Measurement error: Sensitivity and specificity

Measurement is a common practice in epidemiological research. We use different tools to measure exposures and outcomes from questionnaires to clinical examinations, diagnostic tests and instruments. These measurements will not usually provide the exact (the true) value of the variable 100% of the time. For example measuring blood pressure of the same person with the same instrument may provide different results depending how it is measured and read. The same question about a person's characteristic may elicit different responses depending on how it is worded and how it is asked. Different diagnostic tests may give different results for people in the sample.

In order to be aware of the degree of accuracy of our research results (which depends on accuracy of our measurements) we have to be aware of the accuracy of the tools we use. There are two concepts related to measurement accuracy (or errors in measurement) which are validity and reliability. **Validity** means how accurate the test is in measuring the true value of the variable we are interested in. While **reliability** means how consistent (i.e. giving the same result) the test is when it is used by different observers or over different periods of time. In the following paragraphs we explain main methods for measuring validity.

Measuring Validity

Validity means the extent to which the tool (a diagnostic test, a question or any other mean of measurement) measures the true value of the variable that we are interested in. But how we measure validity? For example suppose we want to measure blood glucose of 100 people using a figure prick quick test instrument. What is the validity of this test? How do we decide to what extent this test measures the true value of blood glucose?

Ideally we would measure validity by comparing the results of our test with the true values. But the true value may not actually be known for many variables. The best estimate of the true value is the result of the best available test or the gold standard test. We measure validity of a test by comparing the results of the new test with the results of the gold standard test.

Measuring validity of a continuous variable

This can be measured by estimating how well the results of the new test correlate with the true values i.e. results of the gold standard test. This factor is known as the **validity coefficient**. When a results of the new test of a continuous variable e.g. blood sugar is compared to results of a gold standard, the **Pearson** correlation coefficient can be calculated as an indication of correlation of the two results and thus validity of the new test.

Measuring validity of a categorical variable

In measuring categorical variables, for example using a particular test to diagnose cases of TB (TB, no TB), there is probability of mistakes in diagnosis. We may place people in the wrong category i.e. we diagnose a normal person as a case of TB or we diagnose a case of TB as no TB (misclassification). To measure the accuracy of tests for categorical variables we usually compare our test with the best test (the gold standard) and calculate sensitivity and specificity by tabulating the results in a table as shown below.

Result of our test	The true result (The standard test)			
	Positive	Negative	Total	
Positive	A	B	A+B	Positive predictive value
Negative	C	D	C+D	Negative predictive value
Total	A+C	B+D	A+B+C+D	
	Sensitivity	Specificity		

For example suppose we used a new test on 100 people to diagnose HIV and wanted to know how valid is this test in comparison with the standard test. We use both the new and the standard test on all 100 people and tabulate the results as shown in the table below.

Result of the new test	The true result (The standard test)		
	Positive	Negative	Total
Positive	30	20	50
Negative	10	40	50
Total	40	60	100

From the data presented in this table we can calculate sensitivity, specificity, positive predictive value and negative predictive value of the new test.

Sensitivity

Sensitivity of a test measures how accurate the test is in diagnosing the cases. Sensitivity is the proportion of true positives correctly identified by the new test. i.e.

Sensitivity = A/(A+C) = 30/40=75%

This means that the new test is able to correctly diagnose only 75% of true cases. In other words the test will miss 25% of HIV cases and misclassify them as non-cases (false negatives). When sensitivity is low, the test will miss more case of the disease. Therefore an ideal test would have a 100% sensitivity in which case it will correctly diagnose all cases of the disease and will have zero false negatives. But in practice sensitivity is always less than 100%.

Specificity

The specificity of the new test measures how accurate the test is in diagnosing none cases i.e. people who do not have the disease. In other words, specificity is the proportion of true negatives correctly identified by the test.

Specificity =D/(B+D) = 40/60=67%

This means that the test gives a true result for only 67% of people who are actually clear from the disease i.e. the test will misclassify 33% of clear people as having the disease (false positives). When specificity is low, the test will misdiagnose more normal people as cases. Therefore an ideal test would have a 100% specificity in which case it will correctly diagnose all normal people as normal and will have zero false positives. But in practice specificity is always less than 100%.
There is no test with 100% sensitivity and 100% specificity. A more sensitive test will usually be less specific and a more specific test will usually be less sensitive. Therefore we have to agree on a cutoff level of specificity and sensitivity.

The positive predictive value

The positive predictive value (PPV) of the test measures the extent to which the test predicts the disease. In other words PPV is the proportion of individuals with a positive result who actually have the disease.

PPV = A/(A+B)= 30/50= 60%

This means that only 60% of people who are diagnosed by the test as having the disease do actually have the disease. In other words 40% of the positive test results are wrong.

The negative predictive value

The negative predictive value (NPV) of the test measures the extent to which the test predicts absence of disease. In other words NPV is the proportion of individuals with a negative result who are actually free from the disease.

NPV = D/(C+D)=40/50=80%

This means that 80% of the negative results of the disease are actually true. In other words 20% of the negative results of the test belong to people who actual have the disease.

Relation of these parameters to prevalence of the disease

Sensitivity and specificity are inherent characteristics of the test. The true prevalence of the disease in the population does not affect sensitivity and specificity i.e. whether prevalence of the disease is high or low, this will not affect the sensitivity and specificity of the test. But the predictive values vary with prevalence. Positive predictive values are directly related to the prevalence of the disease i.e. PPV of the same test is more in a population with higher prevalence of the disease than its PPV in another population with lower prevalence. The negative predictive values are inversely related to the prevalence of the disease in the population i.e. NPV of the same test is less in a population with higher prevalence of the disease than its NPV in another population with lower prevalence.

Relation of Sensitivity to specificity

Sensitivity and specify of a test are related. As we cannot have test with 100% sensitivity and 100% specificity, we often have to choose a cut off even if it is not clear. When sensitivity is high the specificity tends to be lower and vice versa. When we use a test which is very sensitive i.e. it misses very few true cases, it tends to lose on specificity i.e. it will have more false positives.

For example suppose we decide to use clinical features (fever, weight loss and cough for 3 weeks) as the case definition for TB. This test (case definition in this example) will have high sensitivity because most TB cases will have these symptoms; therefore it will detect majority of true cases of the disease. But the specificity of the test will be low because these symptoms may be present in other diseases; therefore the test will misclassify as TB many people who are not true cases of TB (many false positives i.e. poor specificity). In the same way if use a test with very high specificity e.g. sputum culture, probably we will not misdiagnose any normal people as TB, instead we will fail to detect many true TB cases (many false negatives i.e. poor sensitivity).

******