REFERENCES

1. Position paper of the American College of Physicians, 16 January 1984. Long-term care of the elderly. J Am Geriatr Soc 1984;32:700-4
2. Calkins E. Residency training in geriatric medicine — 1984. Bull NY Acad Med 1985;61:534-46
3. Ahronheim JC. Pitfalls of the teaching nursing home. A case for balanced geriatric education. N Engl J Med 1983;308:335-6
4. Rabin DL. Physician care in nursing homes. Ann Intern Med 1981;94:126-8
5. Gillick MR. Is the care of the chronically ill a medical prerogative? N Engl J Med 1984;310:190-3

The Dynamic Nature of Sensitivity and Specificity

THE INFORMATION GAINED from laboratory tests can reduce diagnostic uncertainty, but properly interpreting the results can be difficult.[1] Bayesian analysis is a popular quantitative method used to determine the probability of disease given a positive or negative result (the predictive value). This is calculated from the sensitivity and specificity of the laboratory test and probability of disease before testing.[2] It is well recognized that the predictive value will vary in different clinical situations according to changes in the prior probability of disease. What is often not appreciated is that the sensitivity and specificity are also dynamic parameters.

Traditionally, at a given prior probability of disease and cut-off point used to define a "positive" test result, the sensitivity and specificity have been viewed as constant properties of a laboratory test.[3] However, the clinical accuracy of a laboratory test (expressed as sensitivity and specificity or likelihood ratio) is not an intrinsic property but depends on the population tested. The severity of illness among the sample of patients with disease and the spectrum of people contained in the non-diseased sample can greatly influence the sensitivity and specificity.[4] In general, it is easier for a laboratory test to detect advanced disease than mild disease (the sensitivity). Likewise, a laboratory test may more readily distinguish diseased people from normals than from non-diseased people who have clinical features similar to those of the disease (the specificity). Unfortunately only the "fixed" sensitivities and specificities are often used in quantitative clinical decision models. This may lead to gross errors in calculating the predictive value.

With advanced or severe disease, laboratory tests are more likely to be positive. This increases the number of true positive test results, leading to an increase in the sensitivity of the laboratory test (↑ true positives ÷ ↑ true positives + ↓ false negatives = ↑ sensitivity). The likelihood ratio, a function of the sensitivity and specificity, will increase as well (↑ sensitivity ÷ 1 − specificity = ↑ likelihood ratio).

Several studies have documented the phenomenon of shifting sensitivity. Hlatky et al.[5] showed that the exercise electrocardiogram is better at detecting severe coronary artery disease than mild disease. At a fixed prior probability of disease, using the same interpretation criteria, the sensitivity increased with the extent of coronary artery disease (sensitivity 47.9% for one-vessel disease and 84.7% for three-vessel disease). This change in sensitivity, if ignored, could result in inaccurate interpretation and ineffective use. The test may be of little value when trying to "rule out" coronary artery disease because of its inability to detect mild disease. If, for example, the exercise electrocardiogram is used to evaluate a 45-year-old man who has atypical chest pain with a presumed prior probability of coronary artery disease of 30% and a test specificity of 80%, a negative test result effectively "rules out" three-vessel disease but does not exclude the possibility of symptomatic single-vessel coronary artery disease (with negative predictive values for three- and single-vessel disease of 94% and 78%, respectively).

Similarly, Fletcher demonstrated that the sensitivity of the carcinoembryonic antigen (CEA) test to detect colonic carcinoma increases with advancing stages of disease at all levels of specificity.[6] The CEA test is the least sensitive in detecting colonic carcinoma at its early stages when therapy is most effective, and is therefore of little diagnostic value.

The specificity of a laboratory test also varies depending upon the clinical situation in which it is being used. The specificity of a laboratory test used to distinguish disease from normal may be distinctly different than that when the test is used to differentiate disease from related conditions (i.e., diseases affecting the same organ system or with similar clinical manifestations). People who have other conditions that resemble the disease in question may be more likely to have false-positive test results than healthy people. This increase in the number of false-positive results causes a decline in specificity (true negatives ÷ true negatives + ↑ false positives = ↓ specificity).

An example of the changing specificity can be seen when interpreting the results of antinuclear antibody tests. The antinuclear antibody test known as "SS B" is thought to be characteristic of Sjögren's syndrome.[7] The false-positive rate among healthy

people is less than 1%, while the rate among patients with other connective tissue disorders such as systemic lupus erythematosus is 15%.[8] Thus, the specificity ranges from 99% to 85% depending on the population tested. With a prior probability of Sjögren's syndrome of 10%, using the SS B antinuclear antibody test to differentiate Sjögren's syndrome from normal, the positive predictive value is 86.1%. However, with the same prior probability of disease, using the test to distinguish Sjögren's syndrome from other connective tissue disorders the positive predictive value is only 30.8%, a difference of more than 50%.

Rozanski et al.[9] determined the specificity of exercise radionuclide ventriculography during two different time periods at a single institution and documented a profound decrease in specificity from 86% to 21%. The test technique, personnel and interpretation criteria did not change. Initially, the non-diseased group of patients had a very low pretest probability of coronary artery disease, with few false-positive studies, giving a high specificity. In time, the test was used primarily in individuals who had clinical features suggesting coronary artery disease, increasing the number of false-positive results. This change in the population being tested accounted for the temporal decline in specificity.

The sensitivity and specificity of a new laboratory test are typically first evaluated by comparing populations near the two extremes of a response spectrum; a population of patients with definite, severe disease, and a normal population that includes healthy volunteers — the "sickest of sick" versus the "wellest of well."[9] Though the results may be internally valid, they tend to overestimate the performance of the test as it is actually used in clinical practice. Ideally the spectrum of sensitivities and specificities should be determined by examining the results from various well-defined groups of diseased and non-diseased people. The appropriate comparison group depends on the reason for obtaining the test. For the purpose of diagnosis, the comparison group would be patients presenting with complaints similar to the disease in question; for screening, it would be the asymptomatic population-at-large; for monitoring the course of disease, it would be other people with the disease. Unfortunately, few laboratory tests undergo such thorough evaluation.

Clinicians need to be aware that the performance of a test can vary significantly depending on how the test is used as a result of changes in sensitivity and specificity, irrespective of the cut-off point or prior probability of disease. In general, sensitivity tends to increase as the stage or severity of illness in the diseased group increases. The specificity of a laboratory test diminishes as the clinical features of the populations with and without the disease in question become more similar. Failure to appreciate this possible variation in test performance can substantially limit the precision, accuracy, and usefulness of quantitative methods of laboratory interpretation. — *Richard P. Lofgren, MD, Section of General Internal Medicine, Department of Medicine, University of Minnesota; Minneapolis VA Medical Center, Minneapolis, MN.*

REFERENCES

1. Casscells W, Schoenberger A, Graboys TB. Interpretation by physicians of clinical laboratory results. N Engl J Med 1978;299:999-1001
2. Griner PF, Panzer RJ, Greenland P. Selection and interpretation of diagnostic tests and procedures: principles and applications. Ann Intern Med 1981;94:553-600
3. Feinstein AR. Diagnostic and spectral markers. In: Clinical Epidemiology: The Architecture of Clinical Research. Philadelphia: W. B. Saunders; 1985:597-631
4. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926-30
5. Hlatky MA, Pryor DB, Harrell FE Jr, et al. Factors affecting sensitivity and specificity of exercise electrocardiography. Am J Med 1984;77:64-70
6. Fletcher RH. Carcinoembryonic antigen. Ann Intern Med 1986;104:66-70
7. Isenberg DA, Hammond L, Fisher C, et al. Predictive value of SS-B precipitating antibodies in Sjögren's syndrome. Br Med J 1982;284:1738-40
8. Tan EM. Special antibodies for the study of systemic lupus erythematosus: an analysis. Arthritis Rheum 1982;25:753-6
9. Rozanski A, Diamond GA, Berman D, et al. The declining specificity of exercise radionuclide ventriculography. N Engl J Med 1983; 309:518-22

Bon Voyage to David Siscovick and Michael O'Malley

AS WE SEND VOLUME 2, Issue 6 off to press, we bid farewell to two of our associate editors, David Siscovick and Michael O'Malley. They are leaving their UNC posts to pursue the next phases of their respective careers. Over the first two and a half years of the JOURNAL, David and Michael have been enormously helpful — to the editors, to many authors whose submissions were enhanced by their careful work, and most of all, to the JOURNAL, which has shone more brightly because of all their polishing and ideas.

Most people who contribute to causes do so occasionally, but these two gave of their time almost daily. All SGIM members and JGIM subscribers owe them a vote of thanks. Here in the Editorial Office we miss them already. Bon voyage. — *The Editors*