TITLE: Sensitivity, Specificity, and Predictive Values in Diagnostic Testing

PRESENTER: Yachana Kataria

Slide 1:

Hello, my name is Yachana Kataria. I am a Clinical Chemistry Fellow at Boston Children's Hospital. Welcome to this Pearl of Laboratory Medicine on "Sensitivity, Specificity, and Predictive Values in Diagnostic Testing."

Slide 2:

A good diagnostic test should be able to distinguish those who have disease from those who do not have disease because misdiagnosis can either lead to no treatment or over treatment.

Suppose we have values for a serum test for patients. The green curve represents the distribution of results for patients without the disease. The red curve represents the distribution of results without the disease. The blue line represents the threshold for a positive and a negative test. Any result to the right of the blue line will have a positive test, and any result to the left of the blue line will have a negative test. Ideally, we want anyone with disease to test positive and only people with the disease to test positive.

Thus, if we had a perfect test, there would be nobody **without** the disease who tests positive. Likewise, there would be nobody with the disease who does not test positive

Slide 3:

Unfortunately, there is no perfect test. The distribution of result for non-diseased and diseased individuals overlap; this suggests that the test does not achieve perfect discrimination. Anyone with a result to the right of the blue threshold line who has the disease will have a positive test. These will be classified as the true positives in our population.

The figure shows that some of the distribution of results of the disease population falls below the blue threshold line. What does that mean? We know that anyone left of that blue threshold line will test negative, but they have the disease. Therefore, these individuals will be categorized as the false negatives.

Anyone with a result to the left of the blue threshold line who does not have the disease will have a negative test. These will be the true negatives as they don't have the disease and have a negative test.

However, the figure shows that some of the distribution of results for the non-disease population is above the blue threshold line. Non-diseased individuals with results above the cutoff will have a positive test and be classified as a false positive.

Ultimately, we want to minimize the overlap between the results distributions for diseased and non-diseased individuals as much as possible, as the consequences for disease misclassification can incur significant morbidity and mortality.

Slide 4:
As an example, I'll show you how to assess B-type Natriuretic peptide (BNP) as a test for congestive heart failure. BNP is utilized in the emergency department to help establish the diagnosis of heart failure among patients who present with shortness of breath. We will be assessing the diagnostic accuracy, which estimates how accurately the test discriminates between individuals who have the disease and those who do not. Diagnostic accuracy is measured by calculating the tests' sensitivity, specificity, and predictive values; these can be further utilized to look at Receiver Operating Characteristics (ROC) curve.

Slide 5:
A 2013 update from the American Heart Association (AHA) estimated that there were 5.1 million people with heart failure in the United States in 2006. Prevalence and incidence of heart failure increases with age. Despite improvements in therapy, the mortality rate in patients with heart failure has remained unacceptably high, making early detection of susceptible persons who would benefit from preventive measures imperative. 50% of people who develop heart failure die within 5 years of diagnosis, and diagnosis of the disease still remains difficult. For patients with suspected heart failure, BNP has gained attention for diagnosing acutely ill patients who present in the emergency department with shortness of breath.

Slide 6:
BNP was originally called brain-natriuretic peptide because it was found in the brain. They later found it at much higher levels in the heart. proBNP is the pro-hormone, which is cleaved into two fragments upon release from the cardiac muscle cells. The first fragment is BNP, the active hormone. The second fragment is NTproBNP, the N-terminal fragment. Under ventricular stretching of the muscles, they are both released in 1:1 ratio.

Even though NTproBNP is physiologically inactive, its levels are highly correlated with BNP levels. Most experts agree that both peptides offer similar diagnostic information. In this Pearl, however, we'll be focusing on BNP only. BNP is secreted by the ventricles of the heart in response to excessive stretching of heart muscle cells and its levels increase with heart failure symptoms.

Slide 7:

Signs and symptoms for heart failure are nonspecific, and a helpful history is not often obtainable in an acutely ill patient. Shortness of breath is a key symptom of heart failure. However, it may also be a nonspecific finding in the elderly or obese patient in whom comorbidity with respiratory disease and physical deconditioning are common. Routine laboratory values, electrocardiograms, and x-rays are also not accurate enough to always make the appropriate diagnosis.

Multiple studies have established the value of BNP for facilitating the diagnosis of heart failure in patients presenting with shortness of breath. In 2002, a landmark study published in the New Englad Journal of Medicing by Maisel et al that BNP is a strong and independent predictor of heart failure when utilized in the emergency department to determine the cause of death. They reported an odds ratio of 29.60 with a significant 95% confidence interval when assessed with history, symptoms, radiological, and laboratory findings. The diagnostic value of BNP in this setting far exceeds traditional tests.

Slide 8:

Here, we have an example of diagnostic performance of BNP testing for diagnosis of congestive heart failure in dyspneic patients in the emergency room from a large prospective study.

Let's first put together a 2x2 table. The columns are disease status - in this example, it is presence or absence of congestive heart failure. The rows are test result status - whether the patient got a positive or negative BNP test. Remember to always be consistent in how you assign the 2x2 table.
* 672 individuals have congestive heart failure (CGF) and elevated BNP levels. These are the true positives in our population.
* 72 individuals have congestive heart failure but don't have an elevated BNP. These are the false negative individuals.
* There are a total of 744 people with congestive heart failure.
* 198 individuals do not have congestive heart failure but have elevated BNP. These are the false positive individuals.
* 644 individuals don't have heart failure and don't have elevated BNP levels. These individuals represent the true negative in our population.
* There are a total of 842 people without congestive heart failure.

Slide 9:

Now, let's assess the sensitivity of the BNP test. But first, what is sensitivity? Sensitivity is the proportion of patients who have the disease that'll get a positive test result. In other words, it is the probability of a positive test, given the patient has the disease.

In a 2x2 table, sensitivity is estimated as the number of true positives over the summation of true positive and false negatives. For the BNP test, we have 672 true positive divided by a total of 744. Therefore, sensitivity equals 90%. 90% of the patients with congestive heart failure will get the correct diagnosis of congestive heart failure.

Slide 10:
Now, let's assess the specificity of the BNP test. But first, what is specificity? Specificity is the proportion of patients who do not have the disease and will test negative. In other words, it is the probability of a negative test, given the patient does not have the disease.

The equation for specificity is the true negative subjects divided by the sum of true negative and false positive subjects. To determine specificity, we have to divide 644 by 842. Therefore, specificity equals 76%.

Slide 11:
Let's look at the predictive values of BNP.
- The positive predictive value (PPV) is: If the test is positive, what is the probability that the patient does have the disease?
- The negative predictive value (PNV) is: If the test is negative, what is the probability that the patient does not have disease?

To calculate the predictive values, we now look at the 2x2 table going across rather than down.

To calculate the positive predictive value, we look at the proportion of true positive divided by true positive plus the false positive. In this example, we would divide 673 by 870. Thus, the positive predictive value is 77% for BNP. Therefore, there is a 77% probability that a patient with a positive test result has the disease.

To calculate the negative predictive value, we look at the proportion of true negative divided by true negative plus false negative. Here we would divide 644 by 716. Thus, the negative predictive value is 90%. This means that if the subject got a negative test, we are 90% sure that it is a true negative.

Slide 12:
Here is a summary of the diagnostic performance of BNP testing for diagnosis of congestive heart failure. Thus far, I've shown that at a threshold of 100 pg/mL of BNP:
- The sensitivity is 90%
- The specificity is 76%
- The PPV is 77%.
- The NPV is 90%

Taken together, the test does reasonably well at differentiating disease from non-disease.

Slide 13:
Heart failure is fairly common in the population. In fact, the prevalence of the disease in our example is 47%. Prevalence is the percent of patients being tested who have the disease in question. Now, let's suppose that the prevalence of heart disease has decreased from 47% to 5%, and it has become a relatively rare disease.

In our example, the total number of individuals with elevated BNP has decreased from 870 to 433. Here we have nearly a perfect test. 71 of the 79 patients that have congestive heart failure are correctly identified with a positive BNP test. How does that change the performance statistics of BNP?

Well, we see that sensitivity and specificity has remained the same. The positive predictive value plummeted; previously, it was 77% and now it is 16%. However, the negative predictive value increased from 90% to 99%.

Slide 14:
The prevalence of most diseases is low. Thus, positive predictive value, even for a good test with a high sensitivity, can be poor when there are few persons with the disease, and most of the positives will be false positives. Which is exactly what we observed when we artificially changed the prevalence of BNP.

Predictive values are affected by outcome prevalence. Lowering the disease prevalence will lower the positive predictive value and raise the negative predictive value. We can also calculate the positive predictive value using Bayes' Theorem.

Clinicians care about the predictive values. The negative predictive value is the "reassurance number" - when it is very high, the patient can be assured that they don't have or won't have the outcome. Physicians and patients want to know the probability of disease given a positive or negative test result. Raising the outcome prevalence has the **opposite** effects.

Slide 15:
What would happen if you were to lower the threshold for the disease?

Suppose we moved the blue line to the left. Well, we can see that we capture virtually all the true positives. At the same time, however, we capture more true negatives and classify them as false positives, meaning we will diagnose more people who don't have the disease with the disease.

Slide 16:
Now we can examine how the diagnostic accuracy changes.

Suppose we lowered the BNP from 100 pg/mL to 50 pg/mL. We saw that we would capture virtually all the true positives. Accordingly, sensitivity of BNP at 50 pg/mL has increased. On the other hand, specificity goes down from 76% to 62%.

However, if you were to increase the threshold from 100 pg/mL to 150 pg/mL, the sensitivity goes down and the specificity goes up.

There is a clear trade-off between sensitivity and specificity as you change the threshold for the BNP test.

Slide 17:
Sensitivity and specificity are not absolute. They are affected by the distribution of diseased and non-diseased individuals, and they will differ among different populations. Diagnostic accuracy will also vary with the spectrum of the disease as well. Therefore, sensitivity and specificity should be taken into the clinical context for appropriate application of the test.

Gold standard is the reference to which the test is assessed against. Ideally, we want a reference standard that correctly discriminates disease from non-disease. However, that's not usually the case. The test itself requires a definition of positive and negative test results. When we have a continuous scale, like the one observed for BNP, we have to optimize the diagnostic characteristics of the test so we can best identify diseased individuals from non-diseased individuals.

You may want a HIGH sensitivity because missing the disease could have horrible consequences, but at the same time, you'll accept the fact that there will be people who do not have the disease and call them positive as well.

You may want HIGH specificity because it is more important to find patients who do not have disease. The price for that is that you'll have a few false negatives and you'll be sending home patients who do have the disease but tested negative.

Slide 18:
We determine a threshold for a given test by examining the receiver operating characteristic (ROC) curve. In a ROC curve, the sensitivity is plotted as the function of the 1-Specificity for different cut-off points.

Here, you see a ROC curve for different threshold values for BNP for the diagnosis of congestive heart failure. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold.
- A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity).
- A test with no discrimination (the disease and non-disease distributions are the same) has a ROC curve that is the diagonal line from the lower left corner to the upper right corner.
- The closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

The area under the ROC curve reflects the diagnostic ability of a test to differentiate people with and without congestive heart failure. It can be interpreted as the likelihood that a person randomly selected from the disease population will have a more abnormal test result than a person randomly selected from the non-disease population. If the test has perfect diagnostic accuracy there is a 100% chance that a person from the disease population will have a more abnormal result than a person from the non-diseased population. On the other hand, if the test has no diagnostic ability, then there is still a 50% chance (like flipping a coin) that a person randomly selected from the disease population will have a more abnormal result than a person randomly selected from the non-diseased population. Here, the area under the ROC curve is

91%, so there is a 91% chance that a randomly selected diseased person would have a more abnormal result than a randomly selected non-diseased person.

The ROC curve also allows us to see the comparison of the curves generated from two or more tests. This can help us compare the diagnostic accuracy for two or more tests and even test combinations.

With that, I would like to conclude our conversation about sensitivity, specificity, predictive values, and ROC curves, utilizing BNP as an example

Slide 19: References

Slide 20: Disclosures

Slide 21: Thank You from www.TraineeCouncil.org
Thank you for joining me on this Pearl of Laboratory Medicine on "Sensitivity, Specificity, and Predictive Values in Diagnostic Testing."