Slide 1

Sensitivity and Specificity Part I - Introduction and Definitions

Slide 2

This video is designed to accompany pages 81-94 of the workbook “Making Sense of Uncertainty: Activities for Teaching Statistical Reasoning,” a publication of the Van-Griner Publishing Company

Slide 3

While there aren’t any statistics available on exactly how many home pregnancy tests (HTPs) are used each year, it is safe to say that they are used in abundance. An HPT is typically tested using a urine sample and works by using chemistry to bond the glycoprotein hCG to an antibody and an indicator. hCG is presented typically just after fertilization.

The goal of this test is to suggest to the user whether she is pregnant (a “positive” outcome) or is not pregnant (a “negative” outcome).

Of course this can’t be done with certainty. There are two ways the test can be wrong. It can say the user is pregnant when she is not; or it can fail to say the user is pregnant when she is.

Slide 4

All “screening tests” – whether they are screening for depression, concussion, bowel cancer, whatever - are faced with delivering the same kind of decision. Either the test indicates you likely have what the test is designed to look for (a “positive” or “yes” result), or the test indicates that you don’t seem to have what the test is designed to look for (a “negative” or “no” result). And all such tests have some risk of the same two possible mistakes.

Slide 5

These two errors have common names that we need to know.

The test can say you do have what the test is looking for, when you really don’t. So the test has wrongly come back positive. That is called a FALSE POSITIVE.

Or the test can say you don’t have what the test is looking for, when you really do. Here, the test has wrongly come back negative. This is called a FALSE NEGATIVE.

A test that is intended to deliver a dichotomous outcome needs to be evaluated to see how often it is susceptible to false positives – its false positive rate – and how often it is susceptible to false negatives – its false negative rate.

The computation of false negative rates and false positive rates will require two types of information. We will have to have data on the actual classification of subjects using the test (as positives or negatives) and we‘ll have to know the truth about the subjects, whether they really were positives or negatives. The latter information is usually available from some more extensive, perhaps more invasive or expensive testing, often called the “gold standard.”

For example, while and HPT provides an inexpensive screen for pregnancy, a much more definite (and expensive) option is to just go to the doctor for a blood test. Likewise, there are concussion screenings that are important to do on the sideline, but MRIs could provide definitive answers. More on the computation of false positive and false negative rates later.

Slide 6

Let’s develop more language. Your doctor may not use the term “false positive rate” when she talks to you about the reliability of a screening procedure. Instead, she may refer to the “specificity” of the procedure. In generic terms, specificity is the ability of the screening test to correctly identify a negative outcome as a negative outcome. Numerically, specificity is 1 minus the false positive rate.

Similarly, the “sensitivity” of a screening test is the ability of that test to correctly identify positives as positives. Numerically, sensitivity is 1 minus the false negative rate.

Slide 7

Let’s go back to HPTs for a minute. How accurate are they, at least according to their box labels and instructional inserts. The claim made by one brand of HPT is shown here. Notice that it says the test is 99% accurate.

We have two questions to ask. Is this reference to accuracy a reference to sensitivity, specificity or something else entirely? And, secondly, how was this assessment done?

Slide 8

The 99% means that 99% of the time the test will show a positive when a non-pregnant woman’s urine is mixed with commercial hCG. Too much information, perhaps but that is how it is done. Is this a claim about sensitivity or specificity?

Slide 9

Remember, the hCG is the indicator of pregnancy that the test is designed to look for. Since the urine being tested has been mixed with hCG, it is a proxy for urine from a pregnant woman. Hence, the claim is that 99% of the time the test will correctly identify a positive as a positive.

So this is the laboratory claim about the sensitivity of the test. It follows that the false negative rate of this HPT is 1% or 0.01.

That’s quite good for sensitivity.

Slide 10

The tests done in lab to earn that 1% false negative rate were not done using real women trying to use the tests in real urine streams. Turns out that is a problem. A 1998 study in the Archives of Family Medicine found that the actual false negative rates, using real women in real situations, tended to be higher than 1%, sometimes much higher.

Still, the sensitivity of most of these HPTs, tested in the lab in controlled situations, is around 99%.

Slide 11

What about false positive rates and specificity? For an HPT, a false positive is often thought of as the error that is least socially problematic. Hence, kits don’t typically say anything about specificity. However, the study just mentioned also recorded the false positive rates for several different HPT kits in practice and the results are shown here.

Slide 12

So how well does an HPT perform in practice? The chart shown here allows for a head-to-head comparison of several different brands. Keep in mind this study is from 1998 and current brands may produce different results.

There are several interesting things to note.

First, one of the kits (OVA II) was just bad, with a high false negative rate and a high false positive rate. While a few of the kits had similar false positive and false negative rates, by and large those that had good false positive rates had worse false negative rates, and vice versa. This is a common phenomenon for screening tests of all kinds. Ratcheting down the false negative rate typically ratchets up the false positive rate, and the reverse is true as well.

So what is the best test of the ones shown? That depends on what you are looking for. The one with the best balance is probably Predictor with a 3% and 4% false negative and false positive rate, respectively. But there are two with an estimated 0% false negative rate, and if that is the error that one wants to keep the lowest, then those two would be the most desirable.

Slide 13

This concludes our introductory video on sensitivity and specificity. Remember, the false positive rate and false negative rate are common numerical assessments of the risks associated with the results of a screening test.