

Spectrum Bias and Clinical Decision Making

Stuart Spitalnic, MD

Clinical decision making often includes the use of tests to either confirm or exclude conditions from a differential diagnosis. Generally, if a test of adequate specificity is positive in a patient likely to have the condition being tested for, the diagnosis is essentially confirmed. By the same token, if a test of adequate sensitivity is negative in a patient with a low likelihood of having the condition being tested for, the diagnosis is essentially excluded. Unfortunately, the simplicity of the above rules is disturbed by the concept of *spectrum bias*. Spectrum bias results from the fact that a test's sensitivity and specificity are not fixed values but rather vary with the severity or temporal stage of the disease being considered. Failure to take spectrum bias into account might result in using test results to erroneously confirm or exclude a diagnosis.

This article will discuss the concept of spectrum bias, using several hypothetical clinical scenarios. The specific values in the examples are for illustrative purposes only—the goal is to illustrate the concept of spectrum bias without allowing the precise values or complexity of the mathematics to be distracting.

CLINICAL SCENARIO

A patient comes to the emergency department of your hospital reporting a severe headache; neurologic examination reveals no abnormalities. Desiring to practice evidence-based medicine, you consult the medical literature to determine an appropriate strategy for excluding the diagnosis of subarachnoid hemorrhage. The classic teaching is that a computed tomography (CT) scan of the head is inadequate to exclude subarachnoid hemorrhage and that a lumbar puncture is needed for proper exclusion. A search of the literature, however, reveals the following information: (1) The probability of a subarachnoid hemorrhage in a patient with a new, severe headache but normal results on neurologic examination is 1%; (2) the sensitivity of a CT scan for detecting subarachnoid hemorrhage is 95%; and (3) the specificity of a CT scan for detecting subarachnoid hemorrhage is 99%.

You obtain a head CT scan of the patient, and results show no abnormalities. You are left with the following question: In this patient with a new severe head-

ache but normal results on neurologic examination and CT scan, what is the probability of a subarachnoid hemorrhage (ie, what is the posttest probability)?

CONDITIONAL PROBABILITY

You consult the general formula for conditional probability (Bayes' formula), which is listed below.

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

This equation states that the probability of A given B equals the probability of B given A multiplied by the probability of A, divided by the sum of the probability of B given A multiplied by the probability of A and the probability of B given not A multiplied by the probability of not A. The reason for consulting this formula becomes clear after assignments are made for A and B. If A equals a "subarachnoid hemorrhage" and B equals "negative results on a CT scan," the equation will provide the probability of subarachnoid hemorrhage given negative results on a CT scan—the exact information you are seeking. In probability notation, a bar over a term indicates "not" or "the opposite"; in this case, the A with a bar over it means "the probability of not having a subarachnoid hemorrhage." The notation $P[B|A]$ becomes "the probability of negative CT results given a subarachnoid hemorrhage" (1 minus the sensitivity, or the false-negative rate), $P[A]$ becomes "the probability of a subarachnoid hemorrhage" (ie, the prevalence), $P[B|\bar{A}]$ becomes "the probability of negative results on a CT scan given the absence of a subarachnoid hemorrhage" (ie, the specificity), and $P[\bar{A}]$ becomes "the probability of no subarachnoid hemorrhage" (1 minus the prevalence). Thus, the probability of a subarachnoid hemorrhage given negative results on a CT scan can be determined by the following calculation:

Dr. Spitalnic is an Assistant Residency Director, Brown University Emergency Medicine Residency Program, and an Assistant Professor of Medicine, Brown University, Providence, RI; he is also a member of the Hospital Physician Editorial Board.

$$\frac{(1 - \text{sensitivity}) \times (\text{prevalence})}{(1 - \text{sensitivity}) \times (\text{prevalence}) + (\text{specificity}) \times (1 - \text{prevalence})} = \frac{.05 \times .01}{.05 \times .01 + .99 \times .99} = .0005 = .05\%$$

According to this calculation, if a patient has negative results on a CT scan, the probability of there being a subarachnoid hemorrhage is only 1 in 2000, which is considerably lower than the pretest probability of 1 in 100. Incorporating this newly acquired information into your clinical practice, you decide not to perform a lumbar puncture on the patient, because a negative result on a CT scan nearly excludes a subarachnoid hemorrhage from the differential diagnosis. Furthermore, you stop performing lumbar punctures on all patients with severe headache who have similarly negative results on CT scan. You later learn that several of these patients went on to develop subarachnoid hemorrhage. Reviewing your files, you discover that you evaluated 500 people for severe headache, 5 of whom eventually had a subarachnoid hemorrhage (or 1%, as would have been predicted by the prevalence); the diagnosis was missed on all 5 of their CT scans. What went wrong?

THE CONCEPT OF SPECTRUM BIAS

Be assured, especially if you took the time to commit Bayes' formula to memory, that the formula and the logic of the arguments above are all correct. What is missing from the previous determinations is a consideration of the possibility of spectrum bias. As previously stated, spectrum bias implies that the sensitivity and specificity of diagnostic tests vary with the severity or duration of a given medical condition. For example, you can determine the sensitivity of measuring the creatine kinase (myocardial bound) (CKMB) level for detection of myocardial infarction in a patient in your emergency department by calculating the percentage of patients with elevated CKMB levels in whom myocardial infarction has been confirmed. Standard procedure for ruling out myocardial infarction, however, requires the measurement of serial CKMB values over time, because CKMB has different sensitivities at different times. Very early in a myocardial infarction, measurement of CKMB is an insensitive measure but, after several hours, becomes very sensitive. Although a single sensitivity for measurement of CKMB in myocardial infarction can be quoted, it is actually more accurate to express this sensitivity over the spectrum of time.

There are many examples of how spectrum bias affects the performance of a test at different stages of an illness. Consider the use of the leukocyte count in the evaluation of appendicitis. An elevated leukocyte

count is reasonably expected in most patients who have appendicitis; unfortunately, this general assumption does not consider the spectrum of the disease. Although it may be reasonable to expect an elevated leukocyte count in a patient with perforating appendicitis who is developing peritonitis, you might not expect an elevated count in someone experiencing the first twinges of pain associated with an obstructed appendix. Both patients have appendicitis, but they are at opposite ends of the spectrum. Although you can find in the literature a specific percentage listed as the sensitivity of the leukocyte count for detecting appendicitis, you rarely find it reported as different sensitivities for early and late presentations.

The concept of spectrum bias is not limited to numerical values. Consider 2 patients with pneumonia, one who is hypoxic, febrile, and producing purulent sputum and the other who is just beginning to cough. Would you expect a chest radiograph to have the same sensitivity in each of these patients? Both may have pneumonia, but the second patient is less likely to have obvious changes on a chest radiograph.

Subarachnoid hemorrhage (the example used at the outset of this discussion) is a disease that can range from a small sentinel bleed with no symptoms (except headache) to a catastrophic bleed with profound neurologic deterioration, herniation, and death. All cases in this range represent subarachnoid hemorrhages, and all are likely to be included in studies that aim to determine the sensitivity of CT scans for detecting subarachnoid hemorrhage. However, just as a patient who is comatose after having a subarachnoid hemorrhage may have a large bleed (and therefore be likely to have a positive result on a CT scan), a patient who has a headache but no abnormalities detected on neurologic examination may have a very small sentinel bleed, and a small bleed can be below the limits of detection by a CT scan. Consequently, a CT scan will have different sensitivities in patients at different points in the spectrum of the disease. A CT scan might be positive for disease in 100% of patients who are moribund owing to a subarachnoid hemorrhage but might have a sensitivity of only 10% in those who have less impressive symptoms. If, instead of the 95% sensitivity of CT scans for detecting subarachnoid hemorrhage quoted at the beginning of this discussion, you use a figure of 10% (representing the sensitivity of CT scans in detecting subarachnoid hemorrhage in patients who are at the mild end of the spectrum of the disease), you would arrive at a posttest probability (ie, the probability of subarachnoid hemorrhage given negative results on a CT scan) close to 1%, the observed number.

CIRCUMVENTING SPECTRUM BIAS

How can you avoid falling into the evidence trap surrounding spectrum bias when you apply test performance data to your patient population? As with most problems in interpretation of medical literature, the answer can be found by carefully reviewing the methods of the sources of the sensitivity and specificity data. Before incorporating the test into your practice, make sure you know what the sensitivity and specificity of the test is *in patients like the one you are testing*. When, for example, you are evaluating a patient for deep vein thrombosis, determine if the sensitivity and specificity you are quoting was derived from only patients with cancer who had red, swollen, painful legs or whether the study included patients who had no leg symptoms and who were being evaluated for other conditions (eg, possible pulmonary embolism). If you are careful to use only sensitivities and specificities generated by studies of patients

clinically similar to your own patients, you will find that the results of your clinical practice behave more like the mathematical models predict. And isn't that the point of applying mathematical models in the first place? **HP**

SUGGESTED READINGS

Gallagher EJ. Evidence-based evaluation of diagnostic tests. *Ann Emerg Med* 1996;28:347-9.

Lachs MS, Nachamkin I, Edestein PH, et al. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. *Ann Intern Med* 1992;117:135-40.

Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926-30.

Wheeler HB, Hirsh J, Wells P, Anderson FA Jr. Diagnostic tests for deep vein thrombosis. Clinical usefulness depends on probability of disease. *Arch Intern Med* 1994;154:1921-8.

Copyright 2001 by Turner White Communications Inc., Wayne, PA. All rights reserved.