FROM DATA TO DECISIONS

Measures of validity

@ISTOCK.COM/OEZ

BY DANIEL BRONSON-LOWE, PhD, CIC, FAPIC, AND CHRISTINA BRONSON-LOWE, PhD, CCC-SLP, CLD

WELCOME TO THE eighth installment in a series examining statistical concepts relevant to the field of infection prevention. This article continues the discussion from the last two issues around the roles that ratios can play in making data more useful.

The first positive rapid influenza test in the middle of the off-season didn't draw much attention. However, when two more positive results were reported the next day—from long-term inpatients, no less—the questions started coming. At the top of the infection preventionist's list was a deceptively simple-sounding one: How good is the influenza test that is being used?

Figure 1. HAI Lab Test Results

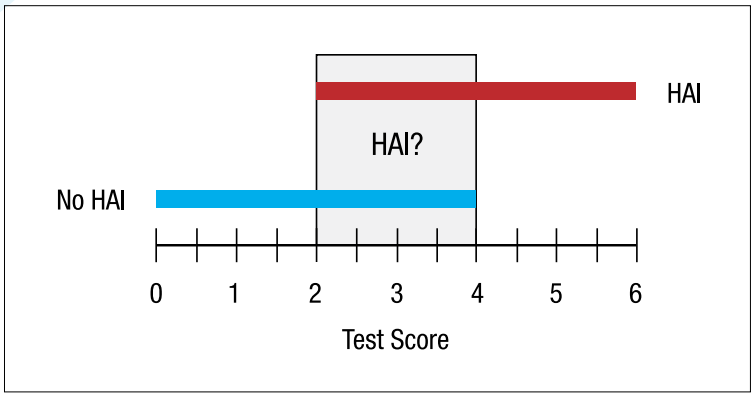


Figure 2. Lab Test Interpretation for 100% Sensitivity

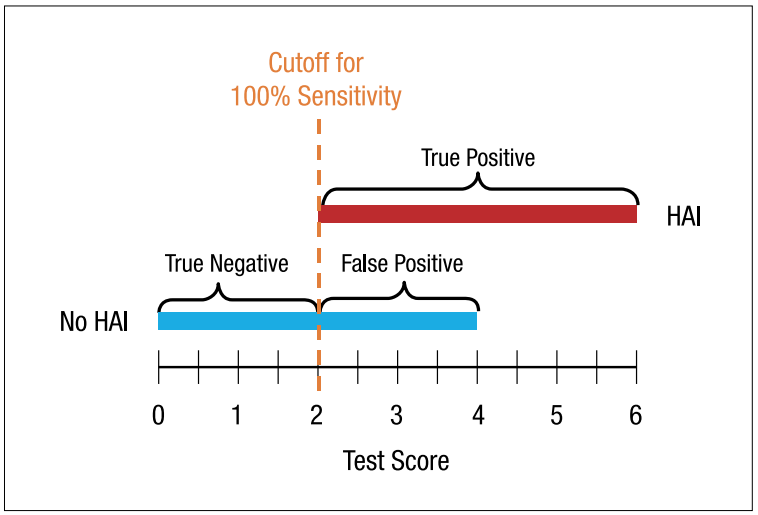
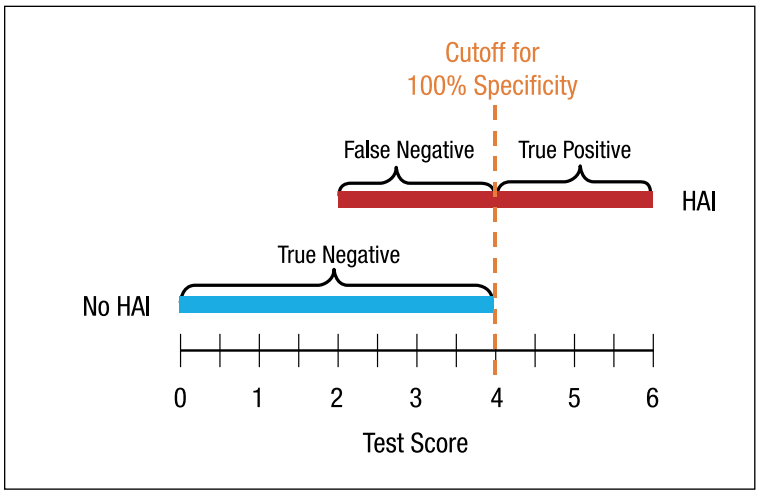


Figure 3. Lab Test Interpretation for 100% Specificity



This question splits into several others. Fortunately, there are established proportions used to answer them.

- **Sensitivity:** If someone *has* the disease, what is the likelihood the test will be *positive*?
- **Specificity:** If someone *does not have* the disease, what is the likelihood the test will be *negative*?
- **Positive predictive value** (PPV): If the test result is *positive*, what is the likelihood that the person truly *has* the disease?
- **Negative predictive value** (NPV): If the test result is *negative*, what is the likelihood that the person truly *does not have* the disease?

SENSITIVITY AND SPECIFICITY

Let's begin with sensitivity and specificity. These are characteristics of the test or tool being used to determine whether someone has a given disease or outcome. They describe aspects of the test's validity, establishing how well it measures what it's supposed to measure. To see how they work, imagine we have developed a new lab test to quickly identify patients with healthcare-associated infections (HAIs). The test result is a score between 0 and 6 (Figure 1). There is just one problem. If someone has an HAI, they will have a score greater than 2. If they do not have an HAI, that person will have a score less than 4. Therefore, between scores 2 and 4, it's unclear whether an HAI is present. We must decide how to handle those scores.

If we set the cutoff score for a positive result at 2, then we'll classify any score of 2 or greater as an HAI. Our test will then successfully identify every HAI because all HAIs have a score of 2 or greater. The test will therefore have 100% sensitivity. However, because some non-HAI patients also have test results greater than 2, using this cutoff will generate some false-positive results (Figure 2).

On the other hand, if we set the cutoff score for a positive result at 4, then we'll only classify scores of 4 or greater as HAIs. Note that we haven't changed the physical analysis our test is using, only the way we interpret its results. With this new cutoff, we'll lose some sensitivity because we'll miss some HAIs—any with a score less than 4—but now the test will have 100% specificity. It will only read positive if we are absolutely certain the patient has an HAI (Figure 3).

If we now imagine moving the cutoff line back and forth between scores of 2 and 4, it becomes clear that the test can never be 100% sensitive and 100% specific for the same cutoff. A test might get close to achieving that if the gray overlapping area is narrow, but any shift toward a given end of the gray area increases one measure while decreasing the other (Figure 1).

Table 1. Test Result versus Actual Disease Status

		Actual Disease Status		Total Test Results
		Positive	Negative	
Test Result	Positive	True Positive (TP)	False Positive (FP)	Number of Individuals with Positive Results (TP + FP)
	Negative	False Negative (FN)	True Negative (TN)	Number of Individuals with Negative Results (FN + TN)
Total People Tested		Number of Individuals with Disease (TP + FN)	Number of Individuals without Disease (FP + TN)	Number of Individuals Tested (TP + FP + TN + FN)

In real life, even if a test had 100% sensitivity or specificity, that wouldn't necessarily make it a good test. To understand why, let's examine how those values are calculated. Table 1 shows the logical framework underlying the math in the following equations. You may find it helpful to use this table and Figures 2 and 3 to track where the numbers for each equation are taken from and how they relate to each other.

Recall that sensitivity answers the question "If someone has the disease, what is the likelihood the test will be positive?" That means we need to know what proportion of the time the test is positive when disease is present:

$$Sensitivity = \frac{Number\ of\ True\text{-}Positive\ Results}{Number\ of\ Individuals\ with\ Disease} \times 100\%$$

To be 100% sensitive, the test must correctly identify all patients with the disease. Note, however, that false-positive results aren't included in this formula. That means that sensitivity isn't impacted by the presence of those errors; therefore, a test that gives a positive result every time it is used will have a sensitivity of 100%. That doesn't mean it's a useful test.

Specificity addresses the question "If someone does not have the disease, what is the likelihood the test will be negative?" It uses the proportion of the time the test is negative when disease is not present:

$$Specificity = \frac{Number\ of\ True\text{-}Negative\ Results}{Number\ of\ Individuals\ without\ Disease} \times 100\%$$

To achieve 100% specificity, the test must correctly identify all patients who are disease-free. This time, false-negative results aren't part of the calculation and therefore have no impact. Consequently, it is possible for a test to have a specificity of 100% if all it does is give a negative result every time it is used.

PPV AND NPV

Now that we've covered two characteristics of the test itself, let's go over two measures that describe how accurate the test results are: PPV and NPV. PPV is the proportion of positive test results that are real and answers the question, "If the test result is positive, what is the likelihood that the person truly has the disease?" In other words, if a patient has a positive result, how worried should he or she be?

$$PPV = \frac{Number\ of\ True\text{-}Positive\ Results}{Number\ of\ Individuals\ with\ Positive\ Results} \times 100\%$$

To achieve a PPV of 100%, every positive test result must be accurate; it must be someone who actually has the disease. The catch with PPV is that it doesn't take into consideration the total number of people

"To achieve 100% specificity, the test must correctly identify all patients who are disease-free. This time, false-negative results aren't part of the calculation and therefore have no impact."

who have disease. It's possible for a test to only identify a few of the total diseased individuals as positive and still end up with a PPV of 100% as long as it doesn't generate any false positive results.

NPV works similarly, providing the proportion of negative test results that are real and answering the question "If the test result is negative, what is the likelihood that the person is really free of disease?" If a patient has a negative test result, how reassured should they be?

$$NPV = \frac{\textit{Number of True-Negative Results}}{\textit{Number of Individuals with Negative Results}} \times 100\%$$

To achieve an NPV of 100%, every negative test result must be a true negative. The total number of people without disease doesn't play a role in this equation; therefore, a test can achieve an NPV of 100% while accurately identifying only a portion of the disease-free individuals. It requires only that there be no false-negative results; false-positive results don't matter here.

Let's take an example of a rapid influenza test with a technical write-up that reads as follows:

- Sensitivity: 80%
- Specificity: 91%
- PPV: 85%
- NPV: 87%

This translates to:

- Sensitivity: If a person has influenza, the test result will be positive 80% of the time.
- Specificity: If a person does not have influenza, the test result will be negative 91% of the time.
- PPV: If the test result is positive, the patient will actually have influenza 85% of the time.
- NPV: If the test result is negative, the patient will truly not have influenza 87% of the time.

With that in mind, let's also say that the technical write-up includes a disclaimer that the documented PPV and NPV are only accurate if the test is used during influenza season. Why might that be the case? As it turns out, while sensitivity and specificity won't change based on the time of year—they are characteristics of the test itself and the test isn't changing—PPV and NPV are impacted by the prevalence of the disease in the population being tested. That means that during the off-season, when influenza is scarce, this flu test could have very different values for PPV and NPV, like so:

- PPV = 8%
- NPV = 99%

This happens because the more disease there is to find, the more likely it is that a positive result will be accurate. Figure 4 depicts the prevalence of influenza over time in an example community in the northern hemisphere, with very little disease present in the summer and a spike in the winter. Now imagine that the two circles in that figure are dartboards. Every time our rapid influenza test is positive during the off-season, we throw a dart at the dartboard on the left. If we hit that very small red bull's-eye, then our positive result is a true positive. If we hit the white area, the result is a false positive. The reason the bull's-eye is so small is that there is very little real influenza to find. The result is that our PPV is very small because most

of the time we get a positive result it is going to be a false positive.

When we use the same influenza test during the peak of the flu season, however, we get to use the dartboard on the right. The bull's-eye is now much larger because there is so much more influenza present in the community that a positive result is much more likely to be real. Thus, our PPV is much higher.

NPV works in reverse (Figure 5). Every time we get a negative result during the off-season, we throw a dart at the left target, and, if we hit the blue bull's-eye, then the result is a true negative. If there is very little disease in the community, then almost every negative result will be a true negative. In the middle of flu season, however, when a larger proportion of the population actually has the disease, the chance of a false negative increases and the bull's-eye shrinks a bit.

CONCLUSION

No test is perfect. Use these measures to better understand tests' strengths and limitations during test selection and interpretation. For example, if the risks associated with a missed diagnosis are significant, it is vital to identify patients with positive results, even if that means having some false positives; therefore, you should look for a test with high sensitivity. Alternately, if a false-positive result can have serious detrimental effects, such as stigmatizing patients or leading them to believe they don't have long to survive, then it may more important to find a test with high specificity and PPV.

If you have any questions or comments, please feel free to contact the authors at IPandEpi@gmail.com. PS

Daniel Bronson-Lowe, PhD, CIC, FAPIC, has been an infection preventionist, an infectious disease epidemiologist, and a statistics lecturer. He has been an instructor for APIC's "Basic Statistics for Infection Preventionists" Virtual Learning Lab and is a senior clinical manager with Baxter Healthcare Corporation.

Christina Bronson-Lowe, PhD, CCC-SLP, CLD, is a speech-language pathologist who has worked in hospitals, inpatient and outpatient rehabilitation, skilled nursing facilities, and home health care.

Additional resources

Potts, A. Use of statistics in infection prevention. In: Grota P, et al., editors. APIC text online. APIC; 2014.

Potts, A. Use of statistics in infection prevention. In: Pogorzelska-Maziarz M, editor. Fundamental statistics & epidemiology in infection prevention. Washington, D.C.: APIC; 2016. p. 18-53.

Figure 4. Variability of Positive Predictive Value

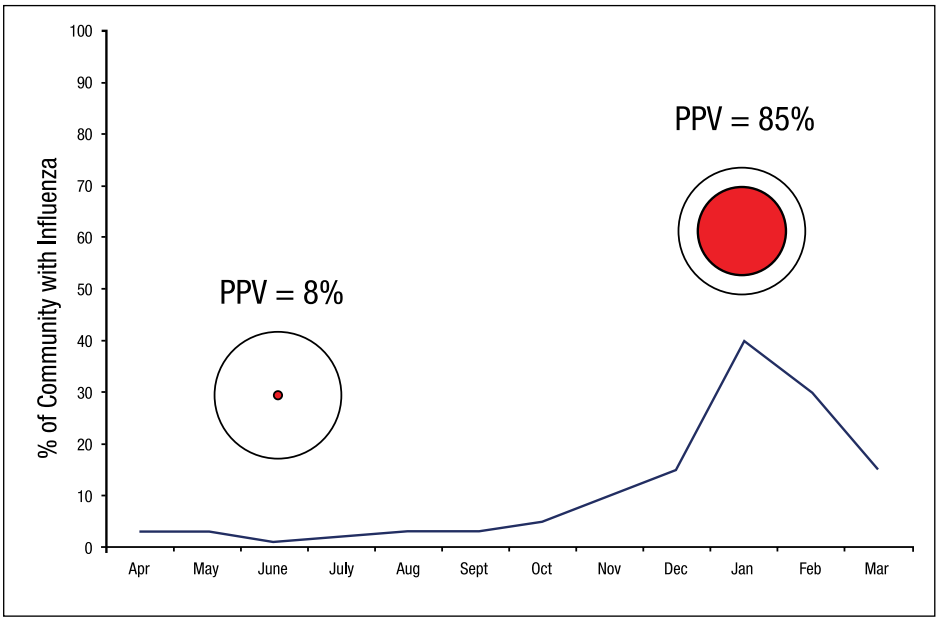
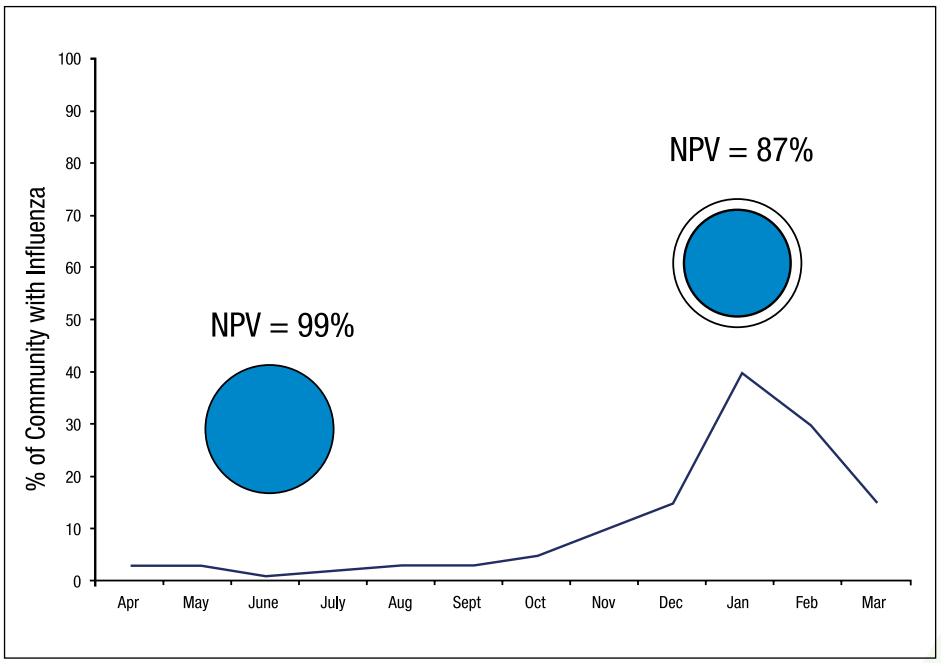


Figure 5. Variability of Negative Predictive Value



"NPV works in reverse (Figure 5). Every time we get a negative result during the off-season, we throw a dart at the left target, and if we hit the blue bull's-eye, then the result is a true negative."