We are indebted to Miss Judith C. Hyman for nursing assistance, to Mrs. Marguerite Sonneborn, Mrs. Jean K. Linz, and Mr. Steven Lietman for technical assistance, to Mrs. Evelyn Fleckenstein and Ms. Gail Otis for assistance, and to members of the Osler medical house staff of the Johns Hopkins Hospital for their cooperation.

References

1. Gilbert DN, Plamp C, Starr P, Bennett WM, Houghton DC, Porter G. Comparative nephrotoxicity of gentamicin and tobramycin in rats. Antimicrob Agents Chemother. 1978; 13:34-40.
2. Luft FC, Patel V, Yum MN, Patel B, Kleit SA. Experimental aminoglycoside nephrotoxicity. J Lab Clin Med. 1975; 86:213-20.
3. Barza M, Pinn V, Tanguay P, Murray T. Nephrotoxicity of newer cephalosporins and aminoglycosides alone and in combination in a rat model. J Antimicrob Chemother. 1978; 4: Suppl A:59-68.
4. Fry TL, Fried FA, Goven BA. Renal toxicity: tobramycin and gentamicin. Invest Urol. 1977; 15:100-3.
5. Akiyoshi M. Evaluation of ototoxicity of tobramycin in guinea pigs. J Antimicrob Chemother. 1978; 4: Suppl A:69-72.
6. Brummett RE, Fox KE, Bendrick TW, Himes DL. Ototoxicity of tobramycin, gentamicin, amikacin and sisomicin in the guinea pig. J Antimicob Chemother. 1978; 4: Suppl A:73-83.
7. Walker BD, Gentry LO. A randomized, comparative study of tobramycin and gentamicin in the treatment of acute urinary tract infections. J Infect Dis. 1976; 134: Suppl:S146-9.
8. Fee WE Jr, Vierra V, Lathrop GR. Clinical evaluation of aminoglycoside toxicity: tobramycin versus gentamicin, a preliminary report. J Antimicrob Chemother. 1978; 4: Suppl A:31-6.
9. Schentag JJ, Plaut ME, Cerra FB, Wels PB, Walczak P, Buckley RJ. Aminoglycoside nephrotoxicity in critically ill surgical patients. J Surg Res. 1979; 26:270-9.
10. Kahlmeter G, Hallberg T, Kamme C. Gentamicin and tobramycin in patients with various infections — concentrations in serum and urinary excretion. J Antimicrob Chemother. 1978; 4: Suppl A:37-45.
11. Gifford RH, Feinstein AR. A critique of methodology in studies of anticoagulant therapy for acute myocardial infarction. N Engl J Med. 1969; 280:351-7.
12. Smith CR, Baughman KL, Edwards CQ, Rogers JF, Lietman PS. Controlled comparison of amikacin and gentamicin. N Engl J Med. 1977; 296:349-53.
13. Chan RA, Benner EJ, Hoeprich PD. Gentamicin therapy in renal failure: a nomogram for dosage. Ann Intern Med. 1972; 76:773-8.
14. Lietman PS, Lipsky JJ, Johannes RS. Gentamicin: a new enzymologic assay. Presented at the 72d annual meeting of the American Society for Microbiology, Philadelphia, April 28, 1972.
15. Smith DH, Van Otto B, Smith AL. A rapid chemical assay for gentamicin. N Engl J Med. 1972; 286:583-6.
16. Madsen PO, Kjaer TB, Mosegaard A. Comparison of tobramycin and gentamicin in the treatment of complicated urinary tract infections. J Infect Dis. 1976; 134: Suppl:S150-2.
17. Klastersky J, Hensgens C, Henri A, Daneau D. Comparative clinical study of tobramycin and gentamicin. Antimicrob Agents Chemother. 1974; 5:133-8.
18. Wade JC, Smith CR, Petty BG, et al. Cephalothin plus an aminoglycoside is more nephrotoxic than methicillin plus an aminoglycoside. Lancet. 1978; 2:604-6.
19. Smith CR, Maxwell RR, Edwards CQ, Rogers JF, Lietman PS. Nephrotoxicity induced by gentamicin and amikacin. Johns Hopkins Med J. 1978; 142:85-90.
20. Smith CR, Lipsky JJ, Lietman PS. Relationship between aminoglycoside-induced nephrotoxicity and auditory toxicity. Antimicrob Agents Chemother. 1979; 15:780-2.
21. Giamarellou H, Metzikoff C, Papachristophorou S, Dontas AS, Daikos GK. Prospective comparative evaluation of gentamicin or gentamicin plus cephalothin in the production of nephrotoxicity in man. J Antimicrob Chemother. 1979; 5:581-90.

SPECIAL ARTICLE

THE THRESHOLD APPROACH TO CLINICAL DECISION MAKING

Stephen G. Pauker, M.D., and Jerome P. Kassirer, M.D.

Abstract The physician's estimate of the probability that a patient has a particular disease is a principal factor in the determination of whether to withhold treatment, obtain more data by testing, or treat without subjecting the patient to the risks of further diagnostic tests.

Using the concepts of decision analysis, we have derived expressions for two threshold probabilities involved in this choice: a "testing" threshold and a "test-treatment" threshold. Values can be assigned to these thresholds from data on the reliability and potential risks of the diagnostic test and the benefits and risks of a specific treatment. Treatment should be withheld if the probability of disease is smaller than the testing threshold, and treatment should be given without further testing if the probability of disease is greater than the test-treatment threshold. The test should be performed (with treatment depending on the test outcome) only if the probability of disease is between the two thresholds. The method exposes important principles of decision making and helps the clinician develop a rational, quantitative approach to the use of diagnostic tests. (N Engl J Med. 1980; 302:1109-17.)

WHEN two different approaches to managing a patient appear to have the same potential value, the decision faced by the physician is often described as a "toss-up." This concept of indifference between strategies is familiar to all experienced clinicians but has received scant attention as a possible reference point in clinical decision making. In a previous study we developed the concept of the therapeutic threshold — a probability of disease that constitutes one such point of indifference.[1] In problems that can be reduced to two choices, i.e., either administering or withholding a specific treatment for a specific disease, this threshold can be used as a guide. It stipulates the action to be taken as a function of the probability of disease in a given patient. The optimal choices are to withhold treatment when the estimated probability of disease is smaller than the therapeutic threshold and to administer treatment when the prob-

From the Clinical Decision Making Group, Department of Medicine, New England Medical Center Hospital, Tufts University School of Medicine (address reprint requests to Dr. Pauker at the New England Medical Center Hospital, 171 Harrison Ave., Boston, MA 02111).

Supported in part by a Research Career Development Award (1 K04 GM 00349) to Dr. Pauker from the National Institute of General Medical Sciences, by a grant (1 P04 LM 03374) from the National Library of Medicine, and by a grant (1 P41 RR 01096) from the Division of Research Resources, National Institutes of Health.

ability is greater than this threshold. The therapeutic threshold is calculated from data describing the benefits and risks of the specific treatment.

In the present study we extend the threshold concept to problems that can be reduced to three choices: to withhold treatment, to treat without testing, or to perform a test that determines the subsequent approach. We derive two thresholds: a "testing" threshold, the probability of disease at which there is no difference between the value of withholding treatment and that of performing the test; and a "test-treatment" threshold, the probability of disease at which there is no difference between the value of performing the test and that of administering treatment. Calculation of these thresholds incorporates data that describe the risks and benefits of treatment and the risks of testing and its reliability in terms of the frequency of false-positive and false-negative results.

According to this approach, the decision to treat, not treat, or perform a test is governed by the estimated probability of disease and the two calculated thresholds. For a patient suspected of having chronic active hepatitis, for example, the best choice is to withhold steroid therapy if the probability of disease is smaller than the "testing" threshold, to administer steroids if the probability of disease is greater than the "test-treatment" threshold, and to perform a liver biopsy only if the probability of disease lies between the two thresholds. Steroids are administered only if the biopsy proves positive.

The thresholds provide a basis for deciding whether to test, treat, or do neither. The relations between tests and treatments identified by the threshold approach also provide a basis for teaching strategies of patient management.

Methods and Results

Definition of the Problem

The problem considered in this analysis has the following characteristics: Only a single disease, disorder, syndrome, or condition is under consideration; a given patient either has it or does not. A well-defined, clearly beneficial treatment is available. A patient who does not have the disease but who nevertheless receives treatment is subjected to some identifiable risk; the patient who has the disease is subjected to the same risk from the treatment, but nevertheless benefits from it. A diagnostic test can provide additional information about the presence of the disease. The test is imperfect — both false-positive and false-negative results occur, and a patient who undergoes the diagnostic test is exposed to some risk.

The physician must decide whether to administer treatment without performing the diagnostic test, to withhold treatment without testing, or to perform the test and then administer or withhold treatment on the basis of the results. The physician's estimate of the likelihood of disease in a given patient, based on information available without use of the diagnostic test, is designated here as the probability of disease. All symbol definitions are summarized in Table 1. Economic costs are not considered in this analysis.

Representation of the Problem

A model with the characteristics described above is shown as a decision tree in Figure 1. One can see from the center branch that treatment is administered if the test is positive and withheld if it is negative. However, because positive results occur in some patients without the disease, some of these patients will nevertheless be subjected to the risk of therapy. Similarly, because negative results occur in some patients with the disease, some of them will not receive treatment and will therefore lose the benefit of therapy.

Each of the eight outcomes in the decision tree is assigned a probability and a relative value, or utility. The expected utility of each of the three available options is then calculated from these probabilities and utilities.[2-13] According to the principles of decision theory, the best choice is the one with the highest expected utility.[2-13]

Derivation of the Thresholds

A decision maker should be indifferent with regard to choosing between two courses of action when the expected utilities of both options are the same. This principle, used previously to derive the therapeutic threshold,[1] can be applied to the more complex problem posed here. The indifference point for the choice between withholding therapy and performing a diagnostic test (Fig. 1) is a probability of disease designated here as the "testing" threshold (T_t). The indifference point for the choice between performing the diagnostic test and administering treatment (Fig. 1) is a probability of disease designated here as the "test-treatment" threshold (T_{trx}). Because the thresholds define these two indifference points, the physician can be guided by the calculated thresholds and estimated probability of disease in a given patient. As illustrated in the top segment of Figure 2, the best choices are to withhold both treatment and the test if the probability of disease is smaller than the testing threshold, to administer treatment without testing if the probability of disease is greater than the test-treatment threshold, and to perform the test only if the probability of disease falls between the two thresholds.

To derive the thresholds, we assigned a probability and utility to each terminal branch of the decision tree (Tables 1 and 2), calculated the expected value of two of the main branches (for example, "administer treatment" and "perform diagnostic test"), set the expected value of the two main branches equal to each other,

Table 1. Definitions of Symbols.

Symbol	Definition
p	Probability of disease before testing
$P_{pos/d}$	Probability of a positive result in patients with disease
$P_{neg/d}$	Probability of a negative result in patients with disease
$P_{pos/nd}$	Probability of a positive result in patients without disease
$P_{neg/nd}$	Probability of a negative result in patients without disease
B_{rx}	Benefit of treatment in patients with disease
R_{rx}	Risk of treatment in patients without disease
R_t	Risk of diagnostic test
T_t	Testing threshold
T_{trx}	Test-treatment threshold

Table 2. Utilities of Outcomes.

Outcome	Notation
Disease, treated	U_{drx}
No disease, treated	U_{ndrx}
Disease, not treated	U_{dno}
No disease, not treated	U_{ndno}
Disease, treated (test performed)	$U_{drx} - R_t$
No disease, treated (test performed)	$U_{ndrx} - R_t$
Disease, not treated (test performed)	$U_{dno} - R_t$
No disease, not treated (test performed)	$U_{ndno} - R_t$

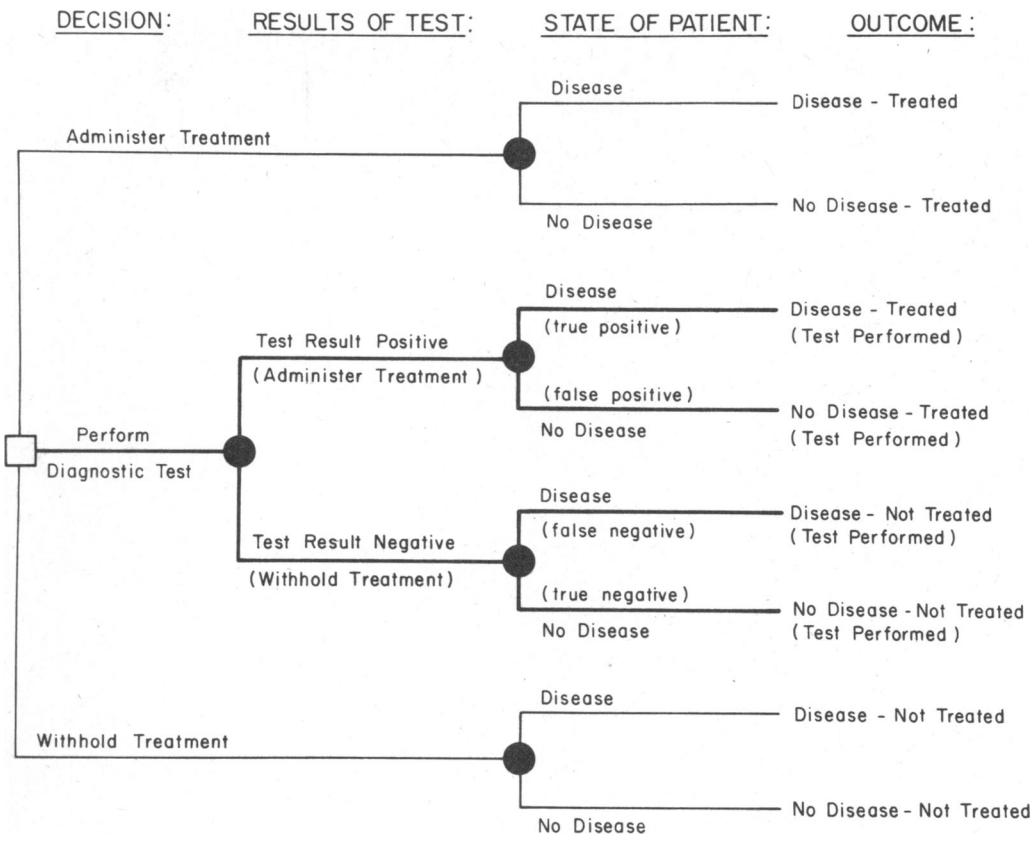

Figure 1. A Decision Tree Demonstrating Three Available Options — Withholding Therapy (Lower Branch), Administering Therapy (Top Branch), and Performing a Diagnostic Test (Middle Branch).
The square denotes the decision, and the circles chance events.

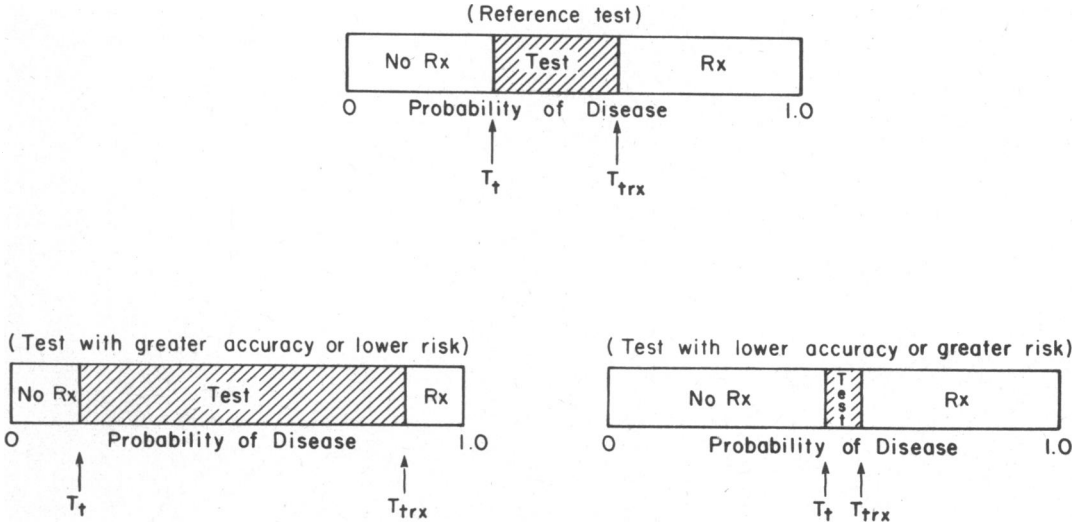

Figure 2. Effect of Differences in the Reliability and Risks of Diagnostic Tests on the Two Thresholds.
In all portions of the figure, the probability of disease (depicted on the horizontal axis) ranges from 0 to 1.

and then solved for the probability of disease. In making these calculations, we simplified the utilities by defining the various benefits and risks of therapy and the risks of testing as differences between the utilities of various outcomes.[1] Thus, the net benefit of treatment (B_{rx})* is the difference in utility, for patients known to have the disease, between treating and withholding therapy. The net risk of treatment (R_{rx})* is the difference in utility, for patients known not to have the disease, between withholding therapy and treating. The net risk of testing (R_t) is the difference in utility between achieving a given outcome with and without exposing the patient to the risks of testing. The reliability of the test is defined in terms of the frequency of true-positive and false-positive results and of true-negative and false-negative results. The derivation of the testing threshold is given in the appendix; all symbols are defined in Table 1. The expressions for the thresholds are:

$$T_t\,(\text{testing threshold}) = \frac{(P_{pos/nd})\times(R_{rx}) + R_t}{(P_{pos/nd})\times(R_{rx}) + (P_{pos/d})\times(B_{rx})}$$

$$T_{trx}\,(\text{test-treatment threshold}) = \frac{(P_{neg/nd})\times(R_{rx}) - R_t}{(P_{neg/nd})\times(R_{rx}) + (P_{neg/d})\times(B_{rx})}$$

Assessment of the Variables

Probability of Disease

Published estimates of disease prevalence can be used as an index of the probability of a given disease in patients with a relatively standard clinical picture. Even in unusual cases, such estimates often form convenient anchor points for subjective estimates.[14] In fact, clinical findings often do not comprise a typical constellation, and when an assessment of the probability of disease is required, it must be tailored to the findings in the individual patient. In this study, the probability of disease is a quantitative assessment by the physician of the likelihood of disease based on the clinical and laboratory findings available before the principal management decision must be made.

Such probability assessments must, of course, be estimates. Studies in nonmedical domains show that people have biases and often make inaccurate estimates[14] and that training improves the reliability of such estimates.[15,16] Of course, most studies of this sort rely on simple tests in which an actual probability is known (for example, the number of various colored balls in an urn), whereas in medicine a probability assessment represents a belief or opinion for which no actual or true value exists.[17] Despite this state of affairs, we believe that subjective probability assessments are useful. At least one study showed that physicians make probability estimates with reasonable reliability.[18] Although physicians may feel uncomfortable when asked to estimate the probability of disease in an individual patient, they often make such judgments more confidently when asked to estimate how many of 100 or 1000 identical patients with a given set of findings will actually be found to have the disease (Pauker SG, Kassirer JP: unpublished observations). Finally, the exact value of a given probability may not be critical if repeated Bayesian analysis is applied to the data[19] or if sensitivity analysis is used to test the limits of this variable.[1,20]

Reliability of Tests

In this study the reliability of a diagnostic test is summarized by four interrelated indexes: the true-positive rate or test sensitivity ($P_{pos/d}$), the true-negative rate or test specificity ($P_{neg/nd}$), the false-positive rate ($P_{pos/nd}$), and the false-negative rate ($P_{neg/d}$).[21] These indexes can often be obtained from published data, but sometimes the values are not specific enough. In such cases, it is necessary to use the subjective probability estimates of experts to obtain assessments of the test reliability tailored to individual clinical problems. If a calculation requires data on the reliability of a liver biopsy in diagnosing chronic active hepatitis, for example, the sensitivity of the biopsy might be lower than that given in published reports if the patient has already received steroid therapy for some time. In such cases adjustments must be based on expert opinion.

*B_{rx} and R_{rx} were designated B and C respectively in our previous description of the therapeutic threshold.[1]

Benefits and Risks

The use of the thresholds requires that the relative value of each potential outcome be assigned on a single, consistent utility scale. Such scales can be expressed in terms of survival, freedom from morbidity (for example, quality-adjusted years of survival or percentage of patients free of major complications), or even in arbitrary units that combine the patient's personal values regarding many factors, including risk.[22] Because the separate utilities are expressed on a single scale, the risks and benefits as defined here will also lie on that scale.

The assessment of risks and benefits is not simple. Even a single attribute such as survival may be expressed in terms of life expectancy, five-year survival, or some other measurement. When several attributes must be considered, when economic costs affect the decision, and when the patient's own values must be included in the decision-making process, the utility assessments are all the more difficult to make and may require special techniques, which are an area of active research. In the examples we present, we use several measures of utility. In one example we make arbitrary assumptions about the relative values of disparate risks and benefits of treatment, because the various outcomes have very different impacts on the patient's welfare. In the other example we base the utility assessments for test and treatment outcomes solely on mortality rates. The necessity for making such assumptions explicit should be viewed as a strength and not as a weakness of this analytic approach: certainly, comparisons of this nature must underlie all clinical decisions, whether they are based on "intuitive" judgment or formal analytical reasoning.

Applications of the Threshold Equations

In the following examples, published data are cited as the source of probabilities and utilities except when such data were not specific enough for use in the calculations. In such cases the opinions of experts were used.

A Patient Suspected of Having Vasculitis with Renal Involvement

Case Description

A 55-year-old man presents with arthralgias, severe hypertension (blood pressure, 240/140 mm Hg), and renal failure (blood urea nitrogen, 90 mg per deciliter; creatinine, 8 mg per deciliter). There is no history of renal disease, but for the past five years the patient has had hypertension treated with methyldopa and diuretics. Six months ago his renal function was normal. The fundi show Grade 3 hypertensive changes, and the skin shows a few scattered petechiae, but the remainder of the physical examination is normal. The white-cell count is 9700 with 2 per cent eosinophils and no left shift; the hemoglobin is 14 g per deciliter and the hematocrit is 45 per cent; a few helmet cells and burr cells are seen on the peripheral smear, and the reticulocyte count is 3.5 per cent. Urinalysis demonstrates 20 to 25 red cells per high-power field and several red-cell casts; there is 3+ proteinuria. Chest x-ray films are normal except for slight cardiomegaly, and the electrocardiogram reveals voltage changes consistent with left ventricular hypertrophy. Stool guaiac shows a 2+ reaction. The patient was begun on corticosteroid therapy 24 hours ago for suspected renal vasculitis. He was also treated with parenteral antihypertensive drugs, and his blood pressure fell to 130/90 mm Hg.

Should a renal biopsy be performed, should steroid therapy be continued for one to two months, or should steroid therapy be discontinued?

Diagnosis

The diagnoses under consideration are renal vasculitis and malignant hypertension. Steroid-responsive vasculitis is regarded here as "disease" in the decision tree, and steroid-unresponsive malignant hypertension as "no disease." On the basis of clinical features, the probability of renal vasculitis is estimated to be between 0.6 and 0.7.

Risks and Benefits of Treatment

The treatment under consideration is short-term corticosteroid therapy (one to two months). Steroid therapy will not benefit patients with malignant hypertension and severe renal insufficiency; it will only subject them to the risk of steroid-induced complications. These serious complications (R_{rx}) will occur in approximately 5 per cent of patients treated.[23-26] It is assumed that steroid therapy will improve renal function in only about 20 per cent of patients with vasculitis complicated by severe renal failure and that the long-term benefits of improving renal function have approximately twice the value of avoiding steroid complications. Thus, the benefit of therapy would be 20% × 2, or 40, but because only 95 per cent of treated patients who improve will avoid serious steroid complications, the net benefit of treating (B_{rx}) is 40 × 95%, or 38.

Risk and Value of the Test

The risk of serious complications (R_t) after needle biopsy of the kidney in a patient with severe hypertension and a creatinine of 8 mg per deciliter is approximately 2 per cent.[27-34] The probability of finding typical vasculitic or glomerular proliferative lesions in vasculitis complicated by severe renal failure is approximately 0.9[35]; the probability of finding glomerular and arteriolar lesions that would be mistaken for vasculitis in patients with malignant hypertension[36-38] is estimated to be 0.05.

Calculations

Benefits and risks:

$$B_{rx} = 38\%; R_{rx} = 5\%; R_t = 2\%$$

Test accuracy:

$$P_{pos/d} = 0.9; P_{neg/d} = 0.1$$
$$P_{pos/nd} = 0.05; P_{neg/nd} = 0.95$$

Testing threshold (T_t):

$$(T_t) = \frac{(P_{pos/nd})\times(R_{rx}) + R_t}{(P_{pos/nd})\times(R_{rx}) + (P_{pos/d})\times(B_{rx})} = \frac{(0.05\times5) + 2}{(0.05\times5) + (0.9\times38)} = 0.07$$

Test-treatment threshold (T_{trx}):

$$(T_{trx}) = \frac{(P_{neg/nd})\times(R_{rx}) - R_t}{(P_{neg/nd})\times(R_{rx}) + (P_{neg/d})\times(B_{rx})} = \frac{(0.95\times5) - 2}{(0.95\times5) + (0.1\times38)} = 0.32$$

Conclusions

If the physician had assessed the probability of vasculitis to be less than 0.07, the best decision would be to discontinue steroid therapy. If the probability of vasculitis was thought to be greater than 0.32, the best choice would be to continue steroid therapy and not to subject the patient to renal biopsy. Only if the probability of vasculitis was between 0.07 and 0.32 would the patient's best interests be served by performing a biopsy. In this particular patient, the probability of vasculitis was thought to be 0.6 to 0.7; for this reason the best decision is to continue steroid therapy and not to perform a renal biopsy.

This particular analysis shows the way in which the threshold concept can be applied to decisions about continuing or discontinuing therapy. It also illustrates the manner in which expert opinion can be integrated into the decision-making process when published data are inadequate. Moreover, it shows that unnecessary diagnostic procedures can be avoided through explicit assessment of the value and risks of such tests.

In the selection of data for this problem, improving renal function was assigned twice the value of avoiding a short-term complication of steroid therapy. The choice of a factor of two was arbitrary but consistent with the impressions of experienced clinicians. If the factor is five rather than two, then T_t is 0.03 and T_{trx} is 0.19; if the factor is one, then T_t is 0.13 and T_{trx} is 0.41. Thus, for any factor between one and five, the decision regarding the patient described here would not be different — the probability of vasculitis (0.6 to 0.7) would far exceed T_{trx}. Extending this sensitivity analysis shows that steroid therapy is the best choice in this patient no matter what value is taken for this factor.

A Patient with an Ulcerating Lesion of the Stomach

Case Description

A 60-year-old man presents with epigastric pain and hematemesis. An upper-gastrointestinal series reveals a 2-cm ulcer on the greater curvature of the stomach.

Should gastroscopy be performed? Should the patient have exploratory surgery? Should the patient be treated medically without being subjected to either gastroscopy or surgery?

Diagnosis

The principal diagnosis under consideration is gastric cancer. On the decision tree, gastric cancer is a "disease," and nonmalignant ulcer "no disease." Given the history and findings on the upper-gastrointestinal series, several gastroenterologists and radiologists concluded that the probability of gastric cancer in this patient is approximately 0.1.

Benefits and Risks of Treatment

The treatment under consideration is laparotomy followed by gastrectomy if the presence of cancer is

confirmed. The risks and benefits of gastrectomy for gastric cancer can be assessed as follows: the operative mortality rate (R_{rx}) for laparotomy in a 60-year-old patient is approximately 2 per cent[39,40]; early gastrectomy in patients with gastric cancer should prevent localized cancer from spreading to regional nodes in some fraction of patients. The relative survival rates for patients with localized cancer of the stomach and for patients with regional spread of the disease are relatively constant when measured five and 10 years after surgery.[41] Patients with localized disease survive at roughly 47 per cent of the rate at which patients without cancer survive, whereas patients with regional involvement survive at only 14 per cent of the survival rate of patients without cancer.[41-43] Thus, the benefit of early surgery (B_{rx}) is a gain of approximately 33 per cent in relative survival rate.

Risk and Value of the Test

The mortality rate of gastroscopy (R_t) is approximately 0.005 per cent.[44] It is assumed that approximately 96 per cent of gastric cancers are detected by gastroscopy if the test is performed with guided lavage and brush cytology (the true-positive rate)[45] and that the false-positive rate is approximately 2 per cent.

Calculations

Benefits and risks:

$$B_{rx} = 33\%; R_{rx} = 2\%; R_t = 0.005\%$$

Test accuracy:

$$P_{pos/d} = 0.96; P_{neg/d} = 0.04$$
$$P_{pos/nd} = 0.02; P_{neg/nd} = 0.98$$

Testing threshold (T_t):

$$(T_t) = \frac{(0.02\times2) + 0.005}{(0.02\times2) + (0.96\times33)} = 0.0014$$

Test-treatment threshold (T_{trx}):

$$(T_{trx}) = \frac{(0.98\times2) - 0.005}{(0.98\times2) + (0.04\times33)} = 0.60$$

Conclusions

If the physician had assessed the probability of gastric cancer to be extremely small (less than 0.0014), then neither gastroscopy nor surgery should be done. If the probability of cancer was thought to be greater than 0.6, then surgery should be done without gastroscopy. If the probability of cancer, however, lies between the two thresholds (in this case, 0.0014 and 0.6), gastroscopy should be performed. In this patient, the probability of disease was estimated at 0.1, and gastroscopy is thus the best choice. If the study is positive, surgery should be performed; if it is negative, the patient should be treated medically. It is of particular interest that gastroscopy is not indicated when the probability of gastric cancer is greater than 0.60 even though the risk of the procedure is very low, simply because the reliability of the test does not warrant its use. This example also serves to illustrate circumstances in which the risk of testing can, for all intents and purposes, be ignored. Thus, if R_t had been 0 instead of the reported figure of 0.005 per cent, the two thresholds would be 0.0013 and 0.60; this trivial change would have had no effect on the decision in this patient.

Discussion

When managing an individual patient, most physicians implicitly consider the trade-offs that influence the decision whether to withhold therapy, order another test, or administer therapy. The factors that must be incorporated into such decisions include the reliability and risk of the test and the value and risk of treatment.[46] Nonetheless, assessment of these factors is often vague, and even if reasonable data on which to base such assessments are available, the factors are frequently integrated tacitly rather than explicitly. In this paper we have shown that decisions regarding the use of diagnostic tests can be made in a rigorous fashion through the calculation of a testing threshold and a test-treatment threshold.

Interpretation of the Thresholds

The two thresholds are represented by the diagrams in Figure 2. The bar at the top represents a clinical situation in which a particular diagnostic test is available. The probability of disease in a given patient and the two thresholds, T_t and T_{trx}, determine which segment of the diagram applies to that patient. When the probability of disease falls within one of the end segments, either the test result will not alter the optimal therapeutic action, or the risk of the test will outweigh the benefit of the diagnostic information that the test may provide. The best choice for probabilities below the testing threshold is to withhold treatment and, for probabilities above the test-treatment threshold, to administer treatment. When the probability of disease before testing lies between the two thresholds, however, the test result could change the likelihood of disease enough to alter the subsequent therapeutic decision, and, thus, the best choice is to perform the test.

The impact of the accuracy and risk of a test is illustrated by the lower two diagrams in Figure 2. The diagram on the lower left illustrates another hypothetical test that is either more accurate (smaller frequency of false-positive and false-negative results) or safer (lower risk) than the reference test shown at the top. In this case, the test has greater diagnostic value as compared with the reference test and should be performed at both lower and higher probabilities than the reference test. By contrast, a third test with lower accuracy or greater risk is illustrated by the diagram on the lower right. In this case, the test has less diagnostic value than the reference test and should be per-

formed only when the probabilities of disease lie in a narrow range.

The Relation between Testing and Treating

Discussions of the indications for diagnostic tests usually consider the virtues and drawbacks of tests in isolation, rather than in relation to their potential therapeutic usefulness. The critical principles that link the benefits and risks of testing with those of treatment, previously obfuscated by the lack of a quantitative approach, are highlighted by reference to the thresholds. First of all, it is apparent that the risk of a diagnostic test must be weighed against the risk of exposing patients without the disease to unnecessary treatment: one should be willing to use a test with high risk if it reliably identifies patients who do not have the disease and who will thus avoid high-risk therapy. If the risk of testing exceeds the risk of treating such patients, however, then it would be unreasonable to perform the test even if the test results could provide perfect diagnostic information. In such cases, all patients with probabilities of disease above the previously described therapeutic threshold[1] should be treated without testing. Similarly, in patients with a small chance of having the disease, the risk of testing must be weighed against the potentially lost benefit of treating those few patients who may actually have the disease. If the risk of testing exceeds the benefit of treating such patients, it would again be unreasonable to perform the test. On the other hand, if the risk of testing is zero, the thresholds are determined by the information that the test might provide in the light of the probability of disease before testing. When the probability of disease is either very high or very low, it might not be altered sufficiently by the test results to influence the therapeutic decision: in such cases testing is superfluous and only adds to the patient's burden of risk and expense.

The risk and the reliability of a diagnostic test are counterbalancing forces. If the risk of the test is sufficiently low relative to the risk of therapy, then substantial inaccuracies of testing might be tolerated. Conversely, if the accuracy of the test is sufficiently high, then substantial risks of testing might be tolerated to obtain needed information. Finally, these observations point out once again the importance of considering the potential impact of a diagnostic test before performing the test.[4,7,20,46] The didactic value of these concepts for training both students and house officers is evident.

Because the risk of most tests currently in use is quite small, the accuracy of the test and the probability of disease are major factors that influence the decision to use the test. It is common, however, to face a decision about diagnostic testing in a critically ill patient in whom the risk of a test is considerably increased. Furthermore, such patients often have underlying diseases that limit their life expectancy and thus limit the relative benefit of therapy. It is precisely these patients in whom the risk and potential of errors of the test must be carefully balanced against the potential benefits of additional information, and it is these complex, individual situations in which the threshold technique may have its greatest impact.

Applications

This analytic approach should prove useful in evaluating new diagnostic procedures. If a new procedure provides highly accurate information and is virtually risk free, its use should be encouraged. If, on the other hand, a new test is associated with substantial risk or provides less than perfect information, these factors must be carefully balanced against its potential advantages. In such cases, consideration of the risks and errors of the test in relation to the risks and benefits of therapy should be critical in determining whether to recommend the test.

The method presented here requires modification in certain settings. Our model assumes that the test under consideration provides the physician with only diagnostic information and that therapy should be administered without performing the test if the likelihood of the disease is sufficiently high. In certain circumstances, however, a test may provide information required for directing therapy. When considering coronary-bypass surgery to treat angina pectoris, for example, the surgeon must study a coronary arteriogram to assess the anatomic site for grafts. Similarly, when evaluating a mass in the kidney, the urologist performs a renal arteriogram to assess tumor size, location, and the presence or absence of vascular invasion, and uses these data to determine the nature of the surgical approach. In these circumstances, the test-treatment threshold does not exist because therapy cannot be administered without testing. Nevertheless, the testing threshold remains valid.

This explicit approach to decision making can be applied both to generic clinical problems and to decisions involving individual patients. Examples of common problems that should be amenable to the threshold approach include the choice among pulmonary arteriography, anticoagulation, or no therapy for suspected pulmonary embolism, and the choices among biopsy, steroid therapy, or no therapy for temporal arteritis, chronic active hepatitis, and the idiopathic nephrotic syndrome. A particular strength of the threshold approach is the ready mechanism that it provides for the combination of data from diverse sources. For example, the clinician might use one source in the literature for information on the sensitivity and specificity of a diagnostic test, another source for data on the complications of a test, and the opinion of one or more consultants for information on the benefits and risks of therapy; the clinician might then use his or her own judgment or the integrated judgment of several physicians (illustrated in both examples) to assess the probability of disease.

The threshold approach to clinical decision making does not necessarily require strict adherence to the model described here. For any clinical problem in which the presence or absence of a disease can be described as a probability, one can derive a threshold by setting the expected utilities of two management options equal to each other and solving for the probability. Of course, problems that do not conform to these conditions can be approached with standard analytic methods of decision making.[2-13]

The threshold concept has the potential of improving the quality and containing the costs of medical care. By identifying situations in which the ultimate therapeutic decision would not be affected by the result of a test, it should be possible to avoid some of the misleading results, the complications, and the financial burden of unneccessary diagnostic tests.

Appendix

The testing threshold (T_t) is the probability of disease (p) at which the expected value of withholding treatment (EV_{no}) is equal to the expected value of testing (EV_t). For other symbols, see Tables 1 and 2.

The expected value of a given strategy is equal to the sum of the pairwise products of probabilities and utilities for each potential outcome of that strategy.[1-4,9] Thus,

$$EV_{no} = (p)\times(U_{dno}) + (1 - p)\times(U_{ndno})$$

$$EV_t = \begin{Bmatrix} (p)\times(P_{pos/d})\times(U_{drx} - R_t) + & (1 - p)\times(P_{pos/nd})\times(U_{ndrx} - R_t) \\ + (p)\times(P_{neg/d})\times(U_{dno} - R_t) + & (1 - p)\times(P_{neg/nd})\times(U_{ndno} - R_t) \end{Bmatrix}$$

Furthermore, since all patients are assumed to have either a positive or negative test result, we have:

$$P_{pos/d} + P_{neg/d} = 1, \text{ and } P_{pos/nd} + P_{neg/nd} = 1$$

Setting $EV_{no} = EV_t$ and transposing,

$$(p)\times\begin{Bmatrix} U_{dno} - U_{ndno} \\ - (P_{pos/d})\times(U_{drx} - R_t) \\ + (P_{pos/nd})\times(U_{ndrx} - R_t) \\ - (P_{neg/d})\times(U_{dno} - R_t) \\ + (P_{neg/nd})\times(U_{ndno} - R_t) \end{Bmatrix} = \begin{Bmatrix} (P_{pos/nd})\times(U_{ndrx} - R_t) \\ - U_{ndno} \\ + (P_{neg/nd})\times(U_{ndno} - R_t) \end{Bmatrix}$$

Substituting $P_{pos/d} = 1 - P_{neg/d}$ and $P_{pos/nd} = 1 - P_{neg/nd}$, and solving for p,

$$p = \frac{(P_{pos/nd})\times(U_{ndno} - U_{ndrx}) + R_t}{(P_{pos/nd})\times(U_{ndno} - U_{ndrx}) + (P_{pos/d})\times(U_{drx} - U_{dno})}$$

Substituting $B_{rx} = U_{drx} - U_{dno}$ and $R_{rx} = U_{ndno} - U_{ndrx}$,

$$p = T_t = \frac{(P_{pos/nd})\times(R_{rx}) + R_t}{(P_{pos/nd})\times(R_{rx}) + (P_{pos/d})\times(B_{rx})}$$

The test-treatment threshold is the probability of disease at which the expected value of testing is equal to the expected value of treating. This threshold is found by setting $EV_t = EV_{rx}$ and solving for p in a fashion similar to that above (calculations are not shown).

References

1. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. N Engl J Med. 1975; 293:229-34.
2. Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. Reading, Mass.: Addison-Wesley, 1968.
3. Ginsberg AS. Decision analysis in the clinical patient management with an application to the pleural-effusion syndrome. Santa Monica, Calif.: Rand, 1971.
4. Schwartz WB, Gorry GA, Kassirer JP, Essig A. Decision analysis and clinical judgment. Am J Med. 1973; 55:459-72.
5. Emerson PA, Teather D, Handley AJ. The application of decision theory to the prevention of deep vein thrombosis following myocardial infarction. Q J Med. 1974; 43:389-98.
6. Schoenbaum SG, McNeil BJ, Kavet J. The swine-influenza decision. N Engl J Med. 1976; 295:759-65.
7. Sisson JC, Schoomaker EB, Ross JC. Clinical decision analysis: the hazard of using additional data. JAMA. 1976; 236:1259-63.
8. McNeil BJ, Hessel SJ, Branch WT, Bjork L, Adelstein SJ. Measures of clinical efficacy. III. The value of the lung scan in the evaluation of young patients with pleuritic chest pain. J Nucl Med. 1976; 17: 163-9.
9. Kassirer JP. The principles of clinical decision making: an introduction to decision analysis. Yale J Biol Med. 1976; 49:149-64.
10. Pauker SG. Coronary artery surgery: the use of decision analysis. Ann Intern Med. 1976; 85:8-18.
11. Pauker SP, Pauker SG. Prenatal diagnosis: a directive approach to genetic counseling using decision analysis. Yale J Biol Med. 1977; 50:275-89.
12. Bunker JP, Barnes BA, Mosteller F, eds. Costs, risks, and benefits of surgery. New York: Oxford University Press, 1977.
13. Safran C, Desforges JF, Tsichlis PN, Bluming AZ. Decision analysis to evaluate lymphangiography in the management of patients with Hodgkin's disease. N Engl J Med. 1977; 296:1088-92.
14. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science. 1974; 185:1124-31.
15. Lichtenstein S, Fischhoff B. Training for calibration (ARI Technical Report TR078-A32). Alexandria, Va.: US Army Research Institute for the Behavioral and Social Sciences, 1978.
16. Kahneman D, Tversky A. Intuitive prediction: biases and corrective procedures. Manage Sci. 1979; 12:313-27.
17. Slovic P, Fischhoff B, Lichtenstein S. Behavioral decision theory. Annu Rev Psychol. 1977; 28:1-39.
18. Lusted LB. A study of the efficacy of diagnostic radiologic procedures: final report on diagnostic efficacy. Chicago: American College of Radiology, 1977.
19. de Dombal FT, Horrocks JC. Use of receiver operating characteristic (ROC) curves to evaluate computer confidence threshold and clinical performance in the diagnosis of appendicitis. Methods Inf Med. 1978; 17:157-61.
20. Pauker SG, Kassirer JP. Clinical application of decision analysis: a detailed illustration. Semin Nucl Med. 1978; 8:324-35.
21. Lusted LB. Introduction to medical decision making. Springfield, Ill.: Charles C Thomas, 1968.
22. Kenney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: John Wiley, 1976.
23. Miller RB, Harrington JT, Ramos CP, Relman AS, Schwartz WB. Long-term results of steroid therapy in adults with idiopathic nephrotic syndrome. Am J Med. 1969; 46:919-29.
24. Black DAK, Rose G, Brewer DB. Controlled trial of prednisone in adult patients with the nephrotic syndrome. Br Med J. 1970; 3:421-6.
25. The Boston Collaborative Drug Surveillance Program. Acute adverse reactions to prednisone in relation to dosage. Clin Pharmacol Ther. 1972; 13:694-8.
26. Jick H. What to expect from prednisone. Drug Ther. August, 1975; 5:85-9.
27. Lee DA, Roger R, Agre KM, Rubini M. Late complications of percutaneous renal biopsy. J Urol. 1967; 97:793-7.
28. Muth RG. The safety of percutaneous renal biopsy: an analysis of 500 consecutive cases. J Urol. 1965; 94:1-3.
29. Kark RM. Renal biopsy. JAMA. 1968; 205:220-6.
30. Slotkin EA, Madsen PO. Complications of renal biopsy: incidence in 5000 reported cases. J Urol. 1962; 87:13-5.
31. Kark RM, Muehrke RC, Pollak VE, et al. An analysis of 500 percutaneous renal biopsies. Arch Intern Med. 1958; 101:439-51.
32. River GL, Dovenbarger WV, Nikolai TF, Moffat NA. Unusual complications of kidney biopsy. J Urol. 1970; 103:15-7.
33. Brun C, Raaschou F. Kidney biopsies. Am J Med. 1958; 24:676-91.
34. Diaz-Buxo JA, Donadio JV Jr. Complications of percutaneous renal biopsy: an analysis of 1000 consecutive biopsies. Clin Nephrol. 1975; 4:223-7.
35. Patalano VJ, Sommers SC. Biopsy diagnosis of periarteritis nodosa: glomerulonephritis and renal arteriolitis as aids. Arch Pathol. 1961; 72:1-7.

36. Heptinstall RH. Pathology of the kidney. 2d ed. Boston: Little, Brown, 1974:121-62.
37. Burkholder PM. Malignant nephrosclerosis: an immunohistopathologic study on localized γ-globulin and fixation of guinea pig complement in human kidneys. Arch Pathol. 1965; 80:583-9.
38. Jones DB. Arterial and glomerular lesions associated with severe hypertension: light and electron microscopic studies. Lab Invest. 1974; 31:303-13.
39. Singer RB, Levinson L. Medical risks: patterns of mortality and survival: a reference volume. Lexington, Mass.: Lexington Books, 1976.
40. Bunker JP, Forrest WH, Mosteller F, Vandam LD. The national halothane study. Bethesda, Md.: National Institute of General Medical Sciences, 1969.
41. Axtell LM, Cutler SJ, Meyers MH, eds. End results in cancer (report no. 4). Washington, D.C.: Government Printing Office, 1972:49. (DHEW publication no. (NIH)73-272).
42. Olearchyk AS. Gastric carcinoma: a critical review of 243 cases. Am J Gastroenterol. 1978; 70:25-45.
43. Serlin O, Keehn RJ, Higgins GA Jr, Harrower HW, Mendeloff GL. Factors related to survival following resection for gastric carcinoma: analysis of 903 cases. Cancer. 1977; 40:1318-29.
44. Hafter E. Ethical problems in the management of gastroenterological patients: the risk of oesophagoscopy and gastroscopy. Scand J Gastroenterol. 1977; 12: Suppl 47:12-4.
45. Halter F, Witzel L, Grétillat PA, Scheurer U, Keller M. Diagnostic value of biopsy, guided lavage, and brush cytology in esophagogastroscopy. Am J Dig Dis. 1977; 22:129-31.
46. Kassirer JP, Pauker SG. Should diagnostic testing be regulated? N Engl J Med. 1978; 299:947-9.

MEDICAL PROGRESS

THERAPEUTIC APPLICATIONS OF ANGIOGRAPHY

(First of Two Parts)

CHRISTOS A. ATHANASOULIS, M.D.

THERAPEUTIC angiography is the field of radiology that employs methods that were formerly used for diagnosis but have been improved or modified to serve as therapeutic measures. For example, the angiographic catheter used to confirm the diagnosis of a suspected liver or kidney lesion may also become the vehicle for selective delivery of vasoactive drugs, chemotherapeutic agents, or embolizing materials, or it may be modified to become a double-lumen balloon catheter for dilatation in a patient with symptomatic arterial narrowing. Angiographic procedures and their applications may be divided into the following groups: procedures used to reduce regional blood flow, procedures used to increase blood flow, and miscellaneous procedures.

REDUCTION OF REGIONAL BLOOD FLOW

Blood flow to an organ, a lesion, or a vascular bed may be reduced by mechanical occlusion of the lumen of the feeding vessel or regional infusion of vasoconstricting drugs. Posterior pituitary extract is the only vasoconstrictor used for this purpose, and its application is discussed in a later section. Mechanical means to occlude the lumen of a vessel based on angiographic catheterization techniques include embolization, insertion of catheters with fixed or detachable balloons, and intravascular electrocoagulation.

Transcatheter Vessel Occlusion

Intravascular occlusion by means of embolization is not similar to surgical ligation of the vessel. The former is more peripheral and depends on the size and nature of the emboli; the latter is proximal and does not prevent collateral vessels from reconstituting the distal branches of the ligated vessel. Depending on the nature of the lesion and the vascular bed, different embolic materials are desirable. An arteriovenous malformation must be completely filled with a liquid embolic substance, whereas an arteriovenous fistula should be occluded not with particulate matter or liquids, which might escape into the venous outflow, but with a mechanical device occluding the fistula itself. Combinations of more than one embolic material are often necessary, for example, peripheral embolization of a renal tumor with particulate matter may be followed by proximal occlusion of the feeding vessel with a mechanical device.

The ideal method of vessel occlusion should adequately reduce blood flow without producing tissue ischemia, and the ideal embolic material should be nontoxic, stable, insoluble in vivo, malleable, radiopaque, capable of being sterilized, and easy to introduce through standard angiographic catheters. The search for the best method has resulted in the use of a large number of embolizing agents, attesting to the fact that the "ideal" remains elusive.

Since autologous blood clots and clots modified with heat or aminocaproic acid (Amicar, Lederle Laboratories, Pearl River, N.Y.) lyse within 48 hours to two weeks, they are not used for embolization for fear of early vessel recanalization.[1] Although muscle and adipose tissue produce longer-lasting occlusions, the need for a separate incision to obtain the material makes them less desirable. The gelatin sponge (Gelfoam, Upjohn, Kalamazoo, Mich.) produces occlusions that last for four to five weeks with no tissue reaction.[2] Its availability and ease of use have made the gelatin sponge the most commonly employed embolic material when temporary occlusion is desired. The method of delivery is simple. Small fragments of gelatin soaked and suspended in normal saline or contrast medium are loaded one or two at a time into a 1-ml or 3-ml syringe and are injected through the angiographic catheter into the vessel to be embolized.

From the departments of Radiology, Massachusetts General Hospital and Harvard Medical School (address reprint requests to Dr. Athanasoulis at the Department of Radiology, Massachusetts General Hospital, Boston, MA 02114).