

Project: Interpreting COVID Test Results- ANSWER KEY

Part 1: How Accurate are Covid Tests Overall?

In 2020-21, the global Covid-19 pandemic has highlighted, more than ever before, **the importance of modern statistical science!** Statisticians and data scientists have worked closely with physicians, epidemiologists, and public health policy makers to study every aspect of the virus's spread and search for effective vaccines and treatments. An essential part of their work has involved the development of accurate rapid Covid tests, so that infected patients can be identified in time to slow the spread of the disease.

At some point during the pandemic, you may have had to get a Covid-19 test; maybe you developed symptoms, or maybe you were tested as a precaution after exposure. Whether your test result was positive or negative, you may have wondered, **“How do I know the test result is correct?”**

Unfortunately, **no medical test can be 100% accurate.** While the Covid tests currently in use do have a high level of accuracy overall, there are many factors that could affect the likelihood of an individual getting an inaccurate Covid test result. These include viral load (the actual amount of virus in the body), how long after exposure someone is tested, and the quality of test design and testing equipment. In this project, we won't consider these factors or many others that affect real world Covid testing; instead, we're going to focus on a single, VERY important concept that is often misunderstood by patients, the media, and sometimes even healthcare workers! The concept is this: **what is the difference between “the probability of a positive Covid test, GIVEN that a person really has Covid” vs. “the probability that a person really has Covid, GIVEN that they have a positive Covid test”?**

At first glance, the difference between those two probabilities might not seem obvious or important; maybe they sound like they're closely related, or that knowing one will tell us the other. As it turns out, and as we'll see in this project, these probabilities are actually very different! Knowing one does not tell us the other, and knowing how to interpret your Covid test result in a meaningful way depends on understanding this distinction.

To start to explore these ideas, let's look at some published data concerning the ID NOW rapid Covid test manufactured by Abbott. In a press release published

Oct. 7, 2020, they state, “In an in-patient care study (hospitals and nursing homes), a total of 518 symptomatic patients were evaluated, including 94 PCR positive subjects. ID NOW demonstrated 79.8% positive agreement (sensitivity) and 94.3% negative agreement (specificity) compared to lab-based molecular PCR tests.”¹ What does all this mean? Let’s start by defining a couple of terms that test manufacturers routinely talk about when publishing data related to the effectiveness of their tests:

- **Sensitivity** is the probability that a test result will be positive when the infection really is present;
- **Specificity** is the probability that a test result will be negative when the infection is NOT present.

Another way to say this is that sensitivity is the proportion of infected patients that will be correctly identified by the test, and specificity is the proportion of non-infected patients that will be correctly identified by the test. In the 518 patients that were tested in this group, 94 actually had Covid (according to the highly reliable PCR testing that was done), and 424 did not have Covid. So a sensitivity of 79.8% and specificity of 94.3% lets us calculate the following numbers:

79.8% of 94 patients (with Covid) = 75 true positive test results.
94.3% of 424 patients (without Covid) = 400 true neg. test results.

	Has Covid	Doesn’t Have Covid	Total
Positive Test Result	75 (true positives)	24 (false positives)	99
Negative Test Result	19 (false negatives)	400 (true negatives)	419
Total	94	424	518

¹ https://abbott.mediaroom.com/2020-10-07-Abbott-Releases-ID-NOW-TM-COVID-19-Interim-Clinical-Study-Results-from-1-003-People-to-Provide-the-Facts-on-Clinical-Performance-and-to-Support-Public-Health#assets_2429_124429-111

Now let's recall what we learned about **conditional probability** to reframe the concepts of “sensitivity” and “specificity” in terms of the formula we learned:

- **Sensitivity** = probability that the test result is positive, GIVEN that the person has Covid = $\frac{\# \text{ true positive test results }}{\# \text{ patients with Covid }} = \frac{75}{94} \approx 79.8\%$
 - **Specificity** = probability that the test result is negative, GIVEN that the person does NOT have Covid = $\frac{\# \text{ true negative test results }}{\# \text{ patients without Covid }} = \frac{400}{424} \approx 94.3\%$
-

Now let's talk about some other commonly used terms in the world of medical testing, and practice computing them from the table above.

- **Accuracy** = percentage of test results that are accurate =
$$\frac{\# \text{ true positives } + \# \text{ true negatives }}{\# \text{ total patients tested }}$$
- **False positive rate** = percentage of non-infected patients receiving false positive test result =
$$\frac{\# \text{ false positive test results }}{\# \text{ patients without Covid }}$$
- **False negative rate** = percentage of infected patients receiving false negative test result =
$$\frac{\# \text{ false negative test results }}{\# \text{ patients with Covid }}$$

Problem 1: Using the numbers from the chart above, calculate the **accuracy**, **false positive rate**, and **false negative rate** for the ID NOW Covid test. Give your answer as both a **fraction** and a **percentage accurate to one decimal place**.

Accuracy = $475/518 \approx 91.7\%$

False positive rate = $24/424 \approx 5.7\%$

False negative rate = $19/94 \approx 20.2\%$

A few things to notice:

- In terms of the formulas, **false positive rate is just $1 - \text{specificity}$** and **false negative rate is just $1 - \text{sensitivity}$** .
- In terms of conditional probability, we can say that **false positive rate** = probability that the test result is a (false) positive, GIVEN that the person does NOT have Covid.
- Similarly, we can say that the **false negative rate** = probability that the test result is a (false) negative, GIVEN that the person DOES have Covid.

Part 2: How Should We Interpret a Specific Covid Test Result?

Notice that all the properties we've talked about so far (**accuracy**, **sensitivity**, **specificity**, **false positive rate**, and **false negative rate**) only depend on the quality of the test itself, and its ability to accurately detect Covid; they DON'T depend on the probability that a person coming in for a test actually has Covid. Specifically, this means that if we tested two groups- one with a very high proportion of Covid-positive patients, and one with a very low proportion- all of these test properties should (on average) be exactly the same in both groups. These test properties are all commonly discussed in the media, and published in medical literature as indicators of the usefulness and quality of a diagnostic test. Notice that accuracy means exactly what it sounds like- "what percentage of the patients tested will get accurate results?"- and sensitivity/ specificity just fill in more details ("what percentage of infected patients will get accurate results?" and "what percentage of non-infected patients will get accurate results?", respectively.)

So... if we know the accuracy, sensitivity, and specificity of a Covid test, we should be well-equipped to interpret a positive or negative test result... right?

WRONG!!

Let's think about why this is wrong. So far the rates we've looked at tell us,

- “**If I have Covid**, what's the chance my test will be accurate [or inaccurate]?”
- “**If I don't have Covid**, what's the chance my test will be accurate [or inaccurate]?”

But the problem is- at the point you take a test- you **DON'T KNOW** whether or not you have Covid. (That's why you're getting tested...!) You only know your test result. **So the questions you REALLY want to answer are:**

- **If my test result is positive**, what's the chance I really do (or don't) have Covid?
- **If my test result is negative**, what's the chance I really don't (or do) have Covid?

To answer **THESE** questions- the ones that are actually practically relevant for interpreting our Covid test results- we need to make use of two more formulas:

- The **positive predictive value (PPV)** (also called precision) is the probability that a patient actually has Covid, **GIVEN** that they have a positive test result:

$$PPV = \frac{\# \text{ true positive test results}}{\# \text{ total positive test results}}$$

- The **negative predictive value (NPV)** is the probability that a patient actually doesn't have Covid, **GIVEN** that they have a negative test result:

$$NPV = \frac{\# \text{ true negative test results}}{\# \text{ total negative test results}}$$

Problem 2: Calculate the PPV and NPV for the ID NOW Covid test, using the data from the sample group given above. Give your answer both as a fraction, and as a percentage accurate to one decimal place:

$$\text{PPV} = 75/99 \approx 75.8\%$$

$$\text{NPV} = 400/419 \approx 95.5\%$$

Here's the catch- unlike the properties we considered earlier, PPV and NPV WILL be different (sometimes very different!) for sample groups with different prevalence of Covid. ("Prevalence" generally means the rate of a health condition in the population, and here we're using it to mean the proportion of a sample group that has Covid.) To see why this is the case, **let's use what we've learned so far to consider what would happen if we administer the ID NOW Covid test to a group of 1000 patients, 500 of whom have Covid:**

- The sensitivity is still the same, at 79.8%, so there would be $(0.798)(500) = 399$ true positive test results. (The other 101 infected patients would get a false negative.)
- The specificity is still the same, at 94.3%, so there would be $(0.943)(500) = 472$ true negative test results. (The other 28 non-infected patients would get a false positive.)

	Has Covid	Doesn't Have Covid	Total
Positive Test Result	399 (true positives)	28 (false positives)	427
Negative Test Result	101 (false negatives)	472 (true negatives)	573
Total	500	500	1000

Problem 3: Calculate PPV and NPV again for the ID NOW test, this time using the hypothetical data in the chart above. Are your answers pretty close to the values you found in Problem 2, or significantly different?

$$\text{PPV} = 399 / 427 \approx 93.4\%$$

$$\text{NPV} = 472 / 573 \approx 82.4\%$$

These values are significantly different from the values in Problem 2.

Problem 4: Take a minute to review the definitions of the terminology we've learned so far: sensitivity, specificity, accuracy, false positive rate, false negative rate, positive predictive value, and negative predictive value. Suppose now that you're not sure whether or not you have Covid, and you take the ID NOW test. If you get a positive test result, which of these statistics would be the most relevant in helping you interpret your test result? How about if you got a negative test result?

If you get a positive test result, the PPV will be most relevant; it will answer "what's the probability I really have Covid?"

If you get a negative test result, the NPV will be most relevant; it will answer "what's the probability I really don't have Covid?"

Problem 5: Now suppose that another group of 1000 patients were given the ID NOW test, but this time, only 5% are actually infected with Covid. Using the same sensitivity and specificity values we used previously, fill in the chart below, and calculate PPV and NPV.

(Note that there are 50 patients with Covid (5% of 1000) and 950 patients without Covid.)

	Has Covid	Doesn't Have Covid	Total
Positive Test Result	40 (true positives)	54 (false positives)	94
Negative Test Result	10 (false negatives)	896 (true negatives)	906
Total	50	950	1000

$$\text{PPV} = 40 / 94 \approx 42.6\%$$

$$\text{NPV} = 896 / 906 \approx 98.9\%$$

ANSWER KEY

Working through these calculations may help you see where the formulas for PPV and NPV come from:

$$PPV = \frac{(prevalence)(sensitivity)}{(prevalence)(sensitivity) + (1 - prevalence)(1 - specificity)}$$

$$NPV = \frac{(1 - prevalence)(specificity)}{(1 - prevalence)(specificity) + (prevalence)(1 - sensitivity)}$$

Problem 6: Substitute sensitivity = 0.798, specificity = 0.943, and prevalence = 0.05 into the formulas above to confirm the values for PPV and NPV you found in Problem 5.

Using prevalence = 0.05, sensitivity = 0.798, specificity = 0.943, we have

$$PPV = \frac{(prevalence)(sensitivity)}{(prevalence)(sensitivity) + (1 - prevalence)(1 - specificity)} = \frac{(0.05)(0.798)}{(0.05)(0.798) + (0.95)(1 - 0.943)}$$

≈ 42.4% (note slight difference from previous answer due to rounding)

$$NPV = \frac{(1 - prevalence)(specificity)}{(1 - prevalence)(specificity) + (prevalence)(1 - sensitivity)} = \frac{(0.95)(0.943)}{(0.95)(0.943) + (0.05)(0.202)}$$

≈ 98.9%

In Problem 4 above, we considered the fact that PPV and NPV are the most relevant statistics in interpreting individual test results; but as we've just seen, the same test can have very different PPV and NPV values, depending on the

prevalence of disease in the population being tested. How can we translate the concept of “disease prevalence” to a single individual getting tested? We can interpret prevalence as the pre-test probability of the individual having Covid- the probability that they have Covid when they first walk in the door, before getting tested. Generally we wouldn’t know this pre-test probability as an exact number, but giving a rough estimate can give us a pretty good idea of PPV or NPV for an individual test result. An individual’s pre-test probability of a disease can depend on many factors, such as recent specific exposures, the overall disease prevalence in the community, and presence or absence of specific symptoms.

Here are the formulas for PPV and NPV again, this time replacing “prevalence” with “pre-test probability”:

$$PPV = \frac{(pre-test\ prob.)(sensitivity)}{(pre-test\ prob.)(sensitivity) + (1-pre-test\ prob.)(1-specificity)}$$

$$NPV = \frac{(1-pre-test\ prob.)(specificity)}{(1-pre-test\ prob.)(specificity) + (pre-test\ prob.)(1-sensitivity)}$$

Problem 7: Using a graphing calculator or spreadsheet, sensitivity = 0.798, and specificity = 0.943, compute PPV and NPV for different values of the pre-test probability. Give your answers as percentages accurate to one decimal place.

Pre-test prob.	PPV	NPV
0.01	12.4%	99.8%
0.02	22.2%	99.6%
0.05	42.4%	98.9%
0.10	60.9%	97.7%
0.25	82.4%	93.3%
0.50	93.3%	82.4%
0.75	97.7%	60.9%
0.90	99.2%	34.2%
0.99	99.9%	4.5%

Problem 8: What do you notice about PPV as the pre-test probability of disease increases? How about NPV?

As the pre-test probability of infection increases, PPV increases and NPV decreases.

Notice that this trend in PPV and NPV lines up with what our intuition tells us about interpreting our test results in light of our pre-test probability.

To use extreme examples to illustrate the point- suppose you attended a crowded indoor meeting four days ago, and later found out that 50 people at the gathering tested positive for Covid. Now you have a fever, severe respiratory symptoms, and have lost your sense of taste and smell. You'd probably be pretty likely to look at a positive Covid test result and say, "Yep, I'm not surprised; this test result is probably accurate." (On the other hand, if the test result were negative, you might be pretty suspicious it's a false negative.) In other words- a high pre-test probability of Covid corresponds to a high positive predictive value [a pos. test result is likely to be true] and a low negative predictive value [a neg. test result is likely to be false.]

Now suppose you've been literally living as a hermit since Feb. 2020, not exposed to any other people until the moment you go in to take a Covid test. You have no symptoms, but as a requirement of starting employment you have to get tested for Covid. If you see a negative Covid test result, how much confidence would you have in the accuracy of that result? (Probably a lot!) And conversely, a positive test result would likely make you think, "Hmmm... I bet this is a false positive.") In other words, a low pre-test probability of Covid corresponds to a low positive predictive value and a high negative predictive value.

Concluding Remarks:

*As you can see, problems involving conditional probability aren't just tedious exercises in your statistics book. These ideas have important real-world

applications, like accurately understanding information about diagnostic testing and health conditions. Mis-information about these ideas is widespread both in journalism and social media. Statistical literacy matters!

*The formulas we used for PPV and NPV are applications of **Bayes' Theorem**, an essential formula that relates conditional probabilities. Specifically, if we know “the probability of B given A” but we want to find “the probability of A given B,” Bayes' Theorem is usually the tool we need! Bayesian Inference is an extremely important area of statistical practice, both in medicine and many other areas of research.

*Covid is currently a very common/ widespread condition, meaning an individual's pre-test probability of disease is not likely to be extremely low. However, there ARE many conditions (like rare genetic disorders, cancers, etc) that do have extremely low pre-test probabilities, and not understanding PPV and NPV can cause a lot of needless concern or further testing. A test for a very rare condition might be 99% accurate overall, but still have an extremely low PPV- meaning that a positive test result is very likely to be a false positive. Now that you've completed this project, you are more empowered to be an informed patient and advocate for your own health, by being better equipped to understand the significance of test results.