EVALUATING DIAGNOSTIC TESTS

Where do the performance numbers come from?

FIRST IN A SERIES

www.biomarketconsulting.com

By Dr. John Alsobrook, PhD, DABCC, FACB

> *"The only math skills you'll need are … grammar school skills."*

IT ONLY SEEMS COMPLICATED. REALLY.

Diagnostic tests are typically marketed based on their performance claims, the numbers that describe how well the test works. You'll hear numbers (usually percentages) quoted for "sensitivity", "specificity", "accuracy", and more. The general idea is very intuitive: larger numbers usually mean better performance. The performance is based on comparing the test results against a gold-standard clinical evaluation. But where do those numbers come from, and what do they mean? Understanding diagnostic test performance statistics can appear to be daunting. Here you'll see that the underlying principles and the necessary arithmetic are straightforward and not very deserving of the anxiety they can generate.

SIMPLE ARITHMETIC

First, let's do a math skills inventory. The only math skills you'll need are counting, adding, and dividing. That's it. Grammar school skills.

SIMPLE DATA

Next, let's consider the kind of information that goes into the calculations. The data come from patients' participation in studies, and there are only two things you need to know for each of the patients: do they have the disease (affected) or not (unaffected or control), and did the test score them as positive or negative. Those are all the skills and data needed to do these performance calculations.

HOW IT'S DONE

Here's how to put it all together. Diagnostic performance is based on counting categories of patients and test results in a 2-by-2 table, in other words a table with 2 rows and 2 columns that has 4 compartments. Refer to the example table below with made-up numbers. The columns are the true disease status of the patient, either affected or not affected. The rows are the results of the test, either positive or negative. Every patient is counted once (and only once) in the correct square of the table. For instance, an affected patient with a positive test result is counted in the upper-left square that is in the 'affected' column and in the 'positive' row.

	Affected	**Unaffected**
Positive	98	5
Negative	2	95

CALCULATING SENSITIVITY

Now that the counting is done, we can do some arithmetic. Let's start with *sensitivity*. Sensitivity is only about the affected patients, in other words every study participant who was counted in the left hand column. A commonly used description of sensitivity is the “true positive” rate, which means the percentage of affected cases for whom the test result was positive. This number is all about the patients with the disease, and how often the test scores them correctly. To calculate the percentage of true positives, the ‘affected & positive test’ count is divided by the total ‘affected’ count. So the arithmetic is: 98 (affected & positive) divided by 98 + 2 (affected & positive PLUS affected & negative) = 98 ÷ 100 = 0.98, or 98%. This means if 100 patients with the disease are evaluated, the test will find 98 of them. The test will miss the other 2 patients with the disease.

CALCULATING THE REST

The other performance numbers typically provided are specificity and accuracy. The principles are similar, they just use different parts of the table. *Specificity* is commonly referred to as the “true negative” rate, which means the percentage of unaffected cases for whom the test result was negative. This number is all about the patients who don’t have the disease, and how often the test scores them correctly. Those are all in the right hand column labeled ‘Unaffected’. To calculate the percentage of true negatives, the ‘unaffected & negative test’ compartment count is divided by the total ‘unaffected’ count. So the arithmetic is 95 (unaffected & negative) divided by 95 + 5 (unaffected & negative PLUS unaffected & positive) = 95 ÷ 100 = 0.95, or 95%. This means if 100 unaffected individuals are evaluated, the test will find 95 of them. The other 5 unaffecteds will have an incorrect positive test result (“false positives”).

Accuracy is an overall number expressing how often the test is right. Look at the table again, the upper-left and lower-right corners are the compartments where the test is right, and the other two (upper-right and lower-left) corners are where the test got it wrong. Accuracy is the number of correct calls (98 + 95) ÷ the total count in all 4 compartments (98 + 95 + 2 + 5) = 193 ÷ 200 = 0.965, or 96.5%. Accuracy is not always a fair way to estimate performance, because it depends on how many affected patients were in the study compared to the number of unaffected patients. A test with good sensitivity (good left-hand column) and poor specificity (poor right-hand column) might have good accuracy if the study just didn’t have very many unaffected patients in the first place.

GOING DEEPER

For the basic performance measures explained here, bigger means better. There are many other ways to slice-and-dice the data from a simple 2 x 2 diagnostic table, such as false positive rate, false negative rate, negative predictive value, and positive predictive value to name just a few. Some of those that are relevant to investors, management, and other non-technical professionals will be outlined in subsequent articles in this series.

BioMarket Consulting, LLC