A Graphical Aid to Medical Decision Making

William A. Benish
Given knowledge of a test's sensitivity and specificity, physicians may use Bayes' theorem to appropriately modify their initial assessment of the likelihood of disease subsequent to obtaining a positive or negative test result. A graphical representation of Bayes' theorem was constructed in order to provide a simple tool to aid in the selection and interpretation of diagnostic tests.

Suppose that after completing a history and physical examination on a patient with chest pain you conclude that there is about a 25% chance that the patient has coronary artery disease. How would you modify this estimate once you learned that the patient had a positive exercise tolerance test? Your answer should depend upon your knowledge of the specificity and sensitivity of the stress test findings in diagnosing coronary artery disease. Your revised estimate will also depend upon the method you use to combine this information with your initial clinical impression (1).

The eighteenth century English clergyman Thomas Bayes has provided us with a rational way of modifying our initial probability (Pi) that a specific disease is present, subsequent to obtaining an abnormal test result (2). The revised probability (Pr) is related to Pi, test sensitivity (SENS), and test specificity (SPEC) by the following expression:

\[
Pr = \frac{(SENS)(Pi)}{(SENS)(Pi) + (100\%-SPEC)(1-Pi)}
\]

To use this formula we must recall that probabilities range from zero to one. An impossible event has zero probability. A certain event corresponds to a probability equal to one. The "25% chance" in the above example translates into Pi = 0.25. It is impossible to practice clinical medicine without some aptitude for estimating probabilities. Clinicians make hundreds of decisions each day that require them to estimate the likelihood of various diagnoses, the utility of diagnostic tests, and the risks and benefits of therapy.

Test sensitivity and specificity define the "operating characteristics" of a test. They are commonly expressed as percents. "Test sensitivity" is defined as the likelihood that a patient will have an abnormal (positive) result, given that he has a certain disease. "Test specificity," on the other hand, offers high specificity but low sensitivity in evaluating possible giardiasis. A procedure with both high sensitivity and high specificity is illustrated by bone marrow biopsy in cases of iron deficiency.

Returning to the above example, if we accept values of 80% and 74%, respectively, for the sensitivity and specificity of exercise testing in diagnosing coronary artery disease, then Pr is calculated to be 0.51 (3). In this case, exercise testing was not helpful in making or excluding the diagnosis of coronary artery disease. Had this calculation been made prior to ordering the stress test, the test might never have been ordered.

As an alternative to performing the above calculation, the Figure provides a rapid graphical method of calculating Pr, given Pi, SPEC, and SENS. It also provides a last refuge for those physicians who want to practice good medicine without walking around with a calculator on their belts. The first step is to locate Pi on the vertical axis. Next, locate the point in the square entitled "specificity" that is directly to the right of Pi and lies on the curve labeled with SPEC. Now locate the point in the square entitled "sensitivity" that is directly below this last point and that lies on the curve labeled with SENS. Pr can be read off the vertical axis by finding the value directly to the right of this last point. Interpolation is required if SPEC or SENS do not lie on one of the constructed curves.

In the present example, Pr is also referred to as the "positive predictive value" of the test. It is the probability of disease, given a positive test result. The Figure can also be used to determine the "negative predictive value"; that is, it can be used to calculate the probability that the disease is absent, given a negative test result. This is accomplished by letting Pi represent our initial probability that disease is absent and then by repeating the above procedure with the exception that the values for sensitivity and specificity are exchanged. For example, if the disease has an observed frequency of 50%; that is, the disease is present in 50% of an elevated serum creatinine, then the respective positive and negative predictive values are found by construction to be the positive and negative predictive values of the disease

Submitted for publication: February 25, 1985
Accepted for publication: March 26, 1985
*Department of Internal Medicine, Henry Ford Hospital
Address reprint requests to Dr Benish, Department of Internal Medicine, Henry Ford Hospital, 2799 W Grand Blvd, Detroit, MI 48202.
Medical Decision Making

for sensitivity and specificity are interchanged. In the case of the patient with chest pain: Pi = 0.75, SPEC = 80%, and SENS = 74%. The negative predictive value is found to be about 0.92.

As a second example, consider the value of measuring the acid phosphatase level in a patient with a prostatic nodule. The likelihood of prostatic carcinoma in an asymptomatic patient with a prostatic nodule is about 50%; that is, Pi = 0.5 (4). The sensitivity and specificity of an elevation in acid phosphatase in identifying prostate cancer have been judged to be 70% and 94%, respectively (5,6). Referring to the Figure, we find the positive predictive value to be about 0.92 and the negative predictive value to be about 0.76. In this setting, serum acid phosphatase values provide a substantial amount of information but probably not enough information to circumvent the need for biopsy. On the other hand, if the initial biopsy is negative, the acid phosphatase level might dictate whether or not a repeat biopsy need be performed (5,6).

It should be clear that a Bayesian approach cannot transform medical decision making into a "cookbook" discipline. Determination of Pi, SENS, and SPEC all require clinical judgment. At the present time, information pertaining to test sensitivity and specificity is frequently difficult to obtain or simply does not exist. Even when these values are available, they should not be accepted uncritically. Most medical tests yield a result that lies on a continuum, and consequently, sensitivity and specificity are influenced by the criteria selected to define normal and abnormal results. Moreover, a given patient is not likely to be a part of the same "normal" or "diseased" population used to define these parameters (7). For example, a textbook value for the specificity of a reactive VDRL in identifying syphilis is of little help in a patient with systemic lupus. Another difficulty lies in the fact that the specificity and sensitivity of a study are often dependent upon the individual performing the test or interpreting the results. Finally, tests are frequently improved or superseded by better studies before their operating characteristics are well defined.

Recognizing that our initial clinical impression as well as our estimates of a test's operating characteristics are largely subjective items of information, our revised (post-test) estimate of the likelihood of disease will further depend upon the method we select to integrate this subjective information. Two options are available for this integrative process: intuition and reason. For those who prefer the latter alternative, it is hoped that this graphical representation of Bayes’ theorem will provide a useful addition to your lab coat pockets.

Figure
Graphic expression of Bayes’ theorem.

As a second example, consider the value of measuring the acid phosphatase level in a patient with a prostatic nodule. The likelihood of prostatic carcinoma in an asymptomatic patient with a prostatic nodule is about 50%; that is, Pi = 0.5 (4). The sensitivity and specificity of an elevation in acid phosphatase in identifying prostate cancer have been judged to be 70% and 94%, respectively (5,6). Referring to the Figure, we find the positive predictive value to be about 0.92 and the negative predictive value to be about 0.76. In this setting, serum acid phosphatase values provide a substantial amount of information but probably not enough information to circumvent the need for biopsy. On the other hand, if the initial biopsy is negative, the acid phosphatase level might dictate whether or not a repeat biopsy need be performed (5,6).

It should be clear that a Bayesian approach cannot transform medical decision making into a “cookbook” discipline. Determination of Pi, SENS, and SPEC all require clinical judgment. At the present time, information pertaining to test sensitivity and specificity is frequently difficult to obtain or simply does not exist. Even when these values are available, they should not be accepted uncritically. Most medical tests yield a result that lies on a continuum, and consequently, sensitivity and specificity are influenced by the criteria selected to define normal and abnormal results. Moreover, a given patient is not likely to be a part of the same “normal” or “diseased” population used to define these parameters (7). For example, a textbook value for the specificity of a reactive VDRL in identifying syphilis is of little help in a patient with systemic lupus. Another difficulty lies in the fact that the specificity and sensitivity of a study are often dependent upon the individual performing the test or interpreting the results. Finally, tests are frequently improved or superseded by better studies before their operating characteristics are well defined.

Recognizing that our initial clinical impression as well as our estimates of a test’s operating characteristics are largely subjective items of information, our revised (post-test) estimate of the likelihood of disease will further depend upon the method we select to integrate this subjective information. Two options are available for this integrative process: intuition and reason. For those who prefer the latter alternative, it is hoped that this graphical representation of Bayes’ theorem will provide a useful addition to your lab coat pockets.

References