Research and Statistics : Sensitivity, Specificity, Predictive Values, and Likelihood Ratios

Diana N. Carvajal and Peter C. Rowe

The online version of this article, along with updated information and services, is located on the World Wide Web at:
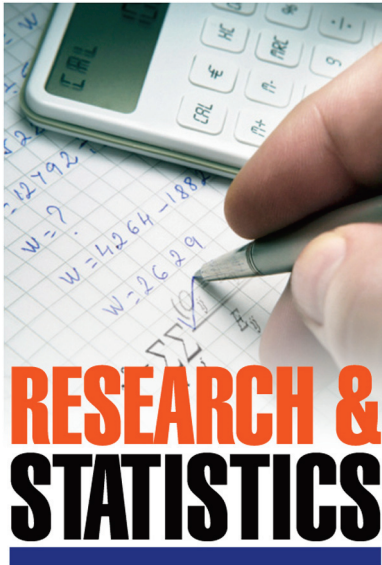http://pedsinreview.aappublications.org/content/31/12/511

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN™

Sensitivity, Specificity, Predictive Values, and Likelihood Ratios

Diana N. Carvajal, MD,* Peter C. Rowe, MD*

Case Presentation

You are seeing a previously healthy 5-year-old girl in your office for a complaint of sore throat and fever. Her mother explains that the child does not have a cough, but she has had a decreased appetite and unusual fatigue over the past 1 to 2 days. Physical examination reveals a slightly ill-appearing child who has an erythematous oropharynx, mild cervical adenopathy, and a temperature of 37.5°C. A white blood cell (WBC) count from the previous evening was $9.2 \times 10^3/$mcL $(9.2 \times 10^9/L)$. According to the mother, her daughter had similar symptoms last year, but she is sure it was not a streptococcal throat infection, and the symptoms resolved without treatment. She asks you whether the current presentation is a streptococcal throat infection that requires antibiotic treatment. Based on the history and physical examination findings, you decide on the basis of your experience that there is about a 30% chance of bacterial pharyngitis. You recall an article discussing the effectiveness of the Breese clinical scoring system for the diagnosis of group A beta-hemolytic Streptococcus (GABHS) pharyngitis. (1) You consider whether the results of this study are valid and relevant to your current patient. (2)

Introduction

For clinicians, it is important to understand and interpret the effectiveness of certain diagnostic tests, especially when such great advancements are being made in medical technology. The usefulness of a test can be determined by how accurately it identifies the target disorder. (3) It is important to know about the properties of individual tests and how each test compares with the gold standard (reference test). (2)(3)

Pretest and Posttest Probability

The pretest probability of a given condition varies by physician experience, season, geography, and the history and physical findings. The pretest probability is the clinician's best estimate of the probability of a specific disease before diagnostic testing and generally has a large impact on the diagnostic process. (3) In any clinical scenario, the diagnostic test serves to modify the pretest probability, which subsequently results in a new posttest probability. The direction and magnitude of this change are determined by the test's properties (eg, sensitivity, specificity, positive/negative predictive values, and likelihood ratios). (3)

Sensitivity and Specificity

Sensitivity and specificity are the two specific elements related to the concept of validity. The validity of a test is defined as its ability to discern between patients who have a certain condition and those who do not. (4) The sensitivity of a test is its ability to recognize correctly persons who *have* a disease or condition. (4) In other words, the sensitivity of a diagnostic test refers to the proportion of patients who have a disorder in whom the results of the test are positive. (3) Sensitivity is calculated by dividing the number of persons who have pos-

*Department of Pediatrics, Division of General and Adolescent Medicine, Johns Hopkins University School of Medicine, Baltimore, Md.

itive test results by the number of persons who have the actual disorder or disease based on the gold standard (Fig. 1). (3)

The specificity of a test is the ability of a test to recognize correctly persons who *do not have* a disease or condition. (4) It is the proportion of patients who do not have a disorder in whom the test result is negative. (3) Specificity is calculated by dividing the number of persons who have negative test results by the number of persons who do not have the disorder or disease (Fig. 1). (4) Calculation of sensitivity and specificity requires knowledge of which patients truly have the condition. Therefore, there must be a gold standard test that provides the true disease status of the patient. (4)

Positive Predictive Value and Negative Predictive Value

Positive predictive value (PPV) is the proportion of patients testing positive who actually have the disease or condition in question. (4) PPV is calculated by dividing the true-positive results by the total number of persons who have positive test results (Fig. 1). Negative predictive value (NPV) is the proportion of patients testing negative who actually *do not* have the condition in question. (4) NPV is calculated by dividing the true-negative results by the total number of persons who have negative test results (Fig. 1). The relationship between predictive value and disease prevalence is important. Although sensitivity and specificity are properties intrinsic to a test and are not affected by the prevalence of a particular disease or condition, the predictive values of a diagnostic test are influenced greatly by prevalence. The higher the disease prevalence, the higher the PPV. (4) This relationship means that knowing the predictive value of a test is most useful and efficient in populations in which the prevalence of a disease is high (high-risk populations). (3)

In the previously mentioned study, the authors attempted to determine the effectiveness of the Breese clinical scoring system for the diagnosis of GABHS pharyngitis by comparing its ability to diagnose the condition accurately against throat-swab cultures (reference standard). The Breese score is calculated on the basis of nine items (month of the year, patient's age, WBC count, fever, sore throat, cough, headache, abnormal pharynx, and cervical lymphadenopathy) and is used for a tentative diagnosis of GABHS pharyngitis. In this study, 416 children had the infection (positive throat cultures) and 441 children did not have the infection (negative cultures). With the Breese scoring system, a score of more than 28 was regarded as a positive test and a score of 28 or less was considered a negative test. The results showed that 286 of the 416 children who had positive throat cultures had positive tests, and 364 of the 441 children who had negative throat cultures had negative tests. Using the previously given formulas (Fig. 1), the sensitivity of the Breese scoring system is: $286/416=0.688$ ($\times 100$)=69% sensitive. The specificity of this test is $364/441=0.825$ ($\times 100$)=83% specific. The PPV of the test is 79% (286/363) and the NPV is 74% (364/494) (Fig. 2).

Although the sensitivity of the test is only modest, the specificity*, PPV, and NPV of the Breese scoring system are relatively high. Of note, PPV and NPV are heavily dependent on the disease prevalence, and the prevalence of GABHS pharyngitis in this study's population is (416/ 857)=48%, which is relatively high. For a rare condition or one that has a much lower prevalence in the population, predictive values are lower and less useful. Further, if the preva-

	Disorder	No Disorder
Positive Test Result	True Positive (TP)	False Positive (FP)
Negative Test Result	False Negative (FN)	True Negative (TN)

Sensitivity = TP/(TP+FN)
Specificity = TN/(TN+FP)
PPV = TP/(TP+FP)
NPV = TN/(FN+TN)

Figure 1. Calculation of sensitivity, specificity, and positive and negative predictive value.

*Satisfactory values for sensitivity and specificity vary, depending on the clinical situation. Serious or life-threatening illnesses require a high sensitivity. Alternatively, for conditions in which labeling a patient as falsely positive may cause a high degree of distress, a higher specificity is preferable. (4)

	Positive Throat Culture	Negative Throat Culture	Total
Breese >28	286	77	363
Breese ≤28	130	364	494
Total	416	441	857

Figure 2. Computation of sensitivity and specificity of the Breese test.

lence of a disease is greater in the sample population than in the target clinical population, the predictive values are overestimated.

Likelihood Ratios

Sensitivity, specificity, and predictive values are important properties of a test, but they have limitations. An alternative approach to the evaluation of a diagnostic test is the examination of the test's likelihood ratio (LR), which is defined as the likelihood that a person who has a target disorder will have a positive test result. (3) The LR indicates by how much the results of a given diagnostic test result will raise (or lower) the pretest probability of the target disorder, effectively yielding a new posttest probability. (3) The pretest probability is patient-specific and exerts a major influence on the diagnostic process, and LRs can be applied to individual patients who have distinct pretest probabilities. Apply-

ing the concept of LR to the Breese score discussion requires the question: How likely is a Breese score of more than 28 to occur in children who actually have GABHS pharyngitis? According to Figure 2, (286/416)=0.688 (sensitivity). Alternatively, the likelihood of a Breese score of more than 28 in children who, although suspected of it, actually do not have GABHS is: (77/441)= 0.175. The ratio of these two likelihoods is called the LR and for a Breese score more than 28, it is (0.688/0.175)=3.93. In other words, a score of more than 28 is about four times more likely to occur in a patient who has GABHS pharyngitis than a patient who does not have GABHS pharyngitis. (3) Use of a simple nomogram initially proposed by Fagan allows easy conversion from pretest to posttest probability using the LR. (3) More information on LRs is presented in a separate article in this series.

Case Discussion

Using clinical and laboratory data, you compute a Breese score of 30 for your 5-year-old patient. After calculating the LR and using the Fagan nomogram, you notice that the probability of GABHS pharyngitis with a positive Breese score has risen from a 30% pretest probability to a 60% posttest probability. You decide that the probability of GABHS is now high enough for you to obtain a throat culture and treat the child with penicillin in the interim.

Conclusion

Sensitivity, specificity, and predictive values are important properties of diagnostic tests, although each has its limitations. LRs are useful for converting from pretest to posttest probability. Articles describing diagnostic tests should report sensitivity, specificity, predictive values, and LRs or provide the reader with the data to calculate them.

References

1. Karacan K, Karakelleoğlu C, Orbak Z. Diagnosis of group A β-hemolytic *Streptococcus* using the Breese Clinical Scoring System. *South Med J.* 2007;100:1192–1197

2. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA.* 1994;271:389–391

3. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA.* 1994;271:703–707

4. Gordis L. *Epidemiology.* 4th ed. Philadelphia, Pa: Saunders Elsevier; 2009

Research and Statistics : Sensitivity, Specificity, Predictive Values, and Likelihood Ratios

Diana N. Carvajal and Peter C. Rowe

Updated Information & Services	including high resolution figures, can be found at: http://pedsinreview.aappublications.org/content/31/12/511
References	This article cites 3 articles, 2 of which you can access for free at: http://pedsinreview.aappublications.org/content/31/12/511#BIBL
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: /site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: /site/misc/reprints.xhtml

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™