

How to think about probabilities with Covid and Covid testing.

Prof. Matt McCormick

Department of Philosophy, CSUS

Answers to questions about what's going on in the world are probabilistic at best. We have imperfect measuring devices: our eyes, our brains, our ideologies, our medical tests, and so on. And the truth is hard to uncover out there. Here's some help on how to think about whether or not you've got covid and what a test results, positive or negative means.

First, the complication here is that we've got two different probabilities that can be working against each other: there is a probability that a person has the disease and there's the probability that is the accuracy of the test. When we combine those, we can get some counterintuitive results. Bayes' Theorem is what we use to clarify.

$$\Pr(H|O) = \frac{\Pr(O|H)\Pr(H)}{\Pr(O|H)\Pr(H) + \Pr(O|\sim H)\Pr(\sim H)}$$

Here, H stands for "hypothesis," O stands for "observation," and $\sim H$ stands for "the hypothesis is false." And the upright line | stands for "given that." So suppose we want to know, what's the probability that you've got covid (the hypothesis) given that you've got a positive test result (the observation): $\Pr(C|+)$. Here's how you'd substitute the variables into Bayes' Theorem.

$$\Pr(C|+) = \frac{\Pr(+|C)\Pr(C)}{\Pr(+|C)\Pr(C) + \Pr(+|\sim C)\Pr(\sim C)}$$

I won't go into an explanation of why Bayes' Theorem is true. Let's just understand that this formula, once you plug in the relevant values, gives you a probabilistic answer to the question "Do I have it?" in the case that you get a positive or negative test result.

Let's try some values. The right side of the numerator is just the base rate: $\Pr(C)$

$$\Pr(C|+) = \frac{\Pr(+|C)\Pr(C)}{\Pr(+|C)\Pr(C) + \Pr(+|\sim C)\Pr(\sim C)}$$

What's the probability that some person chosen from a population has covid? At the moment, in Sacramento county, where I live, there are about 10,000 cases. And there is a population of 500,000. So that's a 2% rate. Those are officially reported cases. The actual rate maybe 10 times higher, but we will deal with that in a moment.

Now, what is $\Pr(+|C)$?

$$\Pr(C|+) = \frac{\Pr(+|C)\Pr(C)}{\Pr(+|C)\Pr(C) + \Pr(+|\sim C)\Pr(\sim C)}$$

That is, what is the probability that you would test positive given that you've got Covid? This is the accuracy of the test. It's also sometimes called the *likelihood*. When someone has it, what percentage of the time does the test give a true positive result, and what percentage of the time is the test wrong. One of the problems we've been having through the pandemic is that our tests aren't very accurate. Let's suppose that we're using a test that is 85% accurate, or that $\Pr(+|C) = .85$.

The only other values we need to complete the formula now are $\Pr(+|\sim C)$ and $\Pr(\sim C)$.

$$\Pr(C|+) = \frac{\Pr(+|C)\Pr(C)}{\Pr(+|C)\Pr(C) + \Pr(+|\sim C)\Pr(\sim C)}$$

For now, since we are assuming that our test is .85 accurate, what is the probability that a person would have a positive test result but not have it? $\Pr(+|\sim C) = .15$. And if the base rate of C in the population is .02, then the $\Pr(\sim C)$, that is the percentage of people who don't have it, is .98. So now, we can fill in all the values:

$$\Pr(C|+) = \frac{(.85)(.02)}{(.85)(.02) + (.15)(.98)}$$

$$\Pr(C|+) = \frac{.017}{.164}$$

$$\Pr(C|+) = .1036$$

That is, with our assumed values, the probability that some random person selected out of the population, given a positive test result with 85% accuracy, is about 10%.

But wait, how can that be? A positive test result means you've got it, right? No, in this case it just means that the probability of someone's having it went from 2% to 10% with the positive test. Think of it this way. The test here is a pretty poor test. It only get the answer right 85% of the time. (If the test got it right 50% of the time, then it would be no better than flipping a coin to find out if you've got it.) And the other problem is that in this example, the base rate, the probability that some random person in the population has it, is pretty low. It's only 2%. It helps to see that the general improbability of having it erodes the confidence of the test result. And in this case, the prior probability or the probability that we started with was just the rate distributed over the whole population. We took the total number of cases and divided by the total number of people in the county. If you had some symptoms, or had an exposure, and went to the doctor and asked for a test, then your prior probability would be higher. We'd have some prior reasons, or some other evidence to think that your base rate is higher than 2%. Put another way, the prior probability of having covid is much higher among people who have a sore throat, a fever, and an exposure, than it is for someone just chosen at random from the population.

Now suppose we give this randomly selected person another test after they get the first positive one--that's what a responsible doctor would do. Now we have a different equation because the base rate, or the prior probability is now 10% from the first test. That is, on the basis of the test we just took, your probability of having covid has gone up from 2% in the general population to 10%. So we insert that new base rate. And let's suppose we are using the same 85% accurate test. And notice that now since the $\Pr(C)$ is .1, then $\Pr(\sim C)$ is .9:

$$\Pr(C|+) = \frac{(.85)(.1)}{(.85)(.02) + (.15)(.9)}$$

$$\Pr(C|+) = \frac{.085}{.22}$$

$$\Pr(C|+) = .38$$

So now, with two positive test results, the probability that you've got it is 38%. That's still not probable; it's more likely that you don't have it now than that you do. But that's not negligible. The evidence is mounting that this random person we selected from the population has it.

Now consider what happens when we use a more accurate test. Suppose we have a test that is 95% accurate, and only give the wrong answer 5% of the time. And now, let's dispose of our contrived example of selecting someone randomly from the population. Suppose that our patient has been exposed to someone who has it, and they've got some symptoms that are consistent with covid. Now, before we've even tested, we've got some evidence to think that this person has it. Suppose we assign a 40% probability to their having it, before the test. Now, what would a positive test result mean?

$$\Pr(C|+) = \frac{(.95)(.4)}{(.95)(.5) + (.05)(.6)}$$

$$\Pr(C|+) = \frac{.38}{.41}$$

$$\Pr(C|+) = .93$$

So now, with a better test, and some other evidence to suspect that this patient has it, it's now 93% probable. This patient very probably has it.

What else can we say about the base rate? In our first examples, we were assuming that official, reported rate of covid in the population is a good measure of the actual rate. But is that a good assumption? Probably not. There are a lot more people out there who have it than just the ones officially reported. Some people get only mildly sick and don't go to the doctor, some people get it and don't get tested, some people get it and never know they've got it. The CDC

has recently said that the real rate of people who have may be *10 times higher* than the official rate. And we've got some reasons to think that the CDC is even understating the truth here. If you talk to someone at a party, the difference between a 2% chance of your being exposed and a 20% of your being exposed is substantial.

So what happens to our calculation when you take the base rate of covid in the population to be 20% instead of 2%? And let's assume a 90% accurate test this time.

$$\Pr(C|+) = \frac{(.95)(.2)}{(.95)(.2) + (.05)(.8)}$$

$$\Pr(C|+) = .32$$

So even now, with a positive test and a higher rate in the population, you probably don't have it, but the evidence is mounting. But if you did another test now and got a positive result, the $\Pr(C|+) = .89$. Now, you've probably got it.

So far, we've dealt with positive test results. And we've seen that the relative rarity of the property in the population erodes a positive test result. But what happens if you get a negative test result? The short answer is that now, with the property rare in the population, a negative result is fortified or increased in probability. Suppose that the base rate is 20% and suppose that this test is 90% accurate. But now we've got a negative test result. What's the probability that this person has it?

$$\Pr(C|-) = \frac{\Pr(-|C)\Pr(C)}{\Pr(-|C)\Pr(C) + \Pr(-|\sim C)(\sim C)}$$

$$\Pr(C|-) = \frac{(.1)(.2)}{(.1)(.2) + (.9)(.8)}$$

$$= \frac{.02}{.74}$$

$$=.027$$

So this person's having covid was at 20% with the rest of the population before the test. But they tested, got a negative result, and now the probability has gone down to .027. So they probably don't have it.

Notice a couple of things here. The left of the numerator, " $\Pr(-|C)$ " is now "what is the probability that you'd get a negative test, given that you've got Covid?" And with a test that's 90% accurate, we will assume that's .1. And the right of the denominator, " $\Pr(-|\sim C)$ " is "what is the probability that you'd get a negative test result, given that you don't have Covid?" That is, what's the probability that the test would get a correct negative result? With a 90% accurate

test, that's .9. And since the rate of C in the population here is .2, then the rate of people not having it is 80% or .8.

So there are a couple of lessons we can extract here. First, when you test and get a positive or negative test result, the answer to the question "Do I have it?" is not as simple as whether it was a positive or negative test result. The answer is modulated by how accurate the test is and how prevalent the disease is. If the disease is rare and the test is inaccurate, then a positive test result doesn't tell us much. More tests, with positive test results, would fortify the answer and support the positive conclusion. If the disease is rare and the test is fairly accurate, then a negative test result is good news; a negative result doesn't give us certainty, but it makes it substantially less probable that you have it. Again, more tests, with more negative results, would fortify that answer. If a person takes multiple tests and gets mixed results, then we are left in a curious epistemic situation. One positive result and one negative result would cancel each other out effectively; you'd have two competing pieces of evidence pushing opposite directions, leaving you just where you started.