How do I interpret a positive diagnostic test?

In this issue of *Gastrointestinal Endoscopy,* Tucci et al[1] report results obtained with a device named ''Mt 21-42'' that can measure pH and ammonium concentration in gastric juice obtained at the time of upper GI (UGI) endoscopy. Gastric pH was evaluated for its ability to diagnose atrophic gastritis of oxyntic mucosa (AGOM), and ammonium concentration was evaluated for its ability to diagnose gastric *Helicobacter pylori* infection.

Subjects for this evaluation were 216 consecutive outpatients referred for diagnostic UGI endoscopy at a single center. The presence or absence of AGOM was based on the histologic appearance of gastric biopsy specimens and gastric *H pylori* infection was based on at least 3 positive results from histology, urease, urea breath test, and serologic tests for *H pylori* antibodies. The authors concluded that a high gastric pH indicated AGOM and a high ammonium concentration indicated *H pylori* infection.

Clinicians who read this article may wonder, ''What does this test mean for me?'' This editorial focuses on issues that clinicians should consider when they read an article that reports results with a diagnostic test. I have focused on the data related to the use of gastric pH to diagnose AGOM to highlight the relevant issues for a clinician who sees one patient at a time. The interested reader can perform similar analyses using the data related to ammonium.

WHAT IS THE STUDY POPULATION?

The current study[1] examined subjects who were referred for UGI endoscopy. Thus, the results for the prevalence of AGOM cannot be extrapolated to the population at large and obviously depend on the reasons that subjects were referred for endoscopy at the single Italian center where the data were collected. It seems likely that indications for UGI endoscopy may differ from one center or clinician to another, and as will be discussed later, the accompanying variation in prevalence is important in interpreting a test result from an individual subject.

0016-5107/$32.00
doi:10.1016/j.gie.2007.04.003

WHAT IS THE ACCURACY OF THE DIAGNOSTIC TEST?

The accuracy of the test depends on the sensitivity and specificity of the test and on the cut point used to determine these measures. Sensitivity is defined as the percentage of subjects with the disease who have a positive test result. Specificity is defined as the percentage of subjects without the disease who have a negative test result. Sensitivity and specificity are properties of the test and have nothing to do with the prevalence of the disease in the population. They do, however, depend on the cut point.

> The critical features in interpreting a result from an individual patient are the sensitivity and specificity of the test and the prevalence of the disease in the population.

Choosing a cut point to define an abnormal result requires the clinician to make a tradeoff between sensitivity and specificity. A cut point that has high sensitivity will have low specificity, and one that has high specificity will have low sensitivity. There is no ''rule'' for choosing a cut point. The authors[1] do not specify how they chose the cut point for gastric pH; however, the results in their Figure 1 suggest that pH 4 was chosen to maximize sensitivity.

Different individuals can appropriately reach different conclusions regarding the cut point. Assuming that high values are abnormal, a low cut point will have high sensitivity (many true positives) and low specificity (few true negatives). A high cut point will have the opposite (ie, few true positives and many true negatives). In most instances, choosing a cut point depends on the use to which the cut point will be put.[2] For example, consider a test for a disease that is fatal if untreated, but is completely treatable; a positive result can be followed by a second test that is completely accurate and without risk. Such a test should have a high sensitivity (low cut point) because the second test will identify the false positives resulting from the first test. In contrast, consider a test for an incurable noncommunicable disease. In this situation, the test should have a high specificity (high cut point) to avoid false positives (incorrectly telling a healthy person that he or she will die

TABLE 1. Results from measuring gastric pH

	AGOM	
Test result	**Yes**	**No**
Positive	26	27
Negative	1	162
Total	27	189

A positive test result is gastric pH >4. Values are from reference 1.

TABLE 2. Expected results from measuring gastric pH in a population with a relatively low prevalence of AGOM

	AGOM	
Test result	**Yes**	**No**
Positive	96	1417
Negative	4	8483
Total	100	9900

A positive test result is gastric pH >4. Results were calculated using the values for sensitivity and specificity from the study by Tucci et al[1] and assuming a prevalence of AGOM of 1%.

soon). When there is no reason a priori to choose a particular cut point, it is common to choose the value that gives optimal sensitivity and specificity (ie, the highest values for combined sensitivity and specificity).

Frequently, the accuracy of a diagnostic test is quantified with receiver-operator characteristic curve (ROC) analyses.[3,4] Tucci et al[1] did not perform ROC analyses; however, such analyses make it possible to quantify the extent to which the chosen cut points discriminate between healthy and AGOM. In particular, ROC analyses make it possible to calculate the probability that a subject selected randomly from the AGOM group will have a higher value for gastric pH than will a subject selected randomly from the healthy group.[3,4]

The results in Table 1 were obtained from the study by Tucci et al[1] and indicate that the sensitivity is 26 of 27, or 96.3%, and that the specificity is 162 of 189, or 85.7%.

HOW DO I INTERPRET THE RESULT FROM AN INDIVIDUAL PATIENT?

Clinicians are usually faced with interpreting a test result from an individual patient, and they need a way to incorporate the results from the study by Tucci et al[1] into their practice. The critical features in interpreting a result from an individual patient are the sensitivity and specificity of the test and the prevalence of the disease, in this case AGOM, in the population.

In Table 1, the prevalence of AGOM is calculated from the total values. That is, the prevalence of AGOM in the population (subjects referred for UGI endoscopy) is the number of subjects with AGOM (27) divided by the sum of the number of subjects with AGOM (27) plus the number without AGOM (189), or 12.5%. With this prevalence, the probability that a subject with a positive test has AGOM is the number of true-positive results (26) divided by the sum of the number of true-positive results (26) plus the number of false-positive results (27), or 49.6%. Thus, the probability that a subject with a positive test result has AGOM is about the same as tossing a coin. This value of 49.6% is also referred to as the ''positive predictive value.'' The probability that a subject with a negative test result does not have AGOM is the number of true-negative results (162) divided by the sum of the number of false-negative results (1) plus the number of true-negative results (162), or 99.4%. This value is also referred to as the ''negative predictive value.''

Table 2 illustrates how the prevalence of the disease in the population influences the interpretation of a test result from an individual subject. For the purposes of illustration, I assumed that the prevalence of AGOM in the population was 1%, instead of the 12.5% in Table 1. Under this assumption, out of 10,000 subjects, 100 will have AGOM and 9900 will not have AGOM. Because sensitivity and specificity are properties of the test and do not depend on the prevalence of the disease, they can be used to calculate the numbers of true and false values for a given prevalence. With a sensitivity of 96.3% and 100 subjects with AGOM, 96 will have a positive test result (true positives) and 4 will have a negative test result (false negatives). With a specificity of 85.7% and 9900 subjects without AGOM, 8483 will have a negative test result (true negatives) and 1417 will have a positive test result (false positives).

By using the same approach to the data in Table 2 that was used for the data in Table 1, the probability that a subject with a positive test has AGOM (positive predictive value) is now 6%, and 94% of the positive test results are false positives. The probability that a subject with a negative test result does not have AGOM (negative predictive value) is 99.9%.

For any condition with a low prevalence and a test with a specificity that is less than 100%, most of the positive test results will represent false positives. This does not mean, however, that the test is of no use for clinicians. The data in Table 2, for example, illustrate that without the test, the clinician could expect to detect 1 case of AGOM out of 100 by using gastric biopsy alone. If the pH test was performed at endoscopy, biopsy would not need to be performed on subjects with a negative test result (84% of endoscoped subjects) because 99.9% of them will not have AGOM. This approach would fail to detect AGOM in only 0.05% of subjects. However, of the subjects with a positive test result (16% of the subjects who undergo endoscopy), only 96 of 1513 (6%) will have AGOM. Although the diagnostic yield is low using the pH test, it still

offers advantages to performing gastric biopsies on all subjects who undergo UGI endoscopy.

A convenient way for clinicians to use sensitivity, specificity, and prevalence to determine the probability that a subject with a positive test result has the disease is to use Bayes' rule,[2,5] which states that

Odds of subject testing positive having AGOM = Odds of AGOM in population × Likelihood ratio

where

Likelihood ratio = Sensitivity/(100 – Specificity) = True positives/False positives

From Table 2,

Odds of AGOM in population = 100/9900 = 1.01%

Likelihood ratio = 96.3/(100 – 85.7) = 6.73

Odds of a subject who tests positive having AGOM = 1.01% × 6.73 = 6.8%

Obtaining a positive test result causes a 6.73-fold increase in the odds of having AGOM. If the odds of AGOM in the population are 14.3%, as was the case in Table 1, the odds for AGOM in a subject with a positive test result are 6.73 × 14.3%, or 96.1%. Because probability equals odds/(100 + odds), the probability of AGOM in a subject with a positive test result is 49.0%.

Bayesian analysis makes it possible to update a prior probability on the basis of new information. Motulsky[2] pointed out that most clinicians do a pretty good job of combining probabilities intuitively without knowing anything about Bayesian thinking. Calculating the likelihood ratio from an article that reports values for sensitivity and specificity for a particular diagnostic test is an explicit and exact way for clinicians to determine how helpful the test might be in their practice, given their estimate of the prevalence of a particular condition in their patient population.

CAN THIS TEST BE USED TO DETERMINE PREVALENCE OF AGOM?

Some may wonder whether measuring gastric pH with Mt 21-42 at the time of UGI endoscopy can be used to estimate the prevalence of AGOM in a given population. The specificity of a diagnostic test is a critical determinant of its usefulness to determine prevalence. Because 100 minus the specificity of a test gives the percentage of false positives, the prevalence determined using the test can never be less than the percentage of false positives. For example, if a test with a specificity of 99% is used to determine prevalence, the prevalence can never be less than 1%, or 1 in 100. Similarly, if the specificity is 99.9%, the prevalence can never be less than 0.1%, or 1 in 1000. The results from Tucci et al[1] indicate that the specificity of gastric juice for diagnosing AGOM is 85.7%; therefore, the prevalence can never be less than 14.3% when it is assessed with Mt 21-4. Failure to appreciate this relationship between specificity and estimated prevalence can result in important overestimates of prevalence.

SYNTHESIS

Because AGOM has a relatively low prevalence in patients undergoing UGI endoscopy, measuring gastric pH with Mt 21-4 can be useful in that a negative result (pH <4) indicates that no biopsy specimens are needed to search for AGOM. A positive test result can indicate the need for gastric biopsies to detect AGOM, bearing in mind, however, that the lower the prevalence of AGOM in the patient population, the more subjects will have normal specimens in spite of a positive pH test.

Although this editorial focuses on gastric pH, similar analyses can performed using the data for gastric ammonium. Measuring the gastric ammonium concentration with Mt 21-4 had a high sensitivity and specificity for detecting *H pylori* infection; however, it is not clear that this provides a clear advantage over a gastric antral biopsy to measure urease.

DISCLOSURE

The author declares that he has no conflict of interest.

Jerry D. Gardner, MD
Science for Organizations, Inc
Mill Valley, California, USA

Abbreviations: AGOM, atrophic gastritis of oxyntic mucosa; ROC, receiver-operator characteristic curve; UGI, upper GI.

REFERENCES

1. Tucci A, Bisceglia M, Rugge M, et al. Clinical usefulness of gastric-juice analysis in 2007: the stone that the builders rejected has become the cornerstone. Gastrointest Endosc 2007;66:881-90.
2. Motulsky H. Interpreting lab tests: introduction to Bayesian thinking. In: Intuitive biostatistics. New York: Oxford University Press; 1995. p. 129-39.
3. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. N Engl J Med 1975;293:211-5.
4. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240:1285-93.
5. Berry DA. Statistics: a Bayesian perspective. Belmont: Duxbury Press; 1996.