Factsheet

Understanding the Accuracy of Diagnostic and Serology Tests: Sensitivity and Specificity

Johns Hopkins Bloomberg School of Public Health | Center for Health Security

Rachel West, PhD and Amanda Kobokovich, MPH December 2020

Diagnostic and serology tests are crucial tools to fight the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and other infectious disease pathogens. Understanding a test's **sensitivity** and **specificity** is crucial for evaluating its accuracy and potential use.

Defining "sensitivity," "specificity," and "false results"

Sensitivity measures the proportion of positive test results out of all truly positive samples. In other words, a test's sensitivity is its ability to correctly identify those with the disease (the true positives) while minimizing the number of false negative results.

Specificity measures the proportion of negative test results out of all truly negative samples. In other words, a test's specificity is its ability to correctly those without the disease (the true negatives) while minimizing false positive results.

False results are also known as testing errors. The consequences of a testing error—a false positive or a false negative—are not equivalent. A false positive may prevent an individual from returning to work, while a false negative might lead to more disease transmission because the patient and their doctor believe the patient to be noninfected. Therefore, understanding sensitivity, specificity, and how test performance is influenced by disease prevalence is important in any testing strategy.

How sensitivity and specificity affect test manufacturing and use

The higher the values of a test's sensitivity and specificity (each out of 100%), the more accurate the test is in diagnosing a disease or condition.

Manufacturers may use other ways to describe sensitivity and specificity:

- Positive percent agreement (sensitivity) and negative percent agreement (specificity): When a nonreference standard—such as samples tested with an emergency use authorization-granted molecular diagnostic—is used when evaluating the test. In other words, this approach measures how many positives/negatives a test identifies that are in agreement with another test used on the same samples.

- Limit of detection: Sensitivity may also be measured by calculating the lowest detectable number of virus copies that will allow the test to return a positive result at least 95% of the time. A lower limit of detection indicates a more sensitive test, with fewer viral copies per sample necessary to produce a positive test result.

Even if a test has high sensitivity and specificity (close to 100%), the prevalence of disease, user-to-user variation, and sample type can impact the accuracy of a test. These values should be one of many measures we use to identify and trace those impacted by COVID-19.

Understanding the Accuracy of Diagnostic and Serology Tests: Sensitivity and Specificity

Johns Hopkins Bloomberg School of Public Health | Center for Health Security

How sensitivity and specificity are used to determine the value of a test

Sensitivity and specificity are calculated to produce a predictive value. The basic matrix is shown below:

	SARS-CoV-2 Positive	SARS-CoV-2 Negative
Positive test result	True Positive (TP)	False Positive (FP)
Negative test result	False Negative (FN)	True Negative (TN)

Sensitivity equation: Sensitivity = (TP) / (TP+FN)
Specificity equation: Specificity = (TN) / (TN+FP)

For instance, for a test with 95% sensitivity, that means that 95% of results will be true positives and 5% will be false negatives. A specificity of 93% means that 93% of all true negatives will test negative, with 7% falsely testing positive. Below, the columns are the count of true positives and negatives, while the rows are the test results for each group. There are 100 true positives, and 100 true negatives, shown here for a group of 200 samples:

	SARS-CoV-2 Positive	SARS-CoV-2 Negative	Total
Positive test result	95 true positives	7 false positives	102
Negative test result	5 false negatives	93 true negatives	98
Total	100	100	200

Sensitivity = (95) / (95+5) = 95%
Specificity = (93) / (93+7) = 93%

Understanding the Accuracy of Diagnostic and Serology Tests: Sensitivity and Specificity

How to interpret predictive value

Positive predictive value (PPV) and *negative predictive value* (NPV) provide insight into how accurate the positive and negative test results are expected to be in a given population, based on test accuracy and existing disease prevalence. PPV is the proportion of true positive tests out of all positive test results, and it is often used as a measure of the accuracy of tests. Mathematically, that is PPV= (TP)/(TP+FP).

The predictive value depends upon the prevalence of disease in a population. As the prevalence of disease increases (that is, true positives are more common), the likelihood of a false positive decreases. Therefore, predictive value can change over time or in different places, while sensitivity and specificity do not change—as these are characteristics of the test itself. Both PPV and NPV are dependent on the proportion of the test population that has the disease (prevalence).

Example:
The same diagnostic test is used in 2 different places -- City A and City B. City A has a low disease prevalence and City B has a high disease prevalence. In this case:

- $PPV_{Pop\ A} < PPV_{Pop\ B}$ because the number of false positives would be a higher percentage of the total number of positive tests in the population of City A.
- $NPV_{Pop\ A} > NPV_{Pop\ B}$ because the number of false negatives would be a higher percentage of the total number of negative tests in the population City B.

Using sensitivity, specificity, and predictive value to make public health decisions

There are 4 categories of test results: true positives, false positives, true negatives, and false negatives. Tests with different sensitivity and specificity values lead to different proportions of these 4 outcomes, which have consequences for making policy and public health decisions.

Example:
In a city with a population of 1 million people, 15% (or 150,000) of the population is infected with SARS CoV-2. The rest of the population (850,000) are not infected. A molecular diagnostic test (see table below) is administered to everyone in the city.

The test has a sensitivity of 95%—meaning the test will accurately provide a positive result for 95% of the infected individuals and 5% will be missed. The test also has a specificity of 95%—meaning that the test will accurately provide a negative result for 95% of the uninfected individuals, and 5% will test positive.

Based on these conditions:

- 185,000 individuals will test positive
  - 142,500 will be true positives (95% of the infected 150,000)
  - 42,500 (5% of the uninfected 850,000) of those will be false positives. Thus 23% (42,500 out of 185,000) of those testing positive are *not actually infected*.
- 815,000 individuals will receive negative test results, but 7,500 (5% of 150,000) of those will be false negatives—these individuals *are actually infected*.

	SARS-CoV-2 Positive	SARS-CoV-2 Negative	Total	
Positive test result	142,500	42,500	185,000	Population: 1 million Prevalence of 15%: 150,000 infected Sensitivity: 95% Specificity: 95%
Negative test result	7,500	807,500	815,000	False positives: 42,500 False negatives: 7500
Total	150,000	850,000	1 million	Percent of positive tests that were inaccurate: 23% (42,500/185,000) Positive Predictive Value: 77%

While the overall accuracy is high (95%), and 95% of the total population will receive accurate results, both positive and negative, 23%—almost 1 in 4—of those testing positive are not truly positive. The PPV of the test is 77%, even though the sensitivity and specificity are both 95%.

This example illustrates the PPV of the test—only 77% of the positive results will be accurate. Furthermore, when the prevalence of the disease is low—for example, 1% to 2%, as in the beginning of an epidemic or outbreak—the PPV is even lower. In other words, the value of sensitivity and specificity on paper is not the final measure of test accuracy. Predictive values must also be considered when developing any testing strategy, whether to inform individual decision making or community reopening efforts. Understanding the meaning of results and limitations of each test is critical to using testing as a tool to make policy or operational decisions.