Len Goff
April 9, 2020

The value of a positive coronavirus antibody test

This is preliminary and the numbers herein are meant to be illustrative. I am also not an epidemiologist! Comments and corrections are welcomed.

The promise of antibody testing to determine whether an individual has developed an immune response to SARS-CoV-2 (the virus that causes COVID-19) is likely to play a central role in restarting economic activity as countries begin easing strict lockdowns. Of course, no test is perfect, and understanding the nature of the uncertainty in available antibody tests will be very important. The purpose of this note is to discuss how two alternative ways of measuring this uncertainty might seem to paint very different pictures of the reliability of a test. In particular, we'll see algebraically why the divergence will be high when the true rate of immunity in the population is still low, as we might expect over the next few months. We'll also see how the ideal test may be tailored to different objectives and to the situation as it changes over time. This note is inspired by a great set of slides shared by my friend Aleeza Gerstein of the University of Manitoba.

It is quite natural to measure such a test by quantifying how often it successfully detects whether an immune response is or is not present. On this count, a test by Cellex that has recently won FDA approval does remarkably well based on clinical results: about 94% of individuals who have had the infection will test positive, and about 96% of individuals who have not had the infection will test negative. The former of these is called the *sensitivity* of the test and the latter is called the *specificity* of the test. These numbers are both close to one, which suggests that the test is very good at producing a result that aligns with one's true immunity status. However, these numbers alone don't let us answer the following question: suppose that I take the test and get a positive result, what's the chance that I actually have had the infection? We'll see that the answer depends on the overall rate in the population, and that there is still important room for improvement in tailoring tests to answer this question.

Index individuals in a population by i, and let D_i indicate that individual i has been infected (I'll use this interchangeably with having developed an immune response). In particular, we let $D_i = 0$ if individual i has not been infected, and $D_i = 1$ indicate that they have. Let p denote the proportion of the population that actually has developed an immune response. We can write $p = P(D_i = 1)$, where the probability is evaluated with respect to the population distribution. Let's say 5% of the population has been infected so far, so $p = 0.05$.

Let T_i indicate that i obtains a positive test result (i.e. $T_i = 1$ if positive, $T_i = 0$ if negative). We now have enough notation to express the concepts of sensitivity and

specificity introduced earlier:

$$se = P(T_i = 1|D_i = 1) \approx 94\% \qquad \text{and} \qquad sp = P(T_i = 0|D_i = 0) \approx 96\%$$

Now consider the question raised above: what's the probability that I've had the infection if I get a positive test result? This is called the *positive predicted value* (PPV). PPV is related to the sensitivity of the test, but flips around the event we are conditioning on and the one we are seeking the probability of. We can get it using Bayes' Rule:

$$PPV = P(D_i = 1|T_i = 1) = \frac{P(D_i = 1, T_i = 1)}{P(T_i = 1)} = \frac{P(D_i = 1, T_i = 1)}{P(D_i = 1)} \cdot \frac{P(D_i = 1)}{P(T_i = 1)} = \frac{p}{P(T_i = 1)} \cdot se$$

We can see from this expression that the PPV will be lower than the sensitivity of the test if p is smaller than $P(T_i = 1)$. Note that $P(T_i = 1)$ is the rate of positive test results we would get by repeating the test on a large number of randomly sampled individuals (more on this below).

What is the value of $P(T_i = 1)$? We can work it out by from the quantities we've already introduced, using the law of total probability:

$$\begin{aligned} P(T_i = 1) &= P(T_i = 1|D_i = 0)P(D_i = 0) + P(T_i = 1|D_i = 1)P(D_i = 1) \\ &= (1 - sp)(1 - p) + se \cdot p \\ &= p \cdot (se + sp - 1) + (1 - sp) \end{aligned}$$

Collecting terms that are multiplied by p allows us to investigate the important ratio

$$\frac{p}{P(T_i = 1)} = \frac{p}{p \cdot (se + sp - 1) + (1 - sp)} = \frac{1}{(se + sp - 1) + \frac{1-sp}{p}}$$

which gives

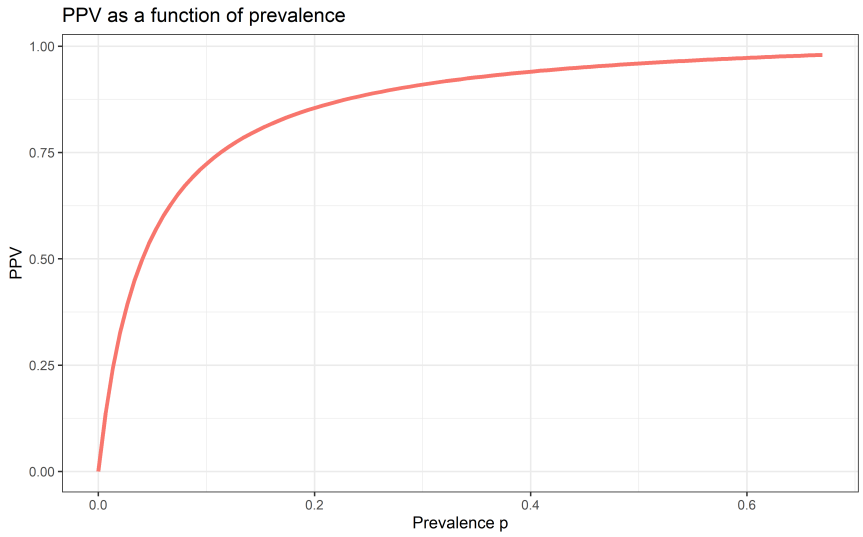
$$PPV = \frac{se}{(se + sp - 1) + \frac{1-sp}{p}} \tag{1}$$

Now consider the two terms in the denominator. They determine the fraction by which the PPV is deflated compared with the sensitivity. The first term does not depend on p, and is: $se + sp - 1 = .94 + .96 - 1 = .9$, which is pretty close to one. The second term, however, can be quite large when the prevalence of infection is low compared with the gap between the specificity and 1. With $p = .05$, the second term is $.04/.05 = 0.8$. Thus, the positive predicted value of the test is $.9 + .8 = 1.7$ times smaller than the sensitivity. That nice high sensitivity of 94% is effectively divided in two: $PPV \approx 55\%$.

This deflation would be even more drastic if say p were equal to 1%. In this case, Equation 1 yields:

$$PPV = \frac{se}{0.9 + .04/.01} = se/4.9$$

making the PPV almost five times lower than the sensitivity! In a world where very few people have had the infection, a positive result from the current test would suggest that the chance you've specifically been infected is still just 20%. That may be reason for pause if we are say deciding whether an individual should return to work. The figure



below shows the PPV for this test as a function of p. We can see that as the true rate of infection gets very high, the PPV approaches the specificity of the test. For example, if about 2/3 of the population were infected (around the suspected "herd immunity" point), the chance that you have been infected given a positive result from the current antibody test would be 94%. But the red curve drops off very sharply for p below .2, which is almost surely where we are right now.

What should we conclude from this? We want to keep p as low as possible while the world lacks a vaccine that is known to be effective and safe. In such a world, the current state of the art antibody test is unlikely to make any individual confident that they are individually immune. But the antibody tests are sure to improve over time. Equation 1 suggests that if the goal is to ensure that individuals have had the infection to issue so-called "immunity certificates" to say return to work, then in the short run increasing the specificity of the test may be more important than increasing its sensitivity. To see this, note that

$$\frac{\partial PPV}{\partial se} = PPV/se - \frac{PPV}{(se + sp - 1) + \frac{1-sp}{p}} \approx 0.27$$

while

$$\frac{\partial PPV}{\partial sp} = (1/p - 1) \cdot \frac{PPV}{(se + sp - 1) + \frac{1-sp}{p}} \approx 6.18$$

With $p = .05$, increasing the specificity by a small amount would give us a more than twenty times larger improvement in the PPV than the same-sized increase in sensitivity would. Increasing specificity means getting rid of false positives, making the test err more on the side of saying people *have not* had the infection.

Note that if the goal of the test were to ensure that people had not been previously infected, we could perform a similar analysis of the so-called *negative predicted value* (NPV) of the test: $P(D_i = 0|T_i = 0)$. With our numbers the NPV of the Cellex test is:

$$NPV = sp \cdot \frac{1-p}{1 - P(T_i = 1)} = \frac{sp}{(se + sp - 1) + \frac{1-se}{1-p}} = \frac{.96}{.9 + .06/.95} \approx .997$$

This suggets that if one tests negative from the antibody test at this point (with $p = .05$), they can be very confident that they have not yet developed an immune response. For the NPV, the relationship with p goes in the other direction: it is large when p is small, and gets smaller as p grows.

Finally, suppose the goal of the test is not to either issue immunity certificates or to suggest that and invidual has not yet been exposed, but rather to estimate the total prevalence of immunity p in the population. This requires performing the test for an adequately large sample of individuals. As this sample size increases, the estimated immunity rate will converge in probability to $P(T_i = 1)$. We've seen that the current test will overstate the true value of p by a factor of about 1.7 in this large sample limit, if p is about 5% (put differently, the estimator has an asymptotic bias of about $.7p$).

What would be required of the sensitivity and specificity for an antibody test to give an accurate estimate of p in large samples? We can get this one of two ways. The first is to have a perfect test: with $sp = 1$ and $se = 1$ (no false positives or false negatives). This would work regardless of the value of p, so in principle that's what we want to aim for. But reality involves a trade-off between specificity and sensitivity. An imperfect test can still be accurate for population surveillance of p if this trade-off is targeted towards a particular range that we expect p to be in. In particular, we'll get $P(T_i = 1) = p$ if:

$$\frac{1 - sp}{1 - se} = \frac{p}{1 - p}$$

Since p is probably low right now, the best test for population-level analysis would have a higher specificity than sensitivity. This is true of the current Cellex test, but estimating p accurately would suggest making specificity higher still, even if it comes at the cost of a lower sensitivity. The current test would be well-suited to estimating the prevalence of immunity if $p/(1 - p)$ were $.04/.06 = 2/3$, which corresponds to about 40% of the population having established an immune response. Hypothetically, which available test it's best to use at the population level might be different at different points along the trajectory of the pandemic, and might differ at any given time from the best test for verifying individual immunity.

References:

"Antibody Testing & The Problem with False Positives" by Aleeza Gerstein (Apr. 7, 2020).