Test Properties 2: Likelihood Ratios, Bayes' Formula, and Receiver Operating Characteristic Curves

Stuart Spitalnic, MD

As much of clinical decision-making involves testing, proper patient care depends on the proper interpretation of test results. Ideally, a positive test would always mean the presence of a condition; a negative result would always mean its absence. Because clinical tests do not work this way, it is essential to understand how test results should be interpreted. An understanding of test properties will allow you to predict, before you even order a test, how much a test might guide your clinical decisions. These principles will help you understand why one testing strategy may provide helpful information whereas another may not be worthwhile at all.

This is the second of a 2-part series reviewing test properties. Part 1, published in the September 2004 issue of *Hospital Physician*, reviewed sensitivity, specificity, and positive and negative predictive values. Part 2 builds on these concepts and discusses likelihood ratios, Bayes' theorem, and receiver operating characteristic (ROC) curves. Both articles include some sample problems for practice in applying the concepts discussed. The definitions and essential formulas from both parts are summarized in the box on page 54.

For most tests, the sensitivity and specificity are readily available. Your laboratory, a textbook, or reference material from a test's manufacturer should be adequate sources for an estimate of a test's sensitivity and specificity. The term *estimate* is important because, although sensitivity and specificity are generally considered fixed for a given test, these values should only be thought of as fixed under specific conditions. For example, the sensitivity and specificity of ultrasound will be different when used to evaluate a swollen, painful leg for deep vein thrombosis than when it is used during the work-up of pulmonary embolism to look for deep vein thrombosis in a patient whose legs are asymptomatic. For now, we will assume that the patients we are testing are similar to those in whom the sensitivity and specificity were determined. So, if a rapid strep test is reported to be 85% sensitive in school children with sore throats, you can expect that the test will be positive in 85% of your school-aged children with sore throats who have strep.

TESTING PATIENTS, NOT TESTS

The real clinical question is not how often a test will be positive in someone known to have a disease, but how well a test can discriminate those with disease from those without. Does a child with sore throat have strep? If I check a test and it is positive, does he have it? How sure am I that the child is free of disease if the result is negative?

Sadly, certainty can rarely be achieved. The best we can do through testing is to improve our level of certainty. The *pre-test probability* is a statement of the likelihood that a patient with a particular presentation has a specific diagnosis. Using the earlier example, let us say that of all school-aged children with sore throat, 30% have strep throat. When seeing an unselected child with sore throat, the pre-test probability of strep throat is 30%. The *post-test probability* is a statement of the likelihood that a patient has a diagnosis after the results of a test. For the strep throat example, let us say that if a child with a 30% pre-test probability of strep throat has a positive strep test, the probability may then become 90% that he actually has strep throat. If his test is negative, there may still be a 5% chance of strep throat. Ninety percent and 5% would be the post-test probabilities with positive and negative tests, respectively. (The post-test probabilities are not 100% or 0% because some of the positive results may be false positives and some of the negative results may be false negatives.)

Testing bring us from a pre-test probability, which is determined by the prevalence of a condition in a given population, to a post-test probability, which ideally

Dr. Spitalnic is an assistant residency director, Brown University Emergency Medicine Residency Program, and an assistant professor of medicine, Brown University, Providence, RI.

TEST PROPERTIES—BASIC DEFINITIONS AND FORMULAS

Sensitivity: The probability a test will be positive given a patient with the condition

Specificity: The probability a test will be negative given a patient without the condition

Positive predictive value (PPV): The probability a patient will have the condition given a positive test result

Negative predictive value (NPV): The probability a patient will not have the condition given a negative test result

The Generic 2 × 2 Table:

	Has condition	**Does not have condition**	
Test positive	A	B	Total positive tests (A + B)
Test negative	C	D	Total negative tests (C + D)
	Number in sample with condition (A + C)	Number in sample without condition (B + D)	Total number of subjects (A + B + C + D)

$$\text{Sensitivity} = \frac{A}{A + C}$$

$$\text{Specificity} = \frac{D}{D + B}$$

$$\text{PPV} = \frac{A}{A + B}$$

$$\text{NPV} = \frac{D}{C + D}$$

Bayes' Formula for Determining Post-Test Probability:

$$P[A|B] = \frac{P[B|A] \times P[A]}{P[B|A] \times P[A] + P[B|\bar{A}] \times P[\bar{A}]}$$

$$\text{The probability of disease given a positive test (PPV)} = \frac{\text{Sensitivity} \times \text{prevalence}}{\text{Sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}$$

$$\text{The probability of no disease given a negative test (NPV)} = \frac{\text{Specificity} \times (1 - \text{prevalence})}{\text{Specificity} \times (1 - \text{prevalence}) + (1 - \text{sensitivitiy}) \times \text{prevalence}}$$

Likelihood Ratios:

Post-test odds = LR × Pre-test odds

Odds = Probability/(1 – Probability)

Probability = Odds/(Odds + 1)

$$\text{LR+} = \frac{\text{Probability of a positive test given the presence of disease}}{\text{Probability of a positive test given the absence of disease}} = \frac{\text{Sensitivity}}{(1 - \text{Specificity})}$$

$$\text{LR–} = \frac{\text{Probability of a negative test given the presence of disease}}{\text{Probability of a negative test given the absence of disease}} = \frac{(1 - \text{Sensitivity})}{\text{Specificity}}$$

Receiver Operating Characteristic Curves:

For various cut-off points for a test, plot as follows: Y axis, sensitivity; X axis, 1 – specificity

would be either high enough to essentially clinch a diagnosis or low enough to rule it out. If a test is not capable of yielding a post-test probability much different than the pre-test probability, the test is not helpful. If the pre-test probability were near 100% (certainty) or 0% (certainly not), there would be no reason to employ any test. The best circumstance for the application of clinical tests is when the pre-test probability leaves a desire for more certainty, and the post-test probability is expected to be much closer to 100% or 0%. Bayes' formula and likelihood ratios can help us quantify the magnitude of the impact a test can have on our diagnostic certainty.

BAYES' FORMULA

In its simplest form, Bayes' formula can be expressed as:

$$P[A|B] = \frac{P[B|A] \times P[A]}{P[B|A] \times P[A] + P[B|\bar{A}] \times P[\bar{A}]}$$

It looks more complicated than it is. The notation was discussed in a previous article in this series, "Clinician's Probability Primer" (*Hosp Physician* 2003;39(2): 39–41); essentially, $P[x]$ is read as *the probability of x*, and with the vertical bar, as in $P[x|y]$, it is read as the conditional *probability of x given y*. The horizontal bar over a term means "not,"' so $P[\bar{A}]$ means *the probability of not A* (or one minus the probability of A). If we substitute "having a disease" for A, and "a positive test" for B, then $P[A]$ is the prevalence, and $P[\bar{A}]$ is 1 minus prevalence. $P[B|A]$ is the probability of a positive test given the presence of disease, which is the same as the sensitivity of the test; $P[B|\bar{A}]$ is the probability of a positive test given no disease, which is the same as 1 minus specificity. So, making all the substitutions and translations, Bayes' formula is as follows:

$$\text{The probability of disease given a positive test} = \frac{\text{Sensitivity} \times \text{prevalence}}{\text{Sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}$$

In last month's article on test properties ("Test Properties 1: Sensitivity, Specificity, and Predictive Values," *Hosp Physician* 2004;40(9):27–31), the following problem was posed:

> A patient comes to your office frantic over the results of a home HIV test. The test touts 99% sensitivity and 99% specificity. On questioning, you determine that this patient is at low risk for HIV; given your assessment of his risk factors, you believe he comes from a population group that has a baseline prevalence of HIV of 1 in 100,000. He now presents to you with a positive result on his home HIV test. Given his baseline risk and the positive home test, what are the chances that this patient is actually HIV positive?

We solved this problem by constructing a 2 × 2 table. Although it seems more cumbersome to use Bayes' formula, once each term is understood, it is simple substitution and calculation. We are interested in the probability of HIV given a positive test. The probability of a positive test in a patient with HIV (ie, sensitivity) is 99% (0.99), the prevalence is 0.00001, the probability of a positive test in a patient without HIV is 1% (1 – specificity) or 0.01, and the pre-test probability of not having HIV (1 – prevalence) is 0.99999. The equation then becomes:

$$\frac{0.99 \times 0.00001}{0.99 \times 0.00001 + 0.01 \times 0.99999} = \frac{0.0000099}{(0.0000099 + 0.0099999)} = 0.00099$$

or, the probability of having HIV given a positive test is now about 1/1000.

In this example, after a positive test, the risk of having the disease is still only 1 in 1000. Even with a positive test, it is still a good bet that the patient is disease free. On the other hand, the probability is 100 times more likely than it was before the test, so the test accomplished *something*. Consider a similar example, but this time the prevalence of disease will be 50%. Say the test is 90% sensitive and 80% specific. Using Bayes' formula, what is the probability of disease if you get a positive test result? (Question 1, answer at end)

It will become apparent as you do more of these types of problems that unless a test's sensitivity and specificity are 100%, the prevalence of disease will determine how much more certain of the presence or absence of disease you may become after a test. The less prevalent a condition, the more likely a positive test is a false positive; the more prevalent a condition, the more likely a negative result is a false negative.

Sensitivity indicates how often a test will be positive in the setting of disease, the *positive predictive value* indicates how often the disease is really present after a positive test. *Specificity* indicates how often a test will be negative in those without a disease, *negative predictive value* indicates how often a disease is absent in someone who tests negative. Post-test probabilities are essentially the positive and negative predictive values and are strongly tied to disease prevalence. How then can we measure the "resolving power" of a test independent of the

disease prevalence, and then compare one test to another? One way is by looking at its likelihood ratios.

LIKELIHOOD RATIO

Recognizing that some positive test results are falsely positive and that some negative test results are falsely negative, a ratio can be made between true positives and false positives; this ratio is called the *likelihood ratio of a positive test* (LR+):

$$\text{LR+} = \frac{\text{Probability of a positive test given the presence of disease}}{\text{Probability of a positive test given the absence of disease}}$$

Remembering the definitions of sensitivity and specificity, substitutions can be made as follows:

$$\text{LR+} = \frac{\text{Sensitivity}}{(1 - \text{Specificity})}$$

For a negative test, the likelihood ratio (LR–) is the ratio of false negatives to true negatives:

$$\text{LR–} = \frac{\text{Probability of a negative test given the presence of disease}}{\text{Probability of a negative test given the absence of disease}}$$

And substituting again:

$$\text{LR–} = \frac{(1 - \text{Sensitivity})}{\text{Specificity}}$$

A likelihood ratio greater than 1 suggests that a test result makes a condition more likely to be actually present; a likelihood ratio less than 1 suggests the test result makes a condition less likely. LR+ values should be greater than 1, LR– values should be positive fractions between 0 and 1. If the LR+ or LR– is equal to 1, that implies that the result is just as likely in those with and without a condition, and therefore, the result adds no information. (Note that if the LR+ was less than 1 and the LR– was greater than 1, the definitions of a positive and negative test result would be reversed.)

The further away from 1 that a LR+ or LR– is, the greater the test's resolving power. A test with a LR+ of 100 will, in the setting of a positive test result, make you more certain of the presence of disease than a positive result in a test that has a LR+ of 10; a test with a LR– of 0.01 will more confidently exclude a condition with a negative result than a test with a LR– of 0.8. Likelihood ratios can be used to compare the resolving power of different tests and to help determine which tests will provide the most information. A benefit of using a test's likelihood ratio over its predictive values is that, unlike predictive values, the likelihood ratio depends only on sensitivity and specificity and not on disease prevalence.

USING LIKELIHOOD RATIOS

In addition to using likelihood ratios to determine how much more information one test is likely to provide over another, likelihood ratios can also be used to calculate post-test probabilities. There are nomograms available to get you from pre-test probability to post-test probability when a test's likelihood ratio is known, but they can be a bit cumbersome. The likelihood ratios relate the pre- and post-test odds as follows:

$$\text{Post-test odds} = \text{LR} \times \text{Pre-test odds}$$

Using this clinically requires converting pre-test probability to odds, performing the calculation, and then converting odds back to probability. Again, it is less complicated than it seems at first. Remember that odds, for any event with a probability of occurring *P*, can be found by:

$$\text{Odds} = \frac{P}{(1 - P)}$$

If you are evaluating a patient with a 20% chance of having a condition, the pre-test probability is 0.20, and the pre-test odds would be 0.2/0.8, or 0.25. Now, if you have a test with a LR+ of 3 and the result is positive, the post-test odds would be equal to the likelihood ratio multiplied by the pre-test odds, or 3 × 0.25 or 0.75. Finally, to convert back to probability,

$$\text{Probability} = \frac{\text{Odds}}{(1 + \text{Odds})}$$

Now we are calculating the post-test probability, so we use the post-test odds, as follows:

$$\text{Post-test probability} = \frac{\text{Post-test odds}}{(1 + \text{Post-test odds})}$$

In the present example, this would be 0.75/1.75 = 0.43.

Consider the following example: You have determined that the patient is at moderate risk for a particular condition and estimate the pre-test probability to be 60%. You do a test that has a reported LR+ of 5. The

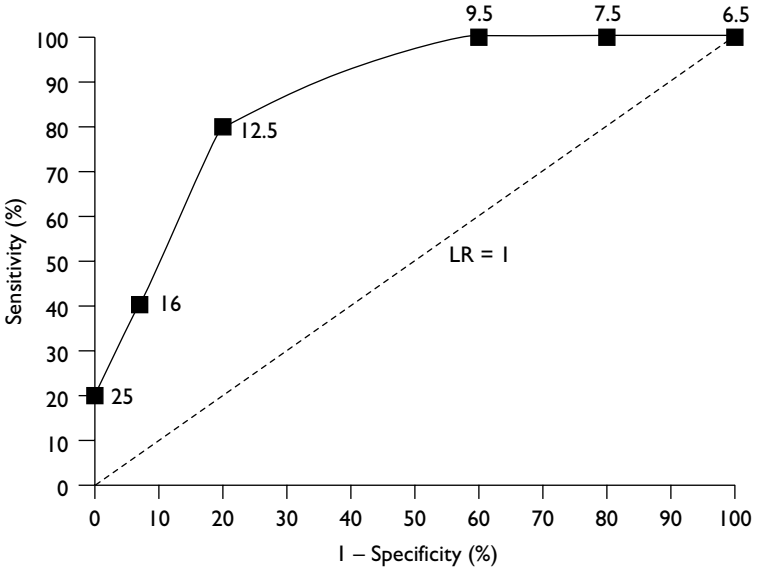

Figure. Receiver operating characteristic curve. Plotted points represent sensitivity and 1 – specificity of indicated leukocyte count cut-off value.

patient's result is positive. The question is, after the positive test result, how much more certain are you that the patient has the condition? That is, what is the post-test probability? If the LR– is 0.6, what is the post-test probability after a negative test? (Question 2, answer at end)

RECEIVER OPERATING CHARACTERISTIC CURVES

A troponin test for myocardial infarction may be considered positive if the result is greater than 1 ng/mL. A D-dimer test for thromboembolism may be considered positive if the result is greater than 0.5 μg/mL. How are these numbers determined? Might other cut-off values be better? One way of visualizing the performance of a test while varying the cut-off values for a positive test is to look at its ROC curve.

Recognize that you could choose a very low cut-off for any test and have 100% sensitivity—the specificity will be 0% if *all* your test results are positive, but everyone with disease will be identified. Similarly, with a very high cut-off you will never identify anyone with a condition (sensitivity = 0%), but you will have no false positives (specificity = 100%). As you adjust a cut-off value, you will always sacrifice some sensitivity or specificity to get improvement in the other. The goal is to find the cut-off point that excludes the most people who do not have a condition without missing an unacceptable number of people who do.

To construct an ROC curve, a graph is made with sensitivity on the Y axis, and 1 minus specificity on the X axis. The curve is created by plotting a test's sensitivity against 1 – specificity while varying the cut-off value used to define a positive test. In last month's article, an example was presented using leukocyte counts as a screen for positive blood cultures. The 15 patients with negative blood cultures and the 5 with positive blood cultures had the following leukocyte counts ($\times 10^3/mm^3$):

(–) Blood Cultures	(+) Blood Cultures
7	11.1
7.2	14.6
7.2	15.5
8.6	18.3
9	26.7
9.1	
10.1	
10.2	
11	
11.1	
12.1	
12.2	
14.7	
15.1	
17	

Let us consider some arbitrary cut-off points for these data and see how the sensitivity and specificity of the leukocyte count vary.

Cut-off Point	Sensitivity (%)	Specificity (%)
6.5	100	0
7.5	100	20
9.5	100	40
12.5	80	80
16	40	93
25	20	100

Note that as the cut-off point for a positive test is increased, the specificity improves (fewer false positives)

but the sensitivity decreases (more false negatives). This phenomenon is true for anything short of a perfect test. Plotting an ROC curve for these cut-off points results in the graph shown (**Figure**).

An ideal test would have a point in the upper left corner of the graph—this point represents a sensitivity and specificity of 100%. If you start at the upper right corner of the curve and trace to the left, the point where the line starts to go down is the point at which the increased specificity is resulting in a meaningful decreases in sensitivity. Similarly, if you start tracing up from the lower left corner, when you pull away from the vertical line (1 – specificity = 0), you are getting noticeable decreases in your specificity as you increase sensitivity. A test that adds no information regarding the likelihood of the presence or absence of a condition (ie, LR+ and LR– both equal to 1) would have an ROC curve that follows the dotted line.

By looking at the characteristics of the curve, you get a sense of what the best cut-off value would be for a given test. The cut-off value that is actually chosen should represent the point where the balance between sensitivity and specificity makes the most sense for the particular condition. When comparing 2 tests that are testing similar conditions, the "better" test will be the one with the curve that comes closest to the upper left corner—this can be more formally represented by measuring and comparing the area under the ROC curve.

SUMMARY

Bayes' formula is a useful way to calculate the post-test probability of a condition when the prevalence and the sensitivity and specificity of the test are known. The post-test probability is strongly tied to the prevalence of disease in a population, and this relationship should always be considered when deciding how much additional information a clinical test may provide.

Likelihood ratios are useful for comparing different tests' ability to resolve those with disease from those without. The further a test's likelihood ratio is from 1, the greater the impact the test will have on the post-test probability.

Receiver operating characteristic curves are used to graphically represent how changes in cut-off values for a test impact the resulting sensitivity and specificity. ROC curves are often encountered in research papers that evaluate a new application of a clinical test.

ANSWERS TO QUESTIONS IN TEXT

Question 1

From Bayes' formula,

$$\text{The probability of disease given a positive test} = \frac{\text{Sensitivity} \times \text{prevalence}}{\text{Sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}$$

Prevalence = 0.5, sensitivity = 0.9, specificity = 0.8

$0.5 \times 0.9 / (0.5 \times 0.9 + 0.2 \times 0.5) = 0.82$

Question 2

Post-test odds = LR × Pre-test odds

Pre-test probability = 0.6, pre-test odds = 1.5

LR+ = 5; after positive test, post-test odds = $5 \times 1.5 = 7.5$;
Post-test probability = $7.5/(7.5 + 1) = 7.5/8.5 = 0.88$

LR– = 0.6; after negative test, post-test odds =
$0.6 \times 1.5 = 0.9$; Post-test probability =
$0.9/(0.9 + 1) = 0.9/1.9 = 0.47$ **HP**

SUGGESTED READING

Guyatt G, Rennie D. Users guide to the medical literature. Chicago: American Medical Association Press; 2002.

Lee TH. Interpretation of data for clinical decisions. In: Goldman L, Bennett JC. Cecil Textbook of medicine. 21st ed. Philadelphia: WB Saunders; 2000:79–84.

Riegelman RK, Hirsch RP. Studying a study and testing a test. 3rd ed. Boston: Little, Brown and Company; 1996.

EDITOR'S NOTE

Articles previously published in the *Primer in Literature Interpretation* series can be downloaded at our web site (www.turner-white.com).