Current Concepts Review

Decision Analysis*

BY JOSEPH BERNSTEIN, M.S., M.D., PHILADELPHIA, PENNSYLVANIA

Of all practitioners who treat back pain, orthopaedic surgeons are the most expensive*. This distinction has been earned not because of high professional fees but rather because orthopaedic surgeons frequently order expensive, and perhaps unnecessary, diagnostic tests. Unnecessary testing may harm patients and upset payers. It also may provoke insurance companies or governmental agencies to deprive doctors of their autonomy by controlling their ability to order expensive diagnostic tests. Thus, unnecessary testing may harm physicians as well.

The correct use of diagnostic tests is an important component of effective medical practice. When tests are used at the wrong time or are interpreted inappropriately, not only are resources wasted on the test itself but also an incorrect course of clinical action might be taken. The ordering of tests only when they are appropriate is a skill that may be attributed to the art of medicine. Nevertheless, basic principles of probability theory, which are given the collective label of decision analysis, can be used to ascertain when a test may be clinically helpful.

Decision analysis can be used to limit excessive testing without limiting clinical power. A decision-analysis approach to back pain, for example, might involve examining the likelihood that magnetic resonance imaging could alter the clinical treatment. Such an assessment would take into account the diagnostic power of the test and the utility values that the patient assigns to the possible outcomes of treatment. Even if the test were to demonstrate an abnormality, there still might be an insufficient likelihood that the identified lesion was responsible for the symptoms. It follows, accordingly, that some patients who have positive findings on magnetic resonance imaging will be managed no differently than those who have negative findings. If the same treatment would be chosen regardless of the result of the test, then the test should be omitted. In decision-analysis terms, the informational contribution of magnetic resonance imaging will be insufficient to cross the treatment threshold. Such explicit consideration of the test’s power may help to dissuade both the physician and the patient from superfluous testing.

Although the decision-analysis approach has been used by physicians to help choose treatment regimens and to identify the best diagnostic test to order, it is not yet prevalent in orthopaedics. The goal of this article, therefore, is to introduce some of the fundamentals of decision analysis. Only by explicitly measuring the amount of information that a diagnostic test provides and by quantifying the level of diagnostic certainty required for action can it be asserted that a given test is indicated. Such a process makes better outcomes more probable, without higher costs.

Tests Are Imperfect

A central tenet of medical decision analysis is that certainty in clinical medicine is often unattainable. Rather, the art of diagnosis is to reduce uncertainty enough so that optimum decisions can be made. For example, if a patient with pain caused by a total hip implant has a positive indium scan, the probability that an infection is present increases but rarely reaches 100 per cent. Therefore, the question for the surgeon is whether the uncertainty that an infection is present has been reduced enough to warrant managing the patient as if an infection were present. The need to think in terms of likelihood arises because the indium scan, like most tests, can be incorrect; that is, its true-positive and true-negative rates are not 100 per cent. These rates define the test’s power. This power is expressed as the informational contribution of the test— that is, the amount of change that the test can effect in the physician’s assessment of the patient’s status.

Sensitivity and Specificity

The sensitivity and specificity of a test define its informational contribution. These two values, termed conditional probabilities, represent the test’s ability to detect or exclude the presence of disease. The sensitivity of a test is the conditional probability that the test will be positive when the patient has the disease. It is measured in a study population by calculating the ratio of the number of subjects who have the disease and for whom the test is positive (the number of true-positive results) to the total number of people who have the disease. Likewise, specificity is the conditional
The negative predictive value is defined similarly. Urban attributes; they depend on sensitivity and specificity but differences between these terms will be discussed later.

Bayes Theorem

The conditional probability that, for example, a magnetic resonance imaging study for the detection of a tear of the glenoid labrum is 88 per cent because the test was positive for 88 per cent of subjects known to have a labral tear. Clinically, however, we need the opposite information. After all, we do not wonder if a patient known to have a labral tear will have a positive test; rather, we wonder if an individual who has a positive test really has a tear. We need to know the conditional probability that a disease is present or absent, given the test result. This conditional probability must be calculated for each patient and cannot be inferred from the sensitivity and the specificity alone. The conditional probability that a disease is present or absent must be calculated with two equations, known collectively as the Bayes theorem.

Bayes Theorem

The conditional probability that, for example, a magnetic resonance imaging study that detects a labral tear has identified an actual tear is termed the post-test probability of the test; the differences between these terms will be discussed later. The negative predictive value is defined similarly. Unlike sensitivity and specificity, which are inherent properties of the test, post-test probabilities are derived attributes; they depend on sensitivity and specificity but also on the prevalence (or the pre-test probability) of disease. The equations of the Bayes theorem are:

Post-test probability of a positive test =

\[
\frac{\text{pre-test probability} \times \text{sensitivity}}{(\text{pre-test probability} \times \text{sensitivity}) + ([1 - \text{pre-test probability}] \times [1 - \text{specificity}])}
\]

Post-test probability of a negative test =

\[
\frac{\text{pre-test probability} \times (1 - \text{specificity})}{(\text{pre-test probability} \times [1 - \text{sensitivity}]) + ([1 - \text{pre-test probability}] \times \text{specificity})}
\]

These equations show that, although sensitivity and specificity are constant for a given test, the post-test probability fluctuates with varying degrees of pre-test probability. Accordingly, it is axiomatic that to know the likelihood of a disease after a test the likelihood before the test must first be known.

Post-test probability is a continuous variable, and its value is based on sensitivity, specificity, and prevalence. (Prevalence is the ratio of individuals in a given population who have a disease to the total number of individuals in that population. The same ratio thus represents the probability that any random individual has the disease.) Predictive values of a test are post-test probabilities found in a given group with a known prevalence. These predictive values should not be extrapolated to groups or individuals who have a different (or an unknown) prevalence. For example, in order to determine if a patient has an infected hip, it should not be stated that the positive predictive value of a positive culture of fluid aspirated from the hip is always 70 per cent, as that value belongs only to a particular setting with its given prevalence. Assuming a sensitivity of 50
per cent and a specificity of 95 per cent for aspiration of the hip, the positive predictive value after an aspiration is 87 per cent if the pre-test probability is 40 per cent and 71 per cent if the pre-test probability is 20 per cent. Thus, as Stratford noted, "stating the predictive values without indicating the [pre-test probability] does not convey useful information," as extremes in values for prevalence may yield deceptively high or low predictive values.

2 x 2 Tables

The values calculated with use of the Bayes theorem can be displayed in a 2 x 2 table (Figs. 1, 2, and 3). Such a table is similar to a spreadsheet: mathematical relationships between the cells allow calculation of the unknown values on the basis of the known ones. Inspection of the Bayesian equations shows that each term is represented by a box in the 2 x 2 table.

The Term Accuracy Is Deceptive

Another term that often is used to describe the power of a test is accuracy. Although this term has intuitive appeal, it often can be deceptive. Accuracy is the percentage of all test results that are correct: the sum of true positives and true negatives divided by the total of all the results. This value can be misleading because it combines the test’s ability to detect as well as its ability to exclude the presence of disease. Accuracy may be thought of as the weighted average of the sensitivity and the specificity: sensitivity denotes accuracy with regard to individuals who have the disease, and specificity designates accuracy with regard to those who do not have it. The use of accuracy to describe the test’s power may be dangerous when, for example, a test that is highly specific but not sensitive is evaluated with regard to a population that has a low prevalence of a disease. Since most of the population does not have the disease, accuracy will be high, more closely resembling specificity. This may induce the clinician to trust the test in all settings; however, given its low sensitivity, not all negative results are true negatives.

An exaggerated example is the use of the level of serum sodium as a test for osteogenic sarcoma. Any value of more than 280 millimoles per liter indicates a malignant bone tumor. Since a serum sodium level of 280 millimoles per liter is incompatible with life, no subject will have a positive test. A test that is always negative by definition has a sensitivity of zero, yet in the general population this admittedly silly test has an accuracy of more than 99 per cent. This is because far fewer than 1 per cent of all people have a bone tumor. In a typical population of 100 subjects, all 100 tests will be negative and all will be correct. This worthless test has nearly perfect accuracy, but its lack of value would be apparent if sensitivity and specificity were considered separately. Those values, accordingly, are better descriptors than is accuracy.

Limitations of Sensitivity and Specificity

Although sensitivity and specificity are the appropriate measures of a test’s power, that power may be overstated. For example, if a test were to be evaluated in a group that included many healthy control subjects, the reported specificity might be spuriously high because healthy individuals not only would not have the disease in question but typically also would not have other conditions that might be mistaken for it. This phenomenon is called spectrum bias. In addition, if the test being evaluated is a criterion for the physician’s ordering the definitive diagnostic test (for example, if only patients who have a positive magnetic resonance imaging study will be sent for arthroscopy), then reports of that test’s sensitivity may be distorted by so-called test-referral bias: it will never be known how many subjects who are deemed to have a negative result and who therefore are not tested with the definitive study in fact represent a false-negative result. Finally, the clinical specificity may differ from the calculated specificity. For example, it has been shown that abnormalities may be detected on imaging studies of the spines of asymptomatic individuals. A magnetic resonance image may show a bulge in the disc that is genuinely present, but that finding could be clinically irrelevant. Thus, this positive test, although anatomically accurate, is indeed a false-positive result. For these reasons, the surgeon must never abdicate clinical judgment, even with the results of apparently excellent tests in hand.

Threshold Model

A test is indicated if and only if its results could lead to a different course of clinical action. Pauker and Kassirer have developed a threshold model for test-
According to the threshold model, a test is indicated only if it provides enough information to lead to a change in the clinical plan. In this diagram, the pre-test probability is represented as a number line (0 to 100). The pre-test probability and the treatment thresholds are points on the line. The informational contribution of a positive test is represented by the arrow, and the tip of the arrow indicates the post-test probability of a positive test. In this diagram, the post-test probability of a positive test crosses the treatment threshold; therefore, the test is indicated. A test is not indicated if the threshold is not crossed. The latter possibility may be due to a pre-test probability that is too low, a threshold that is too high, or a test with an informational contribution that is too small.

Pre-Test Probability

The pre-test probability is the clinician's estimate of the chance that a disease is present. It may be expressed as a fraction ratio or in terms of odds, but more typically the clinician uses phrases such as "doubtful," "possible," and "very likely." These expressions can be roughly translated into mathematical terms. This process of formulating the pre-test probability is not limited to decision analysis; rather, it is a central part of the normal diagnostic process. Even if a physician does not know the term pre-test probability, he or she indeed is estimating it when assessing a patient and expressing a belief about the presence or absence of a given disease.

The pre-test probability is based on an integration of the medical history, the findings on physical examination, the results of all previous diagnostic studies, and the physician's knowledge and experience. By asserting a pre-test probability for a given patient, the physician is in a sense stating the belief that, in a large population with characteristics similar to those of the patient, a similar fraction of the population will have the disease. Thus, the terms pre-test probability and prevalence often are used interchangeably: the former, to describe individuals, and the latter, to denote groups.

In this era of high technology, physical examination may be an undervalued process. Nevertheless, studies have shown that physical examination can be both sensitive and specific. The diagnosis of a tear of the anterior or posterior cruciate ligament often can be made reliably on the basis of physical findings. The rotator cuff also may be evaluated well with physical examination. Additionally, even when the history and the physical examination do not point to a definitive diagnosis, they may allow discrimination between a serious condition and one that needs no immediate treatment.

Establishment of the pre-test probability may be the most important step in deciding whether a test is indicated. When the decision to order a special test hinges on the level of the pre-test probability, accurate interpretation of the history and the findings on physical examination is of paramount importance. It may be efficient, in such instances, for the patient to be first examined by a specialist, whose findings and confidence in asserting the likelihood of a disease being present may obviate unnecessary testing.
Figs. 5-A and 5-B: Decision trees representing the options for treatment and their respective values. \( T^+ = \) a tear is present, \( T^- = \) a tear is absent, \( A^+ = \) arthroscopy is chosen, and \( A^- = \) arthroscopy is not chosen.

**Clinical Prediction Rules**

Assessment of the probability that a disease is present represents, as Sox et al.\(^65\) noted, a "formidable cognitive task." Some investigators have attempted to derive and test rules of clinical prediction\(^73\) in order to help the clinician to formulate a more accurate pre-test probability. For example, Stiell et al.*\(^69\) found that the clinical probability of a fracture of the ankle can be estimated reliably, before radiographs are made, by assessing several clinical parameters, including the presence of osseous tenderness or deformity, the patient’s age, and the ability to bear weight. Similar rules have been derived and validated for fractures about the knee\(^67/8\). These rules are formulated by determining the frequency of a given finding in a population suspected of having a specific disease and then testing all patients for the presence of that disease. The population can be partitioned into groups according to the presence of various findings. In the studies of the ankle and the knee\(^64-68\), it was found that the prevalence of fracture in the group in which all of the predictor parameters were absent was zero; thus, such rules can be used to eliminate unnecessary radiographic examinations.

**Establishment of the Treatment Threshold**

Since diagnoses are rarely certain, the physician must decide whether to administer or to withhold treatment on the basis of probability. If the probability is high enough, treatment should be given; otherwise, it should be withheld. Although it is customary to regard 95 per cent confidence as a fair approximation of true certainty with regard to diagnosis, this is not always the case. For some diseases 95 per cent is far more certainty than is needed, whereas for others it is not enough. Pauker and Kassirer\(^53\) demonstrated that the degree of certainty required is a variable of the clinical setting. They termed this level the treatment threshold.

The treatment threshold is the level of diagnostic probability at which treating the patient becomes the correct strategy (that is, at which the patient is best managed as if the disease were present). The setting of the threshold makes no particular reference to the truth value of the statement "the patient has the disease." Indeed, it may be likely that the patient does not have the disease (a probability of less than 50 per cent), yet, for reasons to be discussed, the treatment threshold may have been crossed nonetheless.

The treatment threshold is, of course, higher when the cost of treatment is high and lower when the benefits of treatment are great\(^53\); safe, inexpensive, and effective remedies may be applied empirically. For example, a young patient who has tenderness of the ankle and normal radiographic findings may be managed with a cast for a possible non-displaced physeal fracture, even without formal proof of such an injury. In contrast, sophisticated tests and expert consultants may be employed to determine whether a lesion is malignant before operative intervention is performed. The negative implications of application of a cast, if it turns out that the ankle is merely sprained, are not very high, at least not when compared with the benefits derived from treatment; however, the adverse consequences of treating a benign lesion as if it were malignant are much higher. The level of diagnostic certainty needed to treat a possible physeal fracture of the ankle with a cast therefore is considerably lower than that re-
The precise establishment of the treatment threshold depends on the values that the patient (assisted by the physician) places on the possible outcomes of treatment. A decision tree can be used to represent a simplified synopsis of the treatment options and the possible outcomes (Fig. 5-A). As a simplified example, a patient who has a possible tear of the meniscus can be given the choice of either arthroscopy or non-operative treatment. Because diagnosis of this lesion is imperfect, the physician typically cannot state with certainty that a tear is present but can only estimate its probability. This estimate ranges from 0 per cent (certainty that there is no tear) to 100 per cent (certainty that a tear is present). Thus, it is possible that arthroscopy will be performed on a knee that does not have a meniscal tear or that non-operative treatment will be offered to a patient who actually has a lesion.

Assuming two treatment options and two diagnostic possibilities (the presence or absence of a tear), four outcomes are possible: a tear is present and arthroscopy is chosen (T+A+); no tear is present and arthroscopy is chosen (T−A+); a tear is present and arthroscopy is not chosen (T+A−); and no tear is present and arthroscopy is not chosen (T−A−). These are the so-called terminal nodes of the decision tree (Fig. 5-A).

The value of the arthroscopic option increases as the probability of a meniscal tear increases because operative treatment typically helps the patient only if a meniscal tear is present. Thus, at very low levels of pre-test probability, non-operative treatment is the best choice, whereas, at high levels, operative treatment is best. The point of pre-test probability at which these two options provide equal value defines the treatment threshold: above the treatment threshold, the potential benefits of arthroscopy exceed those of non-operative care, and vice versa. The potential benefits of each option depend in turn on the value that the patient assigns to each of the four possible outcomes. If these values can be estimated, the total utility value of the two treatment options can be calculated; this process is termed an expected utility analysis.

**Expected Utility Analysis**

The first step of an expected utility analysis is to assign utilities to the possible outcomes. Utility is the technical term for the value of an outcome. If a patient prefers outcome A to outcome B, outcome A is said to have greater utility. For the purpose of comparing outcomes of disparate situations, a more sophisticated measure, such as quality-adjusted life-years, can be used, but any consistent, linear scale can be employed. Utility encompasses the costs or benefits that are deemed important. Such costs could include financial expense, pain, time lost from work, and potential complications; benefits could include pain relief, the ability to work or engage in recreational activities, and time saved. It is important to note that utility values vary from patient to patient. For example, a twenty-eight-year-old professional baseball player who has avascular necrosis of the hip may elect to have a total hip replacement, whereas a similar patient who works as a laborer may choose an arthrodesis. This difference in preference stems from a difference in utility values. The laborer values pain relief and the avoidance of failure of the implant, whereas the athlete deems the utility of an arthrodesis to be very low because a hip that has been treated with such a procedure lacks the range of motion necessary for playing baseball.

Within an expected utility analysis, the value of any treatment option is the weighted average of the outcomes that may result from choosing that option; that is, the expected value is the sum of the utility values of each possible outcome, multiplied by the likelihood that the given outcome will be achieved (Fig. 5-B).

In the example of a suspected meniscal tear, the total utility of the option of arthroscopy is the utility assigned to the outcome T+A+, multiplied by the probability that a tear is present, added to the utility T−A+, multiplied by the probability that a tear is absent. This sum increases as the pre-test likelihood that a tear is present increases (Fig. 6-A). When the probability of a tear is close to zero, the value of the arthroscopy option is nearly equal to the value assigned to T−A+; similarly, when the probability of a tear approaches 100 per cent, the value of arthroscopy is nearly equal to the value assigned to T+A+. Between these extremes, the utility is the weighted average of the two values. A similar analysis may be carried out for the option of non-operative treatment (Fig. 6-B). When the two lines from Figures 6-A and 6-B are combined (Fig. 6-C), it may be evident that there is a point of probability below which non-operative care will yield the best outcome and above which arthroscopy will provide maximum benefit. That point is the treatment threshold. At probability values above this threshold, the option of arthroscopic treatment offers more utility value than the option of non-operative care. Thus, above that point the surgeon should offer arthroscopy even though it is not necessarily certain that a tear will be found.

At times, no treatment threshold can be defined. If a patient were to deem the value of treatment of a definite tear to be less than the value of no treatment at all, the lines of the treatment-threshold graph would never cross and the treatment threshold would be infinite. (Such a situation may arise, for example, when the diagnosis of a meniscal tear is considered for a knee with end-stage osteoarthritis. If the patient is planning to have a knee replacement imminently, there is no benefit to discovering if the meniscus is torn.) When the treatment of a disease has a lower utility value than the omission of treatment, effort should not be wasted in trying to determine whether the disease is present; treatment simply should be withheld.
Figs. 6-A, 6-B, and 6-C: Graphs showing the utilities of arthroscopy and non-operative treatment and comparing the two. T+ = a tear is present, T− = a tear is absent, A+ = arthroscopy is chosen, and A− = arthroscopy is not chosen.

Fig. 6-A: The utility of arthroscopy, plotted as a function of the probability that a tear is present. (The y-axis, representing an arbitrary utility scale, is drawn at both x = 0 and x = 100 to indicate the y-intercept clearly.) The utility of arthroscopy is equal to the weighted average of the utility of the T−A+ state and the utility of the T+A+ state, or [p(tear) x u(T+A+)] + [(1 − p(tear)) x u(T−A+)]. u[state] = the utility of that particular state, and p(event) = the probability that an event will occur. When p(tear) = 0, the total utility is given by u(T−A+). Similarly, when p(tear) = 100 per cent, u(T+A+) defines the total. This weighted average may be drawn as a straight line connecting the points (x = 0, y = u(T−A+)) and (x = 100, y = u(T+A+)). The utility scale is any arbitrary consistent, linear system that indicates preference (see text). Utility values are determined uniquely for each patient.

Applications of the Treatment Threshold

Clinicians do not need to be expert in utility calculations in order to use the threshold model. Often, the differences in utility between the outcomes are so vast that subtle distinctions need not be made. Like pre-test probability, the treatment threshold may be broadly estimated and still contribute to the analysis of indications for testing. A formal derivation of the treatment threshold was provided by Pauker and Kassirer. They showed that the treatment threshold may be expressed as the ratio of the costs of treating normal subjects to the sum of those costs and the benefits of treating those who have the disease — that is, treatment threshold = costs/(costs + benefits).

When costs are much greater than benefits, the ratio is near one and nearly 100 per cent certainty is required. Similarly, when benefits greatly outweigh costs, the threshold approaches zero. I encourage surgeons at least to estimate the threshold for each patient, not only because a formal decision analysis requires it but also because it will serve as a reminder that costs and benefits vary from patient to patient and that different approaches may be indicated. For example, in one study, the choice of treatment for proximal deep-vein thrombosis was dictated by patients’ subjective willingness to accept risk rather than by the predicted outcome alone. (Patients would not accept even a minute risk of hemorrhage involving the central nervous system even if by doing so they were highly likely to avoid a post-phlebotic syndrome in the lower limb.) Similarly, the treatment of symptomatic prostatic hypertrophy may be based on the degree to which the patient is bothered by the symptoms, as the patient who is sufficiently bothered will trade some months of life expectancy to gain relief from the symptoms. Additional studies of patients’ preferences with regard to orthopaedic outcomes may help to provide useful guidelines for testing and treatment.

Testing Threshold

Pauker and Kassirer incorporated the informational contribution of diagnostic tests into their model of the treatment threshold and thus described two additional thresholds: the testing threshold and the no-test/treat threshold. With use of these two thresholds, the 0 to 100 per cent line of pre-test probability is divided into three ranges. In the first range (no-test/no-treat),
the pre-test probability is so low that even if the results of the test are positive the post-test probability will not cross the treatment threshold. At the other extreme lies the no-test/treat group. For patients who have a pre-test probability in this range, the likelihood of disease is so high that even a negative test should not dissuade a physician from providing treatment. Only between the two thresholds, in the test/treat-if-positive range, is treatment dictated by the test result.

This method can be applied to many clinical problems in orthopaedics. For example, it has been shown that, if a patient’s treatment threshold for a suspected meniscal tear is 60 per cent, a magnetic resonance imaging study is indicated only when the pre-test probability of the tear is between approximately 20 and 95 per cent. With use of the equations of the Bayes theorem and the known values for sensitivity and specificity, it is evident that, if the pre-test probability is less than 20 per cent, the post-test probability can never be more than 60 per cent. Similarly, the post-test probability cannot be less than 60 per cent if the pre-test probability is more than 95 per cent. Thus, if the physician believes that there is a 95 per cent chance that the patient’s symptoms are caused by a torn meniscus, then operative treatment, without a confirmatory magnetic resonance imaging study, is indicated. This is because, at such a high level of pre-test probability, the likelihood that a negative result is false is more than 40 per cent.

Sensitivity and Specificity of Tests with Continuous Results

Sensitivity and specificity can easily be defined for tests that yield only positive or negative results. These so-called binary results contrast with continuous results, or those that may assume any value within a given range. Initially, it may seem as if continuous results do not allow for the definition of sensitivity or specificity, but this is not so. Often in orthopaedics, tests with continuous results are used and sensitivity and specificity can be defined. For example, the leukocyte count in synovial fluid can be used to diagnose an infection in a patient in whom septic arthritis is suspected. The result of the count is not positive or negative for an infection; rather, an absolute leukocyte count is reported. The clinician, however, can apply an accepted cutoff threshold to translate the laboratory value into a binary result. In this fashion, continuous data (such as leukocyte counts, which could range from zero to hundreds of thousands) are transformed into positive or negative values.

The sensitivity of the leukocyte count for the diagnosis of an infection depends on the standard criterion that is used. If this criterion is set at a relatively low level, such as 10,000 per milliliter, the sensitivity of the test will improve; that is, fewer true cases of infection will remain undetected. However, such a definition will lead to incorrect labeling of many cases of inflammatory arthritis as infectious, thus lowering the specificity: specificity decreases as the false-positive rate increases. Often, there is an inherent trade-off between sensitivity and specificity, with gains in one being achieved at the expense of the other. In this example, it is because there is a wide natural variation in the absolute numbers of leukocytes in a milliliter of synovial fluid obtained from patients who have an infection; some patients with a true infection perhaps will have only 28,000 leukocytes per milliliter. Similarly, some patients who have pseudogout will have 100,000 cells per milliliter.

The correct cutoff threshold may be calculated with a decision-analysis technique called the receiver operating characteristic curve. Such a curve demonstrates that many published criteria for the definition of abnormal continuous results are merely suggestions. Rigid application of suggested standards is incorrect; a leukocyte count of 40,000 per milliliter in the synovial fluid of a patient who has cellulitis over the affected joint implies septic arthritis, even if 50,000 per milliliter is typically defined as the threshold for infection. In general, if the likelihood of disease is high before the test is performed, a lower threshold is appropriate; similarly, if the treatment is dangerous or costly, a higher threshold is needed. Because of the natural overlap between the findings in individuals who have a disease and the findings in those who do not, no line of demarcation can lead to 100 per cent accuracy.

Clinical Algorithms

The decision-analysis approach does not need to be applied uniquely in every setting in order for gains to be realized. Broad guidelines often can reduce excessive testing and can lead to substantial cost-savings. Nevertheless, the use of such clinical algorithms is not without pitfalls. Even when the algorithms are validated, the data that have been used to drive them may be incorrect.

Guidelines for the treatment of idiopathic scoliosis, for example, have suggested that progression of the curve to 45 degrees may indicate that an operation is needed. However, recent studies have shown that the measurement of progression of the curve may be imprecise. Two individuals interpreting a given radiograph will not necessarily agree about the size of the curve. Similarly, a given curve may seem larger or smaller depending on the patient’s position or on the technique that was used to make the radiograph. Finally, even if the radiography and the measurements are standardized, it is possible that a given curve may fluctuate in size over the course of a day. Accordingly, if a thoracic curve is measured as 26 degrees at the initial consultation and as 31 degrees six months later, it is not certain that true biological progression of the curve has occurred. The astute clinician therefore will use clinical algorithms only as sugges-
tions. Since these suggestions may not fully account for the clinical picture, rigid application is unwarranted.

**Limitations of Decision Analysis**

Decision analysis is a logical process for identification of the best option in the face of uncertainty; however, there are some practical impediments to its application. First, decision-analysis methods require fairly accurate guesses with regard to the likelihood of the outcomes of treatment. In order to make informed choices between options, the ramifications of these options must be well described. In some fields of medicine, these data are available. For example, a cardiologist can inform a patient who has aortic stenosis of the natural history of untreated disease, the therapeutic options, the risks and benefits of various types of valve-replacement procedures, and the frequency of typical complications. Such outcome-study data are rare in orthopaedics, and the validity of some of the available data has been questioned. The conclusions to be drawn from a decision-analysis process can be no more valid than the data.

Another problem with regard to decision analysis is the measurement of utility values. The choice of words used to present the options may introduce strong biases. Tversky and Kahneman found that people make different choices in equivalent situations on the basis of whether the choice is presented as a loss or a gain. For example, the chance that one-third of a group of patients will survive a disease typically is deemed a better contingency than the chance that two of three patients will die, even though the two phrases represent identical outcomes. People tend to take disproportionate risks to avoid losses compared with the chances that they are willing to take for possible gains. Thus, any study that measures utility must be examined closely with regard to how the options were presented.

Some may criticize the decision-analysis approach because its process does not mirror that of human cognition, which makes decision analysis somewhat difficult to teach and which may impede its widespread use. However, the fact that it is different from the standard process of human reasoning renders it less susceptible to the errors that humans are prone to make. For example, physicians may place unwarranted emphasis on a possible diagnosis merely because they recently read an article on that topic or saw a patient who had that particular problem. Also, surgeons who once missed the diagnosis of a particular illness may become excessive in their zeal not to repeat such a mistake; consequently, they may frequently order tests for what is a very rare disease. Finally, without making an explicit effort to the contrary, physicians may fail to accord appropriate respect to the possibility of false-positive and false-negative results. Thus, it may not always be best to mirror the human cognitive process.

The natural cognitive approach to diagnosis is much stronger than a decision-analysis approach in one important area: the generation of hypotheses. If physicians do not consider the correct diagnosis at all, no system of logic or probability can lead to that diagnosis. Decision analysis does not have any means for suggesting possibilities. Decision analysis also demands that a disease be considered as either present or absent without regard to its severity or its stage. Such distinctions may be clinically important, but they are lost in a process that conceives only in binary terms. It is therefore important that decision analysis be used as an adjunct to the diagnostic process but not as a replacement for clear thought.

Other limitations of the decision-analysis approach lie outside of the reasoning process itself but are no less formidable. Many physicians have ordered a test even when they know that it is not medically indicated. This may happen because many patients expect to be tested, because patients rate the quality of care in direct proportion to the number of tests performed, or because patients regard tests as the source of ultimate truth. The legal climate also has been cited as the cause of excessive testing. Finally, physicians themselves, who have been educated to get the right answer, may be obdurate in their need to be 100 per cent correct. As decision analysis becomes more accepted in the medical community, perhaps some of these difficulties will abate.

It is also important to note that decision-analysis and cost-effectiveness studies are not identical. For example, fewer radiographs might be needed if physicians were willing to accept an occasional missed fracture. Conversely, perhaps more fractures could be detected if physicians were willing to pay more for better testing. In order for a plan for the treatment of a fracture to be considered cost-effective, the benefits of detecting the fracture and the cost of missing the diagnosis first must be determined. Often, this analytical step is omitted. For example, Gelb et al. recently found that most magnetic resonance imaging scans made in a referral sports-medicine practice did not change the clinician's diagnosis. Those authors concluded that such scans are not used in a cost-effective manner. However, without establishing the cost of a missed diagnosis (and ignoring, for the moment, the ethical question of whose values are to be used), it cannot be claimed that it would be cost-effective to perform the test less frequently.

Decision analysis can be used to demonstrate when a test has no chance of influencing clinical treatment. Decision analysis cannot, of course, eradicate the uncertainty that is inherent in medicine, but application of its principles can promote the appropriate use of resources. The elimination of useless diagnostic testing will satisfy even the most stringent definition of the term cost-effective. This should please the purchasers of health care. It also should lead to the best outcomes, and that should please both patients and physicians alike.
References


