sphincter. Total palsy of the oculomotor nerve causes ptosis, a dilated pupil, and leaves the eye "down and out" because of the unopposed action of the lateral rectus and superior oblique. This combination of findings is obvious. More challenging is the diagnosis of an early or partial oculomotor nerve palsy. In this setting, any combination of ptosis, pupil dilation, and weakness of the eye muscles supplied by the oculomotor nerve may be encountered. Frequent serial examinations during the evolving phase of the palsy and a high index of suspicion help ensure that the diagnosis is not missed. The advent of an

oculomotor nerve palsy with any degree of pupil involvement in an otherwise healthy patient, especially when accompanied by pain, raises the specter of a circle of Willis aneurysm. If anMR scan shows no compressive lesion, an arteriogram must be performed to rule out an aneurysm of either the posterior communicating artery or the basilar artery. If the pupil is entirely normal, with all other components of an oculomotor palsy present, aneurysm is so rare that an angiogram is seldom indicated.

A lesion of the oculomotor nucleus in the rostral midbrain produces signs that differ from those caused by a lesion of the nerve itself. There is bilateral ptosis because the levator muscle is innervated by a single central subnucleus. There is also weakness of the contralateral superior rectus, because it is supplied by the oculomotor nucleus on the other side. Occasionally both superior recti are weak. Isolated nuclear oculomotor palsy is quite rare. Usually neurologic examination reveals additional signs to suggest brainstem damage from infarction, hemorrhage, tumor, or infection.

Injury to structures surrounding fascicles of the oculomotor nerve descending through the midbrain has given rise to a number of classic eponymic designations. In *Nothnagel's syndrome*, injury to the superior cerebellar peduncle causes ipsilateral oculomotor palsy and contralateral cerebellar ataxia. In *Benedikt's syndrome*, injury to the red nucleus results in ipsilateral oculomotor palsy and contralateral tremor, chorea, and athetosis. *Claude's syndrome* incorporates features of both the aforementioned syndromes, by injury to both the red nucleus and the superior cerebellar peduncle. Finally, in *Weber's syndrome*, injury to the cerebral peduncle causes ipsilateral oculomotor palsy with contralateral hemiparesis.

In the subarachnoid space the oculomotor nerve is vulnerable to aneurysm, meningitis, tumor, infarction, and compression. In cerebral herniation the nerve becomes trapped between the edge of the tentorium and the uncus of the temporal lobe. Oculomotor palsy can also occur from midbrain torsion and hemorrhages during herniation. In the cavernous sinus, oculomotor palsy arises from carotid aneurysm, carotid cavernous fistula, cavernous sinus thrombosis, tumor (pituitary adenoma, meningioma, metastasis), herpes zoster infection, and the Tolosa-Hunt syndrome.

The etiology of an isolated, pupil-sparing oculomotor palsy often remains obscure, even after neuroimaging and extensive laboratory testing. Most cases are thought to result from microvascular infarction of the nerve, somewhere along its course from the brainstem to the orbit. Usually the patient complains of pain. Diabetes, hypertension, and vascular disease are major risk factors. Spontaneous recovery over a period of months is the rule. If this fails to occur, or if new findings develop, the diagnosis of microvascular oculomotor nerve palsy should be reconsidered. Aberrant regeneration is common when the oculomotor nerve is injured by trauma or compression (tumor, aneurysm). Miswiring of sprouting fibers to the levator muscle and the rectus muscles results in elevation of the eyelid upon downgaze or adduction. The pupil also constricts upon attempted adduction, elevation, or depression of the globe. Aberrant regeneration is not seen after oculomotor palsy from microvascular infarct and hence vitiates that diagnosis.

Trochlear Nerve The fourth cranial nerve originates in the midbrain, just caudal to the oculomotor nerve complex. Fibers exit the brainstem dorsally and cross to innervate the

contralateral superior oblique. The principal actions of this muscle are to depress and to intort the globe. A palsy therefore results in hypertropia and excyclotorsion. The cyclotorsion is seldom noticed by patients. Instead, they complain of vertical diplopia, especially upon reading or looking down. The vertical diplopia is also exacerbated by tilting the head toward the side with the muscle palsy, and alleviated by tilting it away. This "head tilt test" is a cardinal diagnostic feature.

Isolated trochlear nerve palsy occurs from all the causes listed above for the oculomotor nerve, except aneurysm. The trochlear nerve is particularly apt to suffer injury after closed head trauma. The mechanism is unknown, but the free edge of the tentorium may impinge upon the nerve during a concussive blow. Most isolated trochlear nerve palsies are idiopathic and hence diagnosed by exclusion as "microvascular." Spontaneous improvement occurs over a period of months in most patients. A base-down prism (conveniently applied to the patient's glasses as a stick-on Fresnel lens) may serve as a temporary measure to alleviate diplopia. If the palsy does not resolve, the eyes can be realigned by surgically adjusting other eye muscles.

Abducens Nerve (Video 28-2) The sixth cranial nerve innervates the lateral rectus muscle. A palsy produces horizontal diplopia, worse on gaze to the side of the lesion. A nuclear lesion has different consequences, because the abducens nucleus contains interneurons that project via the medial longitudinal fasciculus to the medial rectus subnucleus of the contralateral oculomotor complex. Therefore, an abducens nuclear lesion produces a complete lateral gaze palsy, from weakness of both the ipsilateral lateral rectus and the contralateral medial rectus. *Foville's syndrome* following dorsal pontine injury includes lateral gaze palsy, ipsilateral facial palsy, and contralateral hemiparesis incurred by damage to descending corticospinal fibers. *Millard-Gubler syndrome* from ventral pontine injury is similar, except for the eye findings. There is lateral rectus weakness only, instead of gaze palsy, because the abducens fascicle is injured rather than the nucleus. Infarct, tumor, hemorrhage, vascular malformation, and multiple sclerosis are the most common etiologies of brainstem abducens palsy.

After leaving the ventral pons, the abducens nerve runs forward along the clivus to pierce the dura at the petrous apex, where it enters the cavernous sinus. Along its subarachnoid course it is susceptible to meningitis, tumor (meningioma, chordoma, carcinomatous meningitis), subarachnoid hemorrhage, trauma, and compression by aneurysm or dolichoectatic vessels. At the petrous apex, mastoiditis can produce deafness, pain, and ipsilateral abducens palsy (*Gradenigo's syndrome*). In the cavernous sinus, the nerve can be affected by carotid aneurysm, carotid cavernous fistula, tumor (pituitary adenoma, meningioma, nasopharyngeal carcinoma), herpes infection, and Tolosa-Hunt syndrome.

Unilateral or bilateral abducens palsy is a classic sign of raised intracranial pressure. The diagnosis can be confirmed if papilledema is observed on fundus examination. The mechanism is still debated but is probably related to rostral-caudal displacement of the brainstem. The same phenomenon accounts for abducens palsy from low intracranial pressure (e.g., after lumbar puncture, spinal anesthesia, or spontaneous dural cerebrospinal fluid leak).

Treatment of abducens palsy is aimed at prompt correction of the underlying cause.

However, the cause remains obscure in many instances, despite diligent evaluation. As mentioned above for isolated trochlear or oculomotor palsy, most cases are assumed to represent microvascular infarcts because they often occur in the setting of diabetes or other vascular risk factors. Some cases may develop as a postinfectious mononeuritis (e.g., following a viral flu). Patching one eye or applying a temporary prism will provide relief of diplopia until the palsy resolves. If recovery is incomplete, eye muscle surgery can nearly always realign the eyes, at least in primary position. A patient with an abducens palsy that fails to improve should be reevaluated for an occult etiology (e.g., chordoma, carcinomatous meningitis, carotid cavernous fistula, myasthenia gravis).

Multiple Ocular Motor Nerve Palsies These should not be attributed to spontaneous microvascular events affecting more than one cranial nerve at a time. This remarkable coincidence does occur, especially in diabetic patients, but the diagnosis is made only in retrospect after exhausting all other diagnostic alternatives. Neuroimaging should focus on the cavernous sinus, superior orbital fissure, and orbital apex, where all three ocular motor nerves are in close proximity. In the diabetic or compromised host, fungal infection (*Aspergillus*, Mucorales, *Cryptococcus*) is a frequent cause of multiple nerve palsies. In the patient with systemic malignancy, carcinomatous meningitis is a likely diagnosis. Cytologic examination may be negative despite repeated sampling of the cerebrospinal fluid. The cancer-associated Lambert-Eaton myasthenic syndrome can also produce ophthalmoplegia. Giant cell (temporal) arteritis occasionally manifests as diplopia from ischemic palsies of extraocular muscles (Figs. 28-CD1 and28-CD2). Fisher syndrome, an ocular variant of Guillain-Barre, can produce ophthalmoplegia with areflexia and ataxia. Often the ataxia is mild, and the areflexia is overlooked because the physician's attention is focused upon the eyes.

Supranuclear Disorders of Gaze These are often mistaken for multiple ocular motor nerve palsies. For example, Wernicke's encephalopathy can produce nystagmus and a partial deficit of horizontal and vertical gaze that mimics a combined abducens and oculomotor nerve palsy. The disorder occurs in malnourished or alcoholic patients and can be reversed by giving thiamine. Infarct, hemorrhage, tumor, multiple sclerosis, encephalitis, vasculitis, and Whipple's disease are other important causes of supranuclear gaze palsy.

The *frontal eye field* of the cerebral cortex is involved in generation of saccades to the contralateral side. After hemispheric stroke, the eyes usually deviate towards the lesioned side because of the unopposed action of the frontal eye field in the normal hemisphere. With time, this deficit resolves. Seizures generally have the opposite effect: the eyes deviate conjugately away from the irritative focus. *Parietal lesions* disrupt smooth pursuit of targets moving toward the side of the lesion. Bilateral parietal lesions produce *Balint's syndrome*, characterized by impaired eye-hand coordination (optic ataxia), difficulty initiating voluntary eye movements (ocular apraxia), and visuospatial disorientation (simultanagnosia).

Horizontal Gaze Descending cortical inputs mediating horizontal gaze ultimately converge at the level of the pons. Neurons in the paramedian pontine reticular formation are responsible for controlling conjugate gaze toward the same side. They project directly to the ipsilateral abducens nucleus. A lesion of either the paramedian pontine reticular formation or the abducens nucleus causes an ipsilateral conjugate gaze palsy.

Lesions at either locus produce nearly identical clinical syndromes, with the following exception: vestibular stimulation (oculocephalic maneuver or caloric) will succeed in driving the eyes conjugately to the side in a patient with a lesion of the paramedian pontine reticular formation, but not in a patient with a lesion of the abducens nucleus.

Internuclear Ophthalmoplegia This results from damage to the medial longitudinal fasciculus ascending from the abducens nucleus in the pons to the oculomotor nucleus in the midbrain (hence, "internuclear"). Damage to fibers carrying the conjugate signal from abducens interneurons to the contralateral medial rectus motoneurons results in a failure of adduction on attempted lateral gaze. For example, a patient with a left internuclear ophthalmoplegia will have slowed or absent adducting movements of the left eye. A patient with bilateral injury to the medial longitudinal fasciculus will have bilateral internuclear ophthalmoplegia. Multiple sclerosis is the most common cause, although tumor, stroke, trauma, or any brainstem process may be responsible. *One-and-a half syndrome* is due to a combined lesion of the medial longitudinal fasciculus and the abducens nucleus on the same side. The patient's only horizontal eye movement is abduction of the eye on the other side.

Vertical Gaze This is controlled at the level of the midbrain. The neuronal circuits affected in disorders of vertical gaze are not well elucidated, but lesions of the rostral interstitial nucleus of the medial longitudinal fasciculus and the interstitial nucleus of Cajal cause supranuclear paresis of upgaze, downgaze, or all vertical eye movements. Distal basilar artery ischemia is the most common etiology. *Skew deviation* refers to a vertical misalignment of the eyes, usually constant in all positions of gaze. The finding has poor localizing value because skew deviation has been reported after lesions in widespread regions of the brainstem and cerebellum.

Parinaud's Syndrome Also known as dorsal midbrain syndrome, this is a distinct supranuclear vertical gaze disorder from damage to the posterior commissure. It is a classic sign of hydrocephalus from aqueductal stenosis. Pineal region tumors (germinoma, pineoblastoma), cysticercosis, and stroke also cause Parinaud's syndrome. Features include loss of upgaze (and sometimes downgaze), convergence-retraction nystagmus on attempted upgaze, downwards ocular deviation ("setting sun" sign), lid retraction (Collier's sign), skew deviation, pseudoabducens palsy, and light-near dissociation of the pupils. Disorders of vertical gaze, especially downwards saccades, are an early feature of progressive supranuclear palsy. Smooth pursuit is affected later in the course of the disease. Parkinson's disease, Huntington's chorea, and olivopontocerebellar degeneration can also affect vertical gaze.

Nystagmus This is a rhythmical oscillation of the eyes, occurring physiologically from vestibular and optokinetic stimulation or pathologically in a wide variety of diseases. Abnormalities of the eyes or optic nerves, present at birth or acquired in childhood, can produce a complex, searching nystagmus with irregular pendular (sinusoidal) and jerk features. This nystagmus is commonly referred to as *congenital sensory nystagmus*. It is a poor term, because even in children with congenital lesions, the nystagmus does not appear until several months of age. *Congenital motor nystagmus*, which looks similar to congenital sensory nystagmus, develops in the absence of any abnormality of the sensory visual system. Visual acuity is also reduced in congenital motor nystagmus, probably by the nystagmus itself, but seldom below a level of 20/200.

Jerk Nystagmus This is characterized by a slow drift off the target, followed by a fast corrective saccade. By convention, the nystagmus is named after the quick phase. Jerk nystagmus can be downbeat, upbeat, horizontal (left or right), and torsional. The pattern of nystagmus may vary with gaze position. Some patients will be oblivious to their nystagmus. Others will complain of blurred vision, or a subjective, to-and-fro movement of the environment (oscillopsia) corresponding to their nystagmus. Fine nystagmus may be difficult to see upon gross examination of the eyes. Observation of nystagmoid movements of the optic disc on ophthalmoscopy is a sensitive way to detect subtle nystagmus. The slit lamp is also useful.

Gaze-Evoked Nystagmus This is the most common form of jerk nystagmus. When the eyes are held eccentrically in the orbits, they have a natural tendency to drift back to primary position. The subject compensates by making a corrective saccade to maintain the deviated eye position. Many normal patients have mild gaze-evoked nystagmus. Exaggerated gaze-evoked nystagmus can be induced by drugs (sedatives, anticonvulsants, alcohol); muscle paresis; myasthenia gravis; demyelinating disease; and cerebellopontine angle, brainstem, and cerebellar lesions.

Vestibular Nystagmus Vestibular nystagmus results from dysfunction of the labyrinth (Meniere's disease), vestibular nerve, or vestibular nucleus in the brainstem. Peripheral vestibular nystagmus often occurs in discrete attacks, with symptoms of nausea and vertigo. There may be associated tinnitus and hearing loss. Sudden shifts in head position may provoke or exacerbate symptoms.

Downbeat Nystagmus Downbeat nystagmus occurs from lesions near the craniocervical junction (Chiari malformation, basilar invagination). It has also been reported in brainstem or cerebellar stroke, lithium or anticonvulsant intoxication, alcoholism, and multiple sclerosis. *Upbeat nystagmus* is associated with damage to the pontine tegmentum, from stroke, demyelination, or tumor.

Opsoclonus This rare, dramatic disorder of eye movements consists of bursts of consecutive saccades (saccadomania). When the saccades are confined to the horizontal plane, the term *ocular flutter* is preferred. It can occur from viral encephalitis, trauma, or a paraneoplastic effect of neurobastoma, breast carcinoma, and other malignancies. It has also been reported as a benign, transient phenomenon in otherwise healthy patients.

(Bibliography omitted in Palm version)

Back to Table of Contents

29. DISORDERS OF SMELL, TASTE, AND HEARING - *Anil K. Lalwani*, *James B. Snow, Jr.*

SMELL

The sense of smell determines the flavor and palatability of food and drink. It serves, along with the trigeminal system, as a monitor of inhaled chemicals, including dangerous substances such as natural gas, smoke, and air pollutants. Loss of or decreased ability to smell affects approximately 1% of people under age 60 and more than half of the population beyond this age.

DEFINITIONS

Smell is the perception of odor by the nose. *Taste* is the perception of salty, sweet, sour, or bitter by the tongue. Related sensations during eating such as somatic sensations of coolness, warmth, and irritation are mediated through the trigeminal, glossopharyngeal, and vagal afferents in the nose, oral cavity, tongue, pharynx, and larynx. *Flavor* is the complex interaction of taste, smell, and somatic sensation.

Terms relating to disorders of smell include *anosmia*, an absence of the ability to smell; *hyposmia*, a decreased ability to smell; *hyperosmia* (an increased sensitivity to an odorant); *dysosmia* (distortion in the perception of an odor); *phantosmia*, perception of an odorant where none is present; and *agnosia*, inability to classify, contrast, or identify odor sensations verbally, even though the ability to distinguish between odorants or to recognize them may be normal. An odor stimulus is referred to as an *odorant*. Each category of smell dysfunction can be further subclassified as total (applying to all odorants) or partial (dysfunction of only select odorants).

PHYSIOLOGY OF SMELL

The *olfactory neuroepithelium* is located in the superior part of the nasal cavities. It contains an orderly arrangement of bipolar olfactory receptor cells, microvillar cells, sustentacular cells, and basal cells. The dendritic process of the bipolar cell has a bulb-shaped knob, or vesicle, that projects into the mucous layer and bears six to eight cilia. It is the cilia that contain the odorant receptors. The arrangement of cilia increases overall exposure to the environment and translates into each bipolar cell containing 56 cm_2 (9 $in._2$) of surface area to receive stimulus.

The microvillar cells are located adjacent to the receptor cells on the surface of the neuroepithelium. The sustentacular cells, unlike their counterparts in the respiratory epithelium, are not specialized to secrete mucus. Although they form a tight barrier separating the neurons from the outside environment, their complete function is unknown. The basal cells are progenitors of other cell types in the olfactory neuroepithelium, including the bipolar receptor cells. There is a regular turnover of the bipolar receptor cells, which function as the primary sensory neurons. In addition, with injury to the cell body or its axon, the receptor cell is replaced by a differentiated basal cell which reestablishes a central neural connection. Hence these primary sensory neurons are unique among sensory systems in that they are regularly replaced and regenerate after injury.

The unmyelinated axons of the receptor cells form the fila of the olfactory nerve, pass through the cribriform plate, and terminate within spherical masses of neuropil, termed *glomeruli*, in the olfactory bulb. The glomeruli are the focus of a high degree of convergence of information, since many more fibers enter than leave them. The main second-order neurons are the mitral cells. The primary dendrite of each mitral cell extends into a single glomerulus. Axons of the mitral cells project along with the axons of adjacent tufted cells to the limbic system, including the anterior olfactory nucleus, the prepiriform cortex, the periamygdaloid cortex, the olfactory tubercle, the nucleus of the lateral olfactory tract, and the corticomedial nucleus of the amygdala. Cognitive awareness of smell requires stimulation of the prepiriform cortex or the amygdaloid nuclei.

A secondary potential site of olfactory chemosensation is located in the epithelium of the vomeronasal organ, a tubular structure that opens on the ventral aspect of the nasal septum. Sensory neurons located in the vomeronasal organ detect pheromones, nonvolatile chemical signals that in lower mammals trigger innate and stereotyped reproductive and social behaviors, as well as neuroendocrine changes. The neurons from the organ project to the accessory olfactory bulbs and not the main olfactory bulb, as in the olfactory neuroepithelium. Whether humans use the vomeronasal organ to detect and respond to chemical signals from others remains controversial. Recent work in delineating the molecular pathology of Kallman syndrome suggests that development of the olfactory and vomeronasal system is required for normal sexual maturation (Chap. 335).

The sensation of smell begins with introduction of an odorant to the cilia of the bipolar neuron. Most odorants are hydrophobic; as they move from the air phase of the nasal cavity to the aqueous phase of the olfactory mucous, they are transported toward the cilia by small water-soluble proteins called *odorant-binding proteins* and reversibly bind to receptors on the cilia surface. Binding causes conformational changes in the receptor protein, which induces a chain of biochemical events and results in generation of action potentials in the primary neurons. Transduction depends on the activation of G protein-coupled second messengers. Intensity appears to be coded by the amount of firing in the afferent neurons.

Basic elements of the genetic coding involved in smell are now becoming understood. Olfactory receptor proteins belong to the large family of G protein-coupled receptors that also includes rhodopsins; a- andb-adrenergic receptors; muscarinic acetylcholine receptors; and neurotransmitter receptors for dopamine, serotonin, and substance P. Members of the G protein-coupled receptors are characterized by the presence of 7 putative a-helical transmembrane domains composed of 20 to 28 hydrophobic amino acid residues. In mammals, there are probably 300 to 1000 olfactory receptor genes belonging to 20 different families located on various chromosomes in clusters. The receptor genes are present at more than 25 different human chromosomal locations. The gene clusters have likely risen as a result of repeated duplication of individual genes or clusters of genes. Each olfactory neuron seems to express only one or, at most, a few receptor genes thus providing the molecular basis of odor discrimination. While the receptors are expressed in several tissues, including the olfactory neuroepithelium and mammalian germ cells, their primary role appears to be odorant

recognition and discrimination. Bipolar cells that express similar receptors appear to be scattered across discrete spatial zones. These similar cells converge on a select few number of glomeruli in the olfactory bulb. The result is a potential spatial map of how we receive odor stimulus, much like the tonotopic organization of how we perceive sound.

DISORDERS OF THE SENSE OF SMELL

Disorders of the sense of smell are caused by conditions that interfere with the access of the odorant to the olfactory neuroepithelium (transport loss), injure the receptor region (sensory loss), or damage central olfactory pathways (neural loss). At the present time, there are no clinical tests to differentiate these different types of olfactory losses. Fortunately, the history of the disease provides important clues to the cause. The leading causes of olfactory disorders are summarized inTable 29-1; the most common etiologies are head trauma and viral infections. Head trauma is a frequent cause of anosmia in children and young adults, whereas viral etiologies predominate in older adults.

Cranial trauma is followed by unilateral or bilateral impairment of smell in up to 15% of cases; anosmia is more common than hyposmia. Olfactory dysfunction is more common when trauma is associated with loss of consciousness, moderately severe head injury (grades II to V), and skull fracture. Frontal injuries and fractures disrupt the cribriform plate and olfactory axons that perforate it. Sometimes there is an associated cerebrospinal fluid (CSF) rhinorrhea resulting from a tearing of the dura overlying the cribriform plate and paranasal sinuses. Anosmia may also follow blows to the occiput. Once traumatic anosmia develops, it is usually permanent; only 10% of patients ever improve or recover. Perversion of the sense of smell may occur as a transient phase in the recovery process.

Viral infections destroy the olfactory neuroepithelium, which is replaced by respiratory epithelium. Parainfluenza virus type 3 appears to be especially detrimental to human olfaction. HIV infection is associated with subjective distortion of taste and smell, which may become more severe as the disease progresses. The loss of taste and smell may play an important role in the development and progression of HIV-associated wasting. Congenital anosmias are rare but important. Kallmann syndrome is an X-linked disorder characterized by congenital anosmia and hypogonadotropic hypogonadism resulting from a failure of migration from the olfactory placode of olfactory receptor neurons and neurons synthesizing gonadotropin-releasing hormone (Chap. 328). The responsible gene (*KAL*) has been cloned. Anosmia can also occur in albinos. The receptor cells are present but are hypoplastic, lack cilia, and do not project above the surrounding supporting cells.

Meningioma of the inferior frontal region is the most frequent neoplastic cause of anosmia; loss of smell may be the only neurologic abnormality at presentation. Rarely, anosmia can occur with glioma of the frontal lobe. Occasionally, pituitary adenomas, craniopharyngiomas, suprasellar meningiomas, and aneurysms of the anterior part of the circle of Willis extend forward and damage olfactory structures. These tumors and hamartomas may also induce seizures with olfactory hallucinations, indicating involvement of the uncus of the temporal lobe.

Dysosmia, subjective distortions of olfactory perception, may occur with intranasal disease that partially impairs smell or may represent a phase in the recovery from a neurogenic anosmia. Most dysosmic disorders consist of disagreeable or foul odors, and they may be accompanied by distortions of taste. Dysosmia is associated with depression.

Approach to the Patient

The history of the onset and course of the disorder is often paramount in establishing an etiology. Unilateral anosmia is rarely a complaint and is only recognized by separate testing of smell in each nasal cavity. Bilateral anosmia, on the other hand, brings patients to medical attention. Anosmic patients usually complain of a loss of the sense of taste even though their taste thresholds may be within normal limits. In actuality, they are complaining of a loss of flavor detection, which is mainly an olfactory function. The physical examination should include a complete examination of the ears, upper respiratory tract, and head and neck. A neurologic examination emphasizing the cranial nerves and cerebellar and sensorimotor function is essential. The patient's general mood should be assessed, and any signs of depression should be noted.

The sensory evaluation of olfactory function is necessary to corroborate the patient's complaint, evaluate the efficacy of treatment, and assess the degree of permanent impairment. The degree to which qualitative sensations are present can be assessed by any of several methods. The Odor Stix test uses a commercially available odor-producing magic marker-like pen held approximately 8 to 15 cm (3 to 6 in.) from the patient's nose to check for gross perception of the odorant. Another gross perception of odorant test, the 30-cm alcohol test, uses a freshly opened isopropyl alcohol packet held approximately 30 cm (12 in.) from the patient's nose. There is a commercially available scratch-and-sniff card containing three odors available for testing olfaction grossly. A superior test is the University of Pennsylvania Smell Identification Test (UPSIT). This consists of a 40-item, forced choice, microencapsulated odor, scratch-and-sniff paradigm. For example, one of the items reads, "This odor smells most like (a) chocolate, (b) banana, (c) onion, or (d) fruit punch," and the patient is instructed to answer one of the alternatives. The test is highly reliable, is sensitive to age and sex differences, and provides an accurate quantitative determination of the olfactory deficit. Persons with a total loss of smell function score in the range of 7 to 19 out of 40. The average score for total anosmics is slightly higher than that expected on the basis of chance because of the inclusion of some odorants that act by trigeminal stimulation.

The second step is to establish a detection threshold for the odorant phenyl ethyl alcohol, using a graduated stimulus. Sensitivity for each side of the nose is determined with a detection threshold for phenyl ethyl methyl ethyl carbinol. Nasal resistance can also be measured with anterior rhinomanometry for each side of the nose.

Computed tomography (CT) or magnetic resonance imaging (MRI) of the head is required to rule out paranasal sinusitis, neoplasms of the anterior cranial fossa, nasal cavity, or paranasal sinuses and unsuspected fractures of the anterior cranial fossa. Bone abnormalities are best seen with CT. MRI is useful in evaluating olfactory bulbs, ventricles, and other soft tissue of the brain. Coronal CT is optimal for assessing

cribriform plate, anterior cranial fossa, and sinus anatomy.

Techniques have been developed to biopsy the olfactory neuroepithelium, but in view of the widespread degeneration of the olfactory neuroepithelium and intercalation of respiratory epithelium in the olfactory area of adults with no apparent olfactory dysfunction, biopsy material must be interpreted cautiously.

TREATMENT

Therapy for patients with transport olfactory losses due to allergic rhinitis, bacterial rhinitis and sinusitis, polyps, neoplasms, and structural abnormalities of the nasal cavities can be undertaken rationally and with a high likelihood for improvement. Allergy management, antibiotic therapy, topical and systemic glucocorticoid therapy, and surgery for nasal polyps, deviation of the nasal septum, and chronic hyperplastic sinusitis are frequently effective in restoring the sense of smell.

There is no treatment with demonstrated efficacy for sensorineural olfactory losses. Fortunately, spontaneous recovery often occurs. Zinc and vitamin therapy (especially with vitamin A) are advocated by some. Profound zinc deficiency can produce loss and distortion of the sense of smell but is not a clinically important problem except in very limited geographic areas (Chap. 75). The epithelial degeneration associated with vitamin A deficiency can cause anosmia, but in western societies the prevalence of vitamin A deficiency is low. Exposure to cigarette smoke and other airborne toxic chemicals can cause metaplasia of the olfactory epithelium. Spontaneous recovery can occur if the insult is discontinued. Patient counseling is therefore helpful in these cases.

As mentioned above, more than half of people over age 60 suffer from olfactory dysfunction. No effective treatment exists for presbyosmia, but patients are often reassured to learn that this problem is common in their age group. In addition, early recognition and counseling can help patients to compensate for the loss of smell. The incidence of natural gas-related accidents is disproportionately high in the elderly, perhaps due in part to the gradual loss of smell. Mercaptan, the pungent odor in natural gas, is an olfactory stimulant and does not activate taste receptors. Many elderly with olfactory dysfunction experience a decrease in flavor sensation and find it necessary to hyperflavor food, usually by increasing the amount of salt in their diet. The physician can assist patients in developing healthy strategies to deal with the decreased sense of smell.

TASTE

Compared with disorders of smell, gustatory disorders are uncommon and their pathogenesis poorly understood. Many patients with a loss of olfactory sensitivity also complain of a loss of the sense of taste. On testing, most of these patients have normal detection thresholds for taste.

DEFINITIONS

Disturbances of the sense of taste may be categorized as *total ageusia* -- total absence of gustatory function or inability to detect the qualities of sweet, salt, bitter, or sour;

partial ageusia -- ability to detect some of but not all the qualitative gustatory sensations; *specific ageusia* -- inability to detect the taste quality of certain substances; *total hypogeusia* -- decreased sensitivity to all tastants; *partial hypogeusia* -- decreased sensitivity to some tastants; and *dysgeusia* or *phantogeusia* -- distortion in the perception of a tastant, i.e., the perception of the wrong quality when a tastant is presented or the perception of a taste when there has been no tastant ingested. Confusions of sour and bitter are common and, at times, may be semantic misunderstandings. Frequently, however, they have physiologic or pathophysiologic bases. Other taste quality confusions occur between sour and salty and bitter. It may be possible to differentiate between the loss of flavor recognition in patients with olfactory losses who complain of a loss of taste as well as smell by asking if they are able to taste sweetness in sodas, saltiness in potato chips, etc.

PHYSIOLOGY OF TASTE

The taste receptor cells are located in the taste buds, spherical groups of cells arranged in a pattern resembling the segments of a citrus fruit. At the surface, the taste bud has a pore into which microvilli of the receptor cells project. Unlike the olfactory system, the receptor cell is not the primary neuron. Instead, gustatory afferent nerve fibers contact individual taste receptor cells. Transduction depends on activation of G protein-coupled second messengers but differs in details for each taste quality.

The sense of taste is mediated through the facial, glossopharyngeal, and vagal nerves. The gustatory system consists of at least five receptor populations. Taste buds are located in the papillae along the lateral margin and dorsum of the tongue at the junction of the dorsum and the base of the tongue, and in the palate, epiglottis, larynx, and esophagus. The chorda tympani branch of the facial nerve subserves taste from the anterior two-thirds of the tongue. The posterior third of the tongue is supplied by the lingual branch of the glossopharyngeal nerve. Afferents from the palate travel with the greater superficial petrosal nerve to the geniculate ganglion and then via the facial nerve to the brainstem. The internal branch of the superior laryngeal nerve of the vagus nerve contains the taste afferents from the larynx, including the epiglottis and esophagus.

The central connections of the nerves terminate in the brainstem in the nucleus of the tractus solitarius. The central pathway from the nucleus of the tractus solitarius projects to the ipsilateral parabrachial nuclei of the pons. Two divergent pathways project from the parabrachial nuclei. One ascends to the gustatory relay in the dorsal thalamus, synapses, and continues to the cortex of the insula. There is also evidence for a direct pathway from the parabrachial nuclei to the cortex. (Olfaction and gustation appear to be unique among sensory systems in that at least some fibers bypass the thalamus.) The other pathway from the parabrachial nuclei goes to the ventral forebrain, including the lateral hypothalamus, substantia innominata, central nucleus of the amygdala, and the stria terminalis.

Tastants gain access to the receptor cells through the taste pore. Four classes of taste are recognized: sweet, salt, sour, and bitter. Individual gustatory afferent fibers almost always respond to a number of different chemicals. Response patterns of gustatory afferent axons can be grouped into classes based on the stimulus chemical that produces the largest response. For example, for sucrose-best response neurons, the

second-best stimulus is almost always sodium chloride. The fact that individual gustatory afferent fibers respond to a large number of different chemicals led to the *across-fiber-pattern* theory of gustatory coding, while the best-stimulus analysis led to the concept of *labeled* afferents. It appears that labeled fibers are important for establishing gross quality, but the across-fiber pattern within a best-stimulus category, and perhaps among categories, is needed for discriminating chemicals within qualities. For example, sweetness may be carried by sucrose-best neurons, but the differentiation of sucrose and fructose may require a comparison of the relative activity among sucrose-best, salt-best, and quinine-best neurons. As with olfaction and other sensory systems, intensity appears to be encoded by the quantity of neural activity.

DISORDERS OF THE SENSE OF TASTE

Disorders of the sense of taste are caused by conditions that interfere with the access of the tastant to the receptor cells in the taste bud (transport loss), injure receptor cells (sensory loss), or damage gustatory afferent nerves and central gustatory pathways (neural loss) (Table 29-2). *Transport gustatory losses* result from xerostomia due to many causes, including Sjogren's syndrome, radiation therapy, heavy-metal intoxication, and bacterial colonization of the taste pore. *Sensory gustatory losses* are caused by inflammatory and degenerative diseases in the oral cavity; a vast number of drugs, particularly those that interfere with cell turnover such as antithyroid and antineoplastic agents; radiation therapy to the oral cavity and pharynx; viral infections; endocrine disorders; neoplasms; and aging. *Neural gustatory losses* occur with neoplasms, trauma, and surgical procedures in which the gustatory afferents are injured. Taste buds degenerate when their gustatory afferents are transected but remain when their somatosensory afferents are severed. Patients with renal disease have increased thresholds for sweet and sour tastes, which resolves with dialysis.

A side effect of medication is the single most common cause of taste dysfunction in clinical practice. The mechanism may be a change in the composition of saliva, an effect on receptor function or signal transduction, or disruption of the central processing of gustatory input. Unfortunately, the responsible mechanism is not well understood for most medications. Xerostomia, regardless of the etiology, can be associated with taste dysfunction. It is associated with poor oral clearance, poor dental hygeine, and can adversely affect the oral mucosa, all leading to dysgeusia. However, severe salivary gland failure does not necessarily lead to taste complaints. Xerostomia, along with the use of antibiotics or glucocorticoids, and compromised immune function can lead to overgrowth of *Candida*; overgrowth alone, without thrush or overt signs of infection can be associated with bad taste or hypogeusia. When taste dysfunction occurs in a patient at risk for fungal overgrowth, a trial of nystatin or other anti-fungal medication is warranted.

Upper respiratory infections and head trauma can lead to both smell and taste dysfunction; taste is more likely to improve than smell. The mechanism of taste disturbance in these situations is not well understood. Trauma to the chorda tympani branch of the facial nerve during middle ear surgery or third molar extractions is relatively common and can cause dysgeusia. Bilateral chorda tympani injuries are usually associated with hypogeusia, whereas unilateral lesions produce only limited symptoms, perhaps because responses from taste receptors are disinhibited by the

glossopharyngeal nerve.

Finally, aging itself may be associated with reduced taste sensitivity. The taste dysfunction may be limited to a single compound and may be mild. While many older patients may acknowledge loss of taste when asked, they are unlikely to seek medical attention for taste disturbance alone.

Approach to the Patient

Patients who complain of loss of taste should be evaluated for both gustatory and olfactory function. Clinical assessment of taste is not as well developed or standardized as that of smell. The first step is to perform suprathreshold whole-mouth taste testing for quality, intensity, and pleasantness perception of four taste qualities: sweet, salty, sour, and bitter. Most commonly used reagents for taste testing are sucrose, citric acid or hydrochloric acid, caffeine or quinine (sulfate or hydrochloride), and sodium chloride. The taste stimuli should be freshly prepared. For quantification, detection thresholds are obtained by applying graduated dilutions to the tongue quandrants or by whole-mouth sips. Electric taste testing (*electrogustometry*) is used clinically to identify taste deficits in specific quandrants of the tongue, following precise applications of stimuli. Regional gustatory testing may also be performed to assess for the possibility of loss localized to one or more receptor fields as a result of a peripheral or central lesion.

Once there is objective evidence of a disorder of taste, it is important to establish an anatomic diagnosis before proceeding to an etiologic diagnosis. The history of the disease often provides important clues to the cause. For example, absence of taste on the anterior two-thirds of the tongue associated with a facial paralysis indicates that the lesion is proximal to the juncture of the chorda tympani branch with the facial nerve in the mastoid.

TREATMENT

Therapy for gustatory loss is limited. Nonetheless, some etiologies of taste dysfunction are amenable to intervention. Taste disturbance related to drugs can often be resolved by changing the prescribed medication. Xerostomia can be treated with artificial saliva, providing some benefits to patients with a disturbed salivary milieu. Oral pilocarpine may be beneficial for a variety of forms of xerostomia. Appropriate treatment of bacterial and fungal infections of the oral cavity can be of great help in improving taste function. Taste dysfunction following trauma may resolve spontaneously without intervention and is more likely to do so than posttraumatic smell dysfunction. Altered taste due to surgical stretch injury of chorda tympani nerve usually improves within 3 to 4 months, while dysfunction is usually permanent with transection of the nerve. In most patients with idiopathic cases of altered taste sensitivity, the problem either remains stable or worsens. Zinc and vitamin therapy for gustatory losses is advocated by some but lacks demonstrated efficacy. No effective therapeutic strategies exist for the sensorineural disorders of taste.

HEARING

Hearing loss is one of the most common sensory disorders in humans. Nearly 10% of

the adult population has some hearing loss. For many, this impairment presents early in life. However, hearing loss can present at any age. Between 30 and 35% of individuals over the age of 65 have a hearing loss of sufficient magnitude to require a hearing aid.

PHYSIOLOGY OF HEARING (Fig. 29-1)

Hearing occurs by air conduction and bone conduction. In air conduction, sound waves reach the ear by propagation in air, enter the external auditory canal, and set the tympanic membrane in motion, which in turn moves the malleus, incus, and stapes of the middle ear. Movement of the footplate of the stapes causes pressure changes in the fluid-filled inner ear eliciting a traveling wave in the basilar membrane of the cochlea. The tympanic membrane and the ossicular chain in the middle ear serve as an impedence-matching mechanism, improving the efficiency of energy transfer from air to the fluid-filled inner ear. Hearing by bone conduction occurs when the sounding source, in contact with the head, results in vibration of the bones of the skull, including the temporal bone, producing a traveling wave in the basilar membrane.

Stereocilia of the hair cells of the organ of Corti, which rests on the basilar membrane, are in contact with the tectorial membrane and are deformed by the traveling wave. A point of maximal displacement of the basilar membrane is determined by the frequency of the stimulating tone. High-frequency tones cause maximal displacement of the basilar membrane near the base of the cochlea. As the frequency of the stimulating tone decreases, the point of maximal displacement moves toward the apex of the cochlea.

The inner and outer hair cells of the organ of Corti have different innervation patterns, but both are mechanoreceptors. The afferent innervation relates principally to the inner hair cells, and the efferent innervation relates principally to outer hair cells. The motility of the outer hair cells alters the micromechanics of the inner hair cells creating a cochlear amplifier, which explains the exquisite sensitivity and frequency selectivity of the cochlea.

The current concept of cochlear transduction is that displacement of the tips of the stereocilia allows potassium to flow into the cell, resulting in its depolarization. The potassium influx opens calcium channels near the base of the cell, stimulating transmitter release. The neurotransmitter at the hair cell and cochlear nerve dendrite interface is thought to be glutamate. The action potential in the eighth nerve occurs 0.5 ms after the onset of the cochlear microphonic potential. Each of the cochlear nerve neurons can be activated at a frequency and intensity specific for that cell. This specificity is maintained at each point of the central auditory pathway: dorsal and ventral cochlear nuclei, trapezoid body, superior olivary complex, lateral lemniscus, inferior colliculus, medial geniculate body, and auditory cortex. At low frequencies, individual auditory nerve fibers can respond more or less synchronously with the stimulating tone. At higher frequencies, phase-locking occurs so that neurons alternate in response to particular phases of the cycle of the sound wave. Intensity is encoded by the amount of neural activity in individual neurons, the number of neurons that are active, and the specific neurons that are activated.

GENETIC CAUSES OF HEARING LOSS

More than half of childhood hearing impairment is thought to be hereditary; hereditary hearing impairment (HHI) can also manifest later in life. HHI may be classified as either nonsyndromic, when hearing loss is the only clinical abnormality, or syndromic, when hearing loss is associated with anomalies in other organ systems. Nearly two-thirds of HHIs are nonsyndromic and the remaining one-third are syndromic. Between 70 and 80% of nonsyndromic HHI is inherited in an autosomal recessive manner; another 15 to 20% is autosomal dominant. Less than 5% is X-linked or maternally inherited via the mitochondria.

Over 60 loci harboring genes for nonsyndromicHHI have been mapped, with equal numbers of dominant and recessive modes of inheritance; 14 different genes have been cloned (Table 29-3). The hearing genes fall into the categories of structural proteins (MYO7A, MYO15, TECTA, DIAPH1), transcription factors (POU3F4, POU4F3), ion channels (KCNQ4, PDS), and gap junction proteins (Cx26, Cx30, Cx31). Several of these genes, including connexin 26 (Cx26), TECTA, and MYO7A, cause both autosomal dominant and recessive forms of nonsyndromic HHI. In general, the hearing loss associated with dominant genes has its onset in adolescence or adulthood and varies in severity, whereas the hearing loss associated with recessive inheritance is congenital and profound. Connexin 26 is particularly important because it is associated with nearly 20% of cases of childhood deafness; in heterozygotes the onset of hearing loss may be in adolescence or adulthood. Two frame-shift mutations, 30delG and 167delT, account for>50% of the cases, making population screening feasible. The 167delT mutation is highly prevalent in Ashkenazi Jews; it is predicted that 1 in 1765 individuals in this population will be homozygous and affected. The hearing loss can also vary among the members of the same family, suggesting that other genes or factors likely influence the auditory phenotype.

The contribution of genetics to presbycusis (see below) is also becoming better understood. In addition to connexin 26, several other nonsyndromic genes are associated with hearing loss that progresses with age. It is likely that presbycusis has both environmental and genetic components.

Over 200 syndromes are associated with hearing loss. Common syndromic forms of hearing loss include Usher syndrome (retinitis pigmentosa and hearing loss), Waardenburg syndrome (pigmentary abnormality and hearing loss), Pendred syndrome (thyroid organification defect and hearing loss), Alport syndrome (renal disease and hearing loss), Jervell and Lange-Nielsen syndrome (prolonged QT interval and hearing loss), neurofibromatosis type 2 (bilateral acoustic schwannoma), and mitochondrial disorders [mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS); myoclonic epilepsy and ragged red fibers (MERRF); progressive external ophthalmoplegia (PEO)] (Table 29-4).

Rapid progress in understanding the basis of these and related disorders has revealed a fascinating complexity, including evidence for genetic heterogeneity (different genes resulting in a similar clinical phenotype), allelic disorders (distinct phenotypes associated with different mutations in the same gene), and polygenic modifiers (Chap. 65).

DISORDERS OF THE SENSE OF HEARING

Hearing loss can result from disorders of the auricle, external auditory canal, middle ear, inner ear, or central auditory pathways (Fig. 29-2). *In general, lesions in the auricle, external auditory canal, or middle ear cause conductive hearing losses, whereas lesions in the inner ear or eighth nerve cause sensorineural hearing losses*.

Conductive Hearing Loss This results from obstruction of the external auditory canal by cerumen, debris, and foreign bodies; swelling of the lining of the canal; atresia of the ear canal; neoplasms of the canal; perforations of the tympanic membrane; disruption of the ossicular chain, as occurs with necrosis of the long process of the incus in trauma or infection; otosclerosis; or fluid, scarring, or neoplasms in the middle ear.

Cholesteatoma, i.e., the presence of stratified squamous epithelium in the middle ear or mastoid, occurs frequently in adults. A cholesteatoma is a benign, slowly growing lesion that destroys bone and normal ear tissue. Major theories of pathogenesis of acquired cholesteatoma include traumatic implantation and invasion, immigration and invasion through a perforation, and metaplasia following chronic infection and irritation. On examination, there is often a perforation filled with cheesy white squamous debris. A chronically draining ear that fails to respond to appropriate antibiotic therapy should raise the suspicion of a cholesteatoma. Conductive hearing loss secondary to ossicular erosion is common. Surgery is required to remove this insidiously growing and destructive disease process.

Conductive hearing loss in the presence of a normal ear canal and intact tympanic membrane is suggestive of ossicular pathology. Fixation of the stapes from *otosclerosis* is a common cause of low-frequency conductive hearing loss. It occurs with equal frequency in men and women and has a simple autosomal dominant inheritance with incomplete penetrance. Hearing impairment usually presents between the late teens to the forties. In women, the hearing loss is often first noticeable during pregnancy, as the otosclerotic process is accelerated during pregnancy. A hearing aid or a short outpatient surgical procedure (stapedectomy) can provide adequate auditory rehabilitation. Extension of otosclerosis beyond the stapes footplate to involve the cochlea (cochlear otosclerosis) can lead to mixed or sensorineural hearing loss. Fluoride therapy to prevent hearing loss associated with cochlear otosclerosis remains controversial.

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media (SOM). Trauma, AOM, or chronic otitis media are the usual factors responsible for tympanic membrane perforation. While small perforations often heal spontaneously, larger defects usually require surgical intervention. Tympanoplasty is highly effective (>90%) in the repair of tympanic membrane perforations. Otoscopy is usually sufficient to diagnose AOM, SOM, chronic otitis media, cerumen impaction, tympanic membrane perforation, and eustachian tube dysfunction.

Sensorineural Hearing Loss Damage to the hair cells of the organ of Corti may be caused by intense noise, viral infections, ototoxic drugs (e.g., salicylates, quinine and its synthetic analogues, aminoglycoside antibiotics, loop diuretics such as furosemide and ethacrynic acid, and cancer chemotherapeutic agents such as cisplatin), fractures of the temporal bone, meningitis, cochlear otosclerosis (see above), Meniere's disease, and

aging. Congenital malformations of the inner ear may be the cause of hearing loss in some adults. Genetic predisposition alone or in concert with environmental influences may also be responsible.

Presbycusis (age-associated hearing loss) is the most common cause of sensorineural hearing loss in adults. In the early stages, it is characterized by symmetric, gentle to sharply sloping high-frequency hearing loss. With progression, the hearing loss involves all frequencies. More importantly, the hearing impairment is associated with significant loss in clarity. There is a loss of discrimination for phonemes, recruitment (abnormal growth of loudness), and particular difficulty in understanding speech in noisy environments. Hearing aids may provide limited rehabilitation once the word recognition score deteriorates below 50%. Significant advancements and improvements in cochlear implants have made them the treatment of choice when hearing aids prove inadequate (<30% word recognition score with optimal amplification).

Meniere's disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. Tinnitus and/or deafness may be absent during the initial attacks of vertigo, but they invariably appear as the disease progresses and are increased in severity during an acute attack. The annual incidence of Meniere's disease is 0.5 to 7.5 per 1000; onset is most frequently in the fifth decade of life but may also occur in young adults or the elderly. Histologically, there is distention of the endolymphatic system (endolymphatic hydrops) leading to degeneration of vestibular and cochlear hair cells. This may result from endolymphatic sac dysfunction secondary to infection, trauma, autoimmune disease, inflammatory causes, or tumor; an idiopathic etiology constitutes the largest category and is most accurately referred to as Meniere's disease. Although any pattern of hearing loss can be observed, typically, low-frequency, unilateral sensorineural hearing impairment is present.MRIshould be obtained to exclude retrocochlear pathology such as cerebellopontine angle tumors or demyelinating disorders. Therapy is directed towards the control of vertigo. A low-salt diet is the mainstay of treatment of the control of rotatory vertigo. Diuretics, a short course of glucocorticoids, and intratympanic gentamicin may also be useful adjuncts in recalcitrant cases. Surgical therapy of vertigo is reserved for unresponsive cases and includes endolymphatic sac decompression, labyrinthectomy, and vestibular nerve section. Both labyrinthectomy and vestibular nerve section abolish rotatory vertigo in>90% of the cases. Unfortunately, there is no effective therapy for hearing loss, tinnitus, or aural fullness associated with Meniere's disease.

Sensorineural hearing loss may also result from any neoplastic, vascular, demyelinating, infectious, or degenerative disease or trauma affecting the central auditory pathways. Human immunodeficiency virus leads to both peripheral and central auditory system pathology and is associated with sensorineural hearing impairment.

An individual can have both conductive and sensory hearing loss termed *mixed hearing loss*. Mixed hearing losses are due to pathology that can affect the middle and inner ear simultaneously such as otosclerosis involving the ossicles and the cochlea, head trauma, chronic otitis media, cholesteatoma, middle ear tumors, and some inner ear malformations.

Trauma resulting in temporal bone fractures may be associated with conductive,

sensorineural, and mixed hearing loss. If the fracture spares the inner ear, there may simply be conductive hearing loss due to rupture of the tympanic membrane or disruption of the ossicular chain. These abnormalities are amenable to surgical correction. Profound hearing loss and severe vertigo are associated with temporal bone fractures involving the inner ear. A perilymphatic fistula associated with leakage of inner-ear fluid into the middle ear can occur and may require surgical repair. An associated facial nerve injury is not uncommon.CT is best suited to assess fracture of the traumatized temporal bone, evaluate the ear canal, and determine the integrity of the ossicular chain and the involvement of the inner ear.CSFleaks that accompany temporal bone fractures are usually self-limited; the use of prophylactic antibiotics is controversial.

Tinnitus is defined as the perception of a sound when there is no sound in the environment. It may have a buzzing, roaring, or ringing quality and may be pulsatile (synchronous with the heartbeat). Tinnitus is often associated with either a conductive or sensorineural hearing loss. The pathophysiology of tinnitus is not well understood. The cause of the tinnitus can usually be determined by finding the cause of the associated hearing loss. Tinnitus may be the first symptom of a serious condition such as a vestibular schwannoma. Pulsatile tinnitus requires evaluation of the vascular system of the head to exclude vascular tumors such as glomus jugulare tumors, aneurysms, and stenotic arterial lesions; it may also occur withSOM.

Approach to the Patient

The goals in the evaluation of a patient with auditory complaints are to determine: (1) the nature of the hearing impairment (conductive vs. sensorineural), (2) the severity of the impairment (mild, moderate, severe, profound), (3) the anatomy of the impairment (external ear, middle ear, inner ear, or central auditory pathway pathology, and (4) the etiology. Initially, the history and the physical examination are critical in the identification of the underlying pathology leading to the auditory deficit. The history should elicit characteristics of the hearing loss, including the duration of deafness, nature of onset (sudden vs. insidious), rate of progression (rapid vs. slow), and involvement of the ear (unilateral vs. bilateral). The presence or absence of tinnitus, vertigo, imbalance, aural fullness, otorrhea, headache, facial nerve dysfunction, and head and neck paresthesias should be ascertained. Information regarding head trauma, exposure to ototoxins, occupational or recreational noise exposure, and family history of hearing impairment may also be important. A sudden onset of unilateral hearing loss, with or without tinnitus, may represent a viral infection of the inner ear or a vascular accident. Patients with unilateral hearing loss (sensory or conductive) usually complain of reduced hearing, poor sound localization, and difficulty hearing clearly with background noise. Gradual progression of a hearing deficit is common with otosclerosis, noise-induced hearing loss, vestibular schwannoma, or Meniere's disease. Small vestibular schwannomas typically present with asymmetric hearing impairment, tinnitus, and imbalance (rarely vertigo); cranial neuropathy, in particular of the trigeminal or facial nerve, may accompany larger tumors. In addition to hearing loss, Meniere's disease may be associated with episodic vertigo, tinnitus, and aural fullness. Hearing loss with otorrhea is most likely due to chronic otitis media or cholesteatoma.

Family history may be crucial in delineating a genetic basis of hearing impairment. The

history may also help identify environmental risk factors that lead to hearing impairment in a family. Sensitivity to aminoglycoside ototoxicity, maternally transmitted through a mitochondrial mutation, can be ascertained through a careful family history (Chap. 67). Susceptibility to noise-induced hearing loss or age-related hearing loss (presbycusis) may also be genetically determined.

The physical examination should evaluate the auricle, external ear canal, and tympanic membrane. The external ear canal of the elderly is often dry and fragile; it is preferable to clean cerumen with wall-mounted suction and cerumen loops and to avoid irrigation. In examining the eardrum, the topography of the tympanic membrane is more critical than the presence or absence of the highly touted light reflex. In addition to the pars tensa (the lower two-thirds of the eardrum), the pars flaccida above the short process of the malleus should also be examined for retraction pockets that may be evidence of chronic eustachian tube dysfunction or cholesteatoma. Insufflation of the ear canal is necessary to assess tympanic membrane mobility and compliance. Careful inspection of the nose, nasopharynx, and upper respiratory tract is indicated. Unilateral serous effusion in the adult should prompt a fiberoptic examination of the nasopharynx to exclude neoplasms. Cranial nerves should be carefully evaluated with special attention to facial and trigeminal nerves, which are commonly disturbed with tumors involving the cerebellopontine angle.

The Weber and Rinne tuning fork tests are used to differentiate conductive from sensorineural hearing losses and to confirm the findings of audiologic evaluation. Rinne's test compares the ability to hear by air conduction with the ability to hear by bone conduction. The tines of a vibrating tuning fork are held near the opening of the external auditory canal, and then the stem is placed on the mastoid process; for direct contact, it may be placed on teeth or dentures. The patient is asked to indicate whether the tone is louder by air conduction or bone conduction. Normally, and in the presence of sensorineural hearing loss, a tone is heard louder by air conduction than by bone conduction; however, with conductive hearing loss of ^{3}30 dB (see "Audiologic Assessment," below), the bone-conduction stimulus is perceived as louder than the air-conduction stimulus. The Rinne test is most sensitive in detecting mild conductive hearing losses if a 256-Hz tuning fork is used. The Weber test may be performed with a 256- or 512-Hz fork. The stem of a vibrating tuning fork is placed on the head in the midline and the patient asked whether the tone is heard in both ears or better in one ear than in the other. With a unilateral conductive hearing loss, the tone is perceived in the affected ear. With a unilateral sensorineural hearing loss, the tone is perceived in the unaffected ear. As a general rule, a 5-dB difference in hearing between the two ears is required for lateralization. The combined information from the Weber and Rinne tests permits a tentative conclusion as to whether a conductive or sensorineural hearing loss is present; however, these tests are associated with significant false-positive and -negative responses and therefore should be utilized only as screening tools.

LABORATORY ASSESSMENT OF HEARING

Audiologic Assessment The minimum audiologic assessment for hearing loss should include the measurement of pure tone air-conduction and bone-conduction thresholds, speech reception threshold, discrimination score, tympanometry, acoustic reflexes, and acoustic-reflex decay. This test battery provides a comprehensive screening evaluation

of the whole auditory system and allows one to determine whether further differentiation of a sensory (cochlear) from a neural (retrocochlear) hearing loss is indicated.

Pure tone audiometry assesses hearing acuity for pure tones. The test is administered by an audiologist and is performed in a sound-attenuated chamber. The pure tone stimulus is delivered with an audiometer, an electronic device that allows the presentation of specific frequencies (generally between 250 and 8000 Hz) at specific intensities. Air and bone conduction thresholds are established for each ear. Air conduction thresholds are established by presenting the stimulus in air with the use of headphones. Bone conduction thresholds are accomplished by placing the stem of a vibrating tuning fork or an oscillator of an audiometer in contact with the head. In the presence of a hearing loss, broad-spectrum noise is presented to the nontest ear for *masking* purposes so that responses are based on perception from the ear under test.

The responses are measured in decibels. An *audiogram* is a plot of intensity in decibels required to achieve threshold versus frequency. A decibel (dB) is equal to 20 times the logarithm of the ratio of the sound pressure required to achieve threshold in the patient to the sound pressure required to achieve threshold in a normal hearing person. Therefore, a change of 6 dB represents doubling of sound pressure, and a change of 20 dB represents a ten-fold change in sound pressure. Loudness, which depends on the frequency, intensity, and duration of a sound, doubles with approximately each 10-dB increase in sound pressure level. Pitch, on the other hand, does not directly correlate with frequency. The perception of pitch changes slowly in the low and high frequencies. In the middle tones, which are important for human speech, pitch varies more rapidly with changes in frequency.

Pure tone audiometry establishes the presence and severity of hearing impairment, unilateral vs. bilateral involvement, and the type of hearing loss. Conductive hearing losses with a large mass component, as is often seen in middle-ear effusions, produce elevation of thresholds that predominate in the higher frequencies. Conductive hearing losses with a large stiffness component, as in fixation of the footplate of the stapes in early otosclerosis, produce threshold elevations in the lower frequencies. Often, the conductive hearing loss involves all frequencies, suggesting involvement of both stiffness and mass. In general, sensorineural hearing losses such as presbycusis affect higher frequencies more than lower frequencies. An exception is Meniere's disease, which is characteristically associated with low-frequency sensorineural hearing loss. Noise-induced hearing loss has an unusual pattern of hearing impairment in which the loss at 4000 Hz is greater than at higher frequencies. Vestibular schwannomas characteristically affect the higher frequencies, but any pattern of hearing loss can be observed.

Speech recognition requires greater synchronous neural firing than is necessary for appreciation of pure tones. *Speech audiometry* tests the clarity with which one hears. The *speech reception threshold* (SRT) is defined as the intensity at which speech is recognized as a meaningful symbol and is obtained by presenting two-syllable words with an equal accent on each syllable. The intensity at which the patient can repeat 50% of the words correctly is the SRT. Once the SRT is determined, discrimination or word recognition ability is tested by presenting one-syllable words at 25 to 40 dB above the speech reception threshold. The words are phonetically balanced in that the phonemes

(speech sounds) occur in the list of words at the same frequency that they occur in ordinary conversational English. An individual with normal hearing or conductive hearing loss can repeat 88 to 100% of the phonetically balanced words correctly. Patients with a sensorineural hearing loss have variable loss of discrimination depending on the severity of hearing loss and the site of lesion. Further, as a general rule, neural lesions are associated with more deterioration in discrimination ability than are lesions in the inner ear. For example, in a patient with mild asymmetric sensorineural hearing loss, a clue to the diagnosis of vestibular schwannoma is the presence of greater than expected deterioration in discrimination ability. Deterioration in discrimination ability at higher intensities above the SRT also suggests a lesion in the eighth nerve or central auditory pathways.

Tympanometry measures the impedance of the middle ear to sound and is particularly useful in the identification and diagnosis of middle-ear effusions. A sounding source and microphone are introduced into the ear canal with an airtight seal. The amount of sound that is absorbed through the middle ear or reflected from the middle ear is measured at the microphone. In conductive hearing losses, more sound is reflected than in the normal middle ear. The pressure in the ear canal can be increased or decreased from atmospheric pressure. A *tympanogram* is the graphic representation of change in impedance or compliance as the pressure in the ear canal is changed. It provides information about the status of the tympanic membrane and the ossicular chain. Normally, the middle ear is most compliant at atmospheric pressure, and the compliance decreases as the pressure is increased or decreased; this pattern is seen with normal hearing or in the presence of sensorineural hearing loss. Compliance that does not change with change in pressure suggests middle-ear effusion. With a negative pressure in the middle ear, as with eustachian tube obstruction, the point of maximal compliance occurs with negative pressure in the ear canal. A tympanogram in which no point of maximal compliance can be obtained is most commonly seen with discontinuity of the ossicular chain. A reduction in the maximal compliance peak can be seen in otosclerosis.

During tympanometry, an intense tone (80 dB above the hearing threshold) elicits contraction of the stapedius muscle. The change in compliance of the middle ear with contraction of the stapedius muscle can be detected. The presence or absence of this *acoustic reflex* is important in the anatomic localization of facial nerve paralysis as well as hearing loss. Normal or elevated acoustic reflex thresholds in an individual with significant sensorineural hearing impairment suggests a cochlear hearing loss. Assessment of *acoustic reflex decay* helps differentiate sensory from neural hearing losses. In neural hearing loss, the reflex adapts or decays with time.

Otoacoustic emissions (OAE) can be measured with sensitive microphones inserted into the external auditory canal. The emissions may be spontaneous or evoked with sound stimulation. The presence of OAEs indicates that the outer hair cells of the organ of Corti are intact and can be used to assess auditory thresholds and to distinguish sensory from neural hearing losses.

Evoked Responses *Electrocochleography* measures the earliest evoked potentials generated in the cochlea and the auditory nerve. Receptor potentials recorded include the cochlear microphonic, generated by the outer hair cells of the organ of Corti, and the

summating potential, generated by the inner hair cells in response to sound. The whole nerve action potential representing the composite firing of the first-order neurons can also be recorded during electrocochleography. Clinically, the test is useful in the diagnosis of Meniere's disease where an elevation of the ratio of summating potential to action potential is seen.

Brainstem auditory evoked responses (BAERs) are useful in differentiating the site of sensorineural hearing loss (Chap. 356). In response to sound, five distinct electrical potentials arising from different stations along the peripheral and central auditory pathway can be recorded with computer averaging from scalp surface electrodes. BAERs are valuable in situations in which patients cannot or will not give reliable voluntary thresholds. They are also used to assess the integrity of the auditory nerve and brainstem in various clinical situations, including intraoperative monitoring and in determination of brain death.

Imaging Studies The choice of radiologic tests is largely determined by whether the goal is to evaluate the bony anatomy of the external, middle, and inner ear or to image the auditory nerve and brain. Axial and coronalCT of the temporal bone with fine 1-mm cuts is ideal for determining the caliber of the external auditory canal, integrity of the ossicular chain, and presence of middle-ear or mastoid disease; it can also detect inner-ear malformations. CT is also ideal for the detection of bone erosion often seen in the presence of chronic otitis media and cholesteatoma.MRI is superior to CT for imaging of retrocochlear pathology such as vestibular schwannoma, meningioma, other lesions of the cerebellopontine angle, demyelinating lesions of the brainstem, and brain tumors. Recent experience suggests that both CT and MRI are equally capable of identifying inner-ear malformations and assessing cochlear patency for preoperative evaluation of patients for cochlear implantation.

TREATMENT

In general, conductive hearing losses are amenable to surgical intervention and correction, while sensorineural hearing losses are permanent. The diagnosis of conductive hearing loss is usually straightforward, and the etiology of the conductive deficit is often apparent on physical examination. Atresia of the ear canal can be surgically repaired, often with significant improvement in hearing. Tympanic membrane perforations due to chronic otitis media or trauma can be repaired with an outpatient tympanoplasty. Likewise, conductive hearing loss associated with otosclerosis can be treated by stapedectomy, which is successful in 90 to 95% of cases. Tympanostomy tubes allow the prompt return of normal hearing in individuals with middle-ear effusions. Hearing aids are effective and well-tolerated in patients with conductive hearing losses.

Patients with mild, moderate, and severe sensorineural hearing losses are regularly rehabilitated with hearing aids of varying configuration and strength. Hearing aids have been improved to provide greater fidelity and have been miniaturized. The current generation of hearing aids can be placed entirely within the ear canal, thus reducing the stigma associated with their use. In general, the more severe the hearing impairment, the larger the hearing aid required for auditory rehabilitation. Digital hearing aids lend themselves to individual programming, and multiple and directional microphones at the ear level may be helpful in noisy surroundings. Since all hearing aids amplify noise as

well as speech, the only absolute solution to the problem found thus far is to place the microphone closer to the speaker than the noise source. This arrangement is not possible with a self-contained, cosmetically acceptable device. It is cumbersome and requires a user-friendly environment.

In many situations, including lectures and the theater, hearing-impaired persons benefit from assistive devices that are based on the principle of having the speaker closer to the microphone than any source of noise. Assistive devices include infrared and FM transmission as well as an electromagnetic loop around the room for transmission to the individual's hearing aid. Hearing aids with telecoils can also be used with properly equipped telephones in the same way.

In the event that the hearing aid provides inadequate rehabilitation, cochlear implants are appropriate. Criteria for implantation include severe to profound hearing loss with word recognition score£30% under best aided conditions. Children with congenital and acquired profound hearing impairment are also appropriate candidates for cochlear implantation. Worldwide, more than 20,000 deaf individuals (including 4000 children) have received cochlear implants. Cochlear implants are neural prostheses that convert sound energy to electrical energy and can be used to stimulate the auditory division of the eighth nerve directly. In most cases of profound hearing impairment, the auditory hair cells are lost but the ganglionic cells of the auditory division of the eighth nerve are preserved. Cochlear implants consist of electrodes that are inserted into the cochlea through the round window, speech processors that extract acoustical elements of speech for conversion to electrical currents, and a means of transmitting the electrical energy through the skin. Patients with implants experience sound that helps with speech reading, allows open-set word recognition, and helps in modulating the person's own voice. Usually, within 3 months after implantation, adult patients can understand speech without visual cues. With the current generation of multichannel cochlear implants, nearly 75% of patients are able to converse on the telephone. It is anticipated that improvements in the electrode design and speech processors will permit further enhancement in understanding speech, especially in the presence of background noise.

For individuals who have had both eighth nerves destroyed by trauma or bilateral vestibular schwannomas (e.g., neurofibromatosis type 2), brainstem auditory implants placed near the cochlear nucleus may provide auditory rehabilitation. It is hoped that additional advances may provide benefits similar to those with the cochlear implant.

Tinnitus can often accompany hearing loss. The treatment of tinnitus is particularly problematic. Therapy is usually directed towards minimizing the appreciation of tinnitus. Relief of the tinnitus may be obtained by masking it with background music. Hearing aids are also helpful in tinnitus suppression, as are tinnitus maskers, devices that present a sound to the affected ear that is more pleasant to listen to than the tinnitus. The use of a tinnitus masker is often followed by several hours of inhibition of the tinnitus. Antidepressants have also shown beneficial effect in helping patients deal with tinnitus.

Tinnitus and background noise can significantly affect understanding of speech in individuals with hearing impairment. Hard-of-hearing individuals often benefit from a reduction in unnecessary noise (e.g., radio or television) to enhance the signal-to-noise

ratio. Speech comprehension is aided by lip reading; therefore, the impaired listener should be seated so that the face of the speaker is well illuminated and can be seen at all times. Speaking directly into the ear is occasionally helpful, but usually more is lost in communication than gained when the speaker's face cannot be seen. Speech should be slow enough to make each word distinct, but overly slow speech is distracting and loses contextual and speech-reading benefits. Although speech should be in a loud, clear voice, one should be aware that in sensorineural hearing losses in general and in elderly hard-of-hearing persons in particular, recruitment (the ability to hear loud sounds normally loud) may be troublesome. Above all, optimal communication cannot take place without both parties giving it their full and undivided attention.

PREVENTION

Conductive hearing losses may be prevented by prompt and appropriate antibiotic therapy of adequate duration forAOM and by ventilation of the middle ear with tympanostomy tubes in middle-ear effusions lasting 12 weeks or longer. Loss of vestibular function and deafness due to aminoglycoside antibiotics can largely be prevented by careful monitoring of serum peak and trough levels.

Some 10 million Americans have noise-induced hearing loss, and 20 million are exposed to hazardous noise in their employment. Noise-induced hearing loss can be prevented by avoidance of exposure to loud noise or by regular use of ear plugs or fluid-filled ear muffs to attenuate intense sound. Noise-induced hearing loss results from recreational as well as occupational activities and begins in adolescence. High-risk activities for noise-induced hearing loss include wood and metal working with electrical equipment and target practice and hunting with small firearms. All internal-combustion and electric engines, including snow and leaf blowers, snowmobiles, outboard motors, and chain saws, require protection of the user with hearing protectors. Virtually all noise-induced hearing loss is preventable through education, which should begin before the teenage years. Programs of industrial conservation of hearing are required when the exposure over an 8-h period averages 85 dB on the A scale. Workers in such noisy environments can be protected with preemployment audiologic assessment, the mandatory use of hearing protectors, and annual audiologic assessments.

ACKNOWLEDGEMENT

The authors wish to acknowledge Dr. Joseph B. Martin, who was the co-author of this chapter in the 14th edition.

(Bibliography omitted in Palm version)

Back to Table of Contents

30. INFECTIONS OF THE UPPER RESPIRATORY TRACT - *Marlene Durand*, *Michael Joseph*

Infections of the upper respiratory tract include some of the most common infectious diseases encountered by internists and other primary-care physicians. Pharyngitis, laryngitis, rhinitis, sinusitis, otitis externa, and otitis media account for millions of visits to physicians annually. Although these infections are usually mild enough to be treated on an outpatient basis, the primary-care physician must be able to recognize their serious complications, such as peritonsillar abscess from pharyngitis, subperiosteal abscess from frontal sinusitis, and temporal-bone osteomyelitis from invasive otitis externa. The physician must also identify potentially life-threatening infections of the head and neck, such as epiglottitis, Ludwig's angina, and rhinocerebral mucormycosis.

INFECTIONS OF THE NOSE AND FACE

Skin infections that commonly affect the nose and face include folliculitis, furunculosis, impetigo, and erysipelas. These are discussed in detail elsewhere (Chap. 128, in particularTable 128-1).

Infection of the mucosal surface of the nose is most commonly due to respiratory viruses (e.g., rhinovirus) and presents as acute rhinitis. There are several rare, chronic intranasal infections. *Ozena*, or atrophic rhinitis, is characterized by atrophied mucosa overlaid by foul-smelling dry crusts (Greek *ozein*, "stench"). *Klebsiella ozaenae* is often isolated from nasal cultures, but whether it is a cause of illness or merely a colonizer is unclear. Intranasal irrigation with aminoglycosides (e.g., tobramycin ophthalmic solution) or oral administration of ciprofloxacin has resulted in clinical improvement in some cases. *Klebsiella rhinoscleromatis* causes *rhinoscleroma*, a chronic granulomatous disease of the upper respiratory tract mucosa that is seen in inhabitants of parts of Africa, Asia, and Latin America; it has been described in two patients positive for HIV. Mikulicz cells (foamy histiocytes) are seen in the submucosa of biopsy specimens. Rhinoscleroma can be treated with streptomycin, trimethoprim-sulfamethoxazole, a quinolone, or tetracycline for 2 months. *Pseudomonas mallei* causes *glanders*, a respiratory disease of horses. Infection is rare in humans; nasal inoculation may produce a purulent nasal discharge followed by granulomatous intranasal lesions that ulcerate. Treatment is with sulfadiazine.

Neonatal congenital syphilis may present as rhinitis (snuffles), and the generalized osteochondritis that follows may result in a "saddle-nose" deformity. In *leprosy*, *Mycobacterium leprae* infiltrates the nasal mucosa and may cause chronic nasal congestion and nosebleeds. Involvement of the nasal cartilage may also result in a saddle nose.

Rhinosporidium seeberi is a fungus-like organism, not yet cultured, that causes *rhinosporidiosis*. Pedunculated nasal masses that grow over months or years cause obstruction and a foul odor and must be surgically excised. *Blastomyces dermatitidis*, a fungus prevalent in the Mississippi and Ohio River valleys, usually causes pulmonary disease but may cause chronic ulcerative lesions of the skin and nasal mucosa. *Mucormycosis*, a life-threatening fungal illness that occurs primarily in diabetic patients, may present as black eschars in the nasal cavity (see "Fungal Sinusitis").

THE COMMON COLD

The common cold is a mild, self-limited viral infection of the upper respiratory tract. Adults average two to four colds per year and children six to eight. The most common causes are rhinovirus (40% of cases) and coronavirus (at least 10%), but parainfluenza virus, respiratory syncytial virus, influenza virus, and adenoviruses also account for some cases. Rhinovirus alone has more than 100 different immunotypes, and this diversity has hampered efforts to identify an effective therapy or vaccine. There is no specific treatment for the common cold, although antihistamines, decongestants, and ipratropium bromide nasal spray provide some relief of symptoms. One study found that zinc gluconate lozenges, taken every 2 h, may reduce the duration of symptoms but are associated with nausea in 20% of patients. The value of vitamin C in preventing colds has not been proved.

SINUSITIS

The paranasal sinuses are aerated cavities in the bones of the face that develop as outpouches of the nasal cavity and communicate with this cavity throughout life. The maxillary and ethmoid sinuses are present at birth; the frontal and sphenoid sinuses develop after ages 2 and 7, respectively. Like the nose, the sinuses are lined with respiratory epithelium that includes mucus-producing goblet cells and ciliated cells. The mucous blanket is carried toward the sinus openings (ostia) at a speed of up to 1 cm/min by the beating of the cilia. The ostia are small; the ethmoid sinus ostia, for example, are only 1 to 2 mm in diameter. Delay in the mucociliary transport time or -- more important -- obstruction of the ostia may lead to retained secretions and sinusitis.

Sinusitis is a common problem. In the United States, this infection accounts for millions of office visits annually. The most common type, maxillary sinusitis, is followed in frequency by ethmoid, frontal, and sphenoid sinusitis. A viral infection of the upper respiratory tract is the most common precursor of sinusitis, although only about 0.5% of such infections are complicated by clinically evident acute bacterial sinusitis. Sinusitis develops primarily through ostial obstruction due to mucosal edema. Viral upper respiratory infections also increase the amount of mucus produced and may damage ciliated cells, thereby delaying mucus transport time. Allergic rhinitis is another common cause of ostial obstruction, either by mucosal edema or by polyps. Nasotracheal or nasogastric intubation can result in obstruction of the ostia and is a major risk factor for nosocomial sinusitis in intensive care units. Dental infections may cause 5 to 10% of all cases of maxillary sinusitis; the roots of the upper back teeth (second bicuspid, first and second molars) abut the floor of the maxillary sinus. Other causes of sinusitis include barotrauma from deep-sea diving or airplane travel, mucus abnormalities (e.g., cystic fibrosis), and chemical irritants. Foreign bodies, tumors (e.g., midline granuloma, intranasal lymphoma, or squamous cell carcinoma), and granulomatous diseases (e.g., Wegener's granulomatosis or rhinoscleroma) may all cause sinusitis secondary to obstruction.

ACUTE BACTERIAL SINUSITIS

Manifestations Symptoms of acute sinusitis include purulent nasal or postnasal

drainage, nasal congestion, and sinus pain or pressure whose location depends on the sinus involved. Maxillary sinus pain is often perceived as being located in the cheek or upper teeth; ethmoid sinus pain, between the eyes or retroorbital; frontal sinus pain, above the eyebrow; and sphenoid sinus pain, in the upper half of the face or retroorbital with radiation to the occiput. Sinus pain is frequently worse when the patient bends over or is supine. Fever develops in about half of patients with acute maxillary sinusitis.

Diagnosis The diagnosis of bacterial sinusitis may be difficult, as symptoms may resemble those of the inciting viral upper respiratory infection. The persistence of cold symptoms for 7 to 10 days (or longer than usual for a particular patient) is the most consistent clinical feature of bacterial sinusitis, according to some authors. Four-view sinus x-rays are helpful in the diagnosis of acute sinusitis: radiologic opacity, an air-fluid level, or ³4 mm of sinus mucosal thickening correlates well with active bacterial infection. Computed tomography (CT) of the sinus is much more sensitive than routine radiography, particularly for ethmoid and sphenoid disease. Its use should be reserved for complicated cases and for cases in hospitalized patients, however. In light of the finding that sinus CT often shows reversible acute changes in patients with common colds, it is apparent that routine early use of CT would lead to overdiagnosis of bacterial sinusitis.

Etiology The bacteriology of acute community-acquired maxillary sinusitis has been well defined by studies using direct sinus puncture and aspiration. In children and adults, *Streptococcus pneumoniae* and *Haemophilus influenzae* (not type b), the most common pathogens, cause about one-third and one-fourth of cases, respectively. In children, *Moraxella catarrhalis* is also important, accounting for 20% of cases. Rhinoviruses, influenza viruses, and parainfluenza viruses are found alone or with bacteria in one-fifth of adult cases.

TREATMENT

Empirical therapy for acute bacterial sinusitis should be directed against the common bacterial pathogens; sinus puncture is not indicated in routine cases, and cultures of nasal drainage are not very reliable. Amoxicillin (500 mg orally, three times daily for 10 to 14 days) or trimethoprim-sulfamethoxazole may be effective in the treatment of first-time cases.

[1]The dosages and durations of antimicrobial therapy given in this chapter are appropriate for adults with normal renal function and should be adjusted for children, for patients with impaired renal function, and in light of the response to treatment. Other effective but more expensive antibiotics include amoxicillin/clavulanate, cefuroxime axetil, and clarithromycin. Treatment should be given for 1 to 2 weeks. Intravenous administration of antibiotics may be necessary for the treatment of patients with severe disease who appear toxic. In nosocomial sinusitis, *Staphylococcus aureus* and gram-negative bacilli are most common, and sinus cultures are indicated as an aid in tailoring therapy. Initial broad-spectrum intravenous therapy (e.g., with nafcillin and ceftriaxone) should be adjusted on the basis of culture results. Surgery to widen the ostia and drain thick secretions may be essential in severe acute sinusitis, particularly when ethmoid, frontal, or sphenoid disease fails to respond to initial intravenous therapy.

CHRONIC BACTERIAL SINUSITIS

Manifestations Chronic sinusitis is characterized by symptoms of sinus inflammation lasting³3 months. Most experts believe that this condition is caused by dysfunction of the mucociliary blanket, usually as a result of repeated past infections, rather than by the persistence of bacterial infection. Patients report constant sinus pressure, nasal congestion, and postnasal drainage, especially in the morning. A temperature of³38°C (³100.5°F) is rare and may signify a superimposed acute bacterial infection. Many patients also note a change in nasal discharge (to thick and green) with acute exacerbations.

Diagnosis SinusCTshould be used in all cases of chronic sinusitis to define the extent of disease and to help exclude other diagnoses, such as an obstructing tumor. Patients should be evaluated for allergies and immunodeficiencies (e.g., hypogammaglobulinemia). Evaluation by an otolaryngologist is essential, as this specialist will be able to obtain additional information by an office nasal endoscopic examination. Surgery, now usually done endoscopically, may be necessary to correct blockage of the sinus ostia. This blockage occurs most often in the osteomeatal complex that drains the maxillary, frontal, and anterior ethmoid sinuses. Samples of sinus secretions obtained intraoperatively should be cultured for anaerobes, aerobes, and fungi. Fungal sinusitis may mimic chronic bacterial infection (see "Fungal Sinusitis").

Etiology The bacteriology of chronic sinusitis is not well defined. Nearly all patients with chronic disease, especially those who have had prior sinus surgery, have sinus cultures positive for bacteria. Such cultures may represent colonization rather than infection, however. Patients who have received multiple courses of antibiotics may be colonized by *S. aureus* or by *Pseudomonas* and other gram-negative bacilli. Anaerobes have been isolated from 100% of sinus specimens in some studies but from as few as 2% in others.

TREATMENT

The need for antibiotic therapy must be assessed on an individual basis, with antibiotics chosen in light of recent culture results.

COMPLICATIONS OF BACTERIAL SINUSITIS

Orbital complications of sinusitis, such as orbital cellulitis and orbital abscess, usually arise from ethmoid sinusitis, since the ethmoid is separated from the orbit by only a very thin bone (the lamina papyracea). Patients present with fever, unilateral periorbital edema and erythema, conjunctival injection and chemosis, and proptosis. Eye movement may be decreased; with orbital abscess, the eye is often fixed in the "down and out" position.CT or magnetic resonance imaging (MRI) should be used to rule out an orbital abscess. Treatment of orbital infections should include immediate drainage of any abscess, intravenous administration of broad-spectrum antibiotics -- e.g., nafcillin (1.5 to 2.0 g every 4 h) plus ceftriaxone (2 g/d) -- for at least 7 days, and a consideration of sinus drainage surgery.

Another extracranial complication of sinusitis is frontal subperiosteal abscess (Pott's puffy tumor) from frontal sinusitis. Patients present with a tender doughy swelling over the forehead. Treatment consists of surgical drainage of the abscess and the frontal sinus and 6 weeks of intravenous antibiotic therapy directed at the isolated organisms.

Intracranial complications such as epidural abscess, subdural empyema, meningitis, cerebral abscess, and dural-vein thrombophlebitis may result from sinusitis, particularly from frontal or sphenoid infections. Because the sphenoid sinus sits between the two cavernous sinuses, sphenoid sinusitis is a major cause of cavernous sinus thrombophlebitis.

FUNGAL SINUSITIS

Fungal sinusitis is categorized as noninvasive or invasive. *Noninvasive* disease is chronic and occurs in immunocompetent hosts. It has two forms that are analogous to the noninvasive pulmonary diseases of aspergilloma and allergic bronchopulmonary aspergillosis. A fungus ball (aspergilloma) inside a sinus may cause symptoms of obstruction without invading the mucosa. Typically, only one sinus (often maxillary) is affected, and patients have unilateral symptoms and opacification of only that sinus onCT. Treatment is surgical only, unless special fungal stains show tissue invasion on histopathology. Allergic fungal sinusitis was first described in 1983 and is seen mainly in patients with a history of nasal polyposis and asthma. It is characterized by extremely thick sinus mucus ("allergic mucin") that, on histopathologic examination, is found to contain numerous Charcot-Leyden crystals, eosinophils, and rare fungal hyphae. There is no evidence of tissue invasion. Surgical removal of the inspissated mucus is often curative. Antifungal therapy is not indicated.

Invasive fungal sinusitis presents differently in immunocompetent and immunocompromised hosts. In immunocompromised individuals, fungal disease has an acute presentation. Rhinocerebral mucormycosis is a life-threatening infection due to fungi of the order Mucorales (*Rhizopus*, *Rhizomucor*, *Mucor*, *Absidia*, *Cunninghamella*). Mucormycosis usually involves diabetic patients (70% of cases), half of whom are in ketoacidosis at presentation. Other patients at risk include those with hematologic malignancies, transplant recipients, and patients receiving chronic glucocorticoid or iron chelation (deferoxamine) therapy. Mucormycosis in patients taking deferoxamine is generally due to *Cunninghamella* and is almost always fatal. Symptoms and signs of mucormycosis may be explained by the fungal predilection for blood vessels and nerves and for invasion into the orbital apex and cavernous sinus, with consequent compromise of cranial nerves II through VI. Patients most frequently present with unilateral ocular findings of 1 to 5 days' duration that may mimic bacterial orbital cellulitis: lid swelling and erythema (sometimes bluish in appearance), ptosis, proptosis, decreased extraocular movement, and impaired vision. Retroorbital or periorbital pain is a prominent complaint. There may be either increased or decreased sensation in the first division of the fifth cranial nerve on the involved side; facial palsy with involvement of cranial nerve VII has also been described. Patients may be afebrile and appear nontoxic. Individuals in whom mucormycosis is a consideration should undergo an immediate examination by an otolaryngologist, who will look for intranasal black eschars or necrotic turbinates. If found, these sites should be biopsied and frozen tissue sections examined by a pathologist. In rare cases, the nasal passage appears normal, but biopsy of the middle

turbinate reveals invasive fungi. The finding of tissue invasion by broad-based, nonseptate hyphae necessitates extensive surgical debridement and intravenous therapy with amphotericin B or liposomal amphotericin (Chaps. 200 and207). *Aspergillus* and other filamentous fungi may also cause invasive sinus disease.

Immunocompetent hosts with invasive fungal sinusitis, in contrast, have slowly progressive disease. Fungi in ethmoid and sphenoid sinuses may invade the orbital apex, causing proptosis, ptosis, limitation of eye movement, and decreased vision. Patients may mistakenly be treated with glucocorticoids for presumed optic neuritis or orbital pseudotumor until sinus disease is recognized and biopsies are undertaken. Treatment consists of surgical debridement of the involved sinuses and prolonged intravenous therapy with amphotericin B. In all cases of invasive fungal sinusitis, follow-upCT andMRIshould be conducted frequently to evaluate the progression of disease.

Mortality from invasive fungal sinusitis is high, even among immunocompetent hosts.

EAR AND MASTOID INFECTIONS

AURICULAR CELLULITIS AND PERICHONDRITIS

Auricular cellulitis usually presents as a swollen, erythematous, hot, tender ear. The lobule is especially swollen and red. There may be a history of minor trauma to the ear (e.g., involving earrings, cotton swabs, or scratching). Treatment consists of warm compresses and intravenous administration of antibiotics active against *S. aureus* and streptococci -- e.g., cefazolin (1 g every 8 h) or nafcillin.

Perichondritis, an infection of the perichondrium of the ear, is often accompanied by infection of the underlying cartilage of the pinna (chondritis). Associated interruption of the blood supply to the cartilage may lead to ear deformity. Patients present with a swollen, hot, red, and exquisitely tender pinna, usually with sparing of the lobule. The most common antecedents of the infection are burns, significant trauma to the ear (e.g., as a result of boxing), or ear piercing through the pinna. *Pseudomonas aeruginosa* and *S. aureus* are the most common pathogens. Perichondritis should be treated with antibiotics, such as intravenous ticarcillin/clavulanic acid (3.1 g every 4 h) or intravenous nafcillin plus oral ciprofloxacin, for several weeks. Incision and drainage may be helpful for culture and for resolution of infection, which is often slow. This infection must be distinguished from relapsing polychondritis, a rheumatologic condition (Chap. 325).

OTITIS EXTERNA

The external auditory canal is about 2.5 cm long and is lined by skin. Beneath this skin is cartilage in the lateral half of the canal, temporal bone in the medial half. The skin in the bony portion lacks a subcutaneous layer and is attached directly to the periosteum, an important feature in the pathogenesis of invasive otitis externa (see below). Cerumen, secreted by glands, acidifies the canal and suppresses bacterial growth. However, desquamated skin and retained moisture make the canal especially susceptible to the hydrophilic organism *P. aeruginosa*.

Acute otitis externa, or swimmer's ear, occurs mostly in the summer and may be due to a decrease in canal acidity and resulting bacterial overgrowth. The ear is pruritic and painful, and the canal appears swollen and red. The most common pathogens are *P. aeruginosa*, *S. aureus*, and streptococci. Treatment consists of cleansing of the ear with alcohol-acetic acid mixtures and the administration of topical antibiotic ear drops, such as polymyxin-neomycin (4 drops four times daily for 5 days). Herpes zoster in the external canal causes severe otalgia and is often accompanied by ipsilateral facial paralysis due to the involvement of the geniculate ganglion of cranial nerve VII (Ramsay Hunt syndrome). Reports suggest that treatment with intravenous acyclovir decreases the incidence of permanent facial-nerve palsy, but the results of relevant controlled trials have not yet been published.

Chronic otitis externa causes pruritus rather than ear pain and is often due to irritation from either repeated minor trauma to the canal (e.g., scratching or use of cotton swabs) or drainage of a chronic middle-ear infection. In the latter situation, treatment of chronic otitis media with oral antibiotics will also cover this condition.

Invasive ("malignant") otitis externa is a potentially life-threatening infection, almost always due to *P. aeruginosa*, that slowly invades from the external canal into adjacent soft tissues, mastoid, and temporal bone and eventually spreads across the base of the skull. It occurs primarily in diabetic patients whose diabetes, unlike that of patients with mucormycosis, is usually under control. There is a history of weeks to months of ear pain and drainage, often misdiagnosed as chronic otitis media (an entity that is rarely painful). Examination reveals an edematous canal, with granulation tissue in the posterior wall about halfway down the canal (the region of the cartilage-bone junction). Trismus or partial facial paralysis (cranial nerve VII) is evident in some instances. Cranial nerves IX, X, and XI are occasionally affected as well. Fever is rare in invasive otitis externa and, when it does develop, is usually low-grade.

Laboratory studies generally reveal a normal white blood cell count but a high erythrocyte sedimentation rate.CT andMRIstudies are essential for defining the extent of bone and soft-tissue involvement. CT shows bony destruction of the skull base in advanced cases. For culture, biopsies of granulation tissue in the canal or of deeper tissues are preferable to swab specimens of ear drainage, which may be unreliable. In nearly all cases, antibiotics should be withheld until a deep-tissue specimen is obtained for culture and pathologic examination. Once this specimen has been collected, empirical therapy with intravenous antibiotics active against *Pseudomonas* -- e.g., ticarcillin (3 g every 4 h), piperacillin, or ceftazidime, plus an aminoglycoside) -- may be started intraoperatively. To avoid ototoxicity, ciprofloxacin should be substituted for the aminoglycoside if cultures grow a *Pseudomonas* strain that is sensitive to this drug. In more than 95% of cases, *P. aeruginosa* is the pathogen involved; in the remaining cases, the pathogens include *Staphylococcus epidermidis*, *Aspergillus*, *Fusobacterium*, and *Actinomyces*. Systemic antibiotic treatment should be continued for 6 to 8 weeks. In early cases due to sensitive *Pseudomonas* strains, oral ciprofloxacin alone (750 mg twice daily for 6 weeks) may follow the initial 2 weeks of combination intravenous therapy.

ACUTE OTITIS MEDIA

The middle ear is connected to the nasopharynx via the eustachian tube. When this tube is blocked, fluid collects in the middle-ear and mastoid cavities, providing a culture medium for any bacteria present. Acute otitis media (AOM), or middle-ear infection, may result. Viral upper respiratory infections, which can cause edema of the eustachian tube mucosa, often precede or accompany episodes of AOM. Otitis media, like upper respiratory tract infections, is most common in fall, winter, and spring. The incidence of AOM declines with age. More than two-thirds of children under age 3 have had at least one episode of AOM; the prevalence among adults is only 0.25%.

Symptoms include ear pain, fever, and decreased hearing acuity. On examination, the tympanic membrane moves poorly with insufflation and is usually red, opaque, bulging, or retracted. Spontaneous perforations of the tympanic membrane and otorrhea are occasionally documented.

The bacteriology ofAOM has been delineated for pediatric disease: *S. pneumoniae* (35%), *H. influenzae* (25%), and *M. catarrhalis* (15%) are the most common organisms. Viruses, either alone or with bacteria, are found in one-quarter of pediatric cases. Small studies of AOM in adults have also found *S. pneumoniae* (21%) and *H. influenzae* (26%) to be the most common pathogens. More than 90% of *H. influenzae* infections are due to nontypable strains: those due to type b may be accompanied by bacteremia or meningitis.

TREATMENT

Treatment of otitis media is empirical, as diagnostic tympanocentesis is indicated only for patients who appear toxic, who are immunocompromised, or whose infection is refractory to initial therapy. Although about one-third of *H. influenzae* strains and at least three-quarters of *M. catarrhalis* strains areb-lactamase producers, most authorities still find amoxicillin therapy to be successful in routine cases. Other drugs effective against most b-lactamase-positive strains include amoxicillin/clavulanate (875 mg by mouth twice daily for 7 to 10 days), trimethoprim-sulfamethoxazole, erythromycin/sulfisoxazole, clarithromycin, and second-generation oral cephalosporins (e.g., loracarbef, cefpodoxime proxetil, and cefuroxime axetil). Penicillin resistance in pneumococci, now a major problem, is not mediated by b-lactamase (Chap. 138). Strains exhibiting intermediate resistance may respond to therapy with high-dose amoxicillin or to clindamycin, erythromycin, or trimethoprim-sulfamethoxazole. Quinolones such as levofloxacin, although not approved for use in children, may be effective in adults. Serious infections or those due to highly resistant strains require treatment with parenteral ceftriaxone or vancomycin. Adjunctive treatment ofAOM with antihistamines is of no proven benefit.

Recurrent episodes ofAOM in children are due to the same pathogens that cause primary AOM (*S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*). Most early recurrences (75%), however, are not relapses but are due to different organisms or to different strains of the organism that caused the initial episode. The pattern of recurrent AOM in adults is presumably similar but has not been well studied. Treatment for recurrent AOM should include drugs with activity against resistant strains. Patients with frequent recurrences (e.g., three episodes within 6 months) may benefit from antibiotic prophylaxis with once-daily amoxicillin or sulfisoxazole during the winter months,

although this benefit must be weighed against the risk of selecting more antibiotic-resistant strains of bacteria.

SEROUS OTITIS MEDIA

Otitis media with effusion, or serous otitis media, is characterized by the persistence of middle-ear fluid for several months without other signs of infection. This condition is associated with a 25-dB hearing loss in the affected ear. Cultures of middle-ear fluid are usually negative. Although some clinical trials have found that effusions resolve sooner in antibiotic-treated children than in controls, antibiotics are generally not recommended because the risk of increasing antibiotic resistance in the population is thought to outweigh the small benefit observed. Adenoidectomy, myringotomy, or tympanostomy tubes have been shown to decrease the duration of effusion in children.

CHRONIC SUPPURATIVE OTITIS MEDIA

In chronic suppurative otitis media, patients have painless hearing loss and intermittent purulent ear drainage. On examination, there is a central perforation in the tympanic membrane and purulent drainage from the middle ear. If a cholesteatoma is present, the perforation is peripheral. Culture of draining fluid reveals *P. aeruginosa* (40%), *S. aureus* (20%), *Klebsiella* (20%), and other enteric gram-negative bacilli. Anaerobes are found in 50% of cases, usually in mixed culture with aerobes.CTshould be used to help evaluate a surgically treatable nidus of infection, such as a cholesteatoma or mastoid sequestrum. For therapeutic purposes, patients are divided into two groups: those with and without cholesteatoma. Those in the former group are cured with surgical excision of the cholesteatoma. Those without cholesteatoma require repeated courses of topical antibiotic drops for relapse of "active" (draining) disease, and true cures are rare. The role of systemic antibiotics is unclear. In one study, a course of intravenous antibiotics produced long-term success in 78% of children without cholesteatoma who had persistent otorrhea despite topical and oral antibiotic therapy.

Tuberculous otitis media is rare and is frequently misdiagnosed. It mimics nontuberculous chronic suppurative otitis media, but ear drainage fails to respond to routine antibiotics. On examination, the tympanic membrane often has multiple perforations, and "pearly" or "flabby" tissue is seen in the middle ear. Only 30% of patients have evidence of active tuberculosis on chest x-ray. Treatment is the same as for other types of extrapulmonary tuberculosis.

MASTOIDITIS

The mastoid is the portion of the temporal bone posterior to the ear that contains a honeycomb of air cells lined with respiratory epithelium. These air cells connect with the middle ear. Fluid in the middle ear, a prelude to otitis media, is almost always accompanied by fluid in the mastoid. True mastoiditis, however, has become rare in the antibiotic era, probably because of prompt treatment of otitis media.

Mastoiditis is characterized by erosion of the bony partitions between the mastoid air cells. Patients with acute mastoiditis present with pain, tenderness, and swelling over the mastoid. When there is an overlying subperiosteal abscess or cellulitis, the pinna is

pushed out and forward.CT may show bony destruction or a drainable mastoid abscess.

The reported bacteriology of mastoiditis has varied. Some cases involve organisms similar to those implicated in AOM(*S. pneumoniae*, *H. influenzae*); others are attributable to *S. aureus* and gram-negative bacilli, including *Pseudomonas*. Ideally, therapy should be guided by the results of cultures of middle-ear fluid obtained by tympanocentesis. Initial broad-spectrum therapy, such as that with intravenous ticarcillin/clavulanate plus gentamicin or ciprofloxacin, can later be narrowed.

COMPLICATIONS OF OTITIS MEDIA AND MASTOIDITIS

Extracranial complications include hearing loss, labyrinthitis and resulting vertigo, and facial-nerve palsy. Additional complications from mastoiditis develop when infection tracks under the periosteum of the temporal bone to cause a subperiosteal abscess or breaks through the mastoid tip to cause a neck abscess deep to the sternocleidomastoid muscle (Bezold's abscess). Intracranial complications include epidural abscess, dural venous thrombophlebitis (usually sigmoid sinus), meningitis, and brain abscess.

INFECTIONS OF THE ORAL CAVITY AND PHARYNX

ORAL CAVITY INFECTIONS

The oral cavity extends from the lips to the circumvallate papillae of the tongue and is heavily colonized with viridans streptococci and anaerobes. These organisms can cause several infections in this area. *Gingivitis* is an infection of the gums, the earliest form of periodontal disease. Anaerobes residing in the mouth, especially anaerobic gram-negative rods such as *Prevotella intermedia*, are the most common pathogens. Patients with *Vincent's angina*, also called *acute necrotizing ulcerative gingivitis* or *trench mouth*, have halitosis and ulcerations of the interdental papillae. Oral anaerobes are the cause, and therapy with oral penicillin plus metronidazole or with clindamycin alone is effective in both this condition and gingivitis.

Ludwig's angina is a rapidly spreading, life-threatening cellulitis of the sublingual and submandibular spaces that usually starts in an infected lower molar. Patients are febrile and may drool the secretions they cannot swallow. A brawny, boardlike edema in the sublingual area pushes the tongue up and back. Airway obstruction may result as the infection spreads to the supraglottic tissues. Treatment with intravenous antibiotics active against streptococci and oral anaerobes -- e.g., ampicillin/sulbactam (3 g every 6 h) or high-dose penicillin plus metronidazole -- should be followed by oral antibiotic therapy, with a total treatment duration of 14 days. Airway monitoring is also essential. Intubation or tracheostomy may be necessary.

Noma, or *cancrum oris*, is a fulminant gangrenous infection of the oral and facial tissues that occurs in severely malnourished and debilitated patients and is especially common among children. Beginning as a necrotic ulcer in the gingiva of the mandible, noma is caused by oral anaerobes, especially fusospirochetal organisms (e.g., *Fusobacterium nucleatum*). It is treated with high-dose penicillin, debridement, and correction of the underlying malnutrition.

Herpes simplex commonly causes cold sores of the lips but may also cause painful vesicles on the tongue and buccal mucosa. Primary infection may require intravenous hydration and should be treated with acyclovir. *Thrush*, or oropharyngeal candidiasis, is an infection caused by *Candida* spp. such as *C. albicans*. It occurs in neonates, patients who have received prolonged antibiotic therapy, and immunocompromised patients. More than 90% of patients with AIDS develop thrush. Patients with thrush report a "burning" tongue or "raw" throat and, on examination, have white plaques on the tongue and oral mucosa. Treatment consists of topical antifungal agents (clotrimazole, nystatin) or oral fluconazole. Therapy for fluconazole-resistant thrush in patients with AIDS may be difficult; itraconazole oral solution or amphotericin B oral suspension may be effective.

PHARYNGITIS

Most cases of pharyngitis are thought to be viral. Many occur as part of common colds caused by rhinovirus, coronavirus, or parainfluenza virus. Patients have a scratchy or sore throat as well as coryza and cough. The pharynx is inflamed and edematous, but no exudate is evident. Influenza virus and adenovirus may cause a particularly severe sore throat, along with fever and myalgias. In infection with either of the latter viruses, there is pharyngeal erythema and edema; however, adenovirus infection also commonly causes an exudate, thus mimicking streptococcal pharyngitis. *Infectious mononucleosis* due to Epstein-Barr virus often causes a severe sore throat. Exudative pharyngitis or tonsillitis is documented in half of mononucleosis cases and may also mimic streptococcal infection. *Herpangina*, caused by coxsackievirus, is characterized by fever, sore throat, myalgias, and a vesicular enanthem on the soft palate between the uvula and the tonsils. There are usually only two to six lesions, which begin as small papules that vesiculate and then ulcerate. Fever and nonexudative pharyngitis are common symptoms of the acute retroviral syndrome that develops several weeks after infection withHIV.

The most important bacterial cause of pharyngitis is group A *Streptococcus* (*S. pyogenes*). This organism is responsible for about 15% of all cases of pharyngitis and can cause important complications, both suppurative (peritonsillar and retropharyngeal abscess) and nonsuppurative (scarlet fever, streptococcal toxic shock syndrome, rheumatic fever, acute poststreptococcal glomerulonephritis). Fever, severe sore throat, cervical adenopathy, and inflammation of the tonsils and pharynx (which are covered with exudate) are classic findings. However, many cases of streptococcal pharyngitis are mild, with minimal erythema and no exudate, and mimic the pharyngitis of the common cold. Although some patients may in fact have viral pharyngitis and may simply be colonized with group A streptococci, these individuals must nevertheless be treated for presumed streptococcal pharyngitis. Diagnosis is made by culture. Rapid antigen tests are now available. These tests are less sensitive than they are specific: a positive test may be considered equivalent to a positive culture, but a negative test requires culture confirmation. Either a single dose of intramuscular benzathine penicillin (1.2 million units) or a 10-day course of oral penicillin (250 mg four times daily) or erythromycin is necessary to eradicate the organism. Sensitivity to erythromycin should be verified if this agent is used, as an increase in erythromycin resistance has been noted, especially in Europe. Other antibiotics active against streptococci may be used

(e.g., amoxicillin, cefuroxime), and one trial showed that 4 days of cefuroxime therapy was as effective as 10 days of penicillin treatment in eradicating the organism. However, studies of the prevention of rheumatic fever are available only for penicillin (Chap. 235).

Other bacterial causes of pharyngitis include groups C and G *Streptococcus*, *Neisseria gonorrhoeae* (Chap. 147), *Arcanobacterium haemolyticum*, *Yersinia enterocolitica*, and -- very rarely -- *Corynebacterium diphtheriae* (Chap. 141). In addition, *Mycoplasma pneumoniae* (Chap. 178) and *Chlamydia pneumoniae* (Chap. 179) can cause pharyngitis.

A peritonsillar abscess (*quinsy*) may follow untreated streptococcal pharyngitis. Oral anaerobes also play a role in quinsy. Patients have a severe sore throat and speak with a "hot-potato" voice. Examination reveals pronounced unilateral peritonsillar swelling and erythema causing deviation of the uvula. Immediate aspiration by an otolaryngologist is required in conjunction with antibiotic therapy -- e.g., ampicillin/sulbactam (3 g intravenously every 6 h), penicillin plus metronidazole, or clindamycin.

LARYNGITIS, CROUP, AND EPIGLOTTITIS

LARYNGITIS

Laryngitis is characterized by hoarseness. Most cases of acute laryngitis are caused by viruses (rhinovirus, influenza virus, parainfluenza virus, coxsackievirus, adenovirus, or respiratory syncytial virus). Acute laryngitis may also be associated with group A *Streptococcus* and *M. catarrhalis*. Laryngitis must be differentiated from epiglottitis (see below). The goal of treatment is merely the relief of symptoms except when throat cultures are positive for group A *Streptococcus* (in which case penicillin should be used).

Chronic laryngitis due to infection is rare and must be distinguished from hoarseness of neoplastic etiology. *Tuberculous laryngitis* may be mistaken for laryngeal cancer when assessed by direct laryngoscopy. Laryngeal and supraglottic lesions include mucosal hyperemia and thickening, nodules, and ulcerations. In one study, a history of fever and night sweats was rare, and the most common chest radiographic finding was apical thickening and fibrosis. Biopsy reveals granulomas with acid-fast bacilli. Cultures should be performed to confirm the diagnosis and evaluate the sensitivities of the pathogen. Laryngeal tuberculosis is highly contagious and should be managed with the same precautions and therapy used for active pulmonary disease (Chap. 169). Fungal infections causing laryngitis include histoplasmosis (Chap. 201), blastomycosis (Chap. 203), and candidiasis (Chap. 205). *Histoplasma* and *Blastomyces* may cause nodules on the larynx, with or without ulcerations. *Candida* may cause laryngitis, along with thrush, in immunosuppressed patients or in patients with chronic mucocutaneous candidiasis.

CROUP

Croup, or acute laryngotracheobronchitis, is an infection of the upper and lower respiratory tract that causes marked subglottic edema. It mainly affects 2- and

3-year-old children and usually follows the onset of upper respiratory tract infection by 1 to 2 days. Symptoms include fever, hoarseness, a "seal's bark" cough, and inspiratory stridor. The most common etiology is parainfluenza virus, although croup may also be caused by other respiratory viruses (e.g., influenza or respiratory syncytial virus).

Croup must be differentiated from epiglottitis (see below). Epiglottitis usually progresses more rapidly and produces a more toxic appearance. Neck x-rays may be helpful but do not reliably exclude epiglottitis. In croup, the anterior-posterior neck x-ray shows subglottic edema (the "hourglass sign"); in epiglottitis, the lateral neck view shows a thick epiglottis.

Patients with severe croup should be hospitalized, monitored for hypoxemia through pulse oximetry, and watched for airway obstruction requiring intubation. Humidification is commonly prescribed, but few controlled trials have assessed its benefit. Nebulized racemic epinephrine provides temporary (2-h) improvement in patients with marked stridor, but such patients must be observed for rebound edema. Glucocorticoid therapy, either nebulized or parenteral, is clearly beneficial, and its effects are often evident within 1 h. One trial found that treatment with a single dose of either intramuscular dexamethasone or nebulized budesonide reduced the need for hospitalization of children with moderately severe croup by more than 50%.

EPIGLOTTITIS

Acute epiglottitis (supraglottitis) is a life-threatening, rapidly progressive cellulitis of the epiglottis that may cause complete airway obstruction. It begins as a cellulitis between the tongue base and the epiglottis that pushes the epiglottis posteriorly. The epiglottis itself then becomes swollen, threatening the airway. Before the introduction of *H. influenzae* type b (Hib) vaccine, epiglottitis was most common among children 2 to 4 years old. The disease is now rare in children, since the vaccine has reduced the incidence of invasive disease due to Hib by more than 95%. The incidence in adults has not changed.

The typical young child with epiglottitis has a several-hour history of fever, irritability, dysphonia, and dysphagia and presents sitting forward and drooling. Adolescents and adults usually have a less fulminant presentation, with symptoms (especially sore throat) of 1 or 2 days' duration. Adults may present with dyspnea (25%), drooling (15%), and stridor (10%). Epiglottitis constitutes a medical emergency, as airway occlusion may occur suddenly. Lateral neck films showing an enlarged epiglottis (the "thumb sign") are helpful if positive but may be falsely negative. The value of obtaining these films has also been questioned because doing so may cause a critical delay in securing the airway. Direct viewing of the pharynx by use of a tongue blade should not be attempted, as immediate laryngospasm and airway obstruction may result. Instead, a child with suspected epiglottitis should be transported -- while sitting up -- to the operating room for visualization of the epiglottis with a fiberoptic laryngoscope, with preparations made for immediate airway control. If the epiglottis is cherry-red, an uncuffed endotracheal tube should be placed. Diagnosis in adults is also made by direct viewing of the epiglottis with a flexible fiberoptic laryngoscope, again only after preparations are made to secure the airway.

All patients must be closely monitored in an intensive care unit and should be given antibiotics active against *H. influenzae*. BeforeHibvaccine became available, this organism was responsible for nearly all pediatric cases and was isolated from the blood of almost 100% of the affected children. In adults, blood cultures are positive in about 25% of cases, all of which are due to *H. influenzae*. Other pathogens isolated from the pharynx of adults with epiglottitis include *H. parainfluenzae*, *S. pneumoniae*, group A *Streptococcus*, and (rarely) *S. aureus*; the correlation between throat and epiglottis cultures is unclear, however. Children may be treated with intravenous cefuroxime, ceftriaxone, ampicillin/sulbactam, or trimethoprim-sulfamethoxazole. Adults may be treated for at least 7 days with cefuroxime, ampicillin/sulbactam (3 g intravenously every 6 h), or nafcillin plus ceftriaxone; those highly allergic to penicillin may be given clindamycin plus either trimethoprim-sulfamethoxazole or ciprofloxacin. If the patient with *H. influenzae* epiglottitis has household contacts that include an unvaccinated child under age 4, all members of the household and the patientshould receive prophylactic rifampin to eradicate the carriage of *H. influenzae*.

DEEP NECK INFECTIONS

Deep neck infections may be life-threatening because of airway compromise, involvement of the carotid sheath, or spread into the mediastinum.

SUBMANDIBULAR SPACE INFECTIONS

See *Ludwig's angina* above (under "Oral Cavity Infections").

LATERAL PHARYNGEAL SPACE INFECTIONS

The lateral pharyngeal space, also called the parapharyngeal or pharyngomaxillary space, is in the superior lateral portion of the neck and extends from the hyoid bone to the base of the skull. It lies deep to the lateral wall of the pharynx and is lateral to the tonsil and carotid sheath and medial to the parotid gland. Infection in this space may follow tonsillitis, pharyngitis with adenoid involvement, parotitis, mastoiditis, or periodontal infection.

On presentation, most patients appear toxic and have fever, sore throat, pain on swallowing, and leukocytosis. Infection confined to the posterior (retrostyloid) portion of the lateral pharyngeal space causes swelling of the lateral pharyngeal wall, which may be missed because it is behind the palatopharyngeal arch. Involvement of the anterior portion of this space causes medial displacement of the tonsil, swelling over the parotid gland, and trismus. Rigidity of the neck or torticollis toward the opposite side may develop. Diagnosis is confirmed byCT with contrast.

Treatment includes securing of the airway, surgical drainage in the operating room, and administration for at least 10 days of intravenous antibiotics active against streptococci and oral anaerobes (e.g., ampicillin/sulbactam, 3 g every 6 h). Major complications result from involvement of the carotid sheath and the vessels it contains. These complications are frequently fatal and include jugular vein thrombophlebitis, erosion into the carotid artery, and mediastinitis. Jugular vein thrombophlebitis is characterized by high fevers, chills, and neck tenderness at the angle of the mandible. When it is caused

by *Fusobacterium necrophorum*, it may be accompanied by sepsis and septic pulmonary emboli (*Lemierre's syndrome*). Erosion into the carotid artery is usually heralded by repeated small bleeds into the mouth. The involvement of adjacent cranial nerves may result in ipsilateral Horner's syndrome, hoarseness, or unilateral tongue paresis. Extension of infection along the carotid sheath into the posterior mediastinum results in mediastinitis and a mortality of 50%.MRI is useful in delineating carotid and jugular involvement.

RETROPHARYNGEAL SPACE INFECTIONS

The retropharyngeal space lies between the pharynx and the prevertebral fascia and extends from the base of the skull into the mediastinum. Infection in this space may result from the spread of lateral pharyngeal space infection or from the lymphatic spread of infection in more cephalad sites (posterior sinuses, adenoids, nasopharynx) to the retropharyngeal lymph nodes. Retropharyngeal abscess is most common among infants and young children, probably because the retropharyngeal nodes later involute. Retropharyngeal abscess may also follow trauma to the posterior pharynx (e.g., endoscopy in adults, lollipop-stick perforation in children) or may result from anterior extension of infection from cervical osteomyelitis.

Symptoms include fever, marked difficulty and pain with swallowing, and a "hot-potato" voice. Physical examination may document drooling, nuchal rigidity, and bulging of the posterior pharyngeal wall. Advanced cases include dyspnea and stridor. Diagnosis may be confirmed by a lateral neck soft-tissue x-ray orCT scan. Treatment requires securing of the airway and emergency surgical drainage. Intravenous antibiotics should be given; the agents chosen should be active against streptococci, oral anaerobes, *S. aureus*, and *H. influenzae* (e.g., ampicillin/sulbactam alone or clindamycin plus ceftriaxone). Potential complications include airway obstruction, intraoral rupture of the abscess causing aspiration pneumonia, and mediastinitis.

(Bibliography omitted in Palm version)

Back to Table of Contents

31. ORAL MANIFESTATIONS OF DISEASE - *John S. Greenspan*

A thorough oral examination, to include the oral and pharyngeal soft tissues as well as the teeth, is an important part of the physical examination. The common oral diseases are due to infection by bacteria, fungi, or viruses. The complex development of the orofacial structures leads to close interposition of diverse tissues, which are prone to developmental anomalies, growth disturbances, and neoplasia.

DISEASES OF THE TEETH

DENTAL CARIES, PULPAL AND PERIAPICAL DISEASE, AND COMPLICATIONS

Dental caries is a destructive disease of the hard tissues of the teeth due to infection with *Streptococcus mutans* and other bacteria. In the United States, fewer than half of those 17 years and younger now have carious lesions, although in many segments of the population and in developing countries the disease is more common. Artificial fluoridation of water to a level of 1 part per million, fluoride-containing toothpastes, and topical fluoride administration have reduced the incidence. Conversely, retention of teeth and the aging of the population have led to an increase in root caries. Increasing numbers of individuals surviving cancer therapy and other special populations (diabetic patients and those with xerostomia due to Sjogren's syndrome or to medications) may experience severe caries unless appropriate topical fluoride prophylaxis is used. Treatment of caries involves removal of the softened and infected hard tissues, sealing of the exposed dentine, and restoration of the lost tooth structure with silver amalgam, composite plastic, gold, or porcelain.

If the carious lesion progresses, infection of the dental pulp may occur, causing *acute pulpitis*. The tooth may become sensitive to hot or cold. When severe continuous throbbing pain ensues, pulp damage is irreversible, and root canal therapy becomes necessary. The contents of the pulp chamber and root canals are removed, followed by thorough cleaning, antisepsis, and filling with an inert material. Alternatively, extraction of the tooth may be indicated.

If the pulpitis is not treated successfully, infection may spread beyond the tooth apex into the periodontal ligament. Acute inflammation causes pain on chewing or on percussion, and a *periapical abscess* may form. Chronic inflammation may be painless or produce only slight pain, and a *periapical granuloma* may form within the alveolar bone. Proliferation of epithelial cell rests may convert the granuloma into a *periapical cyst*. Periapical radiolucencies may occur with the granuloma or the cyst but not with the abscess, unless it forms as a complication of one or two lesions. The pus in the periapical abscess may track through the alveolar bone into soft tissues, causing cellulitis and bacteremia, or may discharge into the oral cavity (*parulis* or *gumboil*), into the maxillary sinus, or through the skin of the face or submandibular area. A severe form of cellulitis, *Ludwig's angina*, originates from an infected mandibular molar, involves the submandibular space, and extends throughout the floor of the mouth, with elevation of the tongue, dysphagia, and difficulty breathing. Glottal edema may occur, necessitating tracheotomy.

EFFECT OF SYSTEMIC FACTORS ON TEETH

Enamel hypoplasia of the primary and/or permanent teeth, manifested by alterations ranging from white spots to gross defects in the surface structure of the crowns, may be caused by disturbances of calcium and phosphate metabolism such as are found in vitamin D-resistant rickets, hypoparathyroidism, gastroenteritis, and celiac disease. Premature birth or high fevers may also give rise to enamel hypoplasia. Tetracycline, when given during the second half of pregnancy, in infancy, and in childhood up to 8 years of age, causes both a permanent discoloration of the teeth and enamel hypoplasia. Daily ingestion of more than 1.5 mg fluoride can result in enamel discoloration (*mottling*). Larger teeth are associated with maternal diabetes, maternal hypothyroidism, and large birth size. Tooth size is reduced in Down's syndrome. Systemic disease may give rise to pain that simulates pulpal disease. Maxillary sinusitis is frequently manifested as pain in the maxillary teeth, including sensitivity to thermal changes and percussion. Angina pectoris may result in pain referred to the lower jaw, probably through the vagus nerve.

PERIODONTAL DISEASES

In adults, chronic destructive periodontal disease (*pyorrhea*) is responsible for more loss of teeth than caries, particularly in the aged. However, the prevalence and incidence of periodontal disease also appears to be declining in the United States. The most common form of periodontal disease starts as inflammation of the marginal gingiva (*gingivitis*), which is painless, although the gingiva may bleed on brushing. The disease spreads to involve the periodontal ligament, alveolar bone is slowly resorbed, and periodontal ligament attachment between tooth and bone is lost. The soft tissue separates from the tooth surface, causing "pocket" formation with bleeding on probing and during chewing. Acute inflammation may become superimposed on this chronic process, with the production of pus and the formation of a *periodontal abscess*. Ultimately, extreme bone loss, tooth mobility, and recurrent abscess formation lead to tooth exfoliation or may mandate tooth extraction.

Gingivitis and periodontitis are infections associated with the accumulation of *bacterial plaque*, which may become mineralized (*calculus*) and can be prevented by appropriate *oral hygiene* measures, including tooth brushing, flossing, antibacterial mouth rinses, and the removal of impacted food debris. Poorly fabricated or deteriorated restorations may contribute through overextended or inadequate margins. Therapy consists of removal of plaque and calculus, debridement of the pocket lining and superficial infected cementum, and elimination of other contributing factors.

Periodontal disease appears to be a group of conditions, including *adult periodontitis*, associated with *Porphyromonas gingivalis*, *Prevotella intermedia*, and other gram-negative organisms. *Localized juvenile periodontitis* (LJP) causes rapid, severe pocketing and bone loss and is associated with *Actinobacillus actinomycetemcomitans*, *Capnocytophaga*, *Eikenella corrodens*, and other anaerobes. *Acute necrotizing ulcerative gingivitis* (ANUG) involves sudden inflammation of the gingivae with necrosis, tissue loss, pain, bleeding, and halitosis and is associated with *P. intermedia* and spirochetes. ANUG and an aggressive and rapid form of periodontitis (*necrotizing ulcerative periodontitis*) are seen in association with HIV infection. Some of these cases progress to a destructive gangrene-like lesion of oral soft tissues and bone (*necrotizing*

stomatitis) resembling the *noma* seen in severely malnourished populations. Therapy involves local antibacterial measures, debridement, and, in severe cases, systemic antibiotics effective against anaerobes.

Host factors may be involved in the pathogenesis of periodontal disease in other populations as well. Severe periodontal disease may occur in persons with *Down's syndrome* and *diabetes mellitus*. During pregnancy there may be severe gingivitis and the formation of localized *pyogenic granulomas*. Certain drugs, notably the anticonvulsant *phenytoin* and the calcium channel blocker *nifedipine*, cause *fibrous hyperplasia* of the gingiva, which may cover the teeth, interfere with eating, and be unsightly. *Idiopathic familial gingival fibromatosis* may appear similar. Surgery may correct both conditions; change in medication may reverse the drug-induced form. The oral cavity is a significant reservoir for *Helicobacter pylori*. Uncontrolled diabetes mellitus leads to an exacerbation of oral infection, notably periodontal disease. In individuals genetically predisposed to diabetes, periodontal disease may also precipitate or exacerbate the diabetes. Oral infection has been proposed to contribute to coronary atherosclerosis as well as pregnancy outcomes such as premature labor and low birthweight.

Periapical and periodontal bacterial infections can cause transient bacteremia after tooth extraction and even routine dental prophylaxis. Antibiotic coverage is appropriate in patients with heart valves susceptible to infection or those with prosthetic joints.

DISEASES OF THE ORAL MUCOSA

INFECTIONS

Most oral mucosal diseases involve microorganisms (Table 31-1).

PIGMENTED LESIONS See Table 31-2

DERMATOLOGIC DISEASES SeeTables 31-1,31-2, and31-3and Chaps. 55,56,57,58,59, and60

DISEASES OF THE TONGUE See Table 31-4

HALITOSIS SeeTable 31-5

HIV DISEASE AND AIDS (SeeTable 31-6 and alsoChaps. 191 and309)

Immunosuppression induced by HIV infection predisposes to numerous oral infections, neoplasms, and autoimmune and idiopathic lesions. *Oral candidiasis* (Plate IID-43) and *hairy leukoplakia* (Plate IID-42) [a benign epithelial hyperplasia associated with Epstein-Barr virus (EBV)] are common features of HIV disease and often precede or accompany full-blown AIDS. Oral Kaposi's sarcoma and lymphoma are diagnostic of AIDS. Oral candidiasis is easily treated with topical or systemic antifungals: nystatin oral pastilles, clotrimazole oral troches, nystatin vaginal tablets used orally, fluconazole, and ketoconazole. While most oral lesions of HIV disease are also found in the general population. Necrotizing ulcerative periodontal disease and hairy leukoplakia are strongly

associated with HIV infection and are otherwise very rare.

HEMATOLOGIC AND NUTRITIONAL DISEASE

Gingival bleeding, necrotic ulcers, and enlargement due to malignant infiltrates are seen in all forms of leukemia, particularly *monocytic leukemia*. In *agranulocytosis* severe oral mucosal ulcers are seen, while in *thrombocytopenia* oral petechiae, ecchymoses, and gingival bleeding occur. In *Plummer-Vinson syndrome* (Chap. 105), atrophy of oral mucosa, particularly the tongue papillae, causes redness and soreness as well as dysphagia and is associated with increased susceptibility to oral cancer. A smooth tongue can also be seen in *pernicious anemia* (Chap. 107). Severe oral mucositis with ulcers, candidiasis, bacterial infections, and xerostomia complicate radiation therapy for head and neck cancers. Chemotherapy may also cause mucositis. Although now rarely seen in the United States, oral features of vitamin deficiency include oral mucositis and ulcers, glossitis, and burning sensations in the tongue (*B group vitamin deficiency*) and petechiae, gingival swelling, bleeding, and ulceration as well as loosening of teeth (*scurvy* of vitamin C deficiency).

DISEASES OF THE SALIVARY GLANDS

The major and minor salivary glands can be involved in mumps, sarcoidosis, tuberculosis, lymphoma, and Sjogren's syndrome (Chap. 314). The latter may cause dry eyes and dry mouth (*xerostomia*) and be associated with features of connective tissue diseases, including rheumatoid arthritis or systemic lupus erythematosus. Xerostomia may also be due to medications such as diuretics, antihistamines, or tricylic antidepressants as well as radiation therapy for head and neck cancer. Without lysozyme-rich saliva, *cervical or incisal caries* and oral candidiasis may develop. Management includes fluoride mouth rinses and topical applications, saliva substitutes, salivary stimulation with sugarless candies, and the avoidance of sugar-containing drinks or food. Candidiasis is treated with nystatin or other antifungals. Salivary stones (*sialolithiasis*), usually in the duct of a major salivary gland, cause *sialoadenitis* with pain and swelling, often on eating, especially tart foods such as lemons.

The most common neoplasm of the salivary glands is the *pleomorphic adenoma*, which is benign but will recur unless fully resected; malignant tumors include *mucoepidermoid carcinoma*, *adenoid cystic carcinoma*, and *adenocarcinoma*. The pleomorphic adenoma causes a firm, slowly growing mass in the parotid, palate, or cheek, whereas malignant tumors grow faster and can cause ulceration and invade nerves, producing numbness or facial paralysis.

NEUROLOGIC DISTURBANCES AND OROFACIAL PAIN

The mouth and face may be the site of pain from a number of vascular, neurologic, muscle/connective tissue, or joint conditions. Interdisciplinary diagnosis and management programs involving neurologists, restorative dentists, oral surgeons, otorhinolaryngologists, and other specialists, together with new imaging techniques to diagnose or exclude organic lesions, have begun to clarify this complex field. *Temporal arteritis* causes pain in the face, jaws, and tongue and may mimic temporomandibular joint disease. Glucocorticoids may provide relief. *Myofascial pain* is a dull, constant ache

with local tenderness in the muscles of the jaws and difficulty in opening the mouth. Teeth clenching and grinding (*bruxism*) may play a role. *Arthralgia* of the temporomandibular joint causes local pain, which may extend to the face and head. Both myofascial pain and arthralgia can be relieved with heat, rest, and anti-inflammatory agents. Displacement of the meniscus or condyle may cause pain, clenching, or locking of the mandible in the open position. The joint may become involved in *osteoarthritis* with minimal symptoms, whereas *rheumatoid arthritis* causes pain and swelling in the joint and limitation of movement. *Ankylosis* may occur, necessitating condylectomy (Chap. 312).

Trigeminal neuralgia (tic douloureux) causes sudden, severe, unilateral lancinating pain initiated by touching a "trigger zone" or occurring spontaneously. Confusion with pulpal or periapical pain is common, leading to inappropriate endodontic or surgical therapy. Many cases respond to carbamazepine and phenytoin, but for a few, surgical intervention to decompress the trigeminal nerve is indicated. Similar symptoms in the distribution of the ninth cranial nerve (tongue, pharynx, soft palate) are due to *glossopharyngeal neuralgia*, which may be triggered by swallowing and may produce referred pain in the temporomandibular joint. *Postherpetic neuralgia* may follow trigeminal herpes zoster (Chap. 367) and cause burning, aching, and long-lasting pain. *Facial palsy* is usually unilateral and may be due to trauma, surgical intervention, tumor, or infection of the seventh cranial nerve. *Bell's palsy* is a form with acute onset and unknown cause, possibly viral infection such as herpes zoster. The corner of the mouth droops, and there may be difficulty in speech, eating, and in closing the eye. The symptoms usually disappear spontaneously, but residual facial immobility and lip drooping may persist. Abnormal or reduced *taste sensation* may be due to xerostomia, disturbances of the facial and glossopharyngeal nerves or their central connections, aging, or the wearing of dentures. Disease involving the hypoglossal nerve may cause atrophy of the tongue muscles with protrusion, if bilateral, or deviation toward the affected side, if unilateral. Numb chin (mental neuropathy) may be a sign of primary neural disease, but in the cancer patient it is often a harbinger of tumor relapse or progression.

(Bibliography omitted in Palm version)

Back to Table of Contents

SECTION 5 -ALTERATIONS IN CIRCULATORY AND RESPIRATORY FUNCTIONS

32. DYSPNEA AND PULMONARY EDEMA - *Roland H. Ingram, Jr.*, *Eugene Braunwald*

DYSPNEA

Breathing is controlled by central and peripheral mechanisms that adjust ventilation appropriate to increased metabolic demands during physical activity and increase ventilation in excess of metabolic demands in conditions such as anxiety and fear. A normal resting person is unaware of the act of breathing, and while he or she may become conscious of breathing during mild to moderate exertion, no discomfort is experienced. However, during and following exhausting exertion, an individual may become unpleasantly aware of breathing yet feel reasonably assured that the sensation will be transitory and is appropriate to the level of exercise. Therefore, as a cardinal symptom of diseases affecting the cardiorespiratory system, *dyspnea* is defined as an *abnormally uncomfortable awareness of breathing*.

Although dyspnea is not painful in the usual sense of the word, it is, like pain, involved with both the perception of a sensation and the reaction to that perception. Patients experience a number of uncomfortable sensations related to breathing and use an even larger number of verbal expressions to describe these sensations, such as "cannot get enough air," "air does not go all the way down," "smothering feeling or tightness or tiredness in the chest," and a "choking sensation." It may be necessary, therefore, to review the patient's history meticulously in order to ascertain whether the more abstruse descriptions do, in fact, represent dyspnea. Once it is established that a patient does have dyspnea, it is of paramount importance to define the circumstances in which it occurs and to assess associated symptoms. There are situations in which breathing appears labored but in which dyspnea does not occur. For example, the hyperventilation associated with metabolic acidemia is rarely accompanied by dyspnea. On the other hand, patients with apparently normal breathing patterns may complain of shortness of breath.

QUANTITATION OF DYSPNEA

The gradation of dyspnea is based on the amount of physical exertion required to produce the sensation. In assessing the severity of dyspnea, it is important to obtain a clear understanding of the patient's general physical condition, work history, and recreational habits. For example, the development of dyspnea in a trained runner upon running 2 mi may signify a much more serious disturbance than a similar degree of breathlessness in a sedentary person upon running a fraction of this distance. Interindividual variation in perception must also be considered. Some patients with severe disease may complain of only mild dyspnea; others with mild disease may experience more severe shortness of breath. Some patients with lung or heart disease may have such reduced capabilities due to other disease (e.g., peripheral vascular insufficiency or severe osteoarthritis of the hips or knees) that exertional dyspnea is precluded despite serious impairment of pulmonary or cardiac function.

Some patterns of dyspnea are not directly related to physical exertion. Sudden and

unexpected dyspneic episodes at rest can be associated with pulmonary emboli, spontaneous pneumothorax, hypercapnea secondary to breath holding, or anxiety. Nocturnal episodes of severe paroxysmal dyspnea are characteristic of left ventricular failure. Dyspnea upon assuming the supine posture, *orthopnea* (see below and Chap. 232), thought to be mainly characteristic of congestive heart failure, may also occur in some patients with asthma and chronic obstruction of the airways and is a regular finding in the rare occurrence of bilateral diaphragmatic paralysis. *Trepopnea* is used to describe the unusual circumstance in which dyspnea occurs only in a lateral decubitus position, most often in patients with heart disease, while *platypnea* is dyspnea that occurs only in the upright position. Positional alterations in ventilation-perfusion relationships (Chap. 250) have been invoked to explain these patterns.

MECHANISMS OF DYSPNEA (SeeFig. 32-1)

Dyspnea occurs whenever the work of breathing is excessive. Increased force generation is required of the respiratory muscles to produce a given volume change if the chest wall or lungs are less compliant or if resistance to airflow is increased. Increased work of breathing also occurs when the ventilation is excessive for the level of activity. Although an individual is more apt to become dyspneic when the work of breathing is increased, the work theory does not account for the perceptual difference between a deep breath with a normal mechanical load and a normal-sized breath with an increased mechanical load. The work might be the same with both breaths, but the normal one with the increased load will be associated with discomfort. In fact, with respiratory loading, such as adding a resistance at the mouth, there is an increase in respiratory center output that is disproportionate to the increase in the work of breathing. It has been postulated that whenever the force that muscles actually generate during breathing approaches some fraction of their maximal force-generating ability, which may vary among individuals, dyspnea ensues due to transduction of mechanical to neural stimuli.

In all likelihood, several different mechanisms operate to different degrees in the various clinical situations in which dyspnea occurs. In some circumstances, dyspnea is evoked by stimulation of receptors in the upper respiratory tract; in others it may originate from receptors in the lungs, airways, respiratory muscles, chest wall, or some combination of these structures. In any event, dyspnea is characterized by an excessive or abnormal activation of the respiratory centers in the brainstem. This activation comes about from stimuli transmitted from or through a variety of structures and pathways, including (1) intrathoracic receptors via the vagi; (2) afferent somatic nerves, particularly from the respiratory muscles and chest wall, but also from other skeletal muscles and joints; (3) chemoreceptors in the brain, aortic and carotid bodies, and elsewhere in the circulation; (4) higher (cortical) centers; and perhaps (5) afferent fibers in the phrenic nerves. In general, despite the interindividual variations described above, there is a reasonable correlation between the severity of dyspnea and the magnitude of disturbances of pulmonary or cardiac function that are responsible.

The mechanisms responsible for dyspnea may vary in different conditions (Table 32-1).

DIFFERENTIAL DIAGNOSIS

Obstructive Disease of Airways (See also Chaps. 252 and258) Obstruction to airflow can be present anywhere from the extrathoracic airways out to the small airways in the periphery of the lung. Large extrathoracic airway obstruction can occur acutely, as with aspiration of food or a foreign body or with angioedema of the glottis. An allergic history together with a few scattered hives should raise the possibility of glottic edema. Acute upper airway obstruction is a medical emergency. More chronic forms can occur with tumors or with fibrotic stenosis following tracheostomy or prolonged endotracheal intubation. Whether acute or chronic, the cardinal symptom is dyspnea, and the characteristic signs are stridor and retraction of the supraclavicular fossae with *inspiration*.

Obstruction of intrathoracic airways can occur acutely and intermittently or can be present chronically with worsening during respiratory infections. Acute intermittent obstruction with wheezing is typical of *asthma* (Chap. 252). Chronic cough with expectoration is typical of *chronic bronchitis* (Chap. 258) and *bronchiectasis* (Chap. 256). Most often there are prolongation of expiration and coarse rhonchi that are generalized in chronic bronchitis and may be localized in the case of bronchiectasis. Intercurrent infection results in worsening of the cough, increased expectoration of purulent sputum, and more severe dyspnea. During such episodes, the patient may complain of nocturnal paroxysms of dyspnea with wheezing relieved by cough and expectoration of sputum. Despite the fact that severe limitation of expiratory flow and hyperinflation of the lung are characteristic of these diseases, the sensory experience is often that of an inability to take in a sufficiently deep breath rather than difficulty in exhaling.

The patient with predominant *emphysema* is characterized by many years of exertional dyspnea progressing to dyspnea at rest (Chap. 258). Although a parenchymal disease by definition, emphysema is invariably accompanied by obstruction of airways.

Diffuse Parenchymal Lung Diseases (See also Chap. 259) This category includes a large number of diseases ranging from acute pneumonia to chronic disorders such as sarcoidosis and the various forms of *pneumoconiosis* (Chap. 254). History, physical findings, and radiographic abnormalities often provide clues to the diagnosis. The patients are often tachypneic with arterial P_{CO_2} and P_{O_2}values below normal. Exertion often further reduces the arterial P_{O_2}. Lung volumes are decreased, and the lungs are stiffer, i.e., less compliant than normal.

Pulmonary Vascular Occlusive Diseases (See also Chap. 261) Repeated episodes of dyspnea at rest often occur with recurrent pulmonary emboli. Evidence of a source for emboli, such as phlebitis of a lower extremity or the pelvis, is quite helpful in leading the physician to suspect the diagnosis. Arterial blood gases are most often abnormal, but lung volumes are frequently normal or only minimally abnormal.

Diseases of the Chest Wall or Respiratory Muscles (See alsoChap. 263) The physical examination establishes the presence of a chest wall disease such as severe kyphoscoliosis, pectus excavatum, or ankylosing spondylitis. Although all three of these deformities may be associated with dyspnea, only severe kyphoscoliosis regularly interferes with ventilation sufficiently to produce chronic cor pulmonale and respiratory failure.

Both weakness and paralysis of respiratory muscles can lead to respiratory failure and dyspnea (Chap. 263), but most often the signs and symptoms of the neurologic or muscular disorder are more prominently manifested in other systems.

Heart Disease In patients with cardiac disease, exertional dyspnea occurs most commonly as a consequence of an elevated pulmonary capillary pressure, which in turn may be due to left ventricular dysfunction (Chaps. 231 and232), reduced left ventricular compliance, and mitral stenosis. The elevation of hydrostatic pressure in the pulmonary vascular bed tends to upset the Starling equilibrium (see "Pulmonary Edema," below) with resulting transudation of liquid into the interstitial space, reducing the compliance of the lungs and stimulating J (juxtacapillary) receptors in the alveolar interstitial space. When it is prolonged, pulmonary venous hypertension results in thickening of the walls of small pulmonary vessels and an increase in perivascular cells and fibrous tissue, causing a further reduction in compliance. The competition for space among vessels, airways, and increased liquid within the interstitial space compromises the lumina of small airways, increasing the airways' resistance. Diminution in compliance and an increase in the airways' resistance increase the work of breathing. In advanced congestive heart failure, usually involving elevation of both pulmonary and systemic venous pressures, hydrothorax may develop, interfering further with pulmonary function and intensifying dyspnea.

Orthopnea, i.e., dyspnea in the supine position, is the result of the alteration of gravitational forces when this position is assumed, which elevates pulmonary venous and capillary pressures. These, in turn, increase the pulmonary closing volume (Chap. 250) and reduce the vital capacity.

Paroxysmal (Nocturnal) Dyspnea Also known as *cardiac asthma*, this condition is characterized by attacks of severe shortness of breath that generally occur at night and usually awaken the patient from sleep. The attack is precipitated by stimuli that aggravate previously existing pulmonary congestion; frequently, the total blood volume is augmented at night because of the reabsorption of edema from dependent portions of the body during recumbency. A sleeping patient can tolerate relatively severe pulmonary engorgement and may awaken only when actual pulmonary edema and bronchospasm have developed, with the feeling of suffocation and with wheezing respirations.

Two other forms of nocturnal dyspnea must be distinguished from that due to heart failure. Chronic bronchitis is characterized by mucus hypersecretion and, after a few hours sleep, secretions can accumulate and produce dyspnea and wheezing, both of which are relieved by cough and expectoration of sputum. Asthma patients have circadian variations in their degree of airway obstruction. The obstruction becomes most severe between 2 A.M. and 4 A.M. and can be sufficiently severe that the patient awakens with a sense of suffocation, extreme dyspnea, and wheezing. Although there is a prominent inflammatory component to nocturnal asthma, inhaled bronchodilators usually improve symptoms quickly.

*Cheyne-Stokes respiration***See Chap. 232*

Diagnosis The diagnosis of cardiac dyspnea depends on the recognition of heart disease on the basis of the clinical examination supplemented by noninvasive testing. There may be a history of antecedent myocardial infarction; third and fourth heart sounds may be audible; and/or there may be evidence of left ventricular enlargement, jugular neck vein distention, and/or peripheral edema. Often there are radiographic signs of heart failure, with evidence of interstitial edema, pulmonary vascular redistribution, and accumulation of liquid in the septal planes and pleural cavity. Transthoracic echocardiography is particularly useful in establishing the diagnosis of structural heart disease, which can be responsible for dyspnea. Specifically, left atrial and/or left ventricular dilatation, left ventricular hypertrophy, a reduced left ventricular ejection fraction, and disorders of left ventricular wall motion may be clues to the presence of a cardiac etiology of otherwise unexplained dyspnea.

DIFFERENTIATION BETWEEN CARDIAC AND PULMONARY DYSPNEA

In most patients with dyspnea there is obvious clinical evidence of disease of the heart and/or lungs. Like patients with cardiac dyspnea, patients with chronic obstructive lung disease may also waken at night with dyspnea, but, as pointed out above, this is usually associated with sputum production; the dyspnea is relieved after these patients rid themselves of secretions. The difficulty in the distinction between cardiac and pulmonary dyspnea may be compounded by the coexistence of diseases involving both organ systems.

In patients in whom the etiology of dyspnea is not clear, it is desirable to carry out pulmonary function testing, for these tests may be helpful in determining whether dyspnea is produced by heart disease, lung disease, abnormalities of the chest wall, or anxiety (Chap. 250). In addition to the usual means of assessing patients for heart disease, determination of the ejection fraction at rest and during exercise by echocardiography or radionuclide ventriculography is helpful in the differential diagnosis of dyspnea. The left ventricular ejection fraction is depressed in left ventricular failure, while the right ventricular ejection fraction may be low at rest or may decline during exercise in patients with severe lung disease. Both left and right ventricular ejection fractions are normal at rest and during exercise in dyspnea due to anxiety or malingering. Careful observation during the performance of an exercise treadmill test will often help in the identification of the patient who is malingering or whose dyspnea is secondary to anxiety. Under these circumstances, the patient usually complains of severe shortness of breath but appears to be breathing either effortlessly or totally irregularly. Cardiopulmonary testing, in which the patient's maximal functional exercise capacity is assessed while measurements of the electrocardiogram, blood pressure, oxygen consumption, arterial saturation (oximetry), and ventilation are carried out, is useful in the differentiation between cardiac and pulmonary dyspnea (Table 32-2).

ANXIETY NEUROSIS

Dyspnea experienced by a patient with an anxiety neurosis is difficult to evaluate. The signs and symptoms of acute and chronic hyperventilation do not serve to distinguish between anxiety neurosis and other processes, such as recurrent pulmonary emboli. Another potentially confusing situation is seen when chest pain and electrocardiographic changes accompany the hyperventilation syndrome. When present and attributable to

this condition, often referred to as *neurocirculatory asthenia* (Chap. 13), the chest pain is often sharp, fleeting, and in various loci, and the electrocardiographic changes are most often seen during repolarization. Frequent sighing respirations and an irregular breathing pattern point to a psychogenic origin of the dyspnea. Anxiety and depression in association with heart or lung disease can serve to intensify dyspnea symptoms beyond what would be expected for a given degree of dysfunction.

PULMONARY EDEMA (See Table 32-3)

CARDIOGENIC PULMONARY EDEMA (See Table 32-3, IA)

An increase in pulmonary venous pressure, which results initially in engorgement of the pulmonary vasculature, is common in most instances of dyspnea in association with congestive heart failure. The lungs become less compliant, the resistance of small airways increases, and there is an increase in lymphatic flow that apparently serves to maintain a constant pulmonary extravascular liquid volume. Mild tachypnea is present. If the increase in intravascular pressure is sufficient both in magnitude and duration, there is a net gain of liquid in the extravascular space, i.e., *interstitial* edema. At this point symptoms worsen, tachypnea increases, gas exchange deteriorates further, and radiographic changes, such as Kerley B lines and loss of distinct vascular margins, are seen. At this stage, the capillary endothelial intercellular junctions widen and allow passage of macromolecules into the interstices.

Further elevations in intravascular pressure disrupt the tight junctions between alveolar lining cells, and *alveolar* edema ensues, with outpouring of liquid that contains both red blood cells and macromolecules. With yet more severe disruption of the alveolar-capillary membrane, edematous liquid floods the alveoli and airways. At this point, full-blown clinical pulmonary edema with bilateral wet rales and rhonchi occurs, and the chest radiograph may show diffuse haziness of the lung fields with greater density in the more proximal hilar regions. Typically, the patient is anxious and perspires freely, and the sputum is frothy and blood-tinged. Gas exchange is more severely compromised with worsening hypoxia. Without effective treatment (described in Chap. 232), progressive acidemia, hypercapnia, and respiratory arrest ensue.

The earlier sequence of liquid accumulation described above follows the Starling law of capillary-interstitial liquid exchange:

The pressures tending to move liquid out of the vessel are P_c and p_{IF}, which are normally more than offset by pressures tending to move liquid back into the vasculature, i.e., the algebraic sum of P_{IF} and p_{pl}. Implicit in the preceding equation is that lymphatic flow can increase in the case of imbalance of forces and result in no net accumulation of interstitial liquid. Further elevations in P_c not only increase the outward movement of liquid in each capillary region but also recruit more of the capillary bed, which increases K. These two effects lead to liquid filtration that exceeds clearance capability by the lymphatics, and liquid accumulates in the loose interstitial spaces of the lung. Even greater increases in P_c open first the loose endothelial intercellular junctions and later the tight alveolar intercellular junctions with an increase in permeability to

macromolecules. This secondary disruption of both the function and structure of the alveolar-capillary membrane leads to alveolar flooding.

NONCARDIOGENIC PULMONARY EDEMA (See Table 32-3, IB IC, II, III, and IV)

Several clinical conditions are associated with pulmonary edema based on an imbalance of Starling forces other than through primary elevations of pulmonary capillary pressure. Although diminished plasma oncotic pressure in hypoalbuminemic states (e.g., severe liver disease, nephrotic syndrome, protein-losing enteropathy) might be expected to lead to pulmonary edema, the balance of forces normally so strongly favors resorption that even in these conditions some elevation of capillary pressure is usually necessary before interstitial edema develops. Increased negativity of interstitial pressure has been implicated in the genesis of unilateral pulmonary edema following rapid evacuation of a large pneumothorax. In this situation, the findings may be apparent only by radiography, but occasionally the patient experiences dyspnea with physical findings localized to the edematous lung. It has been proposed that large negative intrapleural pressures during acute severe asthma may be associated with the development of interstitial edema. Lymphatic blockade secondary to fibrotic and inflammatory diseases or lymphangitic carcinomatosis may lead to interstitial edema. In such instances, both clinical and radiographic manifestations are dominated by the underlying disease process.

Other conditions characterized by increases in the interstitial liquid content of the lungs appear to be associated primarily with disruption of the alveolar-capillary membranes. Any number of spontaneously occurring or environmental toxic insults, including diffuse pulmonary infections, aspiration, and shock (particularly due to sepsis and hemorrhagic pancreatitis and following cardiopulmonary bypass), are associated with diffuse pulmonary edema that clearly does not have a hemodynamic origin.**These conditions, which may lead to the acute respiratory distress syndrome, are discussed in Chap. 265.*

Other Forms of Pulmonary Edema There are three forms of pulmonary edema whose precise mechanism remains unexplained. *Narcotic overdose* is a well-recognized antecedent to pulmonary edema. Although illicit use of parenteral heroin is the most frequent cause, parenteral and oral overdoses of legitimate preparations of morphine, methadone, and dextropropoxyphene have also been associated with pulmonary edema. The earlier idea that injected impurities lead to the disorder is untenable. Available evidence suggests that there are alterations in the permeability of alveolar and capillary membranes rather than an elevation of pulmonary capillary pressure.

Exposure to high altitude in association with severe physical exertion is a well-recognized setting for pulmonary edema in unacclimatized yet otherwise healthy persons. Acclimatized high-altitude natives also develop this syndrome upon return to high altitude after a relatively brief sojourn at low altitudes. The syndrome is far more common in persons under the age of 25 years. The mechanism for high-altitude pulmonary edema (HAPE) remains obscure, and studies have been conflicting, some suggesting pulmonary venous constriction and others indicating pulmonary arteriolar constriction as the prime mechanisms. A role for hypoxia at high altitude is suggested by the fact that patients respond to the administration of oxygen and/or return to lower altitudes. Hypoxia per se does not alter permeability of the alveolar-capillary membrane.

Hence increased cardiac output and pulmonary arterial pressures with exercise combined with hypoxic pulmonary arteriolar constriction, which is more prominent in young persons, may combine to make this an example of prearteriolar, high-pressure pulmonary edema.

Neurogenic pulmonary edema has been described in patients with central nervous system disorders and without apparent preexisting left ventricular dysfunction. Although most experimental equivalents have implicated sympathetic nervous system activity, the mechanism whereby sympathetic efferent activity leads to pulmonary edema is a matter of speculation. It is known that a massive adrenergic nervous discharge leads to peripheral vasoconstriction with elevation of blood pressure and shifts of blood to the central circulation. In addition, it is probable that a reduction in left ventricular compliance also occurs, and both factors serve to increase left atrial pressures sufficiently to induce pulmonary edema on a hemodynamic basis. Some experimental evidence suggests that stimulation of adrenergic receptors increases capillary permeability directly, but this effect is relatively minor as compared with the imbalance of Starling forces.

TREATMENT OF PULMONARY EDEMA SeeChap. 232

(Bibliography omitted in Palm version)

Back to Table of Contents

33. COUGH AND HEMOPTYSIS - *Steven E. Weinberger, Eugene Braunwald*

COUGH

Cough is an explosive expiration that provides a normal protective mechanism for clearing the tracheobronchial tree of secretions and foreign material. When excessive or bothersome, it is also one of the most common symptoms for which medical attention is sought. Reasons for the latter include discomfort from the cough itself, interference with normal lifestyle, and concern for the cause of the cough, especially fear of cancer or AIDS.

MECHANISM

Coughing may be initiated either voluntarily or reflexively. As a defensive reflex it has both afferent and efferent pathways. The *afferent limb* includes receptors within the sensory distribution of the trigeminal, glossopharyngeal, superior laryngeal, and vagus nerves. The *efferent limb* includes the recurrent laryngeal nerve and the spinal nerves. The cough starts with a deep inspiration followed by glottic closure, relaxation of the diaphragm, and muscle contraction against a closed glottis. The resulting markedly positive intrathoracic pressure causes narrowing of the trachea. Once the glottis opens, the large pressure differential between the airways and the atmosphere coupled with tracheal narrowing produces rapid flow rates through the trachea. The shearing forces that develop aid in the elimination of mucus and foreign materials.

ETIOLOGY

Cough can be initiated by a variety of airway irritants, which enter the tracheobronchial tree by inhalation (smoke, dust, fumes) or by aspiration (upper airway secretions, gastric contents, foreign bodies). When cough is due to irritation by upper airway secretions (as with postnasal drip) or gastric contents (as with gastroesophageal reflux), the initiating factor may go unrecognized and the cough can be persistent. Additionally, prolonged exposure to such irritants may initiate airway inflammation, which can itself trigger cough and sensitize the airway to other irritants. Cough associated with gastroesophageal reflux is due only in part to aspiration of gastric contents, whereas vagally mediated reflex mechanisms appear to be responsible in many patients.

Any disorder resulting in inflammation, constriction, infiltration, or compression of airways can be associated with cough. Inflammation commonly results from airway infections, ranging from viral or bacterial bronchitis to bronchiectasis. In viral bronchitis, airway inflammation sometimes persists long after resolution of the typical acute symptoms, thereby producing a prolonged cough, lasting for weeks. Pertussis infection is also a possible cause of persistent cough in adults; however, diagnosis is generally made on clinical grounds (Chap. 152). Asthma is a common cause of cough. Although the clinical setting commonly suggests when a cough is secondary to asthma, some patients present with cough in the absence of wheezing or dyspnea, thus making the diagnosis more subtle ("cough variant asthma"). A neoplasm infiltrating the airway wall, such as bronchogenic carcinoma or a carcinoid tumor, is commonly associated with cough. Airway infiltration with granulomas may also trigger a cough, as seen with endobronchial sarcoidosis or tuberculosis. Compression of airways results from extrinsic

masses, including lymph nodes, mediastinal tumors, and aortic aneurysms.

Examples of parenchymal lung disease potentially producing cough include interstitial lung disease, pneumonia, and lung abscess. Congestive heart failure may be associated with cough, probably as a consequence of interstitial as well as peribronchial edema. A nonproductive cough complicates the use of angiotensin-converting enzyme (ACE) inhibitors in 5 to 20% of patients taking these agents. Onset is usually within 1 week of starting the drug but can be delayed up to 6 months. Although the mechanism is not known with certainty, it may relate to accumulation of bradykinin or substance P, both of which are degraded by ACE.

The most common causes of cough can be categorized according to the duration of the cough. Acute cough (<3 weeks) is most often due to upper respiratory infection (especially the common cold, acute bacterial sinusitis, and pertussis), but more serious disorders, such as pneumonia, pulmonary embolus, and congestive heart failure, can also present in this fashion. Chronic cough (>3 weeks) in a smoker raises the possibilities of chronic obstructive lung disease or bronchogenic carcinoma. In a nonsmoker who has a normal chest radiograph and is not taking anACEinhibitor, the most common causes of chronic cough are postnasal drip, asthma, and gastroesophageal reflux.

Approach to the Patient

A detailed *history* frequently provides the most valuable clues for etiology of the cough. Particularly important questions include:

1. Is the cough acute or chronic?

2. At its onset, were there associated symptoms suggestive of a respiratory infection?

3. Is it seasonal or associated with wheezing?

4. Is it associated with symptoms suggestive of postnasal drip (nasal discharge, frequent throat clearing, a "tickle in the throat") or gastroesophageal reflux (heartburn or sensation of regurgitation)? (The absence of such suggestive symptoms does not exclude either of these diagnoses, particularly in the case of gastroesophageal reflux.)

5. Is it associated with fever or sputum? If sputum is present, what is its character?

6. Does the patient have any associated diseases or risk factors for disease (e.g., cigarette smoking, risk factors for infection with HIV, environmental exposures)?

7. Is the patient taking anACEinhibitor?

The general *physical examination* may point to a nonpulmonary cause of cough, such as heart failure, primary nonpulmonary neoplasm, or AIDS. Examination of the oropharynx may provide suggestive evidence for postnasal drip, including oropharyngeal mucus or erythema, or a "cobblestone" appearance to the mucosa. Auscultation of the chest may demonstrate inspiratory stridor (indicative of upper airway

disease), rhonchi or expiratory wheezing (indicative of lower airway disease), or inspiratory crackles (suggestive of a process involving the pulmonary parenchyma, such as interstitial lung disease, pneumonia, or pulmonary edema).

Chest radiography may be particularly helpful in suggesting or confirming the cause of the cough. Important potential findings include the presence of an intrathoracic mass lesion, a localized pulmonary parenchymal infiltrate, or diffuse interstitial or alveolar disease. An area of honeycombing or cyst formation may suggest bronchiectasis, while symmetric bilateral hilar adenopathy may suggest sarcoidosis.

Pulmonary function testing (Chap. 250) is useful for assessing the functional abnormalities that accompany certain disorders producing cough. Measurement of forced expiratory flow rates can demonstrate reversible airflow obstruction characteristic of asthma. When asthma is considered but flow rates are normal, bronchoprovocation testing with methacholine or cold-air inhalation can demonstrate hyperreactivity of the airways to a bronchoconstrictive stimulus. Measurement of lung volumes and diffusing capacity is useful primarily for demonstration of a restrictive pattern, often seen with any of the diffuse interstitial lung diseases.

If *sputum* is produced, gross and microscopic examination may provide useful information. Purulent sputum suggests chronic bronchitis, bronchiectasis, pneumonia, or lung abscess. Blood in the sputum may be seen in the same disorders, but its presence also raises the question of an endobronchial tumor. Gram and acid-fast stains and cultures may demonstrate a particular infectious pathogen, while sputum cytology may provide a diagnosis of a pulmonary malignancy.

More specialized studies are helpful in specific circumstances. *Fiberoptic bronchoscopy* is the procedure of choice for visualizing an endobronchial tumor and collecting cytologic and histologic specimens. Inspection of the tracheobronchial mucosa can demonstrate endobronchial granulomas often seen in sarcoidosis, and endobronchial biopsy of such lesions or transbronchial biopsy of the lung interstitium can confirm the diagnosis. Inspection of the airway mucosa by bronchoscopy can also demonstrate the characteristic appearance of endobronchial Kaposi's sarcoma in patients with AIDS. *High-resolution computed tomography* (HRCT) can confirm the presence of interstitial disease and frequently suggests a diagnosis based on the pattern of disease. It is the procedure of choice for demonstrating dilated airways and confirming the diagnosis of bronchiectasis.

A diagnostic algorithm for evaluation of chronic cough is presented inFig. 33-1.

COMPLICATIONS

Common complications of coughing include chest and abdominal wall soreness, urinary incontinence, and exhaustion. On occasion, paroxysms of coughing may precipitate syncope (cough syncope; Chap. 21), consequent to markedly positive intrathoracic and alveolar pressures, diminished venous return, and decreased cardiac output. Although cough fractures of the ribs may occur in otherwise normal patients, their occurrence should at least raise the possibility of pathologic fractures, which are seen with multiple myeloma, osteoporosis, and osteolytic metastases.

TREATMENT

Definitive treatment of cough depends on determining the underlying cause and then initiating specific therapy. Elimination of an exogenous inciting agent (cigarette smoke,ACEinhibitors) or an endogenous trigger (postnasal drip, gastroesophageal reflux) is usually effective when such a precipitant can be identified. Other important management considerations are treatment of specific respiratory tract infections, bronchodilators for potentially reversible airflow obstruction, chest physiotherapy to enhance clearance of secretions in patients with bronchiectasis, and treatment of endobronchial tumors or interstitial lung disease when such therapy is available and appropriate.

Symptomatic or nonspecific therapy of cough should be considered when: (1) the cause of the cough is not known or specific treatment is not possible, and (2) the cough performs no useful function or causes marked discomfort. An irritative, nonproductive cough may be suppressed by an antitussive agent, which increases the latency or threshold of the cough center. Such agents include codeine (15 mg qid) or nonnarcotics such as dextromethorphan (15 mg qid). These drugs provide symptomatic relief by interrupting prolonged, self-perpetuating paroxysms. However, a cough productive of significant quantities of sputum should usually not be suppressed, since retention of sputum in the tracheobronchial tree may interfere with the distribution of ventilation, alveolar aeration, and the ability of the lung to resist infection.

Other agents working by a variety of mechanisms have also been used to control cough, but objective information assessing their benefit is meager. The inhaled anticholinergic agent, ipratropium bromide (2 to 4 puffs qid), has been used with the rationale of inhibiting the efferent limb of the cough reflex. Inhaled glucocorticoids, ideally administered with a spacer and dosed according to the particular agent, have been used for patients in whom airway inflammation is thought to be playing a role in the cough.

HEMOPTYSIS

Hemoptysis is defined as the expectoration of blood from the respiratory tract, a spectrum that varies from blood-streaking of sputum to coughing up large amounts of pure blood. *Massive hemoptysis* is variably defined as the expectoration of>100 to >600 mL over a 24-h period, although the patient's estimation of the amount of blood is notoriously unreliable. Expectoration of even relatively small amounts of blood is a frightening symptom and can be a marker for potentially serious disease, such as bronchogenic carcinoma. Massive hemoptysis, on the other hand, can represent an acutely life-threatening problem. Large amounts of blood can fill the airways and the alveolar spaces, not only seriously disturbing gas exchange but potentially causing the patient to suffocate.

ETIOLOGY

Because blood originating from the nasopharynx or the gastrointestinal tract can mimic blood coming from the lower respiratory tract, it is important to determine initially that the

blood is not coming from one of these alternative sites. Clues that the blood is originating from the gastrointestinal tract include a dark red appearance and an acidic pH, in contrast to the typical bright red appearance and alkaline pH of true hemoptysis.

The bronchial arteries, which are part of the high-pressure systemic circulation, originate either from the aorta or from intercostal arteries and are the source of bleeding in bronchitis or bronchiectasis or with endobronchial tumors.

An etiologic classification of hemoptysis can be based on the site of origin within the lungs (Table 33-1). The most common site of bleeding is the airways, i.e., the tracheobronchial tree, which can be affected by inflammation (acute or chronic bronchitis, bronchiectasis) or by neoplasm (bronchogenic carcinoma, endobronchial metastatic carcinoma, or bronchial carcinoid tumor). Blood originating from the pulmonary parenchyma can be either from a localized source, such as an infection (pneumonia, lung abscess, tuberculosis), or from a process diffusely affecting the parenchyma (as with a coagulopathy or with an autoimmune process such as Goodpasture's syndrome). Disorders primarily affecting the pulmonary vasculature include pulmonary embolic disease and those conditions associated with elevated pulmonary venous and capillary pressures, such as mitral stenosis or left ventricular failure.

Although the relative frequency of the different etiologies of hemoptysis varies from series to series, most recent studies indicate that bronchitis and bronchogenic carcinoma are the two most common causes. Despite the lower frequency of tuberculosis and bronchiectasis seen in recent compared to older series, these two disorders still represent the most common causes of massive hemoptysis in several series. Even after extensive evaluation, a sizable proportion of patients (up to 30% in some series) have no identifiable etiology for their hemoptysis. These patients are classified as having idiopathic or cryptogenic hemoptysis, and subtle airway or parenchymal disease is presumably responsible for the bleeding.

Approach to the Patient

The *history* is extremely valuable. Hemoptysis that is described as blood-streaking of mucopurulent or purulent sputum often suggests bronchitis. Chronic production of sputum with a recent change in quantity or appearance favors an acute exacerbation of chronic bronchitis. Fever or chills accompanying blood-streaked purulent sputum suggests pneumonia, whereas a putrid smell to the sputum raises the possibility of lung abscess. When sputum production has been chronic and copious, the diagnosis of bronchiectasis should be considered. Hemoptysis following the acute onset of pleuritic chest pain and dyspnea is suggestive of pulmonary embolism.

A history of previous or coexisting disorders should be sought, such as renal disease (seen with Goodpasture's syndrome or Wegener's granulomatosis), lupus erythematosus (with associated pulmonary hemorrhage from lupus pneumonitis), or a previous malignancy (either recurrent lung cancer or endobronchial metastasis from a nonpulmonary primary tumor). In a patient with AIDS, endobronchial or pulmonary parenchymal Kaposi's sarcoma should be considered. Risk factors for bronchogenic carcinoma, particularly smoking and asbestos exposure, should be sought. Patients

should be questioned about previous bleeding disorders, treatment with anticoagulants, or use of drugs that can be associated with thrombocytopenia.

The *physical examination* may also provide helpful clues to the diagnosis. For example, examination of the lungs may demonstrate a pleural friction rub (pulmonary embolism), localized or diffuse crackles (parenchymal bleeding or an underlying parenchymal process associated with bleeding), evidence of airflow obstruction (chronic bronchitis), or prominent rhonchi, with or without wheezing or crackles (bronchiectasis). Cardiac examination may demonstrate findings of pulmonary arterial hypertension, mitral stenosis, or heart failure. Skin examination may reveal Kaposi's sarcoma, arteriovenous malformations of Osler-Rendu-Weber disease, or lesions suggestive of systemic lupus erythematosus.

Diagnostic evaluation of hemoptysis starts with a chest radiograph to look for a mass lesion, findings suggestive of bronchiectasis (Chap. 256), or focal or diffuse parenchymal disease (representing either focal or diffuse bleeding or a focal area of pneumonitis). Additional initial screening evaluation often includes a complete blood count, a coagulation profile, and assessment for renal disease with a urinalysis and measurement of blood urea nitrogen and creatinine levels. When sputum is present, examination by Gram and acid-fast stains (along with the corresponding cultures) is indicated.

Fiberoptic bronchoscopy is particularly useful for localizing the site of bleeding and for visualization of endobronchial lesions. When bleeding is massive, rigid bronchoscopy is often preferable to fiberoptic bronchoscopy because of better airway control and greater suction capability. In patients with suspected bronchiectasis,HRCT is now the diagnostic procedure of choice, having replaced bronchography.

A diagnostic algorithm for evaluation of nonmassive hemoptysis is presented inFig. 33-2.

TREATMENT

The rapidity of bleeding and its effect on gas exchange determine the urgency of management. When the bleeding is confined to either blood-streaking of sputum or production of small amounts of pure blood, gas exchange is usually preserved; establishing a diagnosis is the first priority. When hemoptysis is massive, maintaining adequate gas exchange, preventing blood from spilling into unaffected areas of lung, and avoiding asphyxiation are the highest priorities. Keeping the patient at rest and partially suppressing cough may help the bleeding to subside. If the origin of the blood is known and is limited to one lung, the bleeding lung should be placed in the dependent position, so that blood is not aspirated into the unaffected lung.

With massive bleeding, the need to control the airway and maintain adequate gas exchange may necessitate endotracheal intubation and mechanical ventilation. In patients in danger of flooding the lung contralateral to the side of hemorrhage despite proper positioning, isolation of the right and left mainstem bronchi from each other can be achieved by selectively intubating the nonbleeding lung (often with bronchoscopic guidance) or by using specially designed double-lumen endotracheal tubes. Another

option involves inserting a balloon catheter through a bronchoscope by direct visualization and inflating the balloon to occlude the bronchus leading to the bleeding site. This technique not only prevents aspiration of blood into unaffected areas but also may promote tamponade of the bleeding site and cessation of bleeding.

Other available techniques for control of significant bleeding include laser phototherapy, electrocautery, embolotherapy, and surgical resection of the involved area of lung. With bleeding from an endobronchial tumor, the neodymium:yttrium-aluminum-garnet (Nd:YAG) laser can often achieve at least temporary hemostasis by coagulating the bleeding site. Electrocautery, which uses an electric current for thermal destruction of tissue, can be used similarly for management of bleeding from an endobronchial tumor. Embolotherapy involves an arteriographic procedure in which a vessel proximal to the bleeding site is cannulated, and a material such as Gelfoam is injected to occlude the bleeding vessel. Surgical resection is a therapeutic option either for the emergent therapy of life-threatening hemoptysis that fails to respond to other measures or for the elective but definitive management of localized disease subject to recurrent bleeding.

(Bibliography omitted in Palm version)

Back to Table of Contents

34. APPROACH TO THE PATIENT WITH A HEART MURMUR - *Patrick T. O'Gara*, *Eugene Braunwald*

Auscultation of the heart constitutes the final step in the cardiovascular examination and, for many patients with established or suspected cardiac disease, represents a defining moment in the doctor-patient relationship. The examiner must bring to this exercise an integrated approach that incorporates pertinent information from several sources. The auscultatory findings must be interpreted in the context of the history and general physical examination and with the observations made regarding the venous wave forms and major arterial pulses. In this way, abnormalities of heart sounds, adventitious sounds, and murmurs can be placed in their proper perspective.

In many patients, a heart murmur is the only or the most conspicuous finding on physical examination. The recognition of a heart murmur usually leads to additional testing, such as electrocardiography, chest radiography, and echocardiography, and may result in referral to a cardiologist. The differential diagnosis of a heart murmur should begin with an unbiased and systematic evaluation of its major attributes: timing, duration, intensity, quality, frequency, configuration, location, radiation, and response to maneuvers (seeTable 225-1). Laboratory testing can be pursued thereafter to clarify any remaining ambiguity and to provide additional anatomic and physiologic information that will impact on patient management.

Heart murmurs are defined in terms of their timing within the cardiac cycle. *Systolic murmurs* begin with or after the first heart sound (S_1) and terminate at or before the component (A_2 or P_2) of the second heart sound (S_2) that corresponds to their side of origin (left or right). *Diastolic murmurs* begin with or after the associated component of S_2 and end at or before the subsequent S_1. *Continuous murmurs* are not confined to either phase of the cardiac cycle but rather begin in systole and proceed through S_2 into all or part of diastole.

The appropriate timing of heart murmurs is the first critical step in their identification. The distinction between S_1 and S_2, and, therefore, systole and diastole, is usually a straightforward process but can be difficult in the setting of a tachyarrhythmia, in which case the heart sounds can be distinguished by simultaneous palpation of the carotid arterial pulse. The upstroke should closely follow S_1. The principal causes of heart murmurs are shown inTable 34-1, and the critical importance of the timing of heart murmurs in the differential diagnosis is shown inFig. 34-1.

SYSTOLIC HEART MURMURS

Systolic heart murmurs derive from the increased turbulence associated with (1) enhanced or accelerated flow across a normal semilunar valve, through a normal ventricular outflow tract, or into a dilated great vessel, (2) normal flow across a structurally abnormal semilunar valve or through a narrowed ventricular outflow tract, (3) flow across an incompetent atrioventricular valve, and (4) flow across the interventricular septum. One approach to their differential diagnosis further subdivides these murmurs according to their time of onset and duration within the systolic phase of the cardiac cycle.

EARLY SYSTOLIC MURMURS

Early systolic murmurs begin with S_1 and extend for a variable period of time, ending well before S_2. Their causes are relatively few in number. *Acute severe mitral regurgitation* into a normal-sized, relatively noncompliant left atrium results in an early and attenuated systolic murmur that is decrescendo in configuration and usually best heard at or just medial to the apical impulse (Chap. 236). These characteristics reflect the rapid rise in left atrial pressure caused by the sudden volume load into a nondilated chamber and contrast sharply with the auscultatory features of chronic mitral regurgitation. Clinical settings in which this occurs include: (1) papillary muscle rupture complicating acute myocardial infarction, (2) infective endocarditis, (3) rupture of chordae tendineae, and (4) blunt chest wall trauma.

Acute mitral regurgitation from papillary muscle rupture usually accompanies an inferior, posterior, or lateral infarction. The murmur is associated with a precordial thrill in approximately one-half of cases and is to be distinguished from that associated with postinfarction ventricular septal rupture. The latter is more commonly (90%) accompanied by a thrill at the left sternal edge, is holosystolic, and complicates anterior infarctions as often as inferior-posterior damage. The recognition of either of these mechanical defects mandates aggressive medical stabilization and emergent surgical intervention (Chap. 243).

The other potential causes of acute severe mitral regurgitation may be distinguished on the basis of associated findings. Spontaneous chordal rupture usually occurs on a substrate of myxomatous replacement, such as that underlying most forms of mitral valve prolapse (Chap. 236). This lesion may be part of a more generalized process, as can occur with the Marfan or Ehlers-Danlos syndromes, or it may be an isolated phenomenon. Infective endocarditis is associated with fever, peripheral embolic lesions, and positive blood cultures and most commonly occurs on a previously abnormal valvular apparatus (Chap. 126). Trauma is usually self-evident but may be disarmingly trivial (Chap. 240). It can result in papillary muscle contusion and rupture, chordal interruption, or leaflet avulsion or perforation.

Echocardiography should be performed in all cases of suspected acute severe mitral regurgitation to define the responsible mechanism, estimate the severity, and provide a preliminary assessment as to the feasibility of surgical repair (versus replacement).

Other causes of early systolic murmurs include congenital, small muscular ventricular septal defects. The duration of the murmur is attenuated by the closure of the defect during systolic contraction. The murmur is localized to the left sternal edge and is commonly of grade IV/VI or V/VI intensity. Signs of pulmonary hypertension or left ventricular volume overload are absent. Patients with anatomically large, uncorrected ventricular septal defects accompanied by pulmonary hypertension may also have murmurs confined to early systole. The elevated pulmonary vascular resistance attenuates the degree of shunting as pressures within the right and left ventricles equalize during the latter half of systole.

Tricuspid regurgitation with normal pulmonary artery pressures, such as that caused by infective endocarditis in injection drug users, may produce an early systolic murmur.

The murmur is soft, best heard at the lower left sternal edge, and may accentuate with inspiration (Carvallo's sign). Regurgitant *c-v* waves may be visible in the jugular venous pulse.

MIDSYSTOLIC MURMURS

Midsystolic murmurs begin at a short interval following S_1, end before S_2, and are usually crescendo-decrescendo in configuration (Fig. 34-1*C*). Semilunar valve stenosis is the classic prototype. With aortic valve stenosis, the murmur is usually loudest in the second right intercostal space (aortic area) and radiates along the carotid arteries (Chap. 236). The intensity of the murmur varies directly with the cardiac output; aortic valve stenosis with severe heart failure may produce a misleadingly soft systolic murmur. With a normal cardiac output, a systolic thrill is usually indicative of severe stenosis with a peak gradient in excess of 50 to 60 mmHg. An accompanying early systolic ejection click may be audible in younger patients with a bicuspid valve; its presence localizes the obstruction to the valvular (as opposed to the sub- or supravalvular) level. The midsystolic murmur of aortic stenosis may be well transmitted to the apex, especially in older patients, where it becomes less harsh and slightly higher pitched (Gallavardin effect). The murmur of aortic stenosis should increase following a postpremature beat, whereas a mitral regurgitant murmur would not be expected to change in intensity.

Sclerosis of the aortic valve produces a murmur of similar location, radiation, and configuration, albeit without the usual signs of hemodynamic significance. The carotid upstroke is well preserved, the murmur peaks in midsystole and is not accompanied by a thrill, and only a modest gradient is estimated by Doppler echocardiography. Noncritical sclerodegenerative thickening of the aortic valve leaflets is perhaps the most common cause of a midystolic murmur in older adults. The similar midsystolic murmur of pulmonic valve stenosis, usually introduced by an ejection click, is best appreciated in the second and third left intercostal spaces (pulmonic area). The murmur lengthens and the intensity of P_2diminishes with increasing degrees of stenosis. A midsystolic murmur in the aortic position can also be detected in hyperdynamic states (fever, thyrotoxicosis, pregnancy, anemia) and in the presence of isolated aortic regurgitation with the augmented flow into a dilated proximal aorta.

Crescendo-decrescendo midsystolic murmurs usually of grade II/VI intensity heard in the pulmonic area may be innocent if unaccompanied by any other signs of cardiac disease in children or young adults. They may also reflect enhanced flow into a normal pulmonary artery in hyperkinetic states or augmented flow into a dilated pulmonary artery. The latter may occur with an atrial septal defect, in which case splitting of S_2 is usually abnormal (fixed). Still's murmur is a vibratory, medium frequency, mid-systolic murmur heard best between the lower left sternal edge and the apex in normal children and young adults. It is generated by vibrations of the pulmonic valve leaflets at their attachments or by vibrations of a left ventricular false tendon.

The midsystolic murmur of hypertrophic cardiomyopathy (Chap. 238) is usually loudest between the left sternal edge and apex, of grade II/VI to III/VI intensity, and crescendo-decrescendo in configuration. In contrast to aortic valve stenosis, the murmur does *not* radiate into the neck and the carotid upstrokes are brisk and full and may even

be bifid. The intensity of the murmur associated with hypertrophic cardiomyopathy increases following maneuvers that decrease left ventricular volume (strain phase of the Valsalva maneuver, standing, amyl nitrite) or increase myocardial contractility (inotropic therapy). Conversely, the intensity of the systolic murmur decreases with maneuvers that increase ventricular volume (squatting, passive leg raising), impair contractility (beta-adrenoreceptor blockade), or raise preload and systemic afterload (squatting). Among these several maneuvers, auscultation in the standing and squatting positions, if possible, is perhaps the most sensitive technique to elicit a dynamic change in the intensity of the murmur associated with hypertrophic obstructive cardiomyopathy.

LATE SYSTOLIC MURMURS

A late systolic murmur begins well after the onset of ejection and is usually best heard at the left ventricular apex or between the apex and the left sternal edge. When introduced by a nonejection click, it is usually indicative of systolic prolapse of the mitral valve leaflet(s) into the left atrium. The click and murmur move closer to S_1following maneuvers that decrease left ventricular volume (standing, Valsalva) and move oppositely upon increases in volume (leg raising, squatting). The intensity of the murmur augments with increases in systemic afterload (squatting, pressor agents) and decrease with vasodilation (amyl nitrite). Isometric exercise, which also delays the onset of the murmur, accentuates the intensity.

HOLOSYSTOLIC MURMURS

These murmurs, also termed *pansystolic murmurs*, begin with S_1 and continue through systole to S_2(Fig. 34-1*B*). They are, with rare exception, indicative of atrioventricular valve regurgitation or of a ventricular septal defect; the differential diagnosis is shown inFig. 34-2. The murmur of mitral regurgitation is loudest at the left ventricular apex. Its radiation reflects the direction of the regurgitant jet. With a flail posterior mitral leaflet due to ruptured chordae tendineae, for example, the jet is directed anterosuperiorly, and the murmur radiates prominently to the base of the heart, where it might be confused with aortic valve stenosis unless the carotid upstrokes are carefully examined. Conversely, a flail anterior leaflet is associated with a posteriorly directed jet, which radiates into the axilla and the back. It may even strike the spine and be transmitted to the base of the neck. Severe mitral regurgitation is usually associated with a systolic thrill, a soft S_3, and a short diastolic rumbling murmur best appreciated in the left lateral decubitus position.

The holosystolic murmur of tricuspid regurgitation is generally softer (grades I to III/VI) than that of mitral regurgitation, is loudest at the left lower sternal edge, and increases in intensity upon inspiration. Associated signs include prominent "*c-v*" waves in the jugular venous pulse, systolic hepatic pulsations, and peripheral edema. Among the several causes of tricuspid regurgitation, annular dilatation from right ventricular enlargement in the setting of pulmonary artery hypertension is the most common.

Ventricular septal defect (Chap. 234) also produces a holosystolic murmur, the intensity of which varies inversely with the anatomic size of the defect. It is usually accompanied by a palpable thrill along the mid-left sternal border. The murmur of a ventricular septal defect is louder than that due to tricuspid regurgitation and does not share the latter's

inspiratory increase in intensity or associated peripheral signs.

DIASTOLIC HEART MURMURS

Like systolic murmurs, diastolic murmurs also can be subcategorized according to their time of onset.

EARLY DIASTOLIC MURMURS (Fig. 34-1*E*)

Early diastolic murmurs result from semilunar valve incompetence and begin at the valve closure sound (A_2 or P_2), which reflects their site of origin. They are generally high pitched and decrescendo in configuration, especially in states of chronic regurgitation, in which their duration is a crude index of the severity of the lesion. The murmur of aortic regurgitation is generally, but not always, best heard in the second intercostal space at the left sternal edge. There is a tendency for the murmur associated with primary valvular pathology (e.g., rheumatic deformity, congenital bicuspid valve, endocarditis) to radiate more prominently along the *left* sternal border and to be well transmitted to the apex, while the murmur associated with primary aortic root pathology (e.g., annuloaortic ectasia, aortic dissection) radiates more often along the right sternal edge. It is occasionally necessary to examine the patient sitting forward in full expiration to appreciate the murmur, a maneuver that brings the aortic root closer to the anterior chest wall. Severe aortic regurgitation may be accompanied by a lower-pitched mid- to late-diastolic murmur at the apex (Austin Flint murmur), which is generally thought to reflect turbulence at the mitral inflow area from the mixing of the regurgitant (aortic) and forward (mitral) streams, and should be distinguished from mitral stenosis (see above). In the absence of significant heart failure, chronic severe aortic regurgitation is accompanied by several peripheral signs of significant diastolic runoff, including a wide systemic pulse pressure and water-hammer carotid upstrokes (Corrigan's pulse).

The murmur associated with *acute* aortic regurgitation is notably shorter in duration, lower pitched, and can be difficult to appreciate in the presence of tachycardia. Peripheral signs of significant diastolic runoff may be absent. These attributes reflect the abrupt rise in diastolic pressure within the noncompliant left ventricle, with a correspondingly rapid decline in the aortic diastolic-left ventricular pressure gradient.

The murmur of pulmonic valve regurgitation (Graham Steell murmur) begins with a loud (palpable) pulmonic closure sound (P_2) and is best heard in the pulmonic area with radiation along the left sternal border. Typically, it is high pitched, with a decrescendo quality, and is indicative of significant pulmonary artery hypertension with a diastolic pulmonary artery-right ventricular pressure gradient. Its increase in intensity upon inspiration is one means by which to distinguish it from aortic regurgitation. Signs of right ventricular pressure and volume overload are also usually present. With significant mitral stenosis, an early decrescendo diastolic murmur along the left sternal border is not uncommon and is almost always due to aortic rather than pulmonic regurgitation, despite the coexistence of pulmonary artery hypertension.

Pulmonic valve regurgitation in the absence of pulmonary artery hypertension can occur on a congenital basis and rarely with infective endocarditis. In these instances, the early diastolic murmur is softer and lower pitched than the classic Graham Steell murmur. It

begins at or even after P_2, which should be easily separable from A_2 and thus produce appreciation of an early diastolic pause.

MIDDIASTOLIC MURMURS

Middiastolic murmurs usually result from obstruction and/or augmented flow across the antrioventricular valves. The classic example is that of mitral stenosis due to rheumatic deformity (Fig. 34-1*F*). In the absence of extensive calcification, the first heart sound (S_1) is loud and the murmur begins after the opening snap; the time interval between S_2 and the opening snap is inversely related to the left atrial-left ventricular pressure gradient. The murmur is low pitched and best heard with the bell of the stethoscope over the apex, particularly in the left lateral decubitus position. While its intensity does not reflect the severity of the obstruction accurately, the duration of the murmur does provide some indication as to the magnitude of the obstruction. A longer murmur denotes persistence of a left atrioventricular pressure gradient over a greater proportion of the diastolic time interval. Presystolic accentuation of the murmur (Fig. 34-1*A*) is frequently appreciated in the presence of sinus rhythm and reflects a further increase in transmitral flow consequent to mechanical atrial systole.

The murmur associated with tricuspid stenosis shares many of these features, but it is best heard at the lower left sternal border and, like most right-sided events, increases in intensity upon inspiration. The observant examiner may discern a prolonged *y* descent in the jugular venous pulse. Signs of right heart failure may predominate.

There are several other causes of mid-diastolic murmurs that are important to distinguish from mitral stenosis. *Left atrial myxomas* (Chap. 240) may masquerade as mitral stenosis, but the diastolic murmur is not accompanied by an opening snap or pre-systolic accentuation. Augmented flow across the mitral valve in diastole, such as occurs with severe mitral regurgitation or with large left to right intra-cardiac (ventricular septal defect) or great vessel (patent ductus arteriosus) shunts may produce a short, low pitched mid-diastolic apical murmur. The murmur usually follows a soft S_3 that is lower pitched and later in timing than the opening snap (Fig. 34-1*G*). Severe tricuspid regurgitation can also result in enhanced diastolic tricuspid flow and produce a right-sided filling complex similar to that which accompanies severe mitral regurgitation. The Austin Flint murmur of severe aortic regurgitation has been previously described and occurs in the presence of chronic severe aortic regurgitation.

CONTINUOUS MURMURS

Continuous murmurs begin in systole, peak near S_2, and continue into all or part of diastole (Fig. 34-1*H*). Accordingly, they reflect the persistence of flow between two chambers during both phases of the cardiac cycle. The differential diagnosis of continuous murmurs is shown inTable 34-1. Two innocent variants are the cervical venous hum and the mammary souffle. The former is audible in healthy children and young adults in the right supraclavicular fossa and can be abolished by compression over the internal jugular vein. Its diastolic component may be louder than its systolic counterpart. A mammary souffle represents augmented arterial flow through engorged breasts and becomes audible during the late third trimester of pregnancy or in the early postpartum period. Firm pressure with the diaphragm of the stethoscope can eliminate

the diastolic portion of the murmur. The murmur dissipates with time after delivery.

The classic continuous murmur is that due to a patent ductus arteriosus. It is best heard at or just above and to the left of the pulmonic area and may be audible in the back. Over time, a large uncorrected shunt may lead to elevation of the pulmonary vascular resistance, with resultant pulmonary artery hypertension and diminution or elimination of the diastolic component. A continuous murmur can also signify a ruptured congenital sinus of Valsalva aneurysm, which occurs either spontaneously or as a complication of infective endocarditis. Here, a high-pressure fistula is created between the aorta and a cardiac chamber, usually the right atrium or ventricle. The murmur is loudest along the right or left sternal border and is frequently accompanied by a thrill. Notably, the diastolic component is louder than the systolic component. It can be difficult to distinguish continuous murmurs from the temporally separate systolic and diastolic murmurs of mixed aortic valve disease or isolated severe aortic regurgitation. The emphasis is on the envelopment of S_2 by continuous murmurs and a gap between the to-and-fro murmurs of aortic valve disease.

A variety of other lesions can result in continuous murmurs. A coronary arteriovenous fistula sometimes produces a faint, continuous murmur with a louder diastolic component at the left sternal border or left ventricular apex. Severe atherosclerotic disease of a major systemic artery may produce a continuous bruit, the presence of which signifies very high-grade obstruction. Patients with peripheral pulmonary (branch) stenosis or with pulmonary atresia with extensive bronchial collaterals may also have continuous murmurs best heard in the back or along the lateral thoracic cage. Similar findings are present in patients with severe aortic coarctation, a lesion that should be identifiable on the basis of weak and delayed lower extremity pulses and upper extremity hypertension. The continuous murmurs emanate from the enlarged collateral (intercostal) arteries.

Approach to the Patient

It is widely recognized that, despite the importance placed on them by medical schools and training program directors, the auscultatory skills of medical students and residents have declined considerably since the advent of Doppler echocardiography. Recent surveys indicate that trainees fail to correctly identify up to 80% of adventitious sounds and murmurs. Diagnostic errors are as frequent among third-year medical residents as they are for first-year residents. Few training programs provide a dedicated educational curriculum for cardiac auscultation. In the aggregate, these deficiencies lead to an over-reliance on the use of echocardiography and increase the costs of evaluating patients with heart murmurs.

In many patients the cause of a heart murmur can be readily elucidated from careful assessment of the murmur itself, as described in this chapter, when considered in the light of the history, general physical examination, and other features of the cardiac examination, as described in Chap. 225. When the diagnosis is in doubt, or when additional pathoanatomic and physiologic data are necessary in assessing the patient and planning treatment, transthoracic Doppler echocardiography is of great value in identifying not only the etiology of the murmur but also the severity of the responsible lesion (Fig. 34-3).

The majority of heart murmurs are midsystolic and soft (Grades I to II/VI). When such a murmur occurs in an asymptomatic child or young adult *without* other evidence of heart disease on clinical examination, it is usually benign and echocardiography is not generally required. On the other hand, echocardiographic examination is indicated in patients with loud systolic murmurs (³III/VI), especially those that are holosystolic or late systolic, in most patients with diastolic or continuous murmurs, and in patients with additional unexplained abnormal physical findings on cardiac examination.

(Bibliography omitted in Palm version)

Back to Table of Contents

35. APPROACH TO THE PATIENT WITH HYPERTENSION - *Gordon H. Williams*

DEFINITION

Since there is no dividing line between normal and high blood pressure, arbitrary levels have been established to define persons who have an increased risk of developing a morbid cardiovascular event and/or will benefit from medical therapy. These definitions should take into account not only the level of diastolic pressure but also systolic pressure, age, sex, race, and concomitant diseases. For example, patients with a diastolic pressure>90 mmHg have a significant reduction in morbidity and mortality rate if they receive adequate therapy. These, then, are patients who have hypertension and who should be considered for treatment.

The level of *systolic* pressure is also important in assessing the influence of arterial pressure on cardiovascular morbidity. Some data suggest that it may be more important than diastolic pressure. For example, males with normal diastolic pressures (<82 mmHg) but elevated systolic pressures (>158 mmHg) have a cardiovascular mortality rate 2.5 times higher than individuals who have similar diastolic pressures but whose systolic pressures clearly are normal (<130 mmHg). A reduction in mortality and morbidity with treatment, specifically in the elderly, has been documented in these patients. This beneficial effect results mainly from a reduction in strokes and occurs in women as well. Other significant demographic factors that modify the influence of blood pressure on the frequency of morbid cardiovascular events are age, race, and sex, with young black males being most adversely affected by hypertension.

When hypertension is suspected, blood pressure should be measured at least twice during two separate examinations after the initial screening. In adults, a *diastolic* pressure below 85 mmHg is considered to be normal; one between 85 and 89 mmHg is high normal; one of 90 to 99 mmHg represents stage 1 or mild hypertension; one of 100 to 109 mmHg represents stage 2 or moderate hypertension; and one of³110 mmHg represents stage 3 or severe hypertension. A *systolic* pressure below 130 mmHg indicates normal blood pressure; one between 130 and 139 mmHg indicates high normal; one between 140 and 159 mmHg indicates stage 1 or mild hypertension; one between 160 and 179 mmHg indicates stage 2 or moderate hypertension; and one³180 mmHg indicates stage 3 or severe hypertension. Increasing use of 12- or 24-h blood pressure monitoring may provide additional useful information in patients who are difficult to classify. However, normal values for this procedure and its usefulness in relation to therapeutic outcomes are not currently known. A useful classification of hypertension derived from the Joint Committee on Detection, Evaluation, and Treatment of High Blood Pressure is shown inTable 35-1.

Arterial pressure fluctuates in most persons, whether they are normotensive or hypertensive. Patients who are classified as having *labile hypertension* are those who sometimes, but not always, have arterial pressures in the hypertensive range. These patients are often considered to have borderline hypertension.

Sustained hypertension can become accelerated or enter a malignant phase, although that is unusual in treated patients. Though a patient with *malignant hypertension* often has a blood pressure above 200/140, the condition is defined by the presence of

papilledema, usually accompanied by retinal hemorrhages and exudates, rather than by the absolute pressure level. *Accelerated hypertension* is defined as a significant recent increase over previous hypertensive levels associated with evidence of vascular damage on funduscopic examination but without papilledema.

PATIENT EVALUATION

In evaluating patients with hypertension, the initial history, physical examination, and laboratory tests should be directed at (1) uncovering correctable secondary forms of hypertension (Chap. 246), (2) establishing a pretreatment baseline, (3) assessing factors that may influence the type of therapy or be changed adversely by therapy, (4) determining if target organ damage is present, and (5) determining whether other risk factors for the development of arteriosclerotic cardiovascular disease are present (Chap. 241). Ideally, this evaluation would also determine the underlying mechanism(s) in essential hypertension, particularly if such information leads to a more specific therapeutic program. Unfortunately, at present this aspect of the evaluation is limited by lack of knowledge of some of the underlying mechanisms, by uncertainty as to the correct treatment for a distinct subset even if the underlying mechanisms are known, or by the prohibitive cost of defining a subset of hypertensive patients even if specific therapy were available. However, with the accumulation of additional information, this sixth component of the evaluation of patients with hypertension may become increasingly important.

Symptoms and Signs Most patients with hypertension have no specific symptoms referable to their blood pressure elevation and are identified only in the course of a physical examination. When symptoms do bring the patient to the physician, they fall into three categories. They are related to (1) the elevated pressure itself, (2) the hypertensive vascular disease, and (3) the underlying disease, in the case of secondary hypertension. Though popularly considered a symptom of elevated arterial pressure, headache is characteristic only of severe hypertension; most commonly such headaches are localized to the occipital region and are present when the patient awakens in the morning but subside spontaneously after several hours. Other complaints that may be related to elevated blood pressure include dizziness, palpitations, easy fatigability, and impotence. Complaints referable to vascular disease include epistaxis, hematuria, blurring of vision owing to retinal changes, episodes of weakness or dizziness due to transient cerebral ischemia, angina pectoris, and dyspnea due to cardiac failure. Pain due to dissection of the aorta or to a leaking aneurysm is an occasional presenting symptom.

Examples of symptoms related to the underlying disease in secondary hypertension are polyuria, polydipsia, and muscle weakness secondary to hypokalemia in patients with primary aldosteronism or weight gain, and emotional lability in patients with Cushing's syndrome. The patient with a pheochromocytoma may present with episodic headaches, palpitations, diaphoresis, and postural dizziness.

History A strong family history of hypertension, along with the reported finding of intermittent pressure elevation in the past, favors the diagnosis of essential hypertension. Secondary hypertension often develops before the age of 35 or after 55. A history of use of adrenal steroids or estrogens is of obvious significance. A history of

repeated urinary tract infections suggests chronic pyelonephritis, although this condition may occur in the absence of symptoms; nocturia and polydipsia suggest renal or endocrine disease, while trauma to either flank or an episode of acute flank pain may be a clue to the presence of renal injury. A history of weight gain is compatible with Cushing's syndrome, and one of weight loss is compatible with pheochromocytoma. A number of aspects of the history aid in determining whether vascular disease has progressed to a dangerous stage. These include angina pectoris and symptoms of cerebrovascular insufficiency, congestive heart failure, and/or peripheral vascular insufficiency. Other risk factors that should be asked about include cigarette smoking, diabetes mellitus, lipid disorders, and a family history of early deaths due to cardiovascular disease. Finally, aspects of the patient's lifestyle that could contribute to the hypertension or affect its treatment should be assessed, including diet, physical activity, family status, work, and educational level.

Physical Examination The physical examination starts with the patient's general appearance. For instance, are the round face and truncal obesity of Cushing's syndrome present? Is muscular development in the upper extremities out of proportion to that in the lower extremities, suggesting coarctation of the aorta? The next step is to compare the blood pressures and pulses in the two upper extremities and in the supine and standing positions (for at least 2 min). A rise in diastolic pressure when the patient goes from the supine to the standing position is most compatible with essential hypertension; a fall, in the absence of antihypertensive medications, suggests secondary forms of hypertension. The patient's height and weight should be recorded. Detailed examination of the ocular fundi is mandatory, as funduscopic findings provide one of the best indications of the duration of hypertension and of prognosis. A useful guide is the Keith-Wagener-Barker classification of funduscopic changes (Table 35-2); the specific changes in each fundus should be recorded and a grade assigned. Palpation and auscultation of the carotid arteries for evidence of stenosis or occlusion are important; narrowing of a carotid artery may be a manifestation of hypertensive vascular disease, and it may also be a clue to the presence of a renal arterial lesion, since these two lesions may occur together. In examination of the heart and lungs, evidence of left ventricular hypertrophy and cardiac decompensation should be sought. Is there a left ventricular lift? Are third and fourth heart sounds present? Are there pulmonary rales? A third heart sound and pulmonary rales are unusual in uncomplicated hypertension. Their presence suggests ventricular dysfunction. Chest examination also includes a search for extracardiac murmurs and palpable collateral vessels that may result from coarctation of the aorta.

The most important part of the abdominal examination is auscultation for bruits originating in stenotic renal arteries. Bruits due to renal arterial narrowing nearly always have a diastolic component or may be continuous and are best heard just to the right or left of the midline above the umbilicus or in the flanks; they are present in many patients with renal artery stenosis due to fibrous dysplasia and in 40 to 50% of those with functionally significant stenosis due to arteriosclerosis. The abdomen should also be palpated for an abdominal aneurysm and for the enlarged kidneys of polycystic renal disease. The femoral pulses must be carefully felt, and, if they are decreased and/or delayed in comparison with the radial pulse, the blood pressure in the lower extremities must be measured. Even if the femoral pulse is normal to palpation, arterial pressure in the lower extremities should be recorded at least once in patients in whom hypertension

is discovered before the age of 30 years. Finally, examination of the extremities for edema and a search for evidence of a previous cerebrovascular accident and/or other intracranial pathology should be performed.

Laboratory Investigation There is controversy as to what laboratory studies should be performed in patients presenting with hypertension. In general, the disagreement centers on how extensively the patient should be evaluated for secondary forms of hypertension or subsets of essential hypertension. The *basic* laboratory studies that should be performed in all patients with sustained hypertension are described in (Table 35-3).**The secondary studies* that should be added if (1) the initial evaluation indicates a form of secondary hypertension and/or (2) arterial pressure is not controlled after initial therapy are discussed in Chap. 246.

Renal status is evaluated by assessing the presence of protein, blood, and glucose in the urine and measuring serum creatinine and/or blood urea nitrogen. Microscopic examination of the urine is also helpful. The serum potassium level should be measured both as a screen for mineralocorticoid-induced hypertension and to provide a baseline before diuretic therapy is begun. A blood glucose determination is helpful both because diabetes mellitus may be associated with accelerated arteriosclerosis, renal vascular disease, and diabetic nephropathy in patients with hypertension and because primary aldosteronism, Cushing's syndrome, and pheochromocytoma all may be associated with hyperglycemia. Furthermore, since antihypertensive therapy with diuretics, for example, can raise the blood glucose level, it is important to establish a baseline. The possibility of hypercalcemia may also be investigated. Serum cholesterol, high-density lipoprotein cholesterol, and triglyceride levels identify other factors that predispose to the development of arteriosclerosis.

An electrocardiogram should be obtained in all cases to permit assessment of cardiac status, particularly if left ventricular hypertrophy is present, and to provide a baseline. The echocardiogram is more sensitive than either the electrocardiogram or physical examination in determining whether cardiac hypertrophy is present. However, a complete, detailed echocardiographic study is expensive. Thus, in some circumstances, a cheaper, limited echocardiogram may be a useful addition to the *baseline* evaluation of a hypertensive patient, particularly as left ventricular hypertrophy is an independent cardiovascular risk factor and its presence suggests the need for vigorous antihypertensive therapy. Furthermore, while a substantial increase in arterial pressure usually correlates with the presence of left ventricular hypertrophy, a mild increase may not. Thus, one cannot use the blood pressure as a surrogate marker for the presence or absence of left ventricular hypertrophy. On the other hand, because of the cost of an echocardiogram and the uncertainty as to whether the resultant information would modify therapy, it is unclear that routine *follow-up* echocardiograms during therapy are justified. Furthermore, there are no data to suggest that reversal of left ventricular hypertrophy produces benefits beyond that conferred by blood pressure reduction. The chest roentgenogram may also be helpful by providing the opportunity to identify aortic dilation or elongation and the rib notching that occurs in coarctation of the aorta.

Certain clues from the history, physical examination, and basic laboratory studies may suggest an unusual cause for the hypertension and dictate the need for special studies as outlined inChap 246.

TREATMENT

SeeChap. 246.

(Bibliography omitted in Palm version)

Back to Table of Contents

36. HYPOXIA AND CYANOSIS - *Eugene Braunwald*

HYPOXIA

The fundamental purpose of the cardiorespiratory system is to deliver O_2 (and substrates) to the cells and to remove CO_2 (and other metabolic products) from them. Proper maintenance of this function depends on intact cardiovascular and respiratory systems and a supply of inspired gas containing adequate O_2. When hypoxia occurs consequent to respiratory failure, Pa_{CO_2} usually rises (Chap. 250), and the hemoglobin-oxygen (Hb-O_2) dissociation curve (see Fig. 106-2) is displaced to the right. Under these conditions, the Pa_{O_2} declines. Arterial hypoxemia, i.e., a reduction of O_2 saturation of arterial blood (Sa_{O_2}), and consequent cyanosis are likely to be more marked when such depression of Pa_{O_2} results from pulmonary disease than when the depression occurs as the result of a decline in the fraction of oxygen in inspired air (FI_{O_2}). In this situation Pa_{CO_2} falls secondary to anoxia-induced hyperventilation and the Hb-O_2 dissociation curve is displaced to the left, limiting the decline in Sa_{O_2}.

CAUSES OF HYPOXIA

Anemic Hypoxia Any reduction in the hemoglobin concentration of the blood is attended by a corresponding decline in the O_2-carrying capacity of the blood. In anemic hypoxia, the Pa_{O_2} is normal; but as a consequence of the reduction of the hemoglobin concentration, the absolute quantity of O_2 transported per unit volume of blood is diminished. As the anemic blood passes through the capillaries and the usual quantity of O_2 is removed from it, the P_{O_2} in the venous blood declines to a greater degree than would normally be the case.

Carbon Monoxide Intoxication (See also Chap. 396) Hemoglobin that is combined with carbon monoxide (carboxyhemoglobin, COHb) is unavailable for O_2 transport. In addition, the presence of COHb shifts the Hb-O_2 dissociation curve to the left (see Fig. 106-2) so that O_2 is unloaded only at lower tensions. By such formation of COHb, a given degree of reduction in O_2-carrying power produces a far greater degree of tissue hypoxia than the equivalent reduction in hemoglobin due to simple anemia.

Respiratory Hypoxia Arterial unsaturation is a common finding in advanced pulmonary disease. The most common cause of respiratory hypoxia is ventilation-perfusion mismatch, which results from perfusion of poorly ventilated alveoli. As discussed in Chap. 250, it may also be caused by hypoventilation, and it is then associated with an elevation of Pa_{CO_2}. These two forms of respiratory hypoxia may be recognized because they are usually correctable by inspiring 100% O_2 for several minutes. A third cause is shunting of blood across the lung from right to left by perfusion of nonventilated portions of the lung, as in pulmonary atelectasis or through arteriovenous connections in the lung. The low Pa_{O_2} in this situation is correctable only in part by an FI_{O_2} of 100%.

Hypoxia Secondary to High Altitude As one ascends rapidly to 3000 m (approximately 10,000 ft), the alveolar P_{O_2} declines to about 60 mmHg, and impaired memory and other cerebral symptoms of hypoxia may develop. At higher altitudes, arterial saturation declines rapidly and symptoms become more serious; and at 5000 m (approximately 15,000 ft) unacclimatized individuals usually cease to be able to function

normally.

Hypoxia Secondary to Right-to-Left Extrapulmonary Shunting From a physiologic viewpoint, this cause of hypoxia resembles intrapulmonary right-to-left shunting but is caused by congenital cardiac malformations such as tetralogy of Fallot, transposition of the great arteries, and Eisenmenger's syndrome (Chap. 234). As in pulmonary right-to-left shunting, the Pa_{O_2} cannot be restored to normal with inspiration of 100% O_2.

Circulatory Hypoxia As in anemic hypoxia, the Pa_{O_2} is usually normal, but venous and tissue P_{O_2} values are reduced as a consequence of reduced tissue perfusion. Generalized circulatory hypoxia occurs in heart failure (Chap. 232) and in most forms of shock (Chap. 38).

Specific Organ Hypoxia Decreased perfusion of any organ resulting in localized circulatory hypoxia may occur secondary to organic arterial obstruction or as a consequence of vasoconstriction (Chap. 248). The latter is seen in the upper extremities in Raynaud's phenomenon. Ischemic hypoxia with accompanying pallor occurs in organic arterial obliterative disease. Localized hypoxia also may result from venous obstruction and the resultant congestion and reduced arterial inflow. Edema, which increases the distance through which O_2 diffuses before it reaches cells, also can cause localized hypoxia. In an attempt to maintain adequate perfusion to more vital organs, constriction may reduce perfusion in all limbs in patients with heart failure or hypovolemic shock.

Increased O_2 Requirements If the O_2 consumption of the tissues is elevated without a corresponding increase in perfusion, tissue hypoxia ensues and the P_{O_2} in venous blood becomes reduced. Ordinarily, the clinical picture of patients with hypoxia due to an elevated metabolic rate is quite different from that in other types of hypoxia; the skin is warm and flushed, owing to increased cutaneous blood flow that dissipates the excessive heat produced, and cyanosis is usually absent.

Exercise is a classic example of increased tissue O_2 requirements. These increased demands are normally met by several mechanisms operating simultaneously: (1) increasing the cardiac output and ventilation and thus O_2 delivery to the tissues; (2) preferentially directing the blood to the exercising muscles by changing vascular resistances in various circulatory beds, directly and/or reflexly; (3) increasing O_2 extraction from the delivered blood and widening the arteriovenous O_2 difference; and (4) reducing the pH of the tissues and capillary blood, thereby unloading more O_2 from hemoglobin. If the capacity of these mechanisms is exceeded, then hypoxia, especially of the exercising muscles, will result.

Improper Oxygen Utilization Cyanide (Chap. 396) and several other similarly acting poisons cause cellular hypoxia. The tissues are unable to utilize O_2, and as a consequence, the venous blood tends to have a high O_2 tension. This condition has been termed *histotoxic hypoxia*.

EFFECTS OF HYPOXIA

Changes in the central nervous system, particularly the higher centers, are especially

important consequences of hypoxia. Acute hypoxia causes impaired judgment, motor incoordination, and a clinical picture closely resembling that of acute alcoholism. When hypoxia is long-standing, fatigue, drowsiness, apathy, inattentiveness, delayed reaction time, and reduced work capacity occur. As hypoxia becomes more severe, the centers of the brainstem are affected, and death usually results from respiratory failure. With the reduction of Pa_{O_2}, cerebrovascular resistance decreases and cerebral blood flow increases, increasing O_2delivery to the brain as a compensatory mechanism. However, when the reduction of $Pa_{O_2}$is accompanied by hyperventilation and a reduction of Pa_{CO_2}, cerebrovascular resistance rises, cerebral blood flow falls, and hypoxia is intensified. Hypoxia also causes pulmonary arterial constriction, which shunts blood away from poorly ventilated areas toward better-ventilated portions of the lung. However, it also increases pulmonary vascular resistance and right ventricular afterload.

Glucose is normally broken down to pyruvic acid. However, the further breakdown of pyruvate and the generation of adenosine triphosphate (ATP) consequent to it require O_2; and in the presence of hypoxia increasing proportions of pyruvate are reduced to lactic acid, which cannot be broken down further, causing metabolic acidosis. Under these circumstances, the total energy obtained from the breakdown of carbohydrate is greatly reduced, and the quantity of energy available for the production of ATP becomes inadequate.

An important component of the respiratory response to hypoxia originates in special chemosensitive cells in the carotid and aortic bodies, and in the respiratory center in the brainstem. The stimulation of these cells by hypoxia increases ventilation, with a loss of CO_2, and leads to respiratory alkalosis. When combined with the metabolic acidosis resulting from the production of lactic acid, the serum bicarbonate level declines (Chap. 50).

Diminished P_{O_2} in any tissue results in local vasodilatation, and the diffuse vasodilatation that occurs in generalized hypoxia raises the cardiac output. In patients with underlying heart disease, the requirements of the peripheral tissues for an increase of cardiac output with hypoxia may precipitate congestive heart failure. In patients with ischemic heart disease, a reduced Pa_{O_2} may intensify myocardial ischemia and further impair left ventricular function.

One of the important mechanisms of compensation for chronic hypoxia is an increase in the hemoglobin concentration and in the number of red blood cells in the circulating blood, i.e., the development of polycythemia secondary to erythropoietin production (Chap. 110).

CYANOSIS

Cyanosis refers to a bluish color of the skin and mucous membranes resulting from an increased quantity of reduced hemoglobin, or of hemoglobin derivatives, in the small blood vessels of those areas. It is usually most marked in the lips, nail beds, ears, and malar eminences. Cyanosis, especially if developed recently, is more commonly detected by a family member than the patient. The florid skin characteristic of polycythemia vera (Chap. 110) must be distinguished from the true cyanosis discussed here. A cherry-colored flush, rather than cyanosis, is caused by COHb (Chap. 396). The

degree of cyanosis is modified by the color of the cutaneous pigment and the thickness of the skin, as well as by the state of the cutaneous capillaries. The accurate clinical detection of the presence and degree of cyanosis is difficult, as proved by oximetric studies. In some instances, central cyanosis can be detected reliably when the Sa_{O_2} has fallen to 85%; in others, particularly in dark-skinned persons, it may not be detected until it has declined to 75%. In the latter case, examination of the mucous membranes in the oral cavity and the conjunctivae rather than examination of the skin is more helpful in the detection of cyanosis.

The increase in the quantity of reduced hemoglobin in the mucocutaneous vessels that produces cyanosis may be brought about either by an increase in the quantity of venous blood as the result of dilatation of the venules and venous ends of the capillaries or by a reduction in the Sa_{O_2} in the capillary blood. In general, cyanosis becomes apparent when the mean capillary concentration of reduced hemoglobin exceeds 40 g/L (4 g/dL). It is the *absolute* rather than the *relative* quantity of reduced hemoglobin that is important in producing cyanosis. Thus, in a patient with severe anemia, the relative amount of reduced hemoglobin in the venous blood may be very large when considered in relation to the total amount of hemoglobin in the blood. However, since the concentration of the latter is markedly reduced, the *absolute* quantity of reduced hemoglobin may still be small, and therefore patients with severe anemia and even *marked* arterial desaturation do not display cyanosis. Conversely, the higher the total hemoglobin content, the greater is the tendency toward cyanosis; thus, patients with marked polycythemia tend to be cyanotic at higher levels of Sa_{O_2}than patients with normal hematocrit values. Likewise, local passive congestion, which causes an increase in the total amount of reduced hemoglobin in the vessels in a given area, may cause cyanosis. Cyanosis also is observed when nonfunctional hemoglobin such as methemoglobin or sulfhemoglobin (Chap. 106) is present in blood.

Cyanosis may be subdivided into central and peripheral types. In the *central* type, the Sa_{O_2} is reduced or an abnormal hemoglobin derivative is present, and the mucous membranes and skin are both affected. *Peripheral* cyanosis is due to a slowing of blood flow and abnormally great extraction of O_2 from normally saturated arterial blood. It results from vasoconstriction and diminished peripheral blood flow, such as occurs in cold exposure, shock, congestive failure, and peripheral vascular disease. Often in these conditions the mucous membranes of the oral cavity or those beneath the tongue may be spared. Clinical differentiation between central and peripheral cyanosis may not always be simple, and in conditions such as cardiogenic shock with pulmonary edema there may be a mixture of both types.

DIFFERENTIAL DIAGNOSIS

Central Cyanosis (Table 36-1) Decreased Sa_{O_2}results from a marked reduction in the Pa_{O_2}. This reduction may be brought about by a decline in the FI_{O_2}without sufficient compensatory alveolar hyperventilation to maintain alveolar P_{O_2}. Cyanosis does not occur to a significant degree in an ascent to an altitude of 2500 m (8000 ft) but is marked in a further ascent to 5000 m (16,000 ft). The reason for this difference becomes clear on studying the *S* shape of the Hb-O_2dissociation curve (see Fig. 106-1). At 2500 m (8000 ft) the FI_{O_2} is about 120 mmHg, the alveolar P_{O_2} is approximately 80 mmHg, and the Sa_{O_2} is nearly normal. However, at 5000 m (16,000 ft) the FI_{O_2} and

alveolar P_{O_2} are about 85 and 50 mmHg, respectively, and the Sa_{O_2} is only about 75%. This leaves 25% of the hemoglobin in the arterial blood in the reduced form, an amount likely to be associated with cyanosis in the absence of anemia. Similarly, a mutant hemoglobin with a low affinity for O_2(e.g., Hb Kansas) causes lowered Sa_{O_2}saturation and resultant central cyanosis (Chap. 106).

Seriously *impaired pulmonary function*, through perfusion of unventilated or poorly ventilated areas of the lung or alveolar hypoventilation, is a common cause of central cyanosis (Chap. 250). This condition may occur acutely, as in extensive pneumonia or pulmonary edema, or chronically with chronic pulmonary diseases (e.g., emphysema). In the last situation, secondary polycythemia is generally present, and clubbing of the fingers may occur. However, in many types of chronic pulmonary disease with fibrosis and obliteration of the capillary vascular bed, cyanosis does not occur because there is relatively little perfusion of underventilated areas.

Another cause of reduced Sa_{O_2} is *shunting of systemic venous blood into the arterial circuit.* Certain forms of congenital heart disease are associated with cyanosis (Chap. 234). Since blood flows from a higher-pressure to a lower-pressure region, for a cardiac defect to result in a right-to-left shunt, it must ordinarily be combined with an obstructive lesion distal to the defect or with elevated pulmonary vascular resistance. The most common congenital cardiac lesion associated with cyanosis in the adult is the combination of ventricular septal defect and pulmonary outflow tract obstruction (*tetralogy of Fallot*). The more severe the obstruction, the greater the degree of right-to-left shunting and resultant cyanosis. In patients with patent ductus arteriosus, pulmonary hypertension, and right-to-left shunt, *differential cyanosis* results; that is, cyanosis occurs in the lower but not in the upper extremities.**The mechanisms for the elevated pulmonary vascular resistance that may produce cyanosis in the presence of intra- and extracardiac communications without pulmonic stenosis (Eisenmenger syndrome) are discussed in Chap. 234.*

Pulmonary arteriovenous fistulae (Chap. 57) may be congenital or acquired, solitary or multiple, microscopic or massive. The severity of cyanosis produced by these fistulae depends on their size and number. They occur with some frequency in hereditary hemorrhagic telangiectasia. Sa_{O_2}reduction and cyanosis may also occur in some patients with cirrhosis, presumably as a consequence of pulmonary arteriovenous fistulas or portal vein-pulmonary vein anastomoses.

In patients with cardiac or pulmonary right-to-left shunts, the presence and severity of cyanosis depend on the size of the shunt relative to the systemic flow as well as on the Hb-O_2saturation of the venous blood. With increased extraction of O_2 from the blood by the exercising muscles, the venous blood returning to the right side of the heart is more unsaturated than at rest, and shunting of this blood or its passage through lungs incapable of normal oxygenation intensifies the cyanosis. Also, since the systemic vascular resistance falls with exercise, the right-to-left shunt is augmented by exercise in patients with congenital heart disease and communications between the two sides of the heart. Secondary polycythemia occurs frequently in patients with arterial O_2unsaturation and contributes to the cyanosis.

Cyanosis can be caused by small amounts of circulating methemoglobin and by even

smaller amounts of sulfhemoglobin (Chap. 106). Although they are uncommon causes of cyanosis, these abnormal hemoglobin pigments should be sought by spectroscopy when cyanosis is not readily explained by malfunction of the circulatory or respiratory systems. Generally, digital clubbing does not occur with them. The diagnosis of methemoglobinemia can be suspected if the patient's blood remains brown after being mixed in a test tube and exposed to air.

Peripheral Cyanosis Probably the most common cause of peripheral cyanosis is the normal vasoconstriction resulting from exposure to cold air or water. When cardiac output is low, as in severe congestive heart failure or shock, cutaneous vasoconstriction occurs as a compensatory mechanism so that blood is diverted from the skin to more vital areas such as the central nervous system and heart (Chap. 232), and intense cyanosis associated with cool extremities may result. Even though the arterial blood is normally saturated, the reduced volume flow through the skin and the reduced P_{O2} at the venous end of the capillary result in cyanosis.

Arterial obstruction to an extremity, as with an embolus, or arteriolar constriction, as in cold-induced vasospasm (Raynaud's phenomenon, Chap. 248), generally results in pallor and coldness, but there may be associated cyanosis. Venous obstruction, as in thrombophlebitis, dilates the subpapillary venous plexuses and thereby intensifies cyanosis.

Approach to the Patient

Certain features are important in arriving at the cause of cyanosis:

1. The history, particularly the onset (cyanosis present since birth is usually due to congenital heart disease), and possible exposure to drugs or chemicals that may produce abnormal types of hemoglobin.

2. Clinical differentiation of central as opposed to peripheral cyanosis. Objective evidence by physical or radiographic examination of disorders of the respiratory or cardiovascular systems. Massage or gentle warming of a cyanotic extremity will increase peripheral blood flow and abolish peripheral but not central cyanosis.

3. The presence or absence of clubbing of the digits (see below). Clubbing without cyanosis is frequent in patients with infective endocarditis and ulcerative colitis; it may occasionally occur in healthy persons, and in some instances it may be occupational, e.g., in jackhammer operators. The combination of cyanosis and clubbing is frequent in patients with congenital heart disease and right-to-left shunting and is seen occasionally in persons with pulmonary disease such as lung abscess or pulmonary arteriovenous fistulae. In contrast, peripheral cyanosis or acutely developing central cyanosis is *not* associated with clubbed digits.

4. Determination of Pa_{O2}tension and Sa_{O2} and spectroscopic and other examinations of the blood for abnormal types of hemoglobin (critical in the differential diagnosis of cyanosis).

CLUBBING

The selective bullous enlargement of the distal segments of the fingers and toes due to proliferation of connective tissue, particularly on the dorsal surface, is termed *clubbing*; there is increased sponginess of the soft tissue at the base of the nail. Clubbing may be hereditary, idiopathic, or acquired and associated with a variety of disorders, including cyanotic congenital heart disease, infective endocarditis, and a variety of pulmonary conditions (among them primary and metastatic lung cancer, bronchiectasis, lung abscess, cystic fibrosis, and mesothelioma), as well as with some gastrointestinal diseases (including regional enteritis, chronic ulcerative colitis, and hepatic cirrhosis).

Clubbing in patients with primary and metastatic lung cancer, mesothelioma, bronchiectasis, and hepatic cirrhosis may be associated with *hypertrophic osteoarthropathy*. In this condition, the subperiosteal formation of new bone in the distal diaphyses of the long bones of the extremities causes pain and symmetric arthritis-like changes in the shoulders, knees, ankles, wrists, and elbows. The diagnosis of hypertrophic osteoarthropathy may be confirmed by bone radiographs. Although the mechanism of clubbing is unclear, it appears to be secondary to a humoral substance that causes dilation of the vessels of the fingertip.

(Bibliography omitted in Palm version)

Back to Table of Contents

37. EDEMA - *Eugene Braunwald*

Edema is defined as a clinically apparent increase in the interstitial fluid volume, which may expand by several liters before the abnormality is evident. Therefore, a weight gain of several kilograms usually precedes overt manifestations of edema, and a similar weight loss from diuresis can be induced in a slightly edematous patient before "dry weight" is achieved. *Ascites* (Chap. 46) and *hydrothorax* refer to accumulation of excess fluid in the peritoneal and pleural cavities, respectively, and are considered to be special forms of edema. *Anasarca* refers to gross, generalized edema.

Depending on its cause and mechanism, edema may be localized or have a generalized distribution; it is recognized in its generalized form by puffiness of the face, which is most readily apparent in the periorbital areas, and by the persistence of an indentation of the skin following pressure; this is known as "pitting" edema. In its more subtle form, it may be detected by noting that after the stethoscope is removed from the chest wall, the rim of the bell leaves an indentation on the skin of the chest for a few minutes. When the ring on a finger fits more snugly than in the past or when a patient complains of difficulty in putting on shoes, particularly in the evening, edema may be present.

PATHOGENESIS

About one-third of the total-body water is confined to the extracellular space. Approximately 25% of the latter, in turn, is composed of the plasma volume, and the remainder is interstitial fluid.

Starling Forces The forces that regulate the disposition of fluid between these two components of the extracellular compartment are frequently referred to as the *Starling forces* (see p. 202). The hydrostatic pressure within the vascular system and the colloid oncotic pressure in the interstitial fluid tend to promote movement of fluid from the vascular to the extravascular space. In contrast, the colloid oncotic pressure contributed by the plasma proteins and the hydrostatic pressure within the interstitial fluid, referred to as the *tissue tension*, promote the movement of fluid into the vascular compartment. Consequently there is a movement of water and diffusible solutes from the vascular space at the arteriolar end of the capillaries.

Fluid is returned from the interstitial space into the vascular system at the venous end of the capillary and by way of the lymphatics, and unless these channels are obstructed, lymph flow tends to increase with increases in net movement of fluid from the vascular compartment to the interstitium. These flows are usually balanced so that a steady state exists in the sizes of the intravascular and interstitial compartments, and yet a large exchange between them occurs. However, should any one of the hydrostatic or oncotic pressure gradients be altered significantly, a further net movement of fluid between the two components of the extracellular space will take place. The development of edema then depends on one or more alterations in the Starling forces so that there is increased flow of fluid from the vascular system into the interstitium or into a body cavity.

Edema due to increase in capillary pressure may result from an elevation of venous pressure due to obstruction in venous drainage. This increase in capillary pressure may be generalized, as occurs in congestive heart failure. The Starling forces may be

imbalanced when the colloid oncotic pressure of the plasma is reduced, owing to any factor that may induce hypoalbuminemia, such as saline expansion, malnutrition, liver disease, loss of protein into the urine or into the gastrointestinal tract, or a severe catabolic state.

Capillary Damage Edema may also result from damage to the capillary endothelium, which increases its permeability and permits the transfer of protein into the interstitial compartment. Injury to the capillary wall can result from drugs, viral or bacterial agents, and thermal or mechanical trauma. Increased capillary permeability may also be a consequence of a hypersensitivity reaction and is characteristic of immune injury. Damage to the capillary endothelium is presumably responsible for inflammatory edema, which is usually nonpitting, localized, and accompanied by other signs of inflammation -- redness, heat, and tenderness.

To formulate a hypothesis about the pathophysiology of an edematous state, it is important to discriminate between the *primary* events, such as localized or generalized venous or lymphatic obstruction, reduction of cardiac output, hypoalbuminemia, trapping of fluid in spaces such as the pleural or peritoneal cavities, or an increase in capillary permeability, and the predictable *secondary* consequences, which include the renal retention of salt and water in an attempt to restore the plasma volume when the latter has been reduced, as in venous obstruction (see below). Both the primary event and the secondary consequences contribute to the formation of edema. In some cases the primary event is the renal retention of salt and water. Examples are renal failure, nephrotic syndrome, glomerulonephritis, and early hepatic failure.

Reduction of Effective Arterial Volume In many forms of edema the *effective arterial blood volume*, an as yet poorly defined parameter of the filling of the arterial tree, is reduced, and as a consequence a series of physiologic responses designed to restore it to normal are set into motion. A key element of these responses is the retention of salt and therefore of water, principally by the renal proximal tubule (Fig. 37-1), and in many instances this repairs the deficit of the effective arterial blood volume; often this deficit is repaired without the development of overt edema. If, however, the retention of salt and water is insufficient to restore and maintain the effective arterial blood volume, the stimuli are not dissipated, the retention of salt and water continues, and edema may ultimately develop. This sequence of events is operative in dehydration and hemorrhage. Although in these conditions there is a reduction of effective arterial blood volume and activation of the entire sequence shown in the center of Fig. 37-2, including the diminished excretion of salt and water, because the net sodium and water balance is negative rather than positive, edema does *not* occur. In most conditions that lead to edema, the mechanisms responsible for maintaining a normal effective osmolality in the body fluids operate efficiently so that sodium retention promotes thirst and secretion of the antidiuretic hormone. In edematous states, isotonic expansion of the extracellular fluid space may be massive, while the intracellular fluid volume is unchanged.

Reduced Cardiac Output A reduction of cardiac output, whatever the cause, is associated with a lowering of the effective arterial blood volume as well as of renal blood flow, constriction of the efferent renal arterioles, and an elevation of the filtration fraction, i.e., the ratio of glomerular filtration rate to renal plasma flow. In severe heart failure there is a reduction in the glomerular filtration rate. Activation of the sympathetic

nervous system and of the renin-angiotensin systems are responsible for renal vasoconstriction. The finding that a-adrenergic blocking agents and/or angiotensin-converting enzyme (ACE) inhibitors augment renal blood flow and induce diuresis supports the role of these two systems in elevating renal vascular resistance and salt and water retention.

Renal Factors Reduced cardiac output lowers effective arterial blood volume. There is increased tubular reabsorption of glomerular filtrate in both the proximal and distal tubules (Fig. 37-1). Alterations in intrarenal hemodynamics appear to play a significant role. Heart failure and other conditions, such as nephrotic syndrome and cirrhosis that reduce effective arterial blood volume, cause renal efferent arteriolar constriction. This, in turn, reduces the hydrostatic pressure while the increased filtration fraction raises the colloid osmotic pressure in the peritubular capillaries, thus enhancing salt and water reabsorption in the proximal tubule as well as in the ascending limb of the loop of Henle.

In addition, the diminished renal blood flow characteristic of states in which the effective arterial blood volume is reduced is translated by the renal juxtaglomerular cells into a signal for increased renin release (Chap. 331). The mechanisms responsible for this release include a baroreceptor response: reduced renal perfusion results in incomplete filling of the renal arterioles and diminished stretch of the juxtaglomerular cells, a signal that provides for the elaboration or release, or both, of renin. A second mechanism for renin release involves the macula densa; as a result of reduced glomerular filtration, the sodium chloride load reaching the distal renal tubules is reduced. This is sensed by the macula densa, which signals the neighboring juxtaglomerular cells to secrete renin. A third mechanism involves the sympathetic nervous system and circulating catecholamines. Activation of theb-adrenergic receptors in the juxtaglomerular cells stimulates renin release. These three mechanisms generally act in concert.

The Renin-Angiotensin-Aldosterone (RAA) System (See Chap. 331) Renin, an enzyme with a molecular weight of about 40,000, acts on its substrate, angiotensinogen, ana$_2$globulin synthesized by the liver, to release angiotensin I, a decapeptide, which is broken down to angiotensin II (AII), an octapeptide. This has generalized vasoconstrictor properties; it is especially active on the efferent arterioles and independently increases Na^+ reabsorption in the proximal tubule. The RAA system has long been recognized as a hormone system. However, it also operates locally. Both circulating and intrarenally produced AII contribute to renal vasoconstriction and to salt and water retention. These renal effects of AII are mediated by activation of AII type 1 receptors, which can be blocked by specific antagonists such as losartan. AII also enters the circulation and stimulates the production of aldosterone by the zona glomerulosa of the adrenal cortex. In patients with heart failure, not only is aldosterone secretion elevated but the biologic half-life of aldosterone is prolonged, which further increases the plasma level of the hormone. A depression of hepatic blood flow, particularly during exercise, secondary to a reduction in cardiac output, is responsible for the reduced hepatic catabolism of aldosterone. Aldosterone, in turn, enhances Na^+ reabsorption (and K^+ excretion) by the collecting tubule. The activation of the RAA system is most striking in the early phase of acute, severe heart failure and is less intense in patients with chronic, stable, compensated heart failure.

Although increased quantities of aldosterone are secreted in heart failure and in other

edematous states and although blockade of the action of aldosterone by spironolactone (an aldosterone antagonist) or amiloride (a blocker of epithelial Na^+ channels) often induces a moderate diuresis in edematous states, persistent augmented levels of aldosterone (or other mineralocorticoids) alone do not always promote accumulation of edema, as witnessed by the lack of striking fluid retention in most instances of primary aldosteronism (Chap. 331). Furthermore, although normal individuals retain some salt and water with the administration of potent mineralocorticoids, such as deoxycorticosterone acetate or fludrocortisone, this accumulation is self-limiting, despite continued exposure to the steroid, a phenomenon known as *mineralocorticoid escape*. The failure of normal individuals who receive large doses of mineralocorticoids to accumulate large quantities of extracellular fluid and to develop edema is probably a consequence of an increase in glomerular filtration rate (pressure natriuresis) and through the action of natriuretic substance(s) (see below). The continued secretion of aldosterone may be more important in the accumulation of fluid in edematous states because patients with edema secondary to heart failure, nephrotic syndrome, and cirrhosis are generally unable to repair the deficit in effective arterial blood volume. As a consequence they do not develop pressure natriuresis.

Blockade of the RAA system, by blocking AII receptors or inhibiting ACE, reduces efferent arteriolar resistance and increases renal blood flow. This action (combined in patients with heart failure with a rise in cardiac output secondary to afterload reduction) as well as reduction in the secretion of aldosterone cause diuresis. However, in patients with moderate or severe impairment of renal function or with renal artery stenosis, interference with the RAA system can cause paradoxical sodium retention due to intensification of renal failure.

Arginine Vasopressin (AVP) and Endothelin (See also Chap. 329) The secretion of AVP occurs in response to increased intracellular osmolar concentration and by stimulating V_2 receptors increases the reabsorption of free water in the renal distal tubule and collecting duct, thereby increasing total-body water. Circulating AVP is elevated in many patients with heart failure secondary to a nonosmotic stimulus associated with decreased effective arterial volume. Such patients fail to show the normal reduction of AVP with a reduction of osmolality, contributing to hyponatremia and edema formation.

Endothelin This is a potent peptide vasoconstrictor released by endothelial cells; its concentration is elevated in heart failure and contributes to renal vasoconstriction, Na^+ retention, and edema in heart failure.

Natriuretic Peptides Atrial distention and/or a sodium load cause release into the circulation of atrial natriuretic peptide (ANP), a polypeptide; a high-molecular-weight precursor of ANP is stored in secretory granules within atrial myocytes. Release of ANP causes (1) excretion of sodium and water by augmenting glomerular filtration rate, inhibiting sodium reabsorption in the proximal tubule, and inhibiting release of renin and aldosterone; and (2) arteriolar and venous dilatation by antagonizing the vasoconstrictor actions of AII, AVP, and sympathetic stimulation. Thus, ANP has the capacity to oppose sodium retention and arterial pressure elevation in hypervolemic states.

The closely related brain natriuretic peptide (BNP) is stored primarily in cardiac ventricular myocardium and is released when ventricular diastolic pressure rises. Its

actions are similar to those ofANP. Circulating levels of ANP and BNP are elevated in congestive heart failure but obviously not sufficient to prevent edema formation. In addition, in edematous states (particularly heart failure), there is abnormal resistance to the actions of natriuretic peptides.

CLINICAL CAUSES OF EDEMA

Obstruction of Venous (and Lymphatic) Drainage of a Limb In this condition the hydrostatic pressure in the capillary bed upstream to the obstruction increases so that an abnormal quantity of fluid is transferred from the vascular to the interstitial space. Since the alternative route (i.e., the lymphatic channels) may also be obstructed, an increased volume of interstitial fluid in the limb develops, i.e., there is a trapping of fluid in the extremity, causing local edema at the expense of the blood volume in the remainder of the body, thereby reducing effective arterial blood volume and leading to the consequences shown inFig. 37-2.

When venous and lymphatic drainage are obstructed in a limb, fluid accumulates in the interstitium at the expense of plasma volume. The latter stimulates the retention of salt and water until the deficit in plasma volume has been corrected. Tissue tension rises in the affected limb until it counterbalances the primary alterations in the Starling forces, at which time no further fluid accumulates. The net effect is a local increase in the volume of interstitial fluid. This same sequence occurs in ascites and hydrothorax, in which fluid is trapped or accumulates in the cavitary space, depleting the intravascular volume and leading to secondary salt and fluid retention, as already described.

Congestive Heart Failure (See also Chap. 232) In this disorder the defective systolic emptying of the chambers of the heart and/or the impairment of ventricular relaxation promotes an accumulation of blood in the heart and venous circulation at the expense of the effective arterial volume, and the aforementioned sequence of events (Fig. 37-2) is initiated. In mild heart failure, a small increment of total blood volume may repair the deficit of arterial volume and establish a new steady state. Through the operation of Starling's law of the heart, an increase in the volume of blood within the chambers of the heart promotes a more forceful contraction and may thereby increase the cardiac output (See Fig. 232-1). However, if the cardiac disorder is more severe, retention of fluid cannot repair the deficit in effective arterial blood volume. The increment in blood volume accumulates in the venous circulation, and the increase in capillary and lymphatic hydrostatic pressures promotes the formation of edema. In heart failure, a reduction occurs in baroreflex-mediated inhibition of the vasomotor center, which causes activation of renal vasoconstrictor nerves and theRAAsystem, causing sodium and water retention.

Incomplete ventricular emptying (systolic heart failure) and/or inadequate ventricular relaxation (diastolic heart failure) both lead to an elevation of ventricular diastolic pressure. If the impairment of cardiac function involves the right ventricle, pressures in the systemic veins and capillaries may rise, thereby augmenting the transudation of fluid into the interstitial space and enhancing the likelihood of peripheral edema in the presence of the accumulation of sodium and water, as described above. The elevated systemic venous pressure is transmitted to the thoracic duct with consequent reduction of lymph drainage, further increasing the accumulation of edema.

If the impairment of cardiac function (incomplete ventricular emptying and/or inadequate relaxation) involves the left ventricle primarily, then pulmonary venous and capillary pressures rise [leading in some instances to pulmonary edema (Chap. 32)], as does pulmonary artery pressure; this in turn interferes with the emptying of the right ventricle, leading to an elevation of right ventricular diastolic and of central and systemic venous pressures, enhancing the likelihood of formation of peripheral edema. Pulmonary edema impairs gas exchange and may induce hypoxia, which embarrasses cardiac function still further, sometimes causing a vicious cycle.

Nephrotic Syndrome and Other Hypoalbuminemic States (See alsoChap. 274) The primary alteration in this disorder is a diminished colloid oncotic pressure due to massive losses of protein into the urine. This promotes a net movement of fluid into the interstitium, causes hypovolemia, and initiates the edema-forming sequence of events described above, including activation of theRAAsystem. With severe hypoalbuminemia and the consequent reduced colloid osmotic pressure, the salt and water that are retained cannot be restrained within the vascular compartment, total and effective arterial blood volumes decline, and hence the stimuli to retain salt and water are not abated. A similar sequence of events occurs in other conditions that lead to *severe* hypoalbuminemia, including severe nutritional deficiency states, protein-losing enteropathy, congenital hypoalbuminemia, and severe, chronic liver disease. However, in the nephrotic syndrome, impaired renal Na^+ excretion contributes to edema, even in the absence of severe hypoalbuminemia.

Cirrhosis (See alsoChaps. 46 and299) This condition is characterized by hepatic venous outflow blockade, which in turn causes expansion of the splanchnic blood volume and increased hepatic lymph formation. Intrahepatic hypertension acts as a potent stimulus for renal Na^+ retention and perhaps systemic vasodilation and a reduction of effective arterial blood volume as well. These alterations are frequently complicated by hypoalbuminemia secondary to reduced hepatic synthesis and reduce the effective arterial blood volume even further, leading to activation of theRAAsystem, of renal sympathetic nerves, and other salt- and water-retaining mechanisms. The concentration of circulating aldosterone is elevated by the liver's failure to metabolize this hormone. Initially, the excess interstitial fluid is localized preferentially upstream to the congested portal venous system and obstructed hepatic lymphatics, i.e., in the peritoneal cavity. In later stages, particularly when there is severe hypoalbuminemia, peripheral edema may develop. The excess production of prostaglandins ($PGE_2$and PGI_2) in cirrhosis attenuates renal Na^+ retention. When the synthesis of these substances is inhibited by nonsteroidal anti-inflammatory agents, renal function deteriorates and Na^+ retention increases.

Drug-Induced Edema A large number of widely used drugs can cause edema (Table 37-1). Mechanisms include renal vasoconstriction (nonsteroidal anti-inflammatory agents and cyclosporine), arteriolar dilatation (vasodilators), augmented renal sodium reabsorption (steroid hormones) and capillary damage (interleukin 2).

Idiopathic Edema This syndrome, which occurs almost exclusively in women, is characterized by periodic episodes of edema (unrelated to the menstrual cycle), frequently accompanied by abdominal distention. Diurnal alterations in weight occur with

orthostatic retention of sodium and water, so that the patient may weigh several pounds more after having been in the upright posture for several hours. Such large diurnal weight changes suggest an increase in capillary permeability that appears to fluctuate in severity and to be aggravated by hot weather. There is some evidence that a reduction in plasma volume occurs in this condition with secondary activation of theRAAsystem and impaired suppression ofAVPrelease. Idiopathic edema should be distinguished from cyclical or premenstrual edema, in which the sodium and water retention may be secondary to excessive estrogen stimulation. There are also some cases in which the edema appears to be "diuretic-induced." It has been postulated that in these patients, chronic diuretic administration leads to mild blood volume depletion, which causes chronic hyperreninemia and juxtaglomerular hyperplasia. Salt-retaining mechanisms appear to overcompensate for the direct effects of the diuretics. *Acute* withdrawal of diuretics can then leave the sodium-retaining forces unopposed, leading to fluid retention and edema. Decreased dopaminergic activity and reduced urinary kallikrein and kinin excretion have been reported in this condition and may also be of pathogenetic importance.

TREATMENT

The treatment of idiopathic cyclic edema includes a reduction in salt intake, rest in the supine position for several hours each day, the wearing of elastic stockings (which are put on before arising in the morning), and an attempt to understand any underlying emotional problems. A variety of pharmacologic agents includingACEinhibitors, progesterone, the dopamine receptor agonist bromocriptine, and the sympathomimetic amine dextroamphetamine have all been reported to be useful when administered to patients who do not respond to simpler measures. Diuretics may be helpful initially but may lose their effectiveness with continuous administration; accordingly, they should be employed sparingly, if at all. Discontinuation of diuretics paradoxically leads to diuresis in "diuretic-induced" edema, described above.

DIFFERENTIAL DIAGNOSIS

The differences between the three major causes of generalized edema are shown inTable 37-2.

Localized edema can usually be readily differentiated from generalized edema. The great majority of patients with generalized edema suffer from advanced cardiac, renal, hepatic, or nutritional disorders. Consequently, the differential diagnosis of generalized edema should be directed toward identifying or excluding these several conditions.

LOCALIZED EDEMA (See also Chap. 248)

Edema originating from inflammation or hypersensitivity is usually readily identified. Localized edema due to venous or lymphatic obstruction may be caused by thrombophlebitis, chronic lymphangitis, resection of regional lymph nodes, filariasis, etc. Lymphedema is particularly intractable because restriction of lymphatic flow results in increased protein concentration in the interstitial fluid, a circumstance that aggravates retention of fluid.

EDEMA OF HEART FAILURE (See alsoChap. 232)

The presence of heart disease, as manifested by cardiac enlargement and gallop rhythm, together with evidence of cardiac failure, such as dyspnea, basilar rales, venous distention, and hepatomegaly, usually provides an indication on clinical examination that edema results from heart failure. Noninvasive tests such as echocardiography and radionuclide angiography may be helpful in establishing the diagnosis of heart failure.

EDEMA OF THE NEPHROTIC SYNDROME (See also Chap. 274)

Marked proteinuria (>3.5 g/d), hypoalbuminemia (<35 g/L), and in some instances hypercholesterolemia are present. This syndrome may occur during the course of a variety of kidney diseases, which include glomerulonephritis, diabetic glomerulosclerosis, and hypersensitivity reactions. A history of previous renal disease may or may not be elicited.

EDEMA OF ACUTE GLOMERULONEPHRITIS AND OTHER FORMS OF RENAL FAILURE

The edema occurring during the acute phases of glomerulonephritis is characteristically associated with hematuria, proteinuria, and hypertension. Although some evidence supports the view that the fluid retention is due to increased capillary permeability, in most instances the edema in this disease results from primary retention of sodium and water by the kidneys owing to renal insufficiency. This state differs from congestive heart failure in that it is characterized by a normal (or sometimes even increased) cardiac output and a normal arterial-mixed venous oxygen difference. Patients with edema due to renal failure commonly have evidence of pulmonary congestion on chest roentgenograms before cardiac enlargement is significant, but they usually do not develop orthopnea. Patients with chronic impairment of renal function may also develop edema due to primary renal retention of sodium and water.

EDEMA OF CIRRHOSIS (See also Chap. 299)

Ascites and biochemical and clinical evidence of hepatic disease (collateral venous channels, jaundice, and spider angiomas) characterize edema of hepatic origin. The ascites is frequently refractory to treatment because it collects as a result of a combination of obstruction of hepatic lymphatic drainage, portal hypertension, and hypoalbuminemia. Edema may also occur in other parts of the body in these patients as a result of hypoalbuminemia. Furthermore, the sizable accumulation of ascitic fluid may increase intraabdominal pressure and impede venous return from the lower extremities; hence, it tends to promote accumulation of edema in this region as well.

EDEMA OF NUTRITIONAL ORIGIN

A diet grossly deficient in protein over a prolonged period may produce hypoproteinemia and edema. The latter may be intensified by the development of beriberi heart disease, also of nutritional origin, in which multiple peripheral arteriovenous fistulas result in reduced effective systemic perfusion and effective arterial blood volume, thereby enhancing edema formation (Chap. 75). Edema may actually become intensified when

these famished subjects are first provided with an adequate diet. The ingestion of more food may increase the quantity of salt ingested, which is then retained along with water. So-called "refeeding edema" may also be linked to increased release of insulin, which directly increases tubular sodium reabsorption. In addition to hypoalbuminemia, hypokalemia and caloric deficits may be involved in the edema of starvation.

OTHER CAUSES OF EDEMA

These include hypothyroidism, in which the edema (myxedema) may be located typically in the pretibial region and which may also be associated with periorbital puffiness. Exogenous hyperadrenocortism, pregnancy, and administration of estrogens and vasodilators, particularly the calcium antagonist nifedipine, may also all cause edema.

DISTRIBUTION OF EDEMA

The distribution of edema is an important guide to the cause. Thus, edema limited to one leg or to one or both arms is usually the result of venous and/or lymphatic obstruction. Edema resulting from hypoproteinemia characteristically is generalized, but it is especially evident in the very soft tissues of the eyelids and face and tends to be most pronounced in the morning because of the recumbent posture assumed during the night. Less common causes of facial edema include trichinosis, allergic reactions, and myxedema. Edema associated with heart failure, on the other hand, tends to be more extensive in the legs and to be accentuated in the evening, a feature also determined largely by posture. When patients with heart failure have been confined to bed, edema may be most prominent in the presacral region. Unilateral edema occasionally results from lesions in the central nervous system affecting the vasomotor fibers on one side of the body; paralysis also reduces lymphatic and venous drainage on the affected side.

ADDITIONAL FACTORS IN DIAGNOSIS

The color, thickness, and sensitivity of the skin are significant. Local tenderness and increase in temperature suggest inflammation. Local cyanosis may signify a venous obstruction. In individuals who have had repeated episodes of prolonged edema, the skin over the involved areas may be thickened, indurated, and often red.

Measurement or estimation of the venous pressure is of importance in evaluating edema. Elevation in an isolated part of the body usually reflects localized venous obstruction. Generalized elevation of systemic venous pressure usually indicates the presence of congestive heart failure. Ordinarily, a significant generalized increase in venous pressure can be recognized by the level at which cervical veins collapse (Chap. 225). In patients with obstruction of the superior vena cava, edema is confined to the face, neck, and upper extremities, where the venous pressure is elevated compared with that in the lower extremities. Measurement of venous pressure in the upper extremities is also useful in patients with massive edema of the lower extremities and ascites; it is elevated in the upper extremities when the edema is on a cardiac basis (e.g., constrictive pericarditis or tricuspid stenosis) but is normal when it is secondary to cirrhosis. Severe heart failure may cause ascites that may be distinguished from the ascites caused by hepatic cirrhosis by the jugular venous pressure, which usually is

elevated in heart failure and normal in cirrhosis.

Determination of the concentration of serum albumin aids importantly in identifying those patients in whom edema is due, at least in part, to diminished intravascular colloid oncotic pressure. The presence of proteinuria also affords useful clues. The absence of proteinuria excludes nephrotic syndrome but cannot exclude nonproteinuric causes of renal failure. Slight to moderate proteinuria is the rule in patients with heart failure.

Approach to the Patient

An important first question is whether the edema is localized or generalized. If it is localized, those phenomena that may be responsible should be concentrated upon. Hydrothorax and ascites are forms of localized edema. Either may be a consequence of local venous or lymphatic obstruction, as in inflammatory or neoplastic disease.

If the edema is generalized, it should be determined, first, if there is serious hypoalbuminemia, e.g., serum albumin<25 g/L. If so, the history, physical examination, urinalysis, and other laboratory data will help evaluate the question of cirrhosis, severe malnutrition, protein-losing gastroenteropathy, or the nephrotic syndrome as the underlying disorder. If hypoalbuminemia is not present, it should be determined if there is evidence of congestive heart failure of a severity to promote generalized edema. Finally, it should be determined whether the patient has an adequate urine output, or if there is significant oliguria or even anuria.**These abnormalities are discussed in Chaps. 47, 269, and 270.*

(Bibliography omitted in Palm version)

Back to Table of Contents

38. SHOCK - *Ronald V. Maier*

Shock is the clinical syndrome that results from inadequate tissue perfusion. Irrespective of cause, the hypoperfusion-induced imbalance between the delivery of and requirements for oxygen and substrate leads to cellular dysfunction. The cellular injury created by the inadequate delivery of oxygen and substrates also induces the production and release of inflammatory mediators that further compromise perfusion through functional and structural changes within the microvasculature. This leads to a vicious cycle in which impaired perfusion is responsible for cellular injury which causes maldistribution of blood flow, further compromising cellular perfusion; the latter causes multiple organ failure and, if the process is not interrupted, leads to the death of the patient. The clinical manifestations of shock are the result, in part, of sympathetic neuroendocrine responses to hypoperfusion as well as the breakdown in organ function induced by severe cellular dysfunction.

When very severe and/or persistent, inadequate oxygen delivery leads to irreversible cell injury, and only rapid restoration of oxygen delivery can reverse the progression of the shock state. The fundamental approach to management, therefore, is to recognize overt and impending shock in a timely fashion and to intervene emergently to restore perfusion. Except in cases of cardiogenic shock, this requires the expansion or reexpansion of blood volume. Control of any inciting pathologic process, e.g., continued hemorrhage, impairment of cardiac function, or infection, must occur simultaneously.

Clinical shock is usually accompanied by hypotension, i.e., a mean arterial pressure<60 mmHg in previously normotensive persons. Multiple classification schemes have been developed in an attempt to synthesize the seemingly dissimilar processes leading to shock. Strict adherence to a classification scheme may be difficult from a clinical standpoint because of the frequent combination of two or more causes of shock in any individual patient, but the classification shown in Table 38-1 provides a useful reference point from which to discuss and further delineate the underlying processes. The individual classes are discussed below.

PATHOGENESIS AND ORGAN RESPONSE

MICROCIRCULATION

Normally when cardiac output falls, systemic vascular resistance rises to maintain a level of systemic pressure that is adequate to allow perfusion of the heart and brain at the expense of other tissues, especially muscle, skin, and the gastrointestinal tract. Systemic vascular resistance is determined primarily by the luminal diameter of arterioles. The metabolic rates of the heart and brain are high, and their stores of energy substrate are low. These organs are critically dependent on a continuous supply of oxygen and nutrients, and neither tolerates severe ischemia for more than brief periods. Autoregulation, i.e., the maintenance of blood flow over a wide range of perfusion pressures, is critical in sustaining cerebral and coronary perfusion despite significant hypotension. However, when mean arterial pressure drops to 60 mmHg, flow to these organs falls and their function deteriorates.

Arteriolar vascular smooth muscle has both a- and b-adrenergic receptors (Chap. 72).

The a_1 receptors mediate vasoconstriction, while the b_2 receptors mediate vasodilation. Efferent sympathetic fibers release norepinephrine, which acts primarily on a_1 receptors in one of the most fundamental compensatory responses to reduced perfusion pressure. Other constrictor substances that are increased in most forms of shock include angiotensin II, vasopressin, endothelin-1, and thromboxane A_2. Both norepinephrine and epinephrine are released by the adrenal medulla, and the concentrations of these catecholamines in the blood stream rise. Circulating vasodilators in shock include prostacyclin (PGI_2), nitric oxide (NO), and, importantly, products of local metabolism such as adenosine that match flow to the metabolic needs of the tissue. The balance between these various vasoconstrictor and vasodilator influences acting upon the microcirculation determines local perfusion.

Transport to cells depends on microcirculatory flow; capillary permeability; the diffusion of oxygen, carbon dioxide, nutrients, and products of metabolism through the interstitium; and the exchange of these products across cell membranes. Impairment of the microcirculation, which is central to the pathophysiologic responses in the late stages of all forms of shock, results in the derangement of cellular metabolism, which is ultimately responsible for organ failure.

The normal response to mild or moderate hypovolemia is an attempt at restitution of intravascular volume through alterations in hydrostatic pressure and osmolarity. Constriction of arterioles leads to reductions in both the capillary hydrostatic pressure and the number of capillary beds perfused, thereby limiting the capillary surface area across which filtration occurs. When filtration is reduced while intravascular oncotic pressure remains constant or rises, there is net reabsorption of fluid into the vascular bed, in accord with Starling's law of capillary-interstitial liquid exchange (Chap. 32). Metabolic changes (including hyperglycemia and elevations in the products of glycolysis, lipolysis, and proteolysis) raise extracellular osmolarity, leading to an osmotic gradient between cells and interstitium that increases interstitial and intravascular volume at the expense of intracellular volume.

CELLULAR RESPONSES

Interstitial transport of nutrients is impaired, leading to a decline of intracellular high-energy phosphate stores. Mitochondrial dysfunction and uncoupling of oxidative phosphorylation are the most likely causes for decreased amounts of ATP. As a consequence, there is an accumulation of anaerobic metabolites including hydrogen ions, lactate, and other products of anaerobic metabolism. As shock progresses, these vasodilator metabolites override vasomotor tone, causing further hypotension and hypoperfusion. Dysfunction of cell membranes is thought to represent a common end-stage pathophysiologic pathway in the various forms of shock. Normal cellular transmembrane potential falls, and there is an associated increase in intracellular sodium and water, leading to cell swelling, which interferes further with microvascular perfusion.

NEUROENDOCRINE RESPONSE

Hypovolemia, hypotension, and hypoxia are sensed by baroreceptors and chemoreceptors, which contribute further to an autonomic response that attempts to

restore blood volume, maintain central perfusion, and mobilize metabolic substrates. Hypotension disinhibits the vasomotor center, resulting in increased adrenergic output and reduced vagal activity. Release of norepinephrine induces peripheral and splanchnic vasoconstriction, a major contributor to the maintenance of central organ perfusion, while reduced vagal activity increases the heart rate and cardiac output. The effects of circulating epinephrine released by the adrenal medulla in shock are largely metabolic, causing increased glycogenolysis and gluconeogenesis and reduced pancreatic insulin release.

Severe pain and other severe stress cause the hypothalamic release of adrenocorticotropic hormone (ACTH). This stimulates cortisol secretion, which contributes to decreased peripheral uptake of glucose and amino acids, enhances lipolysis, and increases gluconeogenesis. Increased pancreatic secretion of glucagon during stress accelerates hepatic gluconeogenesis and further elevates blood glucose concentration. These hormonal actions act synergistically in the maintenance of blood volume. The importance of the cortisol response to stress is illustrated by the profound circulatory collapse that occurs in hypoadrenal patients (see below).

Renin release is increased in response to adrenergic discharge and reduced perfusion of the juxtaglomerular apparatus in the kidney. Renin induces the formation of angiotensin I, which is then converted to angiotensin II, an extremely potent vasoconstrictor and stimulator of aldosterone release by the adrenal cortex and of vasopressin by the posterior pituitary. Aldosterone contributes to the maintainance of intravascular volume by enhancing renal tubular reabsorption of sodium, resulting in the excretion of a low-volume, concentrated, sodium-free urine. Vasopressin has a direct action on vascular smooth muscle, contributing to vasoconstriction, and acts on the distal renal tubules to enhance water reabsorption.

CARDIOVASCULAR RESPONSE

Three variables -- ventricular filling (preload), the resistance to ventricular ejection (afterload), and myocardial contractility -- are paramount in controlling stroke volume (Chap. 231). Cardiac output, the major determinant of tissue perfusion, is the product of stroke volume and heart rate. Hypovolemia leads to decreased ventricular preload, which in turn reduces the stroke volume. An increase in heart rate is a useful but limited compensatory mechanism to maintain cardiac output. A shock-induced reduction in myocardial compliance is frequent, reducing ventricular end-diastolic volume and hence stroke volume at any given ventricular filling pressure. Restoration of intravascular volume then returns stroke volume to normal but only at elevated filling pressures. In addition, sepsis, ischemia, myocardial infarction, severe tissue trauma, hypothermia, general anesthesia, prolonged hypotension, and acidemia may all impair myocardial contractility and also reduce the stroke volume at any given ventricular end-diastolic volume. The resistance to ventricular ejection is influenced importantly by the systemic vascular resistance, which is elevated in most forms of shock. However, resistance is depressed in the early hyperdynamic stage of septic shock (see below), thereby allowing the cardiac output to be maintained.

The venous system contains nearly two-thirds of the total circulating blood volume, most in the small veins, and serves as a dynamic reservoir for autoinfusion of blood. Active

venoconstriction as a consequence of a-adrenergic activity is an important compensatory mechanism for the maintenance of venous return and therefore of ventricular filling during shock. On the other hand, venous dilatation, as occurs in neurogenic shock, reduces ventricular filling and hence stroke volume and cardiac output (see below).

PULMONARY RESPONSE

The response of the pulmonary vascular bed to shock parallels that of the systemic vascular bed, and the relative increase in pulmonary vascular resistance, particularly in septic shock, may exceed that of the systemic vascular resistance. Shock-induced tachypnea reduces tidal volume and increases both dead space and minute ventilation. Relative hypoxia and the subsequent tachypnea induce a respiratory alkalosis. Recumbency and involuntary restriction of ventilation secondary to pain reduce functional residual capacity and may lead to atelectasis. Shock is recognized as a major cause of acute lung injury and subsequent acute respiratory distress syndrome (ARDS;Chap. 265). These disorders are characterized by noncardiogenic pulmonary edema secondary to diffuse pulmonary capillary endothelial and alveolar epithelial injury, hypoxemia, and bilateral diffuse pulmonary infiltrates. Hypoxemia results from perfusion of underventilated and nonventilated alveoli. Loss of surfactant and lung volume in combination with increased interstitial and alveolar edema reduce lung compliance. The work of breathing and the oxygen requirements of respiratory muscles increase.

RENAL RESPONSE

Acute renal failure (Chap. 269), a serious complication of shock and hypoperfusion, occurs less frequently than heretofore because of early aggressive volume repletion. Acute tubular necrosis is now more frequently seen as a result of the interactions of shock, sepsis, the administration of nephrotoxic agents (such as aminoglycosides and angiographic contrast media), and rhabdomyolysis; the latter may be particularly severe in skeletal muscle trauma. The physiologic response of the kidney to hypoperfusion is to conserve salt and water. In addition to decreased renal blood flow, increased afferent arteriolar resistance accounts for diminished glomerular filtration rate, which together with increased ADH and aldosterone is responsible for reduced urine formation. Toxic injury causes necrosis of tubular epithelium and tubular obstruction by cellular debris with back-leak of filtrate. The depletion of renal ATP stores that occurs with prolonged renal hypoperfusion is related to subsequent impairment of renal function.

METABOLIC DERANGEMENTS

During shock, there is disruption of the normal cycles of carbohydrate, lipid, and protein metabolism. Through the citric acid cycle, alanine in conjunction with lactate (which is converted from pyruvate in the periphery in the presence of oxygen deprivation) enhances the hepatic production of glucose. With reduced availability of oxygen, the breakdown of glucose to pyruvate and ultimately lactate represents an inefficient cycling of substrate with minimal net energy production. An elevated plasma lactate/pyruvate ratio is consistent with anaerobic metabolism and reflects inadequate tissue perfusion. Decreased clearance of exogenous triglycerides coupled with increased hepatic

lipogenesis causes a significant rise in serum triglyceride concentrations. There is increased protein catabolism, a negative nitrogen balance, and, if the process is prolonged, severe muscle wasting.

INFLAMMATORY RESPONSES

Activation of an extensive network of proinflammatory mediator systems plays a significant role in the progression of shock and contributes importantly to the development of organ injury and failure.

Multiple humoral mediators are activated during shock and tissue injury. The complement cascade, activated through both the classic and alternate pathways, generates the anaphylatoxins C3a and C5a. Direct complement fixation to injured tissues can progress to the C5-C9 attack complex, causing further cell damage. Activation of the coagulation cascade causes microvascular thrombosis, with subsequent lysis leading to repeated episodes of ischemia and reperfusion. Components of the coagulation system, such as thrombin, are potent proinflammatory mediators that cause expression of adhesion molecules on endothelial cells and activation of neutrophils, leading to microvascular injury. Coagulation also activates the kallekrein-kininogen cascade, contributing to hypotension.

Eicosanoids are vasoactive and immunomodulatory products of arachidonic acid metabolism that include cyclooxygenase-derived prostaglandins and thromboxane A_2 as well as lipoxygenase-derived leukotrienes and lipoxins. Thromboxane A_2 is a potent vasoconstrictor that contributes to the pulmonary hypertension and acute tubular necrosis of shock.$PGI_2$and prostaglandin E_2 are potent vasodilators that enhance capillary permeability and edema formation. The cysteinyl leukotrienes LTC_4 and LTD_4 are pivotal mediators of the vascular sequelae of anaphylaxis, as well as of shock states resulting from sepsis or tissue injury. LTB_4 is a potent neutrophil chemoattractant and secretagogue that stimulates the formation of reactive oxygen species. Lipoxins are endogenous autocoids that inhibit leukotriene-mediated responses. Platelet-activating factor, an ether-linked, arachidonyl-containing phospholipid mediator, also carries potent bioactivities that include pulmonary vasoconstriction, bronchoconstriction, systemic vasodilation, increased capillary permeability, and the priming of macrophages and neutrophils to produce enhanced levels of inflammatory mediators.

Tumor necrosis factor (TNF)a, produced by activated macrophages, reproduces many components of the shock state including hypotension, lactic acidosis, and respiratory failure. Interleukin (IL) 1 is also produced by tissue-fixed macrophages and is critical to the inflammatory response occurring in hypoperfusion and septic states. Chemokines also participate in the systemic inflammatory response. For example, IL-8 is a potent neutrophil chemoattractant and activator that upregulates adhesion molecules on the neutrophil to enhance aggregation and adherence to the vascular endothelium. The endothelium normally produces nitric oxide (NO), a potent vasodilator. The inflammatory response stimulates the inducible isoform of NO synthase (iNOS), which is thought to overexpress toxic NO and oxygen-derived free radicals and contributes to the hyperdynamic cardiovascular response that occurs in sepsis.

Multiple inflammatory cells, including neutrophils, macrophages, and platelets, are a

major contributor to inflammation-induced injury. Margination of activated neutrophils in the microcirculation is a common pathologic finding in shock, causing secondary injury due to the release of potentially toxic oxygen radicals and proteases. Adhesion molecules are expressed on the surface of the endothelium and on cytokine-stimulated neutrophils. Tissue-fixed macrophages produce virtually all major components of the inflammatory response and orchestrate the progression and duration of the response.

Approach to the Patient

The underlying problem in all forms of shock is inadequate tissue perfusion and an imbalance between delivery and cellular needs of oxygen and metabolic substrate. It is important to recognize the onset of hypoperfusion at the earliest possible time in order to institute aggressive resuscitation and correction of the underlying etiology.

Monitoring Patients in shock require care in an intensive care unit. Careful and continuous assessment of the physiologic status is necessary. Arterial pressure through an indwelling line, pulse, and respiratory rate should be monitored continuously; a Foley catheter should be inserted to follow urine flow; and mental status assessed frequently.

Although there is ongoing debate as to the indications for using the flow-directed pulmonary artery catheter (PAC, Swan-Ganz catheter) in the management of patients in shock, most intensivists believe that the ability to predict the hemodynamic profiles of patients in shock accurately without a PAC is poor. The PAC is placed percutaneously via the subclavian or jugular vein through the central venous circulation and right heart into the pulmonary artery. There are ports both proximal in the right atrium and distal in the pulmonary artery to provide access for infusions and for cardiac output measurements. Right atrial and pulmonary artery pressures are measured, and the pulmonary capillary wedge pressure (PCWP) serves as an approximation of the left atrial pressure. Normal hemodynamic parameters are shown inTable 228-3 andTable 38-2.

Cardiac output is determined by the thermodilution technique, and high-resolution thermistors can also be used to determine right ventricular end-diastolic volume to monitor further the response of the right heart to fluid resuscitation. APAC with an oximeter port offers the additional advantage of on-line monitoring of the mixed venous oxygen saturation, an important index of tissue perfusion. Systemic and pulmonary vascular resistances are calculated as the ratio of the pressure drop across these vascular beds to the cardiac output (Chap. 228). Determinations of oxygen content in arterial and venous blood, together with cardiac output and hemoglobin concentration allow calculation of oxygen delivery, oxygen consumption, and oxygen-extraction ratio (Table 38-3). The hemodynamic patterns associated with the various form of shock are shown inTable 38-4.

In resuscitation from shock, it is critical to restore tissue perfusion and optimize oxygen delivery, hemodynamics, and cardiac function rapidly. A goal of therapy is to achieve normal mixed venous oxygen saturation and arteriovenous oxygen-extraction ratio. To enhance oxygen delivery, red cell mass, arterial oxygen saturation, and cardiac output may be augmented singly or simultaneously. An increase in oxygen delivery not accompanied by an increase in oxygen consumption implies that oxygen availability is

adequate and that oxygen consumption is not flow-dependent. Conversely, an elevation of oxygen consumption with increased cardiac output implies that the oxygen supply is inadequate. A reduction in systemic vascular resistance accompanying an increase in cardiac output indicates that compensatory vasoconstriction is reversing due to improved tissue perfusion. The determination of stepwise expansion of blood volume on cardiac performance allows identification of the optimum preload.

SPECIFIC FORMS OF SHOCK

HYPOVOLEMIC SHOCK

This most common form of shock results either from the loss of red blood cell mass and plasma from hemorrhage or from the loss of plasma volume alone arising from extravascular fluid sequestration or gastrointestinal, urinary, and insensible losses. The signs and symptoms of nonhemorrhagic hypovolemic shock are the same as those of hemorrhagic shock, although they may have a more insidious onset. The normal physiologic response to hypovolemia is to maintain perfusion of the brain and heart while restoring an effective circulating blood volume. There is an increase in sympathetic activity, hyperventilation, collapse of venous capacitance vessels, release of stress hormones, and expansion of intravascular volume through the recruitment of interstitial and intracellular fluid and reduction of urine output.

Mild hypovolemia (£20% of the blood volume) generates mild tachycardia but relatively few external signs, especially in a supine resting young patient (Table 38-5). With moderate hypovolemia (~20 to 40% of the blood volume) the patient becomes increasingly anxious and tachycardic; although normal blood pressure may be maintained in the supine position, there may be significant postural hypotension and tachycardia. If hypovolemia is severe (³~40% of the blood volume), the classic signs of shock appear; the blood pressure declines and becomes unstable even in the supine position, and the patient develops marked tachycardia, oliguria, and agitation or confusion. Perfusion of the central nervous system is well maintained until shock becomes severe. Hence, mental obtundation is an ominous clinical sign. The transition from mild to severe hypovolemic shock can be insidious or extremely rapid. If severe shock is not reversed rapidly, especially in elderly patients and those with comorbid illnesses, death is imminent. A very narrow time frame separates the derangements found in severe shock that can be reversed with aggressive resuscitation from those of progressive decompensation and irreversible cell injury.

Diagnosis Hypovolemic shock is readily diagnosed when there are signs of hemodynamic instability and the source of volume loss is obvious. The diagnosis is more difficult when the source of blood loss is occult, as into the gastrointestinal tract, or when plasma volume alone is depleted. After acute hemorrhage, hemoglobin and hematocrit values do not change until compensatory fluid shifts have occurred or exogenous fluid is administered. Thus, an initial normal hematocrit does not disprove the presence of significant blood loss. Plasma losses cause hemoconcentration, and free water loss leads to hypernatremia. These findings should suggest the presence of hypovolemia.

It is essential to distinguish between hypovolemic and cardiogenic shock (see below)

because definitive therapy differs significantly. Both forms are associated with a reduced cardiac output and a compensatory sympathetic mediated response characterized by tachycardia and elevated systemic vascular resistance. However, the findings in cardiogenic shock of jugular venous distention, rales, and an S_3gallop distinguish it from hypovolemic shock and signify that volume expansion is undesirable.

TREATMENT

Initial resuscitation requires rapid reexpansion of the circulating blood volume along with interventions to control ongoing losses. In accordance with Starling's law (Chap. 231), stroke volume and cardiac output rise with the increase in preload. After resuscitation, the compliance of the ventricles may remain reduced due to increased interstitial fluid in the myocardium. Therefore, elevated filling pressures are required to maintain adequate ventricular performance.

Volume resuscitation is initiated with the rapid infusion of isotonic saline or a balanced salt solution such as Ringer's lactate through large-bore intravenous lines. No distinct benefit from the use of colloid has been demonstrated. The infusion of 2 to 3 L over 10 to 30 min should restore normal hemodynamic parameters. Continued hemodynamic instability implies that shock has not been reversed and/or that there are significant ongoing blood or volume losses. Continuing blood loss, with hemoglobin concentrations declining to 100 g/L (10 g/dL), should initiate blood transfusion, preferably as fully cross-matched blood. In extreme emergencies, type-specific or O-negative packed red cells may be transfused. In the presence of severe and/or prolonged hypovolemia, inotropic support with dopamine or dobutamine (Chap. 72) may be required to maintain adequate ventricular performance, after blood volume has been restored. Infusion of norepinephrine to increase arterial pressure by raising peripheral resistance is inappropriate, other than as a temporizing measure in severe shock while blood volume is reexpanded.

Successful resuscitation also requires support of respiratory function. Supplemental oxygen should be provided, and endotracheal intubation may be necessary to maintain arterial oxygenation. Following resuscitation from isolated hemorrhagic shock, end-organ damage is frequently less than following septic or traumatic shock. This may be due to the absence of the massive activation of inflammatory mediator response systems and the consequent nonspecific organ injury seen in the latter conditions.

TRAUMATIC SHOCK

Shock following trauma is, in large measure, due to hypovolemia. However, even when hemorrhage has been controlled, patients can continue to suffer loss of plasma volume into the interstitium of injured tissues. These fluid losses are compounded by injury-induced inflammatory responses, which contribute to the secondary microcirculatory injury. This causes secondary tissue injury and maldistribution of blood flow, intensifying tissue ischemia and leading to multiple organ system failure. Trauma to the heart, chest, or head can also contribute to the shock. For example, pericardial tamponade or tension pneumothorax impairs ventricular filling, while myocardial contusion depresses myocardial contractility.

TREATMENT

Inability to maintain a systolic blood pressure ³90 mmHg after trauma-inudced hypovolemia is associated with a mortality rate of ~50%. To prevent decompensation of homeostatic mechanisms, therapy must be promptly administered.

The initial management of the seriously injured patient requires attention to the "ABCs" of resuscitation: assurance of an airway (A), adequate ventilation (breathing, B), and establishment of an adequate blood volume to support the circulation (C). Control of hemorrhage requires immediate attention. Early stabilization of fractures, debridement of devitalized or contaminated tissues, and evacuation of hematomata all reduce the subsequent inflammatory response to the initial insult and minimize subsequent organ injury.

INTRINSIC CARDIOGENIC SHOCK

This form of shock is caused by failure, often sudden, of the heart as an effective pump. It occurs most commonly as a complication of acute myocardial infarction (AMI;Chap. 243), but it may also be seen in patients with severe brady- or tachyarrhythmias, valvular heart disease, or in the terminal stage of chronic heart failure of any cause, including ischemic heart disease and dilated cardiomyopathy. Cardiogenic shock is characterized by a low cardiac output, diminished peripheral perfusion, pulmonary congestion, and elevation of systemic vascular resistance and pulmonary vascular pressures. Acute right heart failure can arise as the result of right ventricular infarction or may complicate the acute respiratory distress syndrome and severe pulmonary hypertension of any etiology. As a consequence of right ventricular failure, left ventricular preload falls, and this, in turn, reduces systemic perfusion. In contrast to other forms of shock, absolute or relative hypovolemia is usually not present in cardiogenic shock.

The ineffective contractile activity of either the right or left side of the heart leads to the accumulation of blood in the venous circulation upstream to the failing ventricle. Cardiogenic shock with left-sided heart failure increases fluid in the lungs that can overwhelm the capacity of the pulmonary lymphatics and causes interstitial and sometimes alveolar edema. Interstitial lung edema usually occurs at pulmonary capillary pressures >18 mmHg, and overt pulmonary alveolar edema develops at pressures>24 mmHg (Chap. 32). Pulmonary edema impacts cardiac function further by impairing diffusion of oxygen, setting up a vicious cycle. The increase in interstitial and intraalveolar fluid causes a progressive reduction in lung compliance, thereby increasing the work of ventilation while increasing perfusion of poorly ventilated alveoli.

In establishing the diagnosis of cardiogenic shock, a history of cardiac disease or of AMIis of value. Associated physical findings include those of hemodynamic instability, peripheral vasoconstriction, and pulmonary and/or systemic venous congestion, as well as findings specific to the underlying cardiac abnormalities. An electrocardiogram may provide evidence of AMI or preexisting cardiac disease. The chest x-ray may show pulmonary edema and cardiomegaly. Transthoracic or transesophageal echocardiograms assist in the diagnosis of structural abnormalities and/or functional impairment of contractility. Serum cardiac markers will support the diagnosis of acute

cardiac injury. Hemodynamic monitoring is usually necessary. Placement of aPAC is helpful and will show a reduced cardiac output and an elevatedPCWP, and direct measurement of right atrial pressure allows calculation of systemic vascular resistance which is elevated.

TREATMENT

For all forms of cardiogenic shock, preload, afterload, and contractility should be modified using the information provided by thePAC. APCWP of 15 to 20 mmHg should be the initial goal. If the PCWP is excessively elevated, inotropic agents may provide significant reduction. The goal is to increase contractility without significant increases in heart rate. Dopamine and norepinephrine exert both inotropic and vasoconstrictor actions (Chap. 72) that are useful in the presence of persistent hypotension. Dobutamine, a positive inotropic agent with vasodilator properties, may be substituted when arterial pressure has been restored. Pulmonary congestion may be responsive to intravenous furosemide. Patients with an inadequate response to these measures can be supported by using intraaortic balloon counterpulsation to permit recovery of myocardial function. Additional measures to consider in cases of refractory cardiogenic shock include urgent myocardial revascularization in patients withAMI(Chap. 243), correction of anatomic cardiac defects such as rupture of the papillary muscles of the interventricular septum, the placement of ventricular assist devices, and even urgent cardiac transplantation.

COMPRESSIVE CARDIOGENIC SHOCK

With compression, the heart and surrounding structures are less compliant and, thus, normal filling pressures generate inadequate diastolic filling. Blood or fluid within the poorly distensible pericardial sac may cause tamponade (Chap. 239). Any cause of increased intrathoracic pressure, such as tension pneumothorax, herniation of abdominal viscera through a diaphragmatic hernia, or excessive positive pressure ventilation to support pulmonary function, can also cause compressive cardiogenic shock. Acute right heart failure with a sudden decline in cardiac output can be caused by pulmonary embolism obstructing right ventricular outflow and impairing left ventricular filling. Although initially responsive to increased filling pressures produced by volume expansion, as compression increases, cardiogenic shock occurs.

The diagnosis of compressive cardiogenic shock is most frequently based on clinical findings, the chest radiograph, and an echocardiogram. The diagnosis of compressive cardiac shock may be more difficult to establish in the setting of trauma when hypovolemia and cardiac compression are present simultaneously. The classic findings of pericardial tamponade include the triad of hypotension, neck vein distention, and muffled heart sounds (Chap. 239). Pulsus paradoxus, i.e., an inspiratory reduction in systolic pressure >10 mmHg, may also be noted. The diagnosis is confirmed by echocardiography, and treatment consists of immediate pericardiocentesis. A tension pneumothorax produces ipsilateral decreased breath sounds, tracheal deviation away from the affected thorax, and jugular venous distention. Radiographic findings include increased intrathoracic volume, depression of the diaphragm of the affected hemithorax, and shifting of the mediastinum to the contralateral side. Chest decompression must be carried out immediately. Release of air and restoration of normal cardiovascular

dynamics is both diagnostic and therapeutic.

SEPTIC SHOCK (See alsoChap. 124)

This form of shock is caused by the systemic response to a severe infection. It occurs most frequently in elderly or immunocompromised patients and in those who have undergone an invasive procedure in which bacterial contamination has occurred. Infections of the lung, abdomen, or urinary tract are most common, and approximately half of the patients have bacteremia. Gram-positive and -negative bacteria, viruses, fungi, rickettsiae, and protozoa have all been reported to produce the clinical picture of septic shock, and the overall response is generally independent of the specific type of invading organism. The clinical findings in septic shock are a consequence of the combination of metabolic and circulatory derangements driven by the systemic infection and the release of toxic components of the infectious organisms, e.g., the endotoxin of gram-negative bacteria or the exotoxins and enterotoxins of gram-positive bacteria. Organism toxins lead to the release of cytokines, includingIL-1 andTNF-a, from tissue macrophages. Tissue factor expression and fibrin deposition are increased, and disseminated intravascular coagulation may develop. The inducible form of NO synthase is stimulated, and NO, a powerful vasodilator, is released. Hemodynamic changes in septic shock occur in two characteristic patterns: early, or hyperdynamic, and late, or hypodynamic, septic shock.

Hyperdynamic Response In hyperdynamic septic shock, tachycardia is present, the cardiac output is normal, and the systemic vascular resistance is reduced while the pulmonary vascular resistance is elevated. The extremities are usually warm. However, splanchnic vasoconstriction with decreased visceral flow is present. The venous capacitance is increased, which decreases venous return. With volume expansion cardiac output becomes supranormal. Myocardial contractility is depressed in septic shock by mediators including NO,IL-1, and/orTNF-a. Inflammatory mediator-induced processes include increased capillary permeability and continued loss of intravascular volume.

In septic shock, in contrast to other types of shock, total oxygen delivery may be increased while oxygen extraction is reduced due to maldistribution of microcirculatory perfusion and impaired utilization. In this setting the presence of a normal mixed venous oxygen saturation is not indicative of adequate peripheral perfusion, and even though the cardiac output may be elevated, it is still inadequate to meet the total metabolic needs. The toxicity of the infectious agents and their byproducts and the subsequent metabolic dysfunction drive the progressive deterioration of cellular and organ function. Acute respiratory distress syndrome, thrombocytopenia, and neutropenia are common complications.

Hypodynamic Response As sepsis progresses, vasoconstriction occurs and the cardiac output declines. The patient usually becomes markedly tachypneic, febrile, diaphoretic, and obtunded, with cool, mottled, and often cyanotic extremities. Oliguria, renal failure, and hypothermia develop; there may be striking increases in serum lactate.

TREATMENT

Aggressive volume expansion with a crystalloid solution to aPCWP of approximately 15 mmHg and the restoration of arterial oxygenation with inspired oxygen and frequently with mechanical ventilation are the highest priorities. In the presence of hypodynamic septic shock, augmentation of cardiac output may require inotropic support with dopamine or norepinephrine in the presence of hypotension or with dobutamine if arterial pressure is normal. Antibiotics should be administered, either appropriate for the results of cultures or empirical therapy based on the likely source of infection. Surgical debridement or drainage may also be necessary to control the infection.

NEUROGENIC SHOCK

Interruption of sympathetic vasomotor input after a high cervical spinal cord injury, inadvertent cephalad migration of spinal anesthesia, or severe head injury may result in neurogenic shock. In addition to arteriolar dilatation, venodilation causes pooling in the venous system, which decreases venous return and cardiac output. The extremities are often warm, in contrast to the usual vasoconstriction-induced coolness in hypovolemic or cardiogenic shock. Treatment involves a simultaneous approach to the relative hypovolemia and to the loss of vasomotor tone. Large volumes of fluid may be required to restore normal hemodynamics. Once hemorrhage has been ruled out, norepinephrine may be necessary to augment vascular resistance.

HYPOADRENAL SHOCK (See alsoChap. 331)

The normal host response to the stress of illness, operation, or trauma requires that the adrenal glands hypersecrete cortisol in excess of that normally required. Hypoadrenal shock occurs in settings in which unrecognized adrenal insufficiency complicates the host response to the stress induced by acute illness or major surgery. Adrenocortical insufficiency may occur as a consequence of the chronic administration of high doses of exogenous glucocorticoids. Recent studies have shown that prolonged stays in a critical state in an intensive care setting may also induce a relative hypoadrenal state. Other, less common causes include adrenal insufficiency secondary to idiopathic atrophy, tuberculosis, metastatic disease, bilateral hemorrhage, and amyloidosis. The shock produced by adrenal insufficiency is characterized by reductions in systemic vascular resistance, hypovolemia, and reduced cardiac output. The diagnosis of adrenal insufficiency may be established by means of anACTHstimulation test (Chap. 331).

TREATMENT

In the hemodynamically unstable patient, dexamethasone sodium phosphate, 4 mg, should be given intravenously. This agent is preferred because unlike hydrocortisone it does not interfere with theACTHstimulation test. If the diagnosis of adrenal insufficiency has been established, hydrocortisone, 100 mg every 6 to 8 h, can be given and tapered to a maintenance level as the patient achieves hemodynamic stability. Simultaneous volume resuscitation and pressor support is required.

ADJUNCTIVE THERAPIES

As described above, the sympathomimetic amines dobutamine, dopamine, and norepinephrine are widely used in the treatment of all forms of shock. The clinical

pharmacology of these agents is described in Chap. 72.

POSITIONING

Positioning of the patient may be a valuable adjunct in the initial treatment of hypovolemic shock. Elevating the foot of the bed (i.e., placing it on "shock blocks") and assumption of the Trendelenburg position without flexion at the knees are effective but may increase work of breathing and risk for aspiration. Simply elevating both legs may be the optimal approach.

PNEUMATIC ANTISHOCK GARMENT (PASG)

The PASG and the military antishock trousers (MAST) are inflatable external compression devices that can be wrapped around the legs and abdomen and have been widely used in the prehospital setting as a means of providing temporary support of central hemodynamics in shock. They cause an increase in systemic vascular resistance and blood pressure by arterial compression, without causing a significant change in cardiac output. While the use of PASG has been recommended in noncardiogenic forms of shock, the most appropriate use appears to be as a means to tamponade bleeding and augment hemostasis. Inflation of the suit provides splinting of fractures of the pelvis and lower extremities and arrests hemorrhage from fractures.

REWARMING

Hypothermia is a potential adverse consequence of massive volume resuscitation. The infusion of large volumes of refrigerated blood products and room-temperature crystalloid solutions can rapidly drop core temperatures if fluid is not run through warming devices. Hypothermia may depress cardiac contractility and thereby further impair cardiac output and oxygen delivery. Hypothermia, particularly temperatures <35° C, directly impairs the coagulation pathway, sometimes causing a significant coagulopathy. Rapid rewarming significantly decreases the requirement for blood products and an improvement in cardiac function. The most effective method for rewarming is extracorporeal countercurrent warmers through femoral artery and vein cannulation. This process does not require a pump and can rewarm from 30° to 36°C in<30 min.

(Bibliography omitted in Palm version)

Back to Table of Contents

39. CARDIOVASCULAR COLLAPSE, CARDIAC ARREST, AND SUDDEN CARDIAC DEATH - *Robert J. Myerburg*, *Agustin Castellanos*

OVERVIEW AND DEFINITIONS

The vast majority of naturally occurring sudden deaths are caused by cardiac disorders. The magnitude of sudden *cardiac* death (SCD) as a public health problem is highlighted by estimates that more than 300,000 deaths occur each year in the United States by this mechanism, accounting for 50% of all cardiac deaths. SCD is a direct consequence of cardiac arrest, which is often reversible if responded to promptly. Since resuscitation techniques and emergency rescue systems are available to save patients who have out-of-hospital cardiac arrest, which was uniformly fatal in the past, understanding the SCD problem has practical importance.

SCDmust be defined carefully. In the context of time, "sudden" is defined, for most clinical and epidemiologic purposes, as 1 h or less between the onset of the terminal clinical event, or an abrupt change in clinical status, and death. An exception is unwitnessed deaths in which pathologists may expand the definition of time to 24 h after the victim was last seen to be alive and stable.

Because of community-based interventions, victims may remain biologically alive for days or even weeks after a cardiac arrest that has resulted in irreversible central nervous system damage. Confusion in terms can be avoided by adhering strictly to definitions of death, cardiac arrest, and cardiovascular collapse (Table 39-1). Death is biologically, legally, and literally an absolute and irreversible event. Death may be delayed in a survivor of cardiac arrest, but "survival after sudden death" is an irrational term. Currently, the accepted definition of SCD is *natural death due to cardiac causes*, heralded by abrupt loss of consciousness within *1 h* of the onset of acute symptoms, in an individual who may have known *preexisting* heart disease but in whom the *time* and *mode* of death are *unexpected*. When biologic death of the cardiac arrest victim is delayed because of interventions, the relevant pathophysiologic event remains the sudden and unexpected cardiac arrest that leads ultimately to death, even though delayed by artificial methods. The language used should reflect the fact that the index event was a cardiac arrest and that death was due to its delayed consequences.

ETIOLOGY, INITIATING EVENTS, AND CLINICAL EPIDEMIOLOGY

Clinical and epidemiologic studies have identified populations at high risk forSCD. In addition, a large body of pathologic data provides information on the underlying *structural abnormalities* in victims of SCD, and studies of clinical physiology have begun to identify a group of *transient functional factors* that may convert a long-standing underlying structural abnormality from a stable to an unstable state (Table 39-2). This information is developing into an understanding of the causes and mechanisms of SCD.

Cardiac disorders constitute the most common causes of sudden *natural* death. After an initial peak incidence of sudden death between birth and 6 months of age (the sudden infant death syndrome), the incidence of sudden death declines sharply and remains low through childhood and adolescence. Among adolescents and young adults, the incidence ofSCD is approximately 1 per 100,000 population per year. The incidence

begins to increase in adults over the age of 30 years, reaching a second peak in the age range of 45 to 75 years, when the incidence approximates 1 to 2 per 1000 per year among the unselected adult population. Increasing age within this range is a powerful risk factor for sudden *cardiac* death, and the proportion of cardiac causes among all sudden natural deaths increases dramatically with advancing years. From 1 to 13 years of age, only one of five sudden *natural* deaths is due to cardiac causes. Between 14 and 21 years of age, the proportion increases to 30%, and then to 88% in the middle-aged and elderly.

Young and middle-aged men and women have very different susceptibilities toSCD, but the gender differences decrease with advancing age. In the 45- to 64-year-old age group, the male SCD excess is nearly 7:1. It falls to 2:1 or less in the 65- to 74-year-old age group. The difference in risk for SCD parallels the risks for other manifestations of coronary heart disease in men and women. As the gender gap for manifestations of coronary heart disease closes in the seventh and eighth decades of life, the excess risk of SCD in males also narrows. Despite the lower incidence among younger women, coronary risk factors such as cigarette smoking, diabetes, hyperlipidemia, and hypertension are highly influential, and SCD remains an important clinical and epidemiologic problem.

Hereditary factors contribute to the risk ofSCD, but largely in a nonspecific manner; they represent expressions of the hereditary predisposition to coronary heart disease. A few specific syndromes, such as congenital long QT interval syndromes (Chap. 230), right ventricular dysplasia, and the syndrome of right bundle branch block and non-ischemic ST-segment elevations (Brugada syndrome), are characterized by specific hereditary risk of SCD. There are also recent data suggesting a familial predispositon to SCD as a specific pattern of coronary heart disease expression.

The major categories of structural causes of, and functional factors contributing to, theSCDsyndrome are listed in Table 39-2. Worldwide, and especially in western cultures, coronary atherosclerotic heart disease is the most common structural abnormality associated with SCD. Up to 80% of all SCDs in the United States are due to the consequences of coronary atherosclerosis. The cardiomyopathies (dilated and hypertrophic, collectively;Chap. 239) account for another 10 to 15% of SCDs, and all the remaining diverse etiologies cause only 5 to 10% of these events. Transient ischemia in the previously scarred or hypertrophied heart, hemodynamic and fluid and electrolyte disturbances, fluctuations in autonomic nervous system activity, and transient electrophysiologic changes caused by drugs or other chemicals (e.g., proarrhythmia) have all been implicated as mechanisms responsible for transition from electrophysiologic stability to instability. In addition, reperfusion of ischemic myocardium may cause transient electrophysiologic instability and arrhythmias.

PATHOLOGY

Data from postmortem examinations ofSCDvictims parallel the clinical observations on the prevalence of coronary heart disease as the major structural etiologic factor. More than 80% of SCD victims have pathologic findings of coronary heart disease. The pathologic description often includes a combination of long-standing, extensive atherosclerosis of the epicardial coronary arteries and acute active coronary lesions,

which include a combination of fissured or ruptured plaques, platelet aggregates, hemorrhage, and thombosis. In one study, chronic coronary atherosclerosis involving two or more major vessels with ³75% stenosis was observed in 75% of the victims. In another study, atherosclerotic plaque fissuring, platelet aggregates, and/or acute thrombosis were observed in 95 of 100 individuals who had pathologic studies after SCD.

As many as 70 to 75% of males who die suddenly have prior myocardial infarctions (MIs), but only 20 to 30% have recent acute MIs. A high incidence of left ventricular (LV) hypertrophy coexists with prior MIs.

CLINICAL DEFINITION OF FORMS OF CARDIOVASCULAR COLLAPSE (Table 39-1)

Cardiovascular collapse is a general term connoting loss of effective blood flow due to acute dysfunction of the heart and/or peripheral vasculature. Cardiovascular collapse may be caused by vasodepressor syncope (vasovagal syncope, postural hypotension with syncope, neurocardiogenic syncope -- Chap. 21), a transient severe bradycardia, or cardiac arrest. The latter is distinguished from the transient forms of cardiovascular collapse in that it usually requires an intervention to achieve resuscitation. In contrast, vasodepressor syncope and many primary bradyarrhythmic syncopal events are transient and non-life-threatening, with spontaneous return of consciousness.

The most common electrical mechanism for true cardiac arrest is ventricular fibrillation (VF), which is responsible for 65 to 80% of cardiac arrests. Severe persistent bradyarrhythmias, asystole, and pulseless electrical activity (an organized electrical activity without mechanical response, formerly called electomechanical dissociation) cause another 20 to 30%. Sustained ventricular tachycardia (VT) with hypotension is a less common cause. Acute low cardiac output states, having precipitous onset, also may present clinically as a cardiac arrest. The causes include massive acute pulmonary emboli, internal blood loss from ruptured aortic aneurysm, intense anaphylaxis, cardiac rupture after myocardial infarction, and unexpected fatal arrhythmia due to electrolyte disturbances.

CLINICAL CHARACTERISTICS OF CARDIAC ARREST

PRODROME, ONSET, ARREST, DEATH

SCDmay be presaged by days, weeks, or months of increasing angina, dyspnea, palpitations, easy fatigability, and other nonspecific complaints. However, these *prodromal complaints* are generally predictive of any major cardiac event; they are not specific for predicting SCD.

The *onset of the terminal event*, leading to cardiac arrest, is defined as an acute change in cardiovascular status preceding cardiac arrest by up to 1 h. When the onset is instantaneous or abrupt, the probability that the arrest is cardiac in origin is>95%. Continuous ECG recordings, fortuitously obtained at the onset of a cardiac arrest, commonly demonstrate changes in cardiac electrical activity in the minutes or hours before the event. There is a tendency for the heart rate to increase and for advanced

grades of premature ventricular contractions (PVCs) to evolve. Most cardiac arrests that are caused by VFbegin with a run of sustained or nonsustainedVT, which then degenerates into VF.

Sudden unexpected loss of effective circulation may be separated into "arrhythmic events" and "circulatory failure." Arrhythmic events are characterized by a high likelihood of patients being awake and active immediately prior to the event, are dominated byVF as the electrical mechanism, and have a short duration of terminal illness (<1 h). In contrast, circulatory failure deaths occur in patients who are inactive or comatose, have a higher incidence of asystole than VF, have a tendency to a longer duration of terminal illness, and are dominated by noncardiac events preceding the terminal illness.

The onset of cardiac arrest may be characterized by typical symptoms of an acute cardiac event, such as prolonged angina or the pain of onset ofMI, acute dyspnea or orthopnea, or the sudden onset of palpitations, sustained tachycardia, or light-headedness. However, in many patients, the onset is precipitous, with minimal forewarning.

Cardiac arrest is, by definition, abrupt. Mentation may be impaired in patients with sustainedVTduring the onset of the terminal event. However, complete loss of consciousness is a *sine qua non* in cardiac arrest. Although rare spontaneous reversions occur, it is usual that cardiac arrest progresses to death within minutes (i.e.,SCD has occurred) if active interventions are not undertaken promptly.

The probability of achieving successful resuscitation from cardiac arrest is related to the interval from onset to institution of resuscitative efforts, the setting in which the event occurs, the mechanism (VF,VT, pulseless electrical activity, asystole), and the clinical status of the patient prior to the cardiac arrest. Those settings in which it is possible to institute prompt cardiopulmonary resuscitation (CPR) provide a better chance of a successful outcome. However, the outcome in intensive care units and other in-hospital environments is heavily influenced by the patient's preceding clinical status. The immediate outcome is good for cardiac arrest occurring in the intensive care unit in the presence of an acute cardiac event or transient metabolic disturbance, but the outcome for patients with far-advanced chronic cardiac disease or advanced noncardiac diseases (e.g., renal failure, pneumonia, sepsis, diabetes, cancer) is not much more successful in hospital than in the out-of-hospital setting.

The success rate for initial resuscitation and survival to hospital discharge after an out-of-hospital cardiac arrest depends in part on the mechanism of the event. When the mechanism isVT, the outcome is best;VF is the next most successful; and asystole and pulseless electrical activity generate dismal outcome statistics (Fig. 39-1). Advanced age also influences adversely the chances of successful resuscitation.

Progression to biologic death is a function of the mechanism of cardiac arrest and the length of the delay before interventions.VF or asystole withoutCPRwithin the first 4 to 6 min has a poor outcome, and there are few survivors among patients who had no life support activities for the first 8 min after onset. Outcome statistics are improved by lay bystander intervention (basic life support -- see below) prior to definitive interventions

(advanced life support -- defibrillation) and even more by early defibrillation. In regard to the latter, the notion that deployment of automatic external defibrillators in communities (e.g., police vehicles, large buildings, stadiums, etc.) will result in improved survival is currently being evaluated.

Death during the hospitalization after a successfully resuscitated cardiac arrest relates closely to the severity of central nervous system injury. Anoxic encephalopathy and infections subsequent to prolonged respirator dependence account for 60% of the deaths. Another 30% occur as a consequence of low cardiac output states that fail to respond to interventions. Recurrent arrhythmias are the least common cause of death, accounting for only 10% of in-hospital deaths.

In the setting of acuteMI, it is important to distinguish between primary and secondary cardiac arrests. *Primary* cardiac arrests refer to those that occur in the absence of hemodynamic instability, and *secondary* cardiac arrests are those that occur in patients in whom abnormal hemodynamics dominate the clinical picture before cardiac arrest. The success rate for immediate resuscitation in primary cardiac arrest during acute MI in a monitored setting should approach 100%. In contrast, as many as 70% of patients with secondary cardiac arrest succumb immediately or during the same hospitalization.

IDENTIFICATION OF PATIENTS AT RISK FOR SUDDEN CARDIAC DEATH

Primary prevention of cardiac arrest depends on the ability to identify individual patients at high risk. One must view the problem in the context of the total number of events and the population pools from which they are derived. The annual incidence ofSCDamong an unselected adult population is 1 to 2 per 1000 population (Fig. 39-2*A*), largely reflecting the prevalence of those coronary heart disease patients among whom SCD is the first clinically recognized manifestation (20 to 25% of first coronary events are SCD). The incidence (percent per year) increases progressively with the addition of identified coronary risk factors to populations free of prior coronary events. The most powerful factors are age, elevated blood pressure,LVhypertrophy, cigarette smoking, elevated serum cholesterol level, obesity, and nonspecific electrocardiographic abnormalities. These coronary risk factors are not specific for SCD but rather represent increasing risk for all coronary deaths. The proportion of coronary deaths that are sudden remains at approximately 50% in all risk categories. Despite the marked *relative* increased risk of SCD with addition of multiple risk factors (from 1 to 2 per 1000 population per year in an unselected population to as much as 50 to 60 per 1000 in subgroups having multiple risk factors for coronary artery disease), the *absolute* incidence remains relatively low when viewed as the relationship between the number of individuals who have a preventive intervention and the number of events that can be prevented. Specifically, a 50% reduction in annual SCD risk would be a huge *relative* decrease but would require an intervention in up to 200 unselected individuals to prevent one sudden death. These figures highlight the importance of primary prevention of coronary heart disease. Control of coronary risk factors may be the only practical method to prevent SCD in major segments of the population, because of the paradox that the majority of events occur in the large unselected subgroups rather than in the specific high-risk subgroups (compare "Events/Year" with "Percent/Year" inFig. 39-2*A*). Under most conditions of higher level of risk, particularly those indexed to a recent major cardiovascular event (e.g.,MI, recent onset of heart failure, survival after out-of-hospital cardiac arrest), the highest risk of

sudden death occurs within the initial 6 to 18 months and then decreases toward baseline risk of the underlying disease (Fig. 39-2*B*). Accordingly, preventive interventions are most likely to be effective when initiated early.

For patients with acute or prior clinical manifestations of coronary heart disease, high-risk subgroups having a much higher ratio ofSCD risk to population base can be identified. The acute, convalescent, and chronic phases ofMIprovide large population subsets with more highly focused risk (Chap. 243). The potential risk of cardiac arrest from the onset through the first 72 h after acute MI (the acute phase) may be as high as 15 to 20%. The highest risk of SCD in relation to MI is found in the subgroup that has experienced sustainedVT orVFduring the convalescent phase (3 days to 8 weeks) after MI. A greater than 50% mortality in 6 to 12 months has been observed among these patients, when managed with conservative medical therapy, and at least 50% of the deaths are sudden. Aggressive intervention techniques may reduce this incidence.

After the acute phase ofMI, long-term risk for total mortality andSCD are predicted by a number of factors. The most important for both SCD and non-SCD is the extent of myocardial damage sustained during the acute event. This is measured by the degree of reduction in the ejection fraction (EF), functional capacity, and/or the occurrence of heart failure. Increasing *frequency* of postinfarctionPVCs, with a plateau above the range of 10 to 30 PVCs per hour on 24-h ambulatory monitor recordings, also indicates increased risk, but advanced *forms* (salvos, nonsustainedVT) may be more powerful predictors. PVCs interact strongly with decreased left ventricular EF. The combination of frequent PVCs, salvos or nonsustained VT, and an EF £ 35% identifies patients who have an annual risk of greater than 20%. The risk falls off sharply with decreasing PVC frequency and the absence of advanced forms, as well as with higher EF. Despite the risk implications of postinfarction PVCs, improved outcome as a result of PVC suppression has not been demonstrated (Chap. 230).

The extent of underlying disease due to any cause and/or prior clinical expression of risk of SCD(i.e., survival after out-of-hospital cardiac arrest not associated with acuteMI) identify patients at very high risk for subsequent (recurrent) cardiac arrest. Survival after out-of-hospital cardiac arrest predicts up to a 30% 1-year recurrent cardiac arrest rate in the absence of specific interventions (see below).

A general rule is that the risk ofSCD is approximately one-half the total cardiovascular mortality rate. As shown inFig. 39-2*A*, the very high risk subgroups provide more focused population fractions ("Percent/Year") for predicting cardiac arrest or SCD; but the impact on the overall population, indicated by the absolute number of preventable events ("Events/Year"), is considerably smaller. The requirements for achieving a major population impact are effective prevention of the underlying diseases and/or new epidemiologic probes that will allow better resolution of subgroups within large general populations.

TREATMENT

The individual who collapses suddenly is managed in four stages: (1) the initial response and basic life support; (2) advanced life support; (3) postresuscitation care; and (4) long-term management. The initial response and basic life support can be

carried out by physicians, nurses, paramedical personnel, and trained lay persons. There is a requirement for increasingly specialized skills as the patient moves through the stages of advanced life support, postresuscitation care, and long-term management.

Initial Response and Basic Life Support The initial response will confirm whether a sudden collapse is indeed due to a cardiac arrest. Observations for respiratory movements, skin color, and the presence or absence of pulses in the carotid or femoral arteries will promptly determine whether a life-threatening cardiac arrest has occurred. As soon as a cardiac arrest is suspected or confirmed, contacting an emergency rescue system (e.g., 911) should be the immediate priority.

Agonal respiratory movements may persist for a short time after the onset of cardiac arrest, but it is important to observe for severe stridor with a persistent pulse as a clue to aspiration of a foreign body or food. If this is suspected, a Heimlich maneuver (see below) may dislodge the obstructing body. A precordial blow, or "thump," delivered firmly by the clenched fist to the junction of the middle and lower third of the sternum may occasionally revertVT orVF, but there is concern about converting VT *to* VF. Therefore, it has been recommended to use precordial thumps as an advanced life support technique when monitoring and defibrillation are available. This conservative application of the technique remains controversial.

The third action during the initial response is to clear the airway. The head is tilted back and chin lifted so that the oropharynx can be explored to clear the airway. Dentures or foreign bodies are removed, and the Heimlich maneuver is performed if there is reason to suspect that a foreign body is lodged in the oropharynx. If respiratory arrest precipitating cardiac arrest is suspected, a second precordial thump is delivered after the airway is cleared.

Basic life support, more popularly known asCPR, is intended to maintain organ perfusion until definitive interventions can be instituted. The elements of CPR are the maintenance of ventilation of the lungs and compression of the chest. Mouth-to-mouth respiration may be used if no specific rescue equipment is immediately available (e.g., plastic oropharyngeal airways, esophageal obturators, masked Ambu bag). Conventional ventilation techniques during CPR require the lungs to be inflated 10 to 12 times per minute, i.e., once every fifth chest compression when two persons are performing the resuscitation and twice in succession every 15 chest compressions when one person is carrying out both ventilation and chest wall compression.

Chest compression is based on the assumption that cardiac compression allows the heart to maintain a pump function by sequential filling and emptying of its chambers, with competent valves maintaining forward direction of flow. The palm of one hand is placed over the lower sternum, with the heel of the other resting on the dorsum of the lower hand. The sternum is depressed, with the arms remaining straight, at a rate of approximately 80 to 100 per minute. Sufficient force is applied to depress the sternum 3 to 5 cm, and relaxation is abrupt.

Advanced Life Support Advanced life support is intended to achieve adequate ventilation, control cardiac arrhythmias, stabilize blood pressure and cardiac output, and restore organ perfusion. The activities carried out to achieve these goals include (1)

intubation with an endotracheal tube, (2) defibrillation/cardioversion and/or pacing, and (3) insertion of an intravenous line. Ventilation with O_2 (room air if O_2 is not immediately available) may promptly reverse hypoxemia and acidosis. The speed with which defibrillation/cardioversion is carried out is an important element for successful resuscitation. When possible, immediate defibrillation should precede intubation and insertion of an intravenous line;CPRshould be carried out while the defibrillator is being charged. As soon as a diagnosis ofVT orVF is obtained, a 200-J shock should be delivered. Additional shocks at higher energies, up to a maximum of 360 J, are tried if the initial shock does not successfully abolish VT or VF. Epinephrine, 1 mg intravenously, is given after failed defibrillation, and attempts to defibrillate are repeated. The dose of epinephrine may be repeated after intervals of 3 to 5 min (seeFig. 39-3*A*).

If the patient is less than fully conscious upon reversion, or if two or three attempts fail, prompt intubation, ventilation, and arterial blood gas analysis should be carried out. Intravenous $NaHCO_3$, which was formerly used in large quantities, is no longer considered routinely necessary and may be dangerous in larger quantities. However, the patient who is persistently acidotic after successful defibrillation and intubation should be given 1 meq/kg $NaHCO_3$initially and an additional 50% of the dose repeated every 10 to 15 min.

After initial unsuccessful defibrillation attempts, or with persistent electrical instability, a bolus of 1 mg/kg lidocaine is given intravenously (Chap. 243), and the dose is repeated in 2 min in those patients who have persistent ventricular arrhythmias or remain inVF. This is followed by a continuous infusion at a rate of 1 to 4 mg/min. If lidocaine fails to provide control, other antiarrhythmic therapies should be tried. For persistent, hemodynamically unstable ventricular arrhythmias, intravenous amiodarone has emerged as the treatment of choice (150 mg over 10 min, followed by 1 mg/min for up to 6 h, and 0.5 mg/min thereafter) (Fig. 39-3*A*). Intravenous procainamide (loading infusion of 100 mg/5 min to a total dose of 500 to 800 mg, followed by continuous infusion at 2 to 5 mg/min) may be tried for persisting, hemodynamically stable arrhythmias; or bretylium tosylate (loading dose 5 to 10 mg/kg in 5 min; maintenance dose 0.5 to 2 mg/min) may be tried as an alternative for unstable arrhythmias. Intravenous calcium gluconate is no longer considered safe or necessary for routine administration. It is used only in patients in whom acute hyperkalemia is known to be the triggering event for resistant VF, in the presence of known hypocalcemia, or in patients who have received toxic doses of calcium channel antagonists.

Cardiac arrest secondary to bradyarrhythmias or asystole is managed differently (Fig. 39-3*B*). The patient is promptly intubated,CPR is continued, and an attempt is made to control hypoxemia and acidosis. Epinephrine and/or atropine are given intravenously or by an intracardiac route. External pacing devices are now available to attempt to establish a regular rhythm, but the prognosis is generally very poor in this form of cardiac arrest, even with successful electrical pacing. Pulseless electrical activity (PEA) is treated similarly to bradyarrhythmias, but its outcome is also dismal. The one exception is bradyarrhythmic/asystolic cardiac arrest secondary to airway obstruction. This form of cardiac arrest may respond promptly to removal of foreign bodies by the Heimlich maneuver or, in hospitalized patients, by intubation and suctioning of obstructing secretions in the airway.

Postresuscitation Care This phase of management is determined by the clinical setting of the cardiac arrest. *Primary*VF in acute MI(Chap. 243) is generally very responsive to life-support techniques and easily controlled after the initial event. Patients are maintained on a lidocaine infusion at the rate of 2 to 4 mg/min for 24 to 72 h after the event. In the in-hospital setting, respirator support is usually not necessary or is needed for only a short time, and hemodynamics stabilize promptly after defibrillation or cardioversion. In *secondary* VF in acute MI (those events in which hemodynamic abnormalities predispose to the potentially fatal arrhythmia), resuscitative efforts are less often successful, and in those patients who are successfully resuscitated, the recurrence rate is high. The clinical picture and outcome are dominated by hemodynamic instability and the ability to control hemodynamic dysfunction. Bradyarrhythmias, asystole, and pulseless electrical activity are commonly secondary events in hemodynamically unstable patients.

The outcome after in-hospital cardiac arrest associated with *non-cardiac* diseases is poor, and in the few successfully resuscitated patients, the postresuscitation course is dominated by the nature of the underlying disease. Patients with cancer, renal failure, acute central nervous system disease, and uncontrolled infections, as a group, have a survival rate of less than 10% after in-hospital cardiac arrest. Some major exceptions are patients with transient airway obstruction, electrolyte disturbances, proarrhythmic effects of drugs, and severe metabolic abnormalities, most of whom may have an excellent chance of survival if they can be resuscitated promptly and maintained while the transient abnormalities are being corrected.

Long-Term Management after Survival of Out-of-Hospital Cardiac Arrest Patients who do not suffer irreversible injury of the central nervous system and who achieve hemodynamic stability should have extensive diagnostic and therapeutic testing to guide long-term management. This aggressive approach is driven by the fact that statistics from the 1970s indicated survival after out-of-hospital cardiac arrest was followed by a 30% recurrent cardiac arrest rate at 1 year, 45% at 2 years, and a total mortality rate of almost 60% at 2 years. Historical comparisons suggest that these dismal statistics may be significantly improved by newer interventions, but the magnitude of the improvement is unknown because of the lack of concurrently controlled intervention studies.

Among those patients in whom an acute transmuralMI is the cause of out-of-hospital cardiac arrest, the management is the same as in any other patient who suffers cardiac arrest during the acute phase of a documented MI (Chap. 243). For almost all other categories of patients, however, extensive diagnostic studies are carried out to determine etiology, functional impairment, and electrophysiologic instability as guides to future management. In general, patients who have out-of-hospital cardiac arrest due to chronic ischemic heart disease, without an acute MI, are evaluated to determine whether transient ischemia or chronic electrophysiologic instability was the more likely cause of the event. If there is reason to suspect an ischemic mechanism, coronary revascularization by angioplasty or bypass surgery, plus drugs (most commonly beta blockers), are used to reduce ischemic burden.

Electrophysiologic instability has been identified by the use of programmed electrical stimulation to determine whether sustainedVT orVF can be induced (Chap. 230). If so, this information can be used as a baseline against which to evaluate drug efficacy for

prevention of inducibility. The rationale for this approach is the assumption that suppression of inducibility predicts long-term benefit by the drug that achieves such suppression. For patients for whom successful drug therapy could not be identified by this technique, insertion of an implantable cardioverter-defibrillator (ICD), antiarrhythmic surgery (e.g., coronary bypass surgery, aneurysmectomy, cryoablation), or empiric amiodarone therapy have been recommended (Chap. 230). Primary surgical success, defined as surviving the procedure and reverting to a noninducible status without drug therapy, is better than 90% when patients are selected for ability to be mapped in the operating room. However, only a small fraction of patients meet the criteria. In addition, VT/VF *cannot* be induced in a number of survivors of cardiac arrest (30 to 50%), and inducible arrhythmias can be suppressed by drugs in no more than 20 to 30% of those whose arrhythmias can be induced. Because of these limitations of drug therapy and surgical approaches, ICD therapy has evolved into the most commonly used strategy for cardiac arrest survivors. ICDs have long been recognized to have very good success rates for sensing and reverting life-threatening arrhythmias, but improvement in long-term total survival outcomes remained lacking until a number of studies solidified the benefit of ICD therapy for specific subgroups. After empiric amiodarone therapy had been suggested to be as good as, or better than, conventional antiarrhythmic drug therapy for survivors of cardiac arrest, ICDs were demonstrated to be superior to amiodarone. Moreover, ICDs were also found to be superior for high risk patients with VT after myocardial infarction.

(Bibliography omitted in Palm version)

Back to Table of Contents

SECTION 6 -ALTERATIONS IN GASTROINTESTINAL FUNCTION

40. DYSPHAGIA - *Raj K. Goyal*

Dysphagia is defined as a sensation of "sticking" or obstruction of the passage of food through the mouth, pharynx, or esophagus. It should be distinguished from other symptoms related to swallowing. *Aphagia* signifies complete esophageal obstruction, which is usually due to bolus impaction and represents a medical emergency. *Difficulty in initiating a swallow* occurs in disorders of the voluntary phase of swallowing. However, once initiated, swallowing is completed normally. *Odynophagia* means painful swallowing. Frequently, odynophagia and dysphagia occur together. *Globus pharyngeous* is the sensation of a lump lodged in the throat. However, no difficulty is encountered when swallowing is performed. *Misdirection of food*, resulting in nasal regurgitation and laryngeal and pulmonary aspiration of food during swallowing, is characteristic of oropharyngeal dysphagia. *Phagophobia*, meaning fear of swallowing, and *refusal to swallow* may occur in hysteria, rabies, tetanus, and pharyngeal paralysis due to fear of aspiration. Painful inflammatory lesions that cause odynophagia may also cause refusal to swallow. Some patients may feel the food as it goes down the esophagus. This esophageal sensitivity is not associated with either food sticking or obstruction, however. Similarly, the *feeling of fullness in the epigastrium* that occurs after a meal or after swallowing air should not be confused with dysphagia.

PHYSIOLOGY OF SWALLOWING

The process of swallowing begins with a voluntary (oral) phase during which a bolus of food is pushed into the pharynx by the contraction of the tongue. The bolus then activates oropharyngeal sensory receptors that initiate the involuntary (pharyngeal and esophageal) phase, or deglutition reflex. The deglutition reflex is a complex series of events and serves both to propel food through the pharynx and the esophagus and to prevent its entry into the airway. When the bolus is propelled backward by the tongue, the larynx moves forward and the upper esophageal sphincter opens. As the bolus moves into the pharynx, contraction of the superior pharyngeal constrictor against the contracted soft palate initiates a peristaltic contraction that proceeds rapidly downward to move the bolus through the pharynx and the esophagus. The lower esophageal sphincter opens as the food enters the esophagus and remains open until the peristaltic contraction has swept the bolus into the stomach. Peristaltic contraction in response to a swallow is called *primary peristalsis*. It involves inhibition followed by sequential contraction of muscles along the entire swallowing passage. The inhibition that precedes the peristaltic contraction is called *deglutitive inhibition*. Local distention of the esophagus from food activates intramural reflexes in the smooth muscle and results in *secondary peristalsis*, which is limited to the thoracic esophagus. *Tertiary contractions* are nonperistaltic because they occur simultaneously over a long segment of the esophagus. Tertiary contractions may occur in response to a swallow or esophageal distention, or they may occur spontaneously.

PATHOPHYSIOLOGY OF DYSPHAGIA

The normal transport of an ingested bolus through the swallowing passage depends on the size of the ingested bolus; the luminal diameter of the swallowing passage; the force

of peristaltic contraction; and deglutitive inhibition, including normal relaxation of upper and lower esophageal sphincters during swallowing. Dysphagia caused by a large bolus or luminal narrowing is called *mechanical dysphagia*, whereas dysphagia due to weakness of peristaltic contractions or to impaired deglutitive inhibition causing nonperistaltic contractions and impaired sphincter relaxation is called *motor dysphagia*.

Mechanical Dysphagia Mechanical dysphagia can be caused by a very large food bolus, intrinsic narrowing, or extrinsic compression of the lumen. In an adult, the esophageal lumen can distend up to 4 cm in diameter. When the esophagus cannot dilate beyond 2.5 cm in diameter, dysphagia to normal solid food can occur. Dysphagia is always present when the esophagus cannot distend beyond 1.3 cm. Circumferential lesions produce dysphagia more consistently than do lesions that involve only a portion of circumferences of the esophageal wall, as uninvolved segments retain their distensibility. The causes of mechanical dysphagia are listed in Table 40-1. Common causes include carcinoma, peptic and other benign strictures, and lower esophageal ring.

Motor Dysphagia Motor dysphagia may result from difficulty in initiating a swallow or from abnormalities in peristalsis and deglutitive inhibition due to diseases of the esophageal striated or smooth muscle.

Diseases of the striated muscle involve the pharynx, upper esophageal sphincter, and cervical esophagus. The striated muscle is innervated by a somatic component of the vagus with cell bodies of the lower motor neurons located in the nucleus ambiguus. These neurons are cholinergic and excitatory and are the sole determinant of the muscle activity. Peristalsis in the striated muscle segment is due to sequential central activation of neurons innervating muscles at different levels along the esophagus. Motor dysphagia of the pharynx results from neuromuscular disorders causing muscle paralysis, simultaneous nonperistaltic contraction, or loss of opening of the upper esophageal sphincter. Loss of opening of the upper sphincter is caused by paralysis of geniohyoid and other suprahyoid muscles or loss of deglutitive inhibition of the cricopharyngeus muscle. Because each side of the pharynx is innervated by ipsilateral nerves, a unilateral lesion of motor neurons leads to unilateral pharyngeal paralysis. Although lesions of striated muscle also involve the cervical part of the esophagus, the clinical manifestations of pharyngeal dysfunction usually overshadow those due to esophageal involvement.

Diseases of the smooth-muscle segment involve the thoracic part of the esophagus and the lower esophageal sphincter. The smooth muscle is innervated by the parasympathetic component of the vagal preganglionic fibers and postganglionic neurons in the myenteric ganglia. The vagal pathway consists of parallel excitatory and inhibitory pathways that use acetylcholine and nitric oxide as neurotransmitters, respectively. The activation of inhibitory nerves causes inhibition that is followed by rebound contraction. These pathways are involved in the resting tone of the lower esophageal sphincter as well as swallow-induced lower esophageal sphincter opening and inhibition followed by peristaltic contractions in the esophageal body. Dysphagia results when the peristaltic contractions are weak or nonperistaltic or when the lower sphincter fails to relax normally. Loss of contractile power occurs due to muscle weakness, as in scleroderma. The nonperistaltic contractions and impaired relaxation of

the lower esophageal sphincter result from a defect in inhibitory vagal innervation and account for dysphagia in achalasia.

The causes of motor dysphagia are also listed inTable 40-1. Important causes are pharyngeal paralysis, cricopharyngeal achalasia, scleroderma of the esophagus, achalasia, and diffuse esophageal spasm and related motor disorders.

Approach to the Patient

History The history can provide a presumptive diagnosis in over 80% of patients. The type of food causing dysphagia provides useful information. Difficulty only with solids implies mechanical dysphagia with a lumen that is not severely narrowed. In advanced obstruction, dysphagia occurs with liquids as well as solids. In contrast, motor dysphagia due to achalasia and diffuse esophageal spasm is equally affected by solids and liquids from the very onset. Patients with scleroderma have dysphagia to solids that is unrelated to posture and to liquids while recumbent but not upright. When peptic stricture develops in patients with scleroderma, dysphagia becomes more persistent.

The duration and course of dysphagia are helpful in diagnosis. Transient dysphagia may be due to an inflammatory process. Progressive dysphagia lasting a few weeks to a few months is suggestive of carcinoma of the esophagus. Episodic dysphagia to solids lasting several years indicates a benign disease characteristic of a lower esophageal ring.

The site of dysphagia described by the patient helps to determine the site of esophageal obstruction; the lesion is at or below the perceived location of dysphagia.

Associated symptoms provide important diagnostic clues. Nasal regurgitation and tracheobronchial aspiration with swallowing are hallmarks of pharyngeal paralysis or a tracheoesophageal fistula. Tracheobronchial aspiration unrelated to swallowing may be secondary to achalasia, Zenker's diverticulum, or gastroesophageal reflux.

Severe weight loss that is out of proportion to the degree of dysphagia is highly suggestive of carcinoma. When hoarseness precedes dysphagia, the primary lesion is usually in the larynx. Hoarseness following dysphagia may suggest involvement of the recurrent laryngeal nerve by extension of esophageal carcinoma. Sometimes hoarseness may be due to laryngitis secondary to gastroesophageal reflux. Association of laryngeal symptoms and dysphagia also occurs in various neuromuscular disorders. Hiccups may rarely occur with a lesion in the distal portion of the esophagus. Unilateral wheezing with dysphagia indicates a mediastinal mass involving the esophagus and a large bronchus.

Chest pain with dysphagia occurs in diffuse esophageal spasm and related motor disorders. Chest pain resembling diffuse esophageal spasms may occur in esophageal obstruction due to a large bolus. A prolonged history of heartburn and reflux preceding dysphagia indicates peptic stricture. A history of prolonged nasogastric intubation, ingestion of caustic agents, ingestion of pills without water, previous radiation therapy, or associated mucocutaneous diseases may provide the cause of esophageal stricture. If odynophagia is present, candidal or herpes esophagitis or pill-induced esophagitis

should be suspected.

In patients with AIDS or other immunodeficiency states, esophagitis due to opportunistic infections such as *Candida*, herpes simplex virus, or cytomegalovirus and tumors such as Kaposi's sarcoma and lymphoma should be suspected.

Physical Examination Physical examination is important in motor dysphagia due to skeletal muscle, neurologic, and oropharyngeal diseases. Signs of bulbar or pseudobulbar palsy, including dysarthria, dysphonia, ptosis, tongue atrophy, and hyperactive jaw jerk, in addition to evidence of generalized neuromuscular disease, should be sought. The neck should be examined for thyromegaly or a spinal abnormality. A careful inspection of the mouth and pharynx should disclose lesions that may interfere with passage of food because of pain or obstruction. Changes in the skin and extremities may suggest a diagnosis of scleroderma and other collagen-vascular diseases or mucocutaneous diseases such as pemphigoid or epidermolysis bullosa, which may involve the esophagus. Cancer spread to lymph nodes and liver may be evident. Pulmonary complications of acute aspiration pneumonia or chronic aspiration may be present.

Diagnostic Procedures Dysphagia is nearly always a symptom of organic disease rather than a functional compaint. If oropharyngeal dysphagia is suspected, videofluoroscopy of oropharyngeal swallowing should be obtained. If mechanical dysphagia is suspected on clinical history, barium swallow, esophagogastroscopy and endoscopic biopsies are the diagnostic procedures of choice. Barium swallow and esophageal motility studies are diagnostic tests for motor dysphagia. Esophagogastroscopy may be needed in patients with motor dysphagia to exclude an associated structural abnormality (Chap. 284).

(Bibliography omitted in Palm version)

Back to Table of Contents

41. NAUSEA, VOMITING, AND INDIGESTION - *William L. Hasler*

Nausea is the subjective feeling of a need to vomit. Vomiting (emesis) is the oral expulsion of upper gastrointestinal contents resulting from contractions of gut and thoracoabdominal wall musculature. *Vomiting* is contrasted with regurgitation, the effortless passage of gastric contents into the mouth. *Rumination* is the repeated regurgitation of stomach contents, which are often rechewed and then reswallowed. In contrast to vomiting, these phenomena often exhibit some volitional control. *Indigestion* is a nonspecific term that encompasses a variety of upper abdominal complaints including nausea, vomiting, heartburn, regurgitation, and dyspepsia (upper abdominal discomfort or pain). Individuals with ulcer-like dyspepsia report epigastric burning or gnawing discomfort. Dysmotility-like dyspepsia is characterized by postprandial fullness, bloating, eructation (belching), anorexia (loss of appetite), and early satiety (an inability to complete a meal due to premature fullness).

NAUSEA AND VOMITING

MECHANISMS

Vomiting is coordinated by the brain stem and is effected by neuromuscular responses in the gut, pharynx, and thoracoabdominal wall. The mechanisms underlying nausea are poorly understood. Because nausea requires conscious perception, the sensation is probably mediated by the cerebral cortex. Electroencephalographic studies show activation of temporofrontal cortical regions with induction of nausea.

Coordination of Emesis Animal studies suggested that vomiting was coordinated by a single locus in the medullary reticular formation. However, further work has shown that no one "vomiting center" exists and that several brain stem nuclei initiate emesis, including the nucleus tractus solitarius; the dorsal vagal and phrenic nuclei; medullary nuclei that regulate respiration; and nuclei that control pharyngeal, facial, and tongue movements. The neurotransmitters involved in coordinating emesis are uncertain; however, neurokinin NK_1, serotonin, and vasopressin pathways are postulated.

Somatic and visceral muscles exhibit stereotypic responses during emesis. Inspiratory thoracic and abdominal wall muscles contract, producing high intrathoracic and intraabdominal pressures that facilitate expulsion of gastric contents. The gastric cardia herniates across the diaphragm, and the larynx moves upward to promote oral propulsion of the vomitus. Under normal conditions, distally migrating upper gut contractions are regulated by an electrical phenomenon, the slow wave, which cycles at 3 cycles per minute in the stomach and 11 cycles per minute in the duodenum. With emesis, slow waves are replaced by orally propagating spike activity, which induces retrograde contractions that assist in the oral expulsion of small intestinal contents.

Activators of Emesis Emetic stimuli act at several anatomic sites. Emesis provoked by noxious thoughts or smells originates in the cerebral cortex, whereas cranial nerves mediate vomiting after gag reflex activation. Motion sickness and inner ear disorders act on the labyrinthine apparatus, while gastric irritants and emetogenic anticancer agents such as cisplatin stimulate gastroduodenal vagal afferent nerves. Nongastric visceral afferents are activated by small intestinal and colonic obstruction and mesenteric

ischemia. The area postrema, a medullary nucleus, responds to bloodborne emetic stimuli and is termed the *chemoreceptor trigger zone*. Many emetic drugs act on the area postrema as do bacterial toxins and metabolic disorders such as uremia, hypoxia, and ketoacidosis.

Neurotransmitters that mediate induction of vomiting are selective for these anatomic sites. Labyrinthine disorders stimulate vestibular cholinergic muscarinic M_1 and histaminergic H_1receptors, whereas gastroduodenal vagal afferent stimuli activate serotonin 5-HT_3receptors. The area postrema is richly served by nerve fibers acting on diverse receptor subtypes including 5-HT_3, M_1, H_1, and dopamine D_2. Optimal pharmacologic management of the patient with vomiting requires an understanding of these pathways.

DIFFERENTIAL DIAGNOSIS

Nausea and vomiting are caused by conditions within and outside the gut as well as by drugs and circulating toxins (Table 41-1).

Intraperitoneal Disorders Visceral obstruction and inflammation of hollow and solid viscera may produce vomiting as the main symptom. Gastric obstruction results from ulcer disease and malignancy, whereas small bowel and colonic obstructions occur as a consequence of adhesions, benign or malignant tumors, volvulus, intussusception, or inflammatory diseases such as Crohn's disease. The superior mesenteric artery syndrome, occurring after weight loss or prolonged bed rest, results when the duodenum is compressed by the overlying superior mesenteric artery. Abdominal irradiation evokes emesis by impairing intestinal contractile function and by inducing strictures. Biliary colic causes nausea likely by action on visceral afferent nerves. Vomiting with pancreatitis, cholecystitis, and appendicitis is due to localized visceral irritation and induction of ileus. Enteric infections with viruses or bacteria such as *Staphylococcus aureus* and *Bacillus cereus* are among the most common causes of acute vomiting, especially in children. Opportunistic infections such as cytomegalovirus or herpes simplex induce emesis in immunocompromised individuals.

Disorders of gastrointestinal motor function also commonly cause nausea and vomiting. Gastroparesis is defined as a delay in emptying of food from the stomach and occurs after vagotomy for peptic ulcer, with pancreatic adenocarcinoma, or in systemic diseases such as diabetes, scleroderma, and amyloidosis. Idiopathic gastroparesis develops in the absence of systemic illness and may follow a viral prodrome suggesting an infectious etiology. Intestinal pseudoobstruction is characterized by disruption of intestinal and colonic motor activity and leads to intestinal retention of food residue and secretions, bacterial overgrowth, nutrient malabsorption, and development of nausea, vomiting, bloating, pain, and alteration of bowel pattern. Intestinal pseudoobstruction may be idiopathic, may be inherited as a familial visceral myopathy or neuropathy, or may result from systemic disease or be a paraneoplastic consequence of malignancy (especially small cell lung carcinoma).

Extraperitoneal Disorders Myocardial infarction and congestive heart failure are cardiac causes of nausea and vomiting. Nausea and vomiting occur after 25% of surgical operations, both within and outside the peritoneum. Postoperative emesis is

more common after laparotomy and orthopedic surgery than after laparoscopy and is more prevalent in women. Increased intracranial pressure from tumors, bleeding, abscess, or obstruction to cerebrospinal fluid outflow produces prominent vomiting with or without concurrent nausea. Motion sickness, labyrinthitis, and Meniere's disease evoke symptoms via labyrinthine pathways. Cyclic vomiting syndrome is a rare disorder of unknown etiology that produces episodes of intractable nausea and vomiting, usually in children. The syndrome shows a strong association with migraine headaches, suggesting that some cases may be migraine variants. Patients with psychiatric illnesses, including anorexia nervosa, bulimia nervosa, and depression, may report significant nausea. Psychogenic vomiting occurs most commonly in women with other emotional problems.

Medications and Metabolic Disorders Drugs are frequent causes of vomiting and may act on the stomach (analgesics, erythromycin) or area postrema (digoxin, opiates, anti-Parkinsonian drugs). Agents that cause emesis include antibiotics, antiarrhythmics, antihypertensives, oral hypoglycemics, and contraceptives. Cancer chemotherapy causes vomiting that is acute (within hours of administration), delayed (after 1 or more days), or anticipatory. Acute emesis resulting from highly emetogenic agents such as cisplatin is mediated by 5-HT_3pathways, whereas delayed emesis is independent of 5-HT_3. Anticipatory nausea often responds better to anxiolytic therapy than to antiemetics.

Metabolic disorders are in the differential diagnosis in certain settings. Pregnancy is the most prevalent endocrinologic cause of nausea, occurring in 70% of women in the first trimester. Hyperemesis gravidarum is a severe form of nausea of pregnancy that can produce significant fluid loss and electrolyte disturbances. Uremia, ketoacidosis, adrenal insufficiency, as well as parathyroid and thyroid disease are other metabolic causes of emesis.

Circulating toxins evoke symptoms through effects on the area postrema. Endogenous toxins are generated in fulminant liver failure, whereas exogenous enterotoxins may be produced by enteric bacterial infection. Ethanol intoxication is a common toxic cause of nausea and vomiting.

Approach to the Patient

History and Physical Examination The history helps determine the etiology of unexplained nausea and vomiting. Drugs and toxins often cause acute symptoms, while established illnesses evoke chronic complaints. Vomiting within 1 h of eating characterizes pyloric obstruction, whereas emesis in the late postprandial period is reported with intestinal obstruction. Gastroparesis can produce nausea within minutes of food consumption but, in severe cases, leads to vomiting of meal residue ingested hours or days previously. Blood in the vomitus raises suspicion of an ulcer or malignancy; feculent emesis is noted with distal intestinal or colonic obstruction. Bilious vomiting excludes gastric obstruction, whereas emesis of undigested food is consistent with a pharyngoesophageal process such as Zenker's diverticulum or achalasia. Relief of abdominal pain by emesis characterizes small bowel obstruction, but vomiting has no effect on pancreatitis or cholecystitis pain. Pronounced weight loss raises concern about malignancy or obstruction. Fevers suggest inflammation, while an intracranial source is

considered if there are headaches or visual field changes. Vertigo or tinnitus indicate labyrinthine disease.

The physical examination complements the history. Abdominal auscultation may reveal absent bowel sounds with ileus. High-pitched rushes suggest bowel obstruction, while a succession splash on abrupt lateral movement of the patient is found with gastroparesis or pyloric obstruction. Tenderness or involuntary guarding raises suspicion of inflammation, whereas fecal blood suggests mucosal injury from ulcer, ischemia, or tumor. Neurologic etiologies present with papilledema, visual field loss, or focal neural abnormalities. Neoplasm is suggested by palpable masses or adenopathy.

The history and examination can characterize complications of emesis. Reports of lightheadedness with demonstration of orthostatic hypotension and reduced skin turgor indicate intravascular fluid loss. Hematemesis, especially with repeated vomiting, suggests a Mallory-Weiss tear of the gastroesophageal junction, while pulmonary abnormalities raise concern for aspiration of vomitus.

Diagnostic Testing With intractable symptoms or an elusive diagnosis, selected diagnostic tests can direct clinical management. Electrolyte replenishment is indicated if hypokalemia or metabolic alkalosis is found. Detection of iron-deficiency anemia mandates a search for mucosal injury. Pancreaticobiliary disease is suggested by abnormal pancreatic enzymes or liver biochemistries, whereas endocrinologic or rheumatologic etiologies are diagnosed by specific hormone or serologic testing. If lumenal obstruction is considered, supine and upright abdominal radiographs may show intestinal air-fluid levels with reduced colonic air. Ileus is characterized by diffusely dilated air-filled bowel loops.

If initial testing is unrevealing, additional anatomic studies may be indicated. Upper endoscopy detects ulcer disease or gastroesophageal malignancy, and small bowel barium radiography diagnoses partial small bowel obstruction. Colonoscopy or barium enema can detect colonic obstruction. Ultrasound or computed tomography of the abdomen defines intraperitoneal inflammatory processes, while computed tomography or magnetic resonance imaging of the head can delineate intracranial sources of nausea and vomiting.

Gastrointestinal motility testing may detect a functional gastrointestinal disorder responsible for symptoms when investigation of anatomic abnormalities is negative. Gastroparesis is most commonly diagnosed with gastric scintigraphy, by which emptying of a radiolabeled meal is measured. A noninvasive means of quantitating gastric slow wave activity with cutaneous electrodes placed over the stomach, electrogastrography, has been proposed as an alternate means of diagnosing abnormal gastric emptying. With intestinal pseudoobstruction, small bowel barium radiography often suggests the diagnosis. Manometry of the small intestine may provide confirmation of the diagnosis as well as complementary information by characterizing the motor abnormality as neuropathic or myopathic based on contractile patterns. Such investigation can obviate the need for open biopsy of the intestine to evaluate for smooth muscle or neuronal degeneration.

TREATMENT

General Principles Therapy of vomiting is tailored to the underlying disease, with the medical or surgical correction of abnormalities if possible. Hospitalization is considered for severe dehydration, especially if oral fluid replenishment cannot be sustained. Once oral intake is tolerated, nutrients are restarted as liquids that are low in fat, as lipids delay gastric emptying. Foods high in indigestible residues are avoided because these also prolong gastric retention.

Antiemetic Medications Drugs that act on the central nervous system serve as antiemetic agents (Table 41-2). Antihistamines such as meclizine and dimenhydrinate and anticholinergic drugs such as scopolamine act on labyrinthine-activated pathways and are useful in the treatment of motion sickness and inner ear disorders. Phenothiazine and butyrophenone dopamine D_2antagonists are used to treat emesis evoked by area postrema stimuli and are effective for many medication, toxic, and metabolic etiologies. Dopamine antagonists freely cross the blood-brain barrier and may cause anxiety, dystonic reactions, hyperprolactinemic effects (galactorrhea and sexual dysfunction), and irreversible tardive dyskinesia.

Other drug classes have antiemetic properties. Serotonin 5-HT_3antagonists such as ondansetron and granisetron are useful in the treatment of postoperative vomiting and after radiation therapy but are mainly used to prevent cancer chemotherapy-induced emesis. The usefulness of 5-HT_3antagonists to control other causes of refractory emesis is less well established. Antidepressant drugs are established therapeutic options for patients with functional bowel disorders such as irritable bowel syndrome. Low-dose tricyclic antidepressants provide moderate symptomatic benefit in patients with unexplained nausea of a functional nature.

Gastrointestinal Motor Stimulants Drugs that stimulate gastric emptying are indicated for gastroparesis (Table 41-2). Cisapride, a serotonin 5-HT_4agonist that stimulates cholinergic nerves in the stomach, has become the preferred drug for outpatient management of gastroparesis. The drug is well tolerated but exhibits very rare drug interactions with selected antibiotics, antifungals, and other agents that predispose to fatal cardiac arrhythmias. Metoclopramide, a combined 5-HT_4agonist and D_2antagonist, is efficacious in the treatment of gastroparesis, but anti-dopaminergic side effects limit its use in 20% of patients. Erythromycin, a macrolide antibiotic, potently increases gastroduodenal motility by action on receptors for motilin, an endogenous stimulant of fasting motor activity. Erythromycin may be most useful when given intravenously to inpatients with refractory gastroparesis; however, oral forms of the drug also have some effect. Domperidone, a D_2antagonist not available in the United States, has prokinetic and antiemetic effects but does not cross into most other brain regions; thus, anxiety and dystonic reactions are rare. The main side effects of domperidone are induction of hyperprolactinemia through effects on pituitary regions served by a porous blood-brain barrier.

Patients with refractory upper gut motility disorders pose significant therapeutic challenges. Liquid suspensions of prokinetic drugs may be beneficial inasmuch as liquids empty from the stomach more rapidly than pills. Metoclopramide can be administered subcutaneously in patients who do not respond to oral drugs. Intestinal pseudoobstruction may respond to the somatostatin analogue octreotide, which induces

propagative small intestinal motor complexes. Placement of a feeding jejunostomy reduces hospitalizations and improves overall health in some patients with gastroparesis who do not respond to drug therapy. Surgical options are limited for refractory cases, but postvagotomy gastroparesis may improve with near-total resection of the stomach. Electrical pacing of the stomach may also be useful.

Selected Clinical Settings Cancer chemotherapeutic agents such as cisplatin are intensely emetogenic. Given prophylactically, 5-HT_3antagonists prevent chemotherapy-induced acute vomiting in most cases (Table 41-2). Optimal antiemetic effects often are obtained with a 5-HT_3antagonist in combination with a glucocorticoid. In high doses, metoclopramide is effective in controlling chemotherapy-evoked emesis, whereas benzodiazepines such as lorazepam are most useful in reducing anticipatory nausea and vomiting. In contrast, delayed emesis 1 to 5 days after chemotherapy is more refractory to treatment. Agents that act as neurokinin NK_1antagonists in the brain stem may be potent antiemetic and antinausea drugs during both the acute and the delayed periods after chemotherapy. Cannabinoids such as tetrahydrocannabinol have been advocated for cancer-associated emesis, but these drugs produce significant side effects and are no more effective than antidopaminergic agents.

The clinician should exercise caution in the management of the patient with nausea of pregnancy. Studies of the teratogenic effects of available antiemetic agents have provided conflicting results. Few controlled trials have been performed in the nausea of pregnancy, although antihistamines such as meclizine and antidopaminergics such as prochlorperazine are more efficacious than placebo. As a consequence, alternative therapies such as pyridoxine or ginger have been recommended.

INDIGESTION

MECHANISMS

Most patients with indigestion have symptoms of a functional nature that result from gastroesophageal acid reflux or from gastric abnormalities including dysfunctional motor activity and afferent hypersensitivity; these symptoms comprise the syndrome functional dyspepsia. Some cases are a consequence of a more serious organic illness.

Gastroesophageal Acid Reflux Acid reflux results from selected physiologic defects. In scleroderma and pregnancy, lower esophageal sphincter (LES) tone is low, but most patients with acid reflux have normal LES pressures. Many individuals show frequent transient LES relaxations during which acid bathes the esophagus. The role of hiatal hernias is controversial -- although most reflux patients exhibit hiatal hernias, most individuals with hiatal hernias do not have excess heartburn.

Gastric Motor Dysfunction Disturbed gastric motility is purported to cause acid reflux in some patients with indigestion. Delayed gastric emptying also is found in 25 to 50% of individuals with functional dyspepsia. The relation of these defects to symptom induction is uncertain as many studies show poor correlation between symptom severity and the degree of motor dysfunction. Abnormal gastric fundic relaxation may cause dyspeptic symptoms such as bloating, fullness, nausea, and early satiety.

Visceral Afferent Hypersensitivity Disturbed gastric sensory function also may cause functional dyspepsia. Visceral afferent hypersensitivity was first demonstrated in patients with irritable bowel syndrome who had heightened perception of rectal balloon inflation without changes in rectal compliance. Patients with dyspepsia may experience discomfort with fundic distention to lower pressures than healthy control subjects.

Other Factors *Helicobacter pylori* has a clear etiologic role in peptic ulcer disease, but ulcers cause only a minority of cases of dyspepsia. The importance of *H. pylori* in the genesis of functional dyspepsia is controversial, but most investigators believe it is of minor importance. Analgesics cause dyspepsia; nitrates, calcium channel blockers, theophylline, and progesterone promote acid reflux. Other exogenous factors that induce acid reflux include ethanol, tobacco, and caffeine viaLESrelaxation. Finally, functional dyspepsia is exacerbated by stress, suggesting a pathogenic role for psychological factors.

DIFFERENTIAL DIAGNOSIS

Functional Causes Gastroesophageal reflux disease (GERD) is prevalent in Western society. Heartburn is reported once monthly by 40% of Americans and daily by 7%. Functional dyspepsia, defined as ³3 months of dyspepsia without an organic cause, also is common. Nearly 25% of the populace has abdominal discomfort at least six times yearly, consistent with functional dyspepsia, but only 10 to 20% consult physicians. The clinician must distinguish these illnesses, which have a benign course, from conditions that have deleterious consequences.

Ulcer Disease In most cases ofGERD, the esophagus is not damaged. However, 5% of patients develop esophageal ulcers, and some form esophageal strictures. Functional dyspepsia is the cause of symptoms in 60% of individuals with dyspepsia. However, 15 to 25% of cases stem from ulcers of the stomach or duodenum. The most common causes of ulcer disease are gastric infection with *H. pylori* and use of nonsteroidal anti-inflammatory drugs. Other rare causes of gastroduodenal ulcer include Crohn's disease and Zollinger-Ellison syndrome, a condition resulting from gastrin overproduction by an endocrine tumor (Chap. 285).

Malignancy Patients with dyspepsia often seek care because of fear of cancer. However,<2% of cases result from gastroesophageal malignancy. Esophageal squamous cell carcinoma occurs most often in those patients with histories of tobacco or ethanol intake. Other risk factors include prior caustic ingestion, achalasia, and the hereditary disorder tylosis. Esophageal adenocarcinoma usually complicates long-standing acid reflux. Eight to 20% of patients withGERDexhibit glandular mucosal metaplasia of the squamous epithelium in the lower esophagus, termed *Barrett's metaplasia*. This condition predisposes to esophageal adenocarcinoma. Gastric malignancies include adenocarcinoma, which is more prevalent in certain Asian societies, and lymphoma (seeChap. 90).

Other Causes Alkaline reflux esophagitis producesGERD-like symptoms in patients who have had surgery for peptic ulcer disease. Opportunistic fungal or viral esophageal infections may produce heartburn or chest discomfort but more often cause painful swallowing. Although biliary colic is in the differential diagnosis of dyspepsia, most

patients with true biliary colic report discrete episodes of right upper quadrant or epigastric pain rather than chronic burning discomfort, nausea, and bloating. Lactose intolerance resulting from intestinal lactase deficiency produces gas, bloating, discomfort, and diarrhea. Lactase deficiency occurs in 15% of Caucasians of northern European descent but is more common in African Americans and Asians. Pancreatic disease (chronic pancreatitis and malignancy), hepatocellular carcinoma, celiac sprue, Menetrier's disease, infiltrative diseases (sarcoidosis and eosinophilic gastroenteritis), mesenteric ischemia, thyroid and parathyroid disease, and abdominal wall strain cause dyspepsia. Extraperitoneal etiologies of indigestion include congestive heart failure and tuberculosis.

Approach to the Patient

History and Physical Examination GERDclassically produces heartburn, a substernal warmth beginning in the epigastrium that moves toward the neck. Heartburn often is exacerbated by meals and may awaken the patient. Associated symptoms include regurgitation of acid and water brash, the reflex release of salty salivary secretions into the mouth. Atypical symptoms include pharyngitis, asthma, cough, bronchitis, hoarseness, and chest pain that mimics angina. Some patients with acid reflux on esophageal pH testing do not report heartburn and instead note abdominal pain or other symptoms.

Individuals with ulcer-like dyspepsia have epigastric gnawing or burning that is relieved by meals or acid suppression. Dysmotility-like dyspepsia is a fullness or pain that is aggravated by eating and associated with nausea, bloating, eructation, and early satiety. There is overlap among the different dyspepsia subclasses and with other functional disorders such as irritable bowel syndrome.

The physical examination of individuals with functional causes of indigestion is usually normal. In atypicalGERD, pharyngeal erythema and wheezing over the lung fields may be present. Poor dentition may occur with prolonged acid regurgitation. Patients with functional dyspepsia may have epigastric tenderness or abdominal distension.

Discrimination between functional and organic causes of indigestion mandates exclusion of selected historical and examination features. Odynophagia suggests esophageal infection, while dysphagia promotes concern about a benign or malignant esophageal blockage. Other features that raise alarm include unexplained weight loss, recurrent vomiting with evidence of dehydration, occult or gross gastrointestinal bleeding, and a palpable mass or adenopathy.

Diagnostic Testing Because indigestion is prevalent in the community and because most cases result from functional illness, a general principle of diagnostic testing is to perform only limited and directed testing of selected individuals.

Once alarm factors are excluded, patients with typicalGERD do not need further evaluation and are treated empirically. Upper endoscopy is indicated to exclude mucosal injury in patients with atypical symptoms, symptoms unresponsive to acid-suppressing drugs, or alarm factors. In patients with>5 years of heartburn, endoscopy is performed to screen for Barrett's metaplasia. Upper gastrointestinal

barium radiography has a slightly higher sensitivity for detecting strictures and rings than endoscopy; however, benign esophageal obstructions may be dilated with an endoscopic approach. Ambulatory esophageal pH testing is considered for atypical symptoms such as unexplained chest pain and for the symptoms that are unresponsive to appropriate medications. Esophageal manometry is most commonly ordered when surgical treatment of GERD is considered. A lowLESpressure may predict failure with drug therapy and identify patients who may require surgery. The demonstration of disordered esophageal body peristalsis may affect the decision to operate or modify the type of operation chosen. Manometry with provocative testing may clarify the diagnosis in patients with atypical symptoms. Blinded perfusion of saline and then acid into the esophagus, known as the Bernstein test, can delineate whether unexplained chest discomfort results from acid reflux.

The approach to unexplained dyspepsia is dependent on the patient's age, symptom profile, and findings on examination. In individuals <45 years of age without alarm factors, blood serology for *H. pylori* is obtained to exclude the organism as a cause of ulcer disease. Upper endoscopy in this patient subset is reserved for those who fail to respond to treatment of *H. pylori*-positive or -negative dyspepsia. Upper endoscopy is performed as the initial diagnostic test in any individual with alarm factors or in patients>45 years of age because of the elevated risk of gastroesophageal malignancy with advancing age.

Further testing is indicated only if other factors are present. If there is blood loss, a blood count is obtained to exclude anemia. Thyroid chemistries or calcium levels screen for metabolic disease. With suspected pancreaticobiliary causes, blood is obtained for amylase, lipase, and liver chemistry determination. If biochemical abnormalities are found, abdominal ultrasound or computed tomography may give important information. Patients with dysmotility-like dyspepsia may selectively exhibit delayed gastric emptying; thus, gastric scintigraphy can be considered when drug treatment fails. Hydrogen breath testing after lactose ingestion may be performed for suspected lactase deficiency.

TREATMENT

General Principles In mild dyspepsia, reassurance that a careful evaluation revealed no serious organic disease may be the only intervention required. Drugs that cause acid reflux or dyspepsia should be stopped if possible. Patients withGERDshould limit ethanol, caffeine, chocolate, and tobacco use because of their effects on theLES. Other measures with efficacy in GERD include ingestion of a low-fat diet, avoidance of snacks before bedtime, and elevation of the head of the bed.

Specific therapy for organic diseases should be offered when possible. In disorders such as biliary colic, surgery is appropriate; whereas lactase deficiency and celiac sprue respond to special diets. Some illnesses such as peptic ulcer disease require specific medical regimens to effect cure. However, as most patients present with functional causes of indigestion, medications that reduce gastric acid, stimulate upper gut motility, or blunt gastric sensitivity are indicated.

Acid Suppressing or Neutralizing Medications Drugs that reduce or neutralize

gastric acid are the most prescribed agents forGERD. Histamine H_2receptor antagonists such as cimetidine, ranitidine, famotidine, and nizatidine are useful in the treatment of mild to moderate GERD. For uncomplicated heartburn, H_2receptor antagonists are given for 4 weeks before endoscopy is considered. For severe symptoms or for many cases of erosive or ulcerative esophagitis, proton pump inhibitors such as omeprazole and lansoprazole are needed. These drugs, which inhibit gastric H^+, K^+-ATPase, are more potent than H_2receptor antagonists. Liquid antacids are useful for short-term control of mild GERD but are less effective for severe disease unless given at high doses that produce side effects (diarrhea with magnesium-containing agents and constipation with aluminum-containing agents). Sucralfate is a salt of aluminum hydroxide and sucrose octasulfate and buffers acid and binds pepsin and bile salts. Its efficacy in GERD and functional dyspepsia is unproven.

Acid suppressing drugs are advocated for first-line therapy of *H. pylori* negative dyspepsia, especially with ulcer-like symptoms. Ranitidine is of benefit in the treatment of functional dyspepsia versus placebo. In young patients without alarm symptoms, a 4-week trial of an H_2receptor antagonist or proton pump inhibitor is given. Endoscopy is performed only if symptoms do not improve.

***Helicobacter pylori* Eradication** Regimens to eradicate *H. pylori* are recommended for young patients with dyspepsia without alarm symptoms in whom the bacterium has been detected by serology. Several drug combinations show efficacy, but most include 10 to 14 days of a proton pump inhibitor or bismuth subsalicylate in concert with two antibiotics. If symptoms resolve, no further intervention is required. Most patients who respond to this "treatment-first" approach have underlying ulcer disease. The usefulness of *Helicobacter* eradication in patients with functional dyspepsia is unproven, but evidence suggests that<15% of cases relate to *H. pylori*. No evidence demonstrates that *H. pylori* eradication is useful in the treatment ofGERD.

Gastrointestinal Motor Stimulants Cisapride is superior to placebo in treatingGERD and can be prescribed as sole therapy or as an adjunct to an acid-suppressing drug. Other motor stimulants such as metoclopramide, erythromycin, and domperidone are of limited use in the treatment of GERD.

Prokinetic agents are frequently used for treatment of functional dyspepsia. Cisapride and domperidone relieve symptoms more effectively than placebo. In general, these drugs are more potent than acid-reducing agents in the treatment of functional dyspepsia and may be given instead of acid suppressants as the initial empirical treatment of young patients with dyspepsia without alarm symptoms who are not infected with *H. pylori*. Patients with dysmotility-like dyspepsia may respond preferentially to motor-stimulating drugs.

Other Options In patients withGERD who do not respond to drug therapy, antireflux surgery may be offered. Operations include the Nissen fundoplication, in which the proximal stomach is wrapped completely around theLES to increase LES pressure, and the Belsey procedure, in which the wrap encircles 270° of the circumference of the LES. The latter is selected if esophageal peristalsis is suboptimal when a 360° wrap might cause dysphagia. Fundoplications can be performed laparoscopically, thereby reducing the morbidity and the postoperative recuperation period.

Some patients with functional dyspepsia do not respond to acid suppressants or prokinetic drugs but may respond to low-dose tricyclic antidepressant therapy. The mechanism of action of these agents in functional dyspepsia is unknown but may involve blunting of visceral pain processing in the brain. Gas and bloating may be the most troubling symptoms in some patients with indigestion and can be difficult to treat. Successes with dietary exclusion of gas-producing foods such as legumes and use of the surface-active compound simethicone or the gas-absorptive agent activated charcoal have been reported. Psychological treatments have been proposed for functional dyspepsia; however, convincing data on their efficacy are lacking.

(Bibliography omitted in Palm version)

Back to Table of Contents

42. DIARRHEA AND CONSTIPATION - *David A. Ahlquist, Michael Camilleri*

Diarrhea and constipation are exceedingly common and together exact an enormous toll in terms of morbidity, loss of work productivity, and consumption of medical resources. Worldwide, more than one billion people suffer one or more episodes of acute diarrhea each year. Among the 100 million persons affected annually by acute diarrhea in the United States, nearly half must restrict activities, 10% consult physicians, 250,000 require hospitalization, and roughly 3000 die (primarily the elderly). The economic burden to society is estimated at more than $20 billion. Because of poor sanitation and more limited access to health care, acute infectious diarrhea remains one of the most common causes of mortality in developing countries, particularly among children, accounting for 5 to 8 million deaths per year. Population statistics on chronic diarrhea and constipation are more uncertain, perhaps due to variable definitions and reporting, but the frequency of these conditions is also high. Based on United States population surveys, prevalence rates for chronic diarrhea range from 2 to 7% and for chronic constipation from 3 to 17%. Diarrhea and constipation are among the most common patient complaints faced by internists and primary care physicians, and they account for nearly 50% of referrals to gastroenterologists.

Although diarrhea and constipation may present as mere nuisance symptoms at one extreme, they can be severe or life-threatening at the other. Even mild symptoms may signal a serious underlying gastrointestinal lesion, like colorectal cancer, or systemic disorder, like thyroid disease. Given the heterogeneous causes and potential severity of these common complaints, it is imperative for clinicians to appreciate the pathophysiology, etiologic classification, diagnostic strategies, and therapeutic principles of diarrhea and constipation so that rational and cost-effective care can be delivered.

NORMAL PHYSIOLOGY

The human small intestine and colon perform important functions including the secretion and absorption of water and electrolytes, the storage and subsequent transport of intraluminal contents aborally, and the salvage of some nutrients after bacterial metabolism of carbohydrate that are not absorbed in the small intestine. The main motor functions are summarized in Table 42-1. Alterations in fluid and electrolyte handling contribute significantly to diarrhea. Alterations in motor and sensory functions of the human colon result in highly prevalent syndromes such as irritable bowel syndrome, chronic diarrhea, and chronic constipation.

NEURAL CONTROL

The small intestine and colon have intrinsic and extrinsic innervation. The *intrinsic innervation* also called the enteric nervous system, comprises myenteric, submucosal, and mucosal neuronal layers. The function of these layers is modulated by interneurons through the actions of neurotransmitter amines or peptides, including acetylcholine, opioids, norepinephrine, serotonin, ATP, and nitric oxide. The myenteric plexus regulates smooth muscle function, and the submucosal plexus affects secretion and absorption.

The *extrinsic innervations* of the small intestine and colon are part of the autonomic