By Barbara Scherokman, MD

Selecting and Interpreting Diagnostic Tests

This article describes basic principles of selecting and interpreting diagnostic tests. Before selecting a diagnostic test, the clinician must first determine probability of disease by evaluating the patient's medical history and results of physical examination. To then decide whether to order a particular diagnostic test, the clinician considers the characteristics (sensitivity, specificity, and false-positive rate) of the proposed test to determine whether results of the test could show a different probability of disease than would be estimated without testing. Clinicians learn the most from a test when probability of disease as estimated without testing (i.e., the "pretest probability") is between 40% and 60%. The "threshold" approach can also aid clinicians in deciding whether to withhold treatment, order more tests, or administer treatment without subjecting patients to risks of further testing.

> *"The most important concept for clinicians to understand is that tests should be selected and interpreted in a way that allows them to influence the clinician's estimate of disease probability."*

Introduction

The high cost of medical care has led to interest in encouraging physicians to make more accurate, cost-effective clinical decisions. Diagnostic tests can aid clinical assessment of disease probability so that therapeutic decisions can be made in patients' best interest, but testing contributes substantially to the cost of medical care, and tests ordered or used improperly can also cause diagnostic error and increase the risk of improper treatment. This paper therefore discusses basic principles of selecting and interpreting diagnostic tests to maximize both diagnostic accuracy and cost effectiveness. These principles can be applied to laboratory testing, radiologic testing, and other diagnostic procedures. The most important concept for clinicians to understand is that tests should be selected and interpreted in a way that allows them to influence the clinician's estimate of disease probability.

Before ordering a diagnostic test, clinicians should remember a major principle discussed by Sackett et al,[1] who stated that clinical data obtained by history and examination are far more powerful than data obtained from diagnostic laboratory tests and are usually sufficient to establish a definitive diagnosis. In addition, we should remember that absolute diagnostic certainty is impossible to attain, regardless of how much laboratory data are obtained.[2] A diagnosis is a hypothesis which test results cause to appear more or less likely to be true. As Kassirer stated, "our task is not to attain certainty, but rather to reduce the level of diagnostic uncertainty enough to make optimal therapeutic decisions."[2] More tests do not necessarily lead to more certainty, however. Extensive testing may give clinicians and patients a false sense of security which may not be justified, given the possibility of false-positive and false-negative test results. False-positive test results may increase the risk that more invasive or inappropriate testing will be done or that unnecessary, even dangerous therapy will be given. False-negative results may increase the risk that appropriate treatment will be withheld.[3]

Another aspect of test utilization discussed here is the decision whether to withhold therapy, order another test, or administer therapy. This is referred to as the "threshold approach" to making clinical decisions.[4,5] Using this approach, clinicians must take into account the reliability, value, and risks of both testing and treatment.

Pretest Probabilities of Disease

The degree of diagnostic certainty needed in making clinical decisions is a function of the degree of risk presented by the therapeutic options and the clinician's estimate of disease probability. When considering administering a specific therapy which is highly efficacious and has a low level of risk, few tests are needed because clinicians can accept substantial diagnostic uncertainty. On the other hand, in situations where treatment options are less effective and more risky, clinicians often need a higher degree of diagnostic certainty.[2]

The second essential aspect of making clinical decisions is that the likelihood or probability that a disease is present must be determined *before* the clinician orders diagnostic tests. To avoid ambiguity, the clinician could assign a number (e.g., between 0 and 1) to the probability of disease presence instead of using a word such as "unlikely" or "possible."[3]

As shown in Figure 1, the probability of disease presence as estimated before diagnostic testing (i.e., the pretest probability) can be depicted as a point on a continuum ranging from absent (number = 0) to present (number = 1).[3] For example, pretest probability of 0.95 indicates a high degree of confidence that a disease is present, whereas a pretest probability of 0.01 indicates the clinician's belief that the disease is almost certainly absent. Positive test results (T+) increase the probability that a disease is present, and negative test results (T-) decrease that probability. The probability of a disease being present after application of a test is called "posttest probability." Tests vary with respect to their ability to influence the pretest probability of disease.[3]

BARBARA SCHEROKMAN, MD, is Assistant Chief of Neurology for the Mid-Atlantic Permanente Medical Group. Barbara has been active in teaching critical reading and lecture skills at the American Academy of Neurology. She is Clinical Associate Professor of Neurology at Georgetown University Medical Center.

Determining the Probability of Disease

To help determine the probability that a specific patient has a certain disease, clinicians rely on known prevalence of that disease in the patient population. For example, a patient seen in a medical clinic in Atlanta for fever and chills is more likely to have a urinary tract infection than malaria, which would be a much more probable diagnosis in central Africa. To determine the pretest probability of disease, prevalence can be adjusted upward or downward, depending on findings from the medical history and physical examination. For example, the probability of hyperthyroidism (e.g., 0.8) would be much higher than the probability of Wilson's disease (e.g., 0.01) when considered as a possible cause of tremor in a 25-year-old man seen for an action tremor of the hands and tachycardia, but Wilson's disease would be much more probable (e.g., 0.95) if the patient's brother had been diagnosed with Wilson's disease.

To help clinicians determine pretest probabilities, clinical guidelines ("clinical prediction rules") have been developed. These rules use signs and symptoms of disease from far more patients than could ever be seen by an individual physician. Accordingly, Billewicz et al[6] have developed a clinical prediction rule for hypothyroidism: Points are assigned to various signs and symptoms, and the pretest probability of disease is determined by adding up the points. Caution must be used in applying these rules, however, because the population from which the rule was derived may have different demographics and spectrum of disease than the population which includes the patient being seen. These rules therefore permit only rough estimation of pretest disease probability.

Major Characteristics of Tests

An ideal test could distinguish absolutely between patients who do and who do not have disease. The clinical usefulness of a test is determined by how much it deviates from this ideal. Data on test characteristics are derived from studying the test against a "gold standard" test, the test that definitively determines the presence or absence of disease. An example of a "gold standard test" would be a biopsy. Patients whom biopsy has shown to have the disease and patients shown not to have the disease are given the diagnostic test in question, after which a two-by-two table (Fig. 2) is used to compare results of biopsy and diagnostic test.

The first two elements of the comparison show how well the diagnostic test correctly identifies patients with and without the disease: *Sensitivity* describes the ability of a test to correctly detect disease; *specificity* describes the ability of a test to correctly identify absence of disease. The *false-positive rate* (cell "b" in Fig. 2) is the
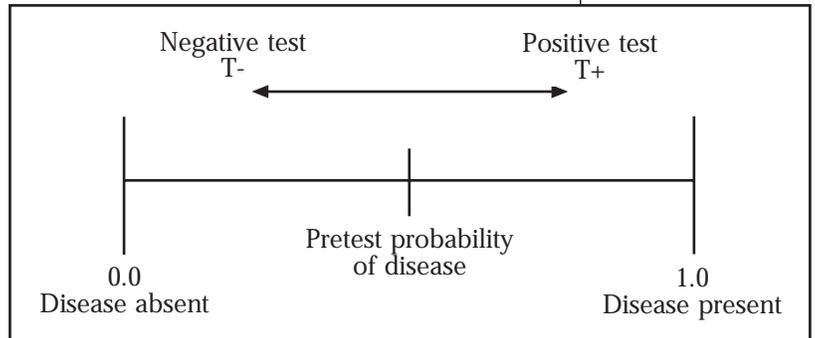

Fig 1. Pretest probability of disease

Pretest Probability %	99	80	50	20	1
Positive Predictive Value %	99.9	99	95	83	16
Negative Predictive Value %	16	83	95	99	99.9

Table 1. Relation between pretest probability, positive predictive value, and negative predictive value of medical diagnostic tests.

tendency of a test to incorrectly classify a patient as having a disease, whereas the *false-negative rate* (cell "c" in Fig. 2) is the tendency of a test to incorrectly classify a patient as not having a disease.[3]

Sensitivity and specificity are said to be "stable" properties of a test because they do not vary with pretest probability of disease. Unfortunately, these test properties are not clinically useful, because in a clinical situation the physician does not know the results of the gold standard. It is much more useful to know the probability of disease in a patient who has a positive test result (the positive predictive value) and the probability of nondisease in a patient with a negative test result (negative predictive value) (Fig. 2). For example, several weeks after a 35-year-old man from rural Virginia awoke with unilateral Bell's palsy, results of ELISA serologic test for Lyme disease were positive at 1:10. Figure 3 describes a study of 289 patients in which sensitivity, specificity, and positive and negative predictive values were derived for results of the Lyme disease serologic test.[7]

These test characteristics are usefully clinically, but unfortunately they are not stable properties (i.e., they vary with the pretest probability of disease). When testing a patient who has a low probability of having the

> *"An ideal test could distinguish absolutely between patients who do and who do not have disease. The clinical usefulness of a test is determined by how much it deviates from this ideal."*

disease, most positive test results will be proved false. In other words, as the pretest probability of disease falls, the predictive value of a positive test also falls and the predictive value of a negative test rises. As the pretest probability of disease falls, a negative test result is more informative than a positive result. Even a laboratory test with 95% sensitivity and 95% specificity loses positive predictive value and gains negative predictive value as pretest probability of disease falls (Table 1).

A test is most informative when the pretest probability of disease is between 40% and 60%. In other words, a diagnostic test is most useful and changes the pretest probability of disease if the patient is believed to have a 50:50 chance of having the disease. At this level of pretest probability, a positive test result essentially confirms the diagnosis, whereas a negative diagnostic test result essentially eliminates the disease from the differential diagnosis. This effect can be seen in Table 1, which shows that when pretest probability of disease is 50%, a positive test result raises the pretest probability to 95% and a negative result lowers the pretest probability to 5%. Thus, a test is more helpful clinically if it changes the pretest probability of disease greatly; and this occurs at the midportion of the table when the clinician is equivocal about the diagnosis.

"In other words, a diagnostic test is most useful and changes the pretest probability of disease if the patient is believed to have a 50:50 chance of having the disease."

Determining the Reliability of Tests

An index—the "likelihood ratio"[8]— has been developed to describe how reliably a diagnostic test detects disease. The likelihood ratio compares the proportions of patients with and without the disease who have been given the diagnostic test and divides the true-positive rate by the false-positive rate (Fig. 2). Thus, the likelihood ratio represents the odds that a given diagnostic test result would be expected in a patient who has the disease.

In other words, the likelihood ratio for disease if test result is positive represents the odds that a positive test result actually came from a patient with the disease. For example, in the Bell's palsy patient mentioned above (Fig. 3), a likelihood ratio of 7 assigned to positive serologic test results for Lyme disease means that a positive test result is 7 times as likely to have come from a patient with Lyme disease as without the disease. The likelihood ratio for absence of disease when the test result is negative represents the odds that a negative result actually came from a patient with the disease.

Likelihood ratios are clinically useful because they are more stable than the positive and negative predictive values and do not vary with change in disease prevalence (pretest probability). They are clinically useful also because they can be calculated for several levels of test result. A normogram (Fig. 4) has been developed for use with likelihood ratios to determine the posttest probability of disease if the pretest probability and the likelihood ratio for the specific test are known.[8] For example (using Fig. 4), if the pretest probability of

Criterion Standard

	Disease	No disease
Positive test results	a	b
Negative test results	c	d

$$\text{Sensitivity} = \frac{a}{a + c} \qquad \text{Specificity} = \frac{d}{b + d}$$

$$\text{Positive predictive value} = \frac{a}{a + b}$$

$$\text{Negative predictive value} = \frac{d}{c + d}$$

$$\text{Likelihood ratio for disease if test positive} = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

$$\text{Likelihood ratio for disease if test negative} = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

Fig 2. Algorithm used for comparing results of diagnostic test and biopsy.

	Lyme disease	No disease	
Positive serologic test results	75 (a)	25 (b)	100
Negative serologic test results	13 (c)	176 (d)	189
	88	201	289 patients

Sensitivity = 0.85   Specificity = 0.88

Positive predictive value = 0.75

Negative predictive value = 0.75

Likelihood ratio for disease if test positive = 7.08

Likelihood ratio for disease if test negative = 0.17

Fig 3. Quantification of test characteristics as derived (using algorithm in Fig 2) for results of Lyme disease seriologic test.

Lyme disease in the Bell's palsy patient is estimated to be about 2%, the clinician would anchor a ruler at 2% on the left (pretest probability of disease) scale, then rotate the ruler to align it with the likelihood ratio of a positive serologic test result of 7 for Lyme disease on the center (likelihood ratio) scale; the posttest probability, 12%, would be found by following the ruler along to the scale at right. The serologic test result for Lyme disease would not raise enough suspicion of Lyme disease in the Bell's palsy patient to treat the patient. In other words, for this patient, in whom the clinical disease probability is low, a positive result for Lyme disease is most likely to be false.

The "Threshold Model": Evaluating the Need for Tests

In clinical practice, physicians are faced with three choices: to withhold treatment, to order a diagnostic test, or to treat without testing. Pauker and Kassirer[4] have described a model which uses two thresholds to aid clinicians in making clinical decisions: 1) a "no treatment/test" threshold, which is the disease probability at which the value of withholding treatment is the same as that of performing a test; and 2) a "test/treatment" threshold, which is the disease probability at which the value of performing the test is the same as that of ad-

ministering treatment. The decision not to treat, to test, or to treat is determined by pretest disease probability and both thresholds. If the probability of disease falls within one of the end segments, testing will not prompt a different clinical action. The best clinical decision for probabilities below the "no treatment/test" threshold is to refrain from treatment; for probabilities above the "test/treatment" threshold, the best decision is to administer treatment. When the pretest disease probability lies between the thresholds, the test result could change the probability of disease enough to alter the therapeutic decision, so the best decision would be to administer a test. Tests that do not change the probability of disease enough to cross the threshold probability are not useful and should not be ordered. For example, the pretest probability of Lyme disease estimated for the Bell's palsy patient (i.e., 2%) is a smaller disease probability than the "no treatment/test" threshold so would not indicate treatment or serologic testing for Lyme disease. If probability of Lyme disease were 50%, serologic testing for Lyme disease should be done. Alternatively, if the probability of Lyme disease is 95%—a figure higher than the "test/treatment" threshold—the patient should be treated.

Conclusion

Diagnostic tests should be selected and administered in a way that allows them to influence the clinician's estimate of pretest disease probability. This estimate is the major factor in determining whether to either withhold treatment, order more tests, or treat without subjecting the patient to the risks of further testing.[4] Laboratory tests are of greatest diagnostic use to clinicians who find themselves in a "50:50 dilemma" and cannot decide whether the patient does or does not have the disease in question. ❖

"In clinical practice, physicians are faced with three choices: to withhold treatment, to order a diagnostic test, or to treat without testing."
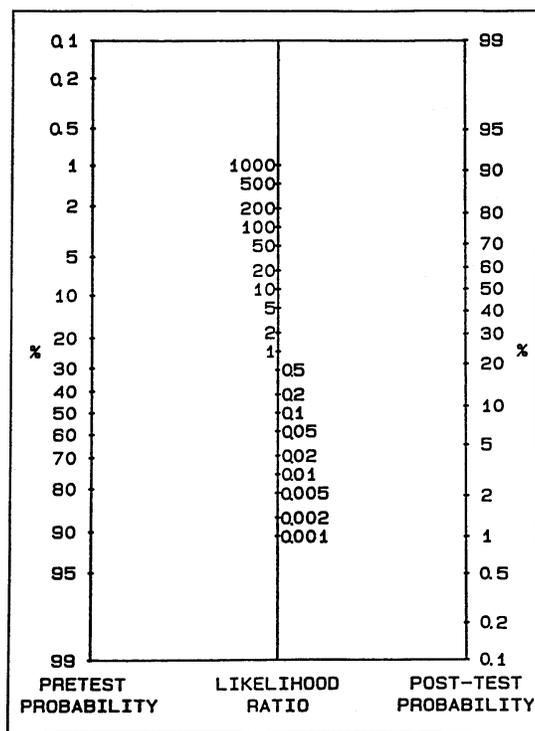

Fig. 4. Normogram[8] for use with likelihood ratio to determine pretest probability of disease if pretest probability and likelihood ratio for test are known.

References
1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd edition. Boston: Little, Brown, 1991.
2. Kassirer JP. Our stubborn quest for diagnostic certainty: a cause of excessive testing. N Engl J Med 1989;320:1489-1491.
3. Herbers JE, Noel GL. Diagnostic tests and clinical decisions. Diagn Endocrinol 1992;1-15.
4. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980;302:1109-1117.
5. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. N Engl J Med 1975;293:229-234.
6. Billewicz WZ, Chapman RS, Crooks J, Day ME, Gossage J, Wayne E, et al. Statistical methods applied to the diagnosis of hypothyroidism. Q J Med 1969;38:255-266.
7. Johnson BJB, Robbins KE, Bailey RE, Cao BL, Sviat SL, Craven RB, et al. Serodiagnosis of Lyme disease: accuracy of a two-step approach using a flagella-based ELISA and immunoblotting. J Infect Dis 1996;174:346-353.
8. Pagan TJ. Normogram for Bayes's theorem. N Eng J Med 1975;293:257.