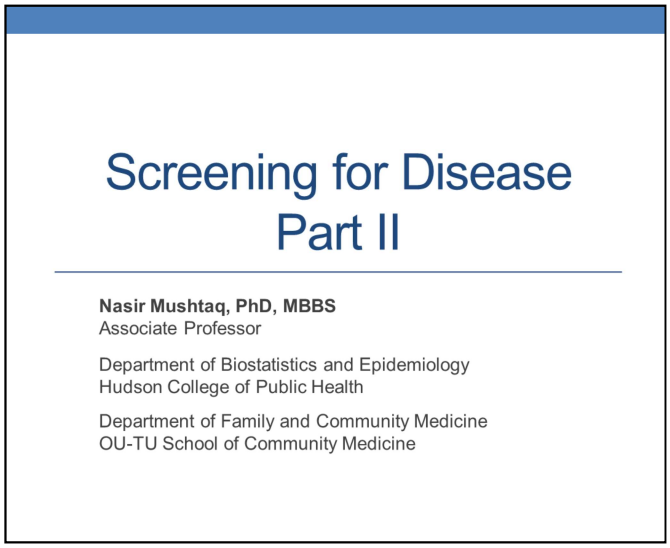
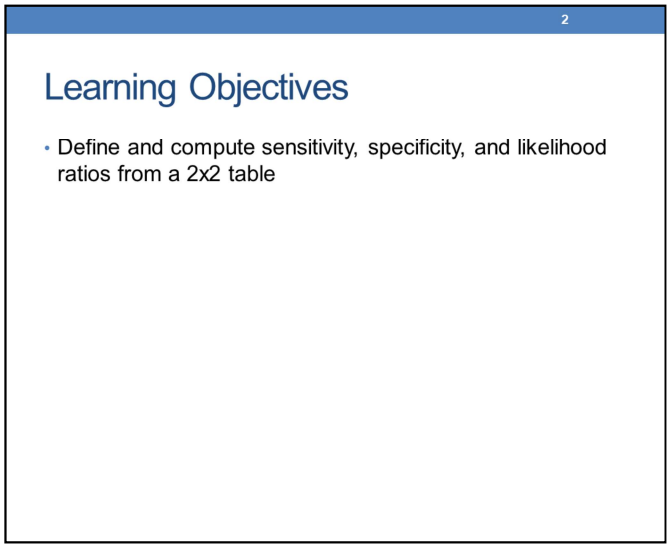
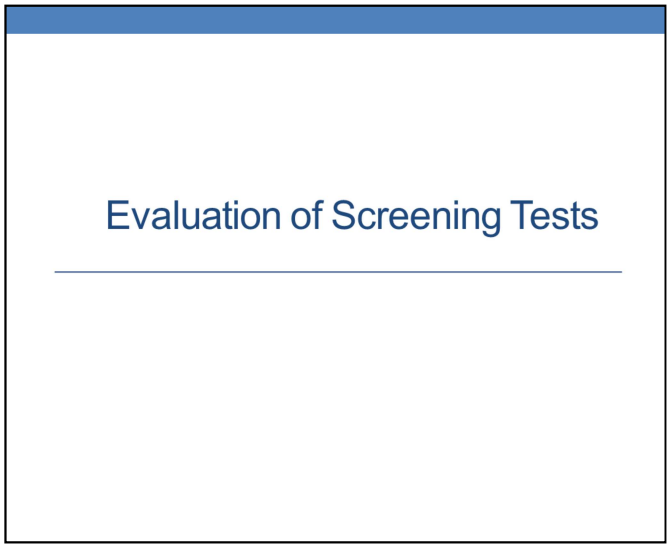


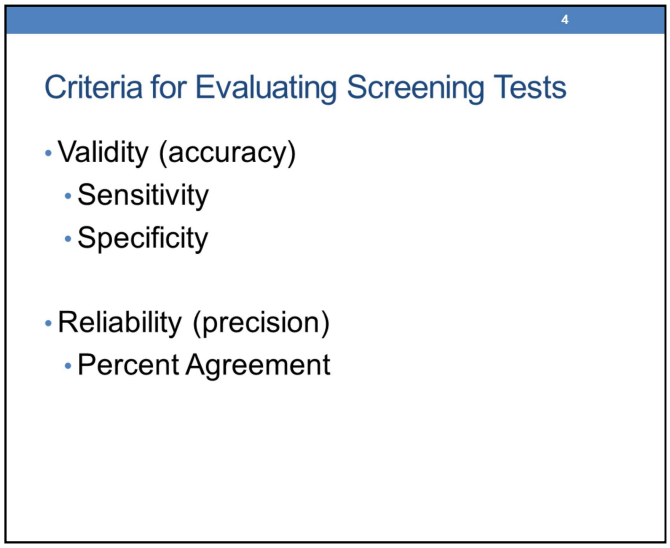
In this second part of lecture series focused on epidemiologic and biostatistical methods related to disease screening, we will define and learn how to compute quantitative measures of accuracy, including sensitivity, specificity, and likelihood ratios.



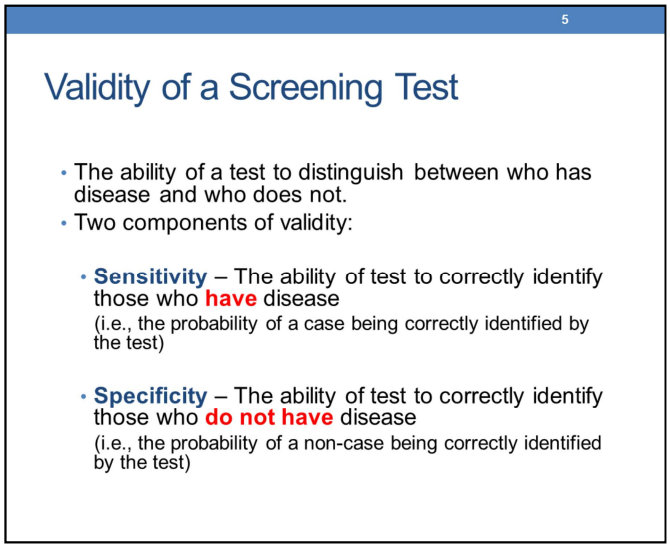
After viewing this segment, you will be able to define and compute quantitative measures of accuracy, including sensitivity, specificity, and likelihood ratios.



Let’s now discuss some of the primary measures that are used to quantify the accuracy of screening tests.



In this module, we will discuss measures of accuracy or validity, including sensitivity and specificity. In addition, we will discuss measures of precision or reliability, including percent agreement.

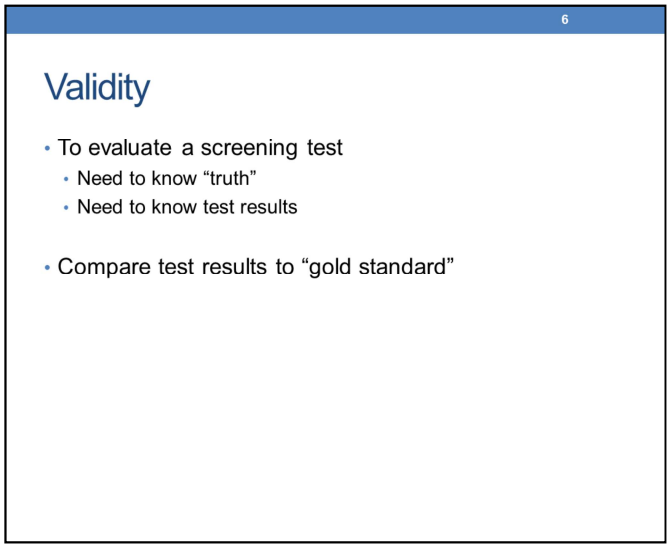


We will first focus on validity, or accuracy, of the screening test.

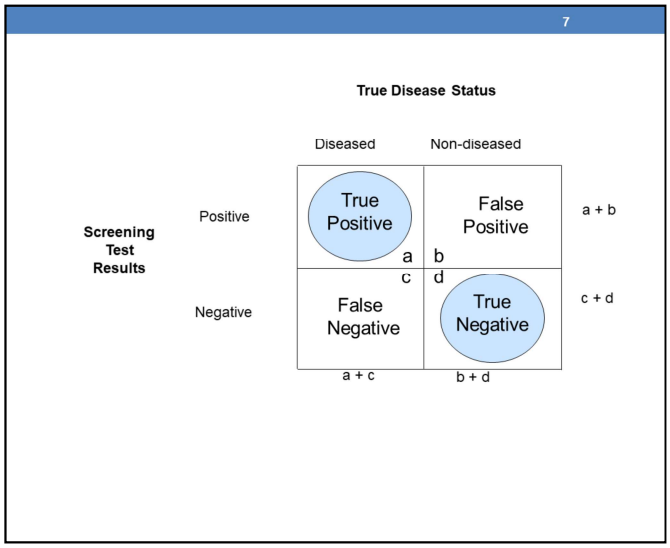
Accuracy relates to the ability of the test to distinguish between those who have the disease and those who do not have the disease.

Sensitivity quantifies the ability of the test to correctly identify those who have the disease. Sensitivity is the probability of a disease case being correctly identified by the test. Or in other words, the probability that a person with the disease will test positive based on the screening test.

Specificity quantifies the ability of the test to correctly identify those who do NOT have the disease. Specificity is the probability of a person without the disease being correctly identified by the test. Or in other words, the probability that a person without the disease will test negative based on the screening test.



In order to quantify the accuracy or validity of a screening test, we need to know the "true" disease state of the patient. The "true" state is typically indicated by the gold standard test. Then, we need to know the test results from the index screening test. To quantify the accuracy, we will compare the results of the index screening test to the results from the gold standard test.



This 2 by 2 table is helpful for summarizing our measures of accuracy.

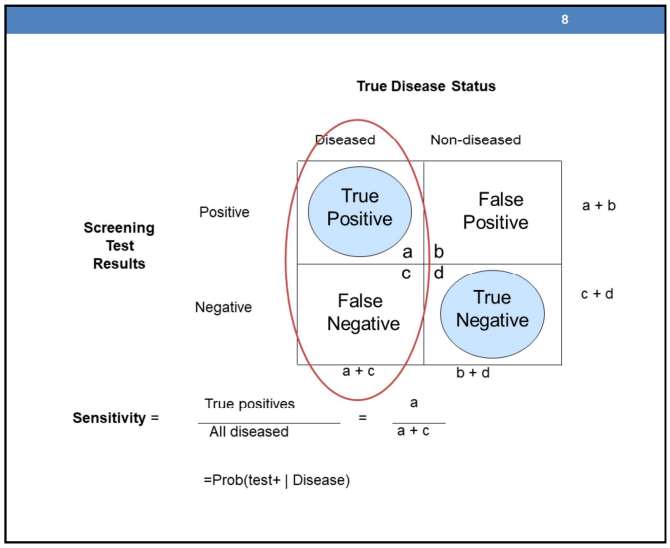
The columns correspond to the true disease status of the patient based on the gold standard test. There are a total of (a+c) diseased participants and (b+d) participants without disease. In addition to the true status of the patients, we also know the index test results. The index screening test classifies participants as positive or negative based on the screening test results. In terms of our notation, we have (a+b) participants with positive screening test results and (c+d) participants with negative screening test results.

For some participants, there is agreement between the test results and the true disease status of the participant.

- Participants who have the disease and have a positive screening test result are true positives.
- Participants who do not have the disease and have a negative screening test result are true negatives.

For some participants, the screening test result does not agree with the true disease status of the participant.

- Participants who do not have the disease but have a positive screening test result are false positives.
- Participants who have the disease but have a negative screening test result are false negatives.

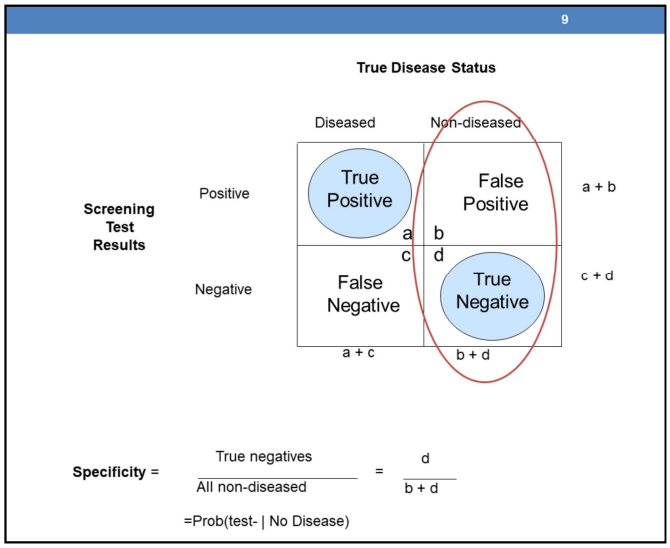


Now, let's use this 2 by 2 grid to define our quantitative measures of accuracy.

We defined sensitivity as the probability of a positive screening test among those who truly have the disease.

Based on our notation, there are (a+c) participants who truly have the disease. The denominator for sensitivity is therefore (a+c). Then, the numerator is the number of diseased participants who screen positive. In terms of our notation, this is (a). Therefore, sensitivity is calculated as (a)/(a+c).

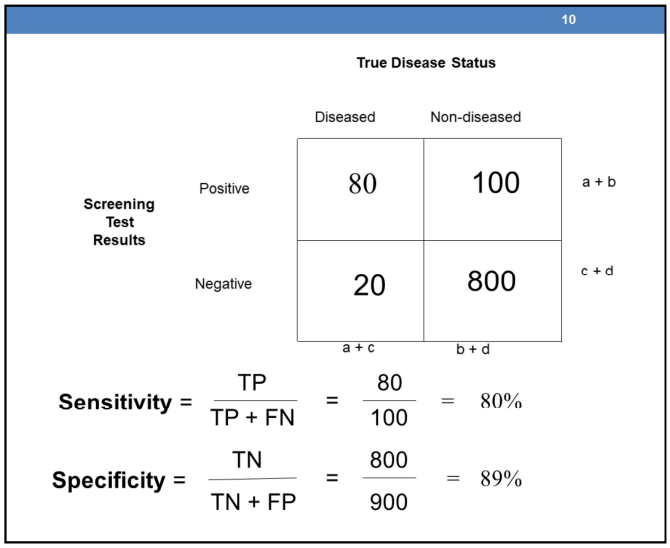
In terms of probability notation, sensitivity is the probability of a positive test among, or conditioned, on those truly having the disease. The (|) notation is a symbol for a conditional probability or a probability calculated among a subgroup or among participants who meet a certain condition, in this case, conditioning on those who truly have disease.



We defined specificity as the probability of a negative screening test among those who truly do NOT have the disease.

Based on our notation, there are (b+d) participants who truly do NOT have the disease. The denominator for specificity is therefore (b+d). Then, the numerator is the number of nondiseased participants who screen negative. In terms of our notation, this is (d). Therefore, specificity is calculated as (d)/(b+d).

In terms of probability notation, specificity is the probability of a negative test among, or conditioned, on those truly without the disease. The (|) notation is a symbol for a conditional probability or a probability calculated among a subgroup or among participants who meet a certain condition, in this case, conditioning on those who truly do not have disease.

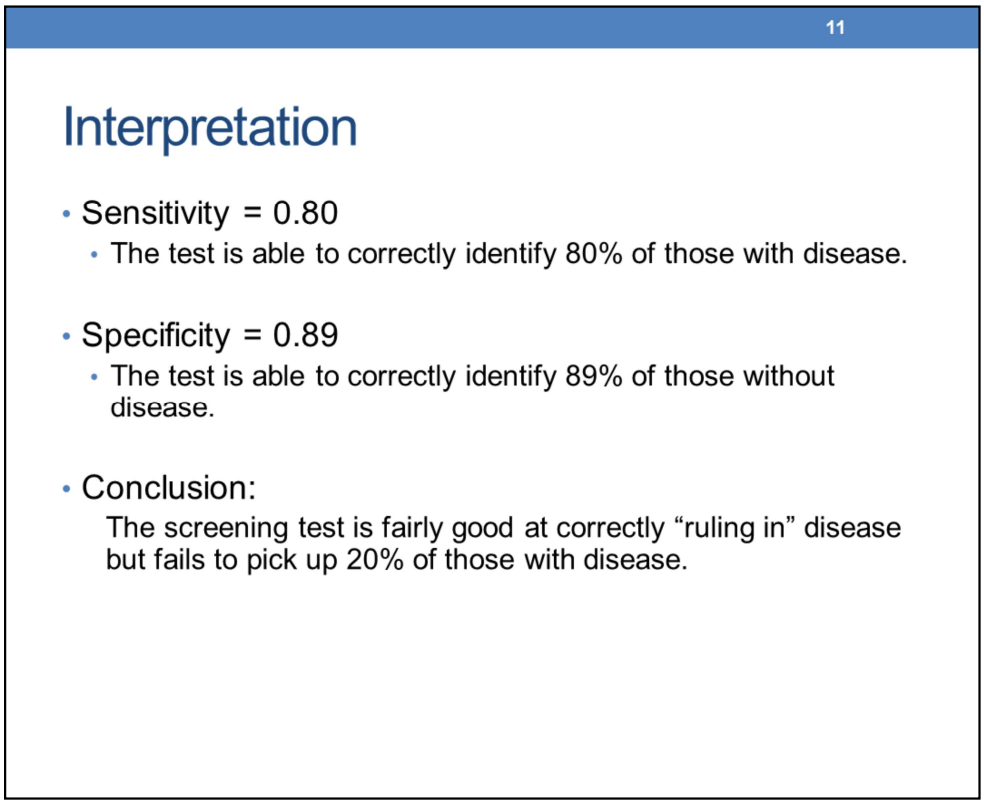


Now, let's consider a data example.

In this case, we have 1000 participants, among whom, 900 are non-diseased and 100 are diseased based on the gold standard assessment. Our screening test identifies 180 participants as positive and 820 participants as negative.

To calculate sensitivity, we focus on the 100 participants who truly are diseased. Among these, 80, or 80%, screened positive based on our test. The sensitivity is 80%.

To calculate specificity, we focus on the 900 participants who truly are non-diseased. Among these, 800, or 89%, screened negative based on our test. The specificity is 89%.



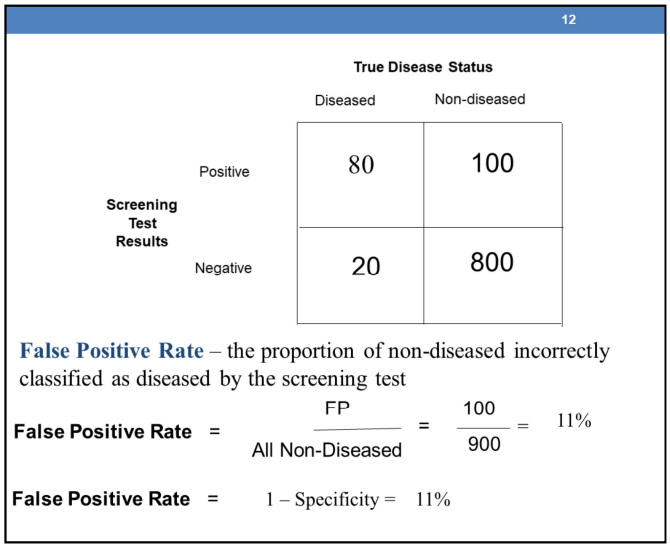
After calculating the values, we can interpret the results.

The calculated sensitivity is 80%. This means that the test is able to correctly identify 80% of those with the disease.

The calculated specificity is 89%. This means that the test is able to correctly identify 89% of those without the disease.

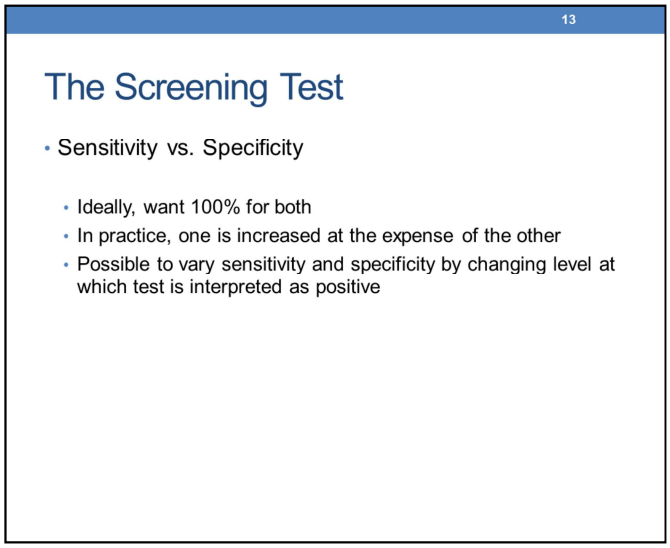
In conclusion, the screening test is fairly good at correctly identifying as negative those without disease (false positive is low relative to true negative); however, the screening test fails to pick up 20% of those with disease.

Given that the false positive error is low, we can state that this test is fairly good at "ruling in" disease.



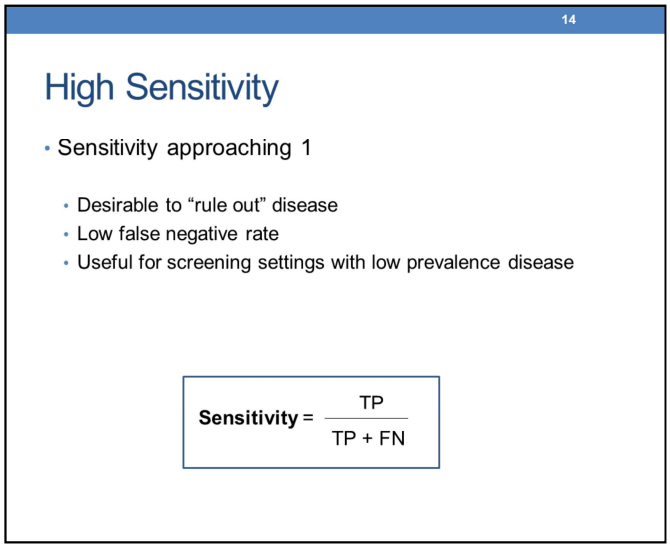
We can also calculate the false positive rate, which is defined as the proportion of the truly non-diseased who are incorrectly classified as diseased by the screening test.

The denominator is the total number of non-diseased participants (b+d) and the numerator is the number of false positives (b). We can show that the false positive rate is equal to 1 minus the specificity. In our data example, specificity was 89% so the false positive rate is 11%. Or, calculating this quantity directly, the false positive rate is 100 false positives divided by 900 participants without disease.

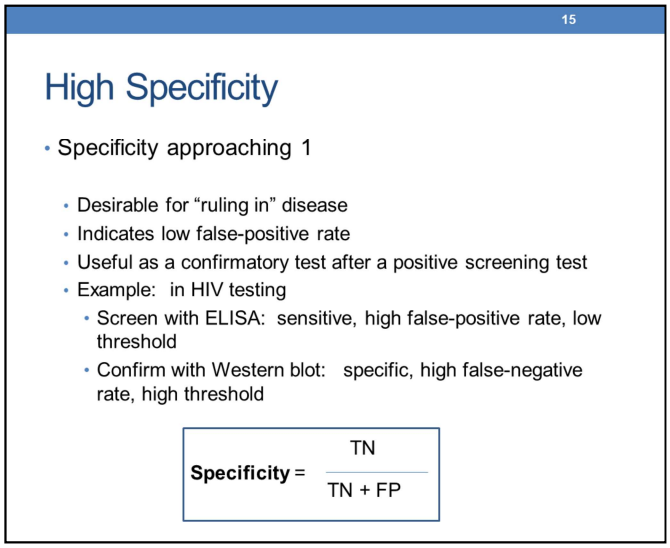


In practice, we want the sensitivity and the specificity both to be high; however, we are often faced with a situation where one characteristic is increased at the expense of the other. For example, if we are particularly concerned about false positive results, that lead to unnecessary, expensive diagnostic testing and cause a large amount of stress to the patient, we may want to target a test that is more specific at the expense of having a lower level of sensitivity (a higher false negative rate).

In practice, it is possible to vary sensitivity and specificity by changing the level at which a test is interpreted as positive. For example, changing the cut-point of an oral glucose tolerance test that is used to screen for gestational diabetes.

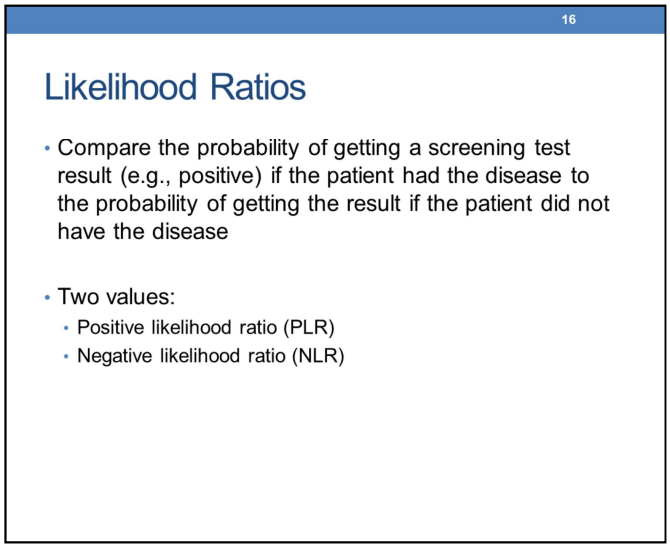


A test with high sensitivity is one that has sensitivity that approaches 1. Because sensitivity is high, the false negative fraction will be low compared to the true positive fraction and therefore, a highly sensitive test is useful for ruling out disease (i.e., the false negative fraction is low so a negative test is likely a true negative and not a false negative). A sensitive test is particularly useful for screening tests in settings with a low disease prevalence.



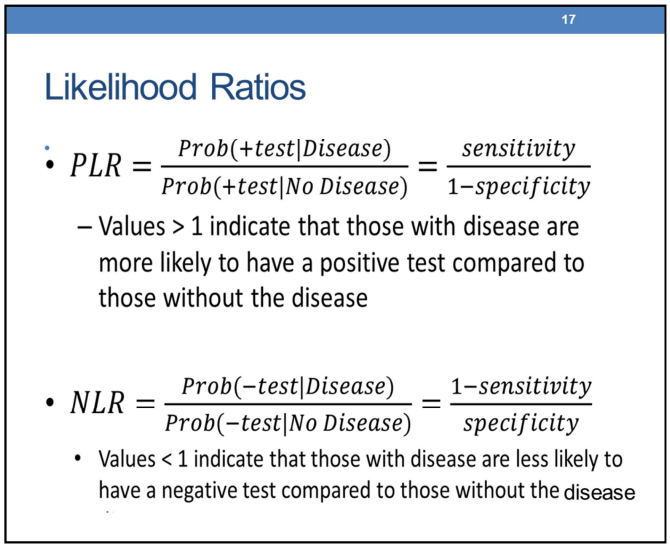
A test with high specificity is one that has specificity that approaches 1. Because specificity is high, the false positive fraction will be low compared to the true negative fraction and therefore, a highly specific test is useful for ruling in disease (i.e., the false positive fraction is low so a positive test is likely a true positive and not a false positive). A specific test is particularly useful as a confirmatory test after a positive screening test.

For example, in HIV testing, we initially use a screening ELISA test that is sensitive but has a high false-positive rate and therefore, low specificity. Then, this low specificity screening ELISA test is followed by a confirmatory Western blot test, useful for ruling in disease, that has high specificity and a low false positive rate, but is lower in terms of sensitivity and has a high false negative rate.



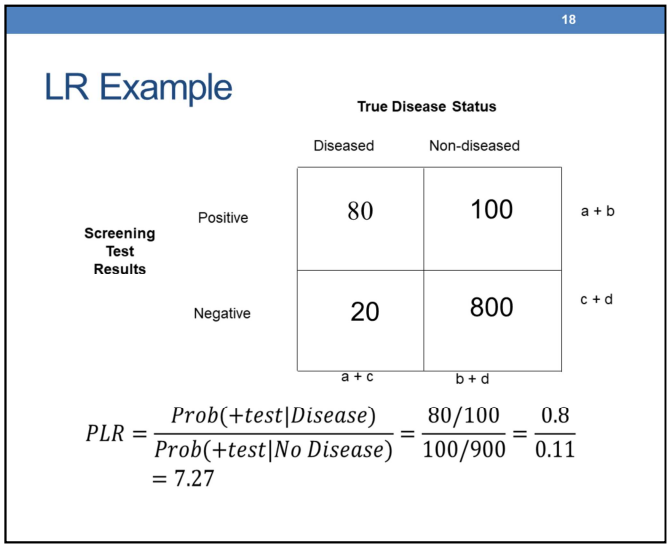
Now, let's discuss a new series of calculations that can be used to define the probability of receiving a particular screening test result, say a positive result, if the patient has the disease compared to the probability of receiving a particular screening test result, again using a positive result as an example, if the patient does NOT have the disease.

We will discuss two such values, a positive likelihood ratio and a negative likelihood ratio value.



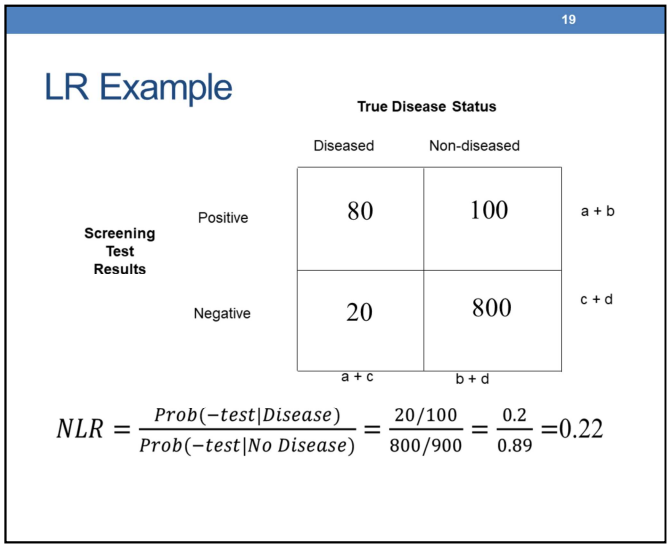
The positive likelihood ratio (PLR) value is the probability of a positive test among the disease participants divided by the probability of a positive test among non-diseased participants. A value greater than 1indicates that those with the disease are more likely to have a positive test compared to those without the disease.

The negative likelihood ratio (NLR) value is the probability of a negative test among the disease participants divided by the probability of a negative test among non-diseased participants. A value less than 1indicates that those with the disease are less likely to have a negative test compared to those without the disease.

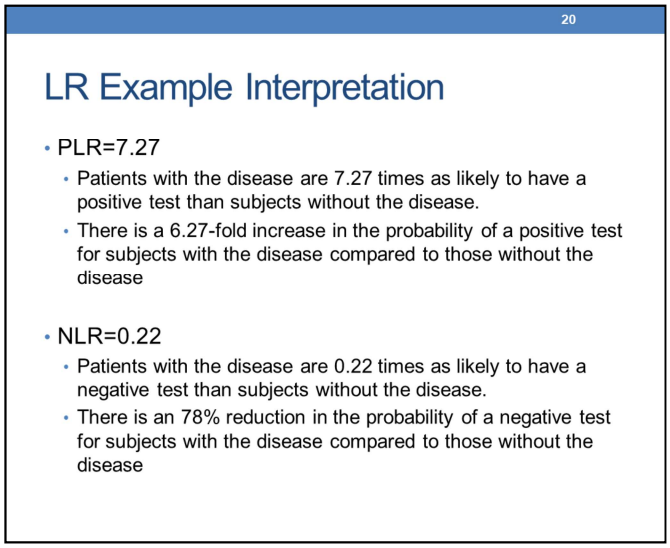


Let's consider our same data example.

In this case, the positive likelihood ratio value is the probability of a positive test among those with disease (0.8) divided by the probability of a positive test among those without disease (0.11) resulting in a value of 7.27.



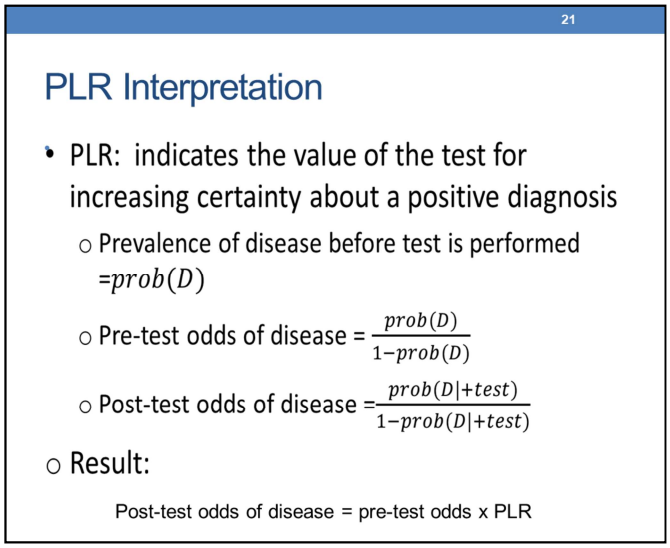
Using the same data example, the negative likelihood ratio value is the probability of a negative test among those with disease (0.2) divided by the probability of a negative test among those without disease (0.89) resulting in a value of 0.22.



Now, let's interpret the calculated values.

A positive likelihood ratio value of 7.27 means that patients with the disease are 7.27 times as likely to have a positive test result compared to participants without the disease. In other words, there is a 6.27-fold increase in the probability of a positive test result for participants with the disease compared to those without the disease.

A negative likelihood ratio value of 0.22 means that patients with the disease are 0.22 times as likely to have a negative test result compared to participants without the disease. In other words, there is a 78% reduction in the probability of a negative test for subjects with the disease compared to those without the disease.



It is helpful to note that the positive likelihood ratio is related to the post-test odds of disease.

The positive likelihood ratio indicates the value of the test for increasing certainty about a positive diagnosis.

If we consider a particular disease with prevalence in the population of Prob(D), we would predict that a given patient has the disease with the probability equal to the prevalence of disease. With no screening test or diagnostic testing information available, we would predict that everyone has the same probability of disease, which is equal to the overall prevalence of disease. If a screening test is useful, however, we can improve our prediction of disease using information from the screening test.

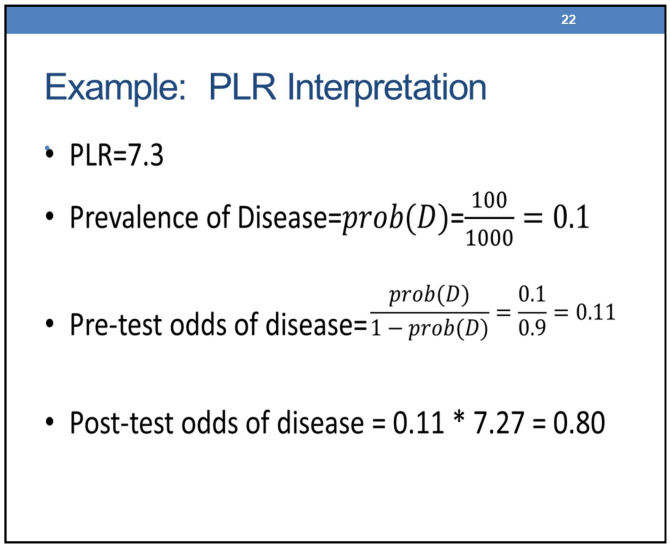
We will define the pre-test odds of disease as the probability of disease divided by (1 minus the probability of disease). This utilizes the definition of odds (p divided by [1-p]) and the estimated probability of disease based on the disease prevalence alone.

Then, if we have a useful screening test result, we can improve our prediction of disease and calculate the post-test odds of disease as the odds of disease conditioned on, or among those, with a positive test result.

An algebraic result is that the post-test odds of disease is equal to the pre-test odds of disease multiplied by the positive likelihood ratio.

If the test is useful, the post-test odds of disease, determined among those who screen positive, should be higher than the pre-test odds of disease, which is

calculated for the entire population without regard to the test result.



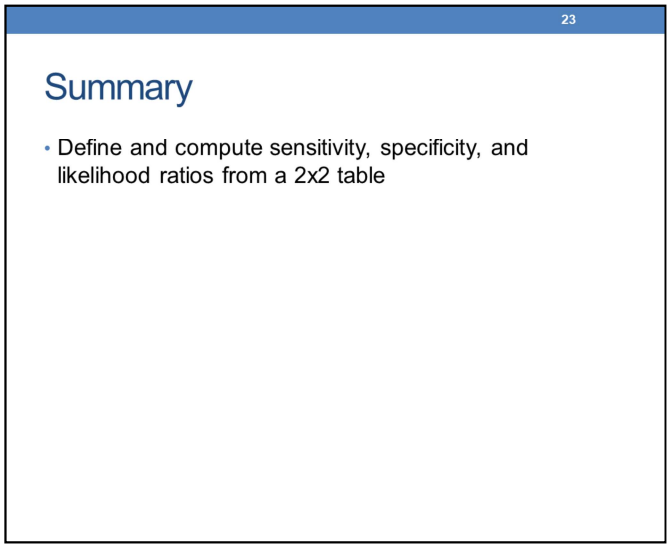
If we return to our data example, the positive likelihood ratio value is 7.3.

The prevalence of disease is 100 (true disease cases) divided by 1000 participants and is equal to 0.10.

The pre-test odds of disease is 0.1/0.9 = 0.11. The pre-test odds of disease reflects only the overall prevalence of disease and no additional information.

The post-test odds of disease is found by multiplying the pretest odds of disease (0.11) by the positive likelihood ratio (7.3), resulting in a value of 0.80.

Given a positive test result, the odds of disease are much higher than what we would predict based on prevalence alone.



In conclusion, we have learned how to calculate sensitivity, specificity, and the positive and negative likelihood ratio values as measures of accuracy. In the next section of this module, we will discuss approaches for calculating accuracy measures for a diagnostic test with a continuous result, for example, the value from an oral glucose tolerance test, and will learn about positive and negative predictive values that can also aid in the evaluation of a diagnostic or screening test.

References

- Wilson, J.M.G. & Jungner, G. (1968). Principles and practice of screening for disease. *WHO Chronicle, 22(11),*473.
- Altman DG & Bland JM (1994). Diagnostic Tests 1: sensitivity and specificity. *BMJ.* 308: 1552
- Altman DG & Bland JM (1994). Diagnostic Tests 2: predictive values. *BMJ.* 309: 102
- Altman DG & Bland JM (1994). Diagnostic Tests 3: receive operating characteristic plots. *BMJ.* 309: 188
- Deeks JJ & Altman DG (2004). Diagnostic Tests 4: likelihood ratios. *BMJ.* 329: 168-169