The Value of Testing

George Rebane
24 March 2020

Trigger warning: The contents of this commentary may not be accessible to the innumerate reader.

As explained in 'Testing, testing, …', the value of testing comes to the fore when you have a sufficiently reliable test that can be administered on a regular and scientific basis over a suspect population. In that case, it could be used to identify concentrations of infections that can be used for a more satisficing allocation of limited healthcare resources. In an update to the referenced commentary, I illustrated how to determine the reliability of test results given the reliability parameters of the test administered to a walk-in member of a suspect population (country, state, county, city, region, …).

In this piece I will answer the questions many reasonable people have about testing – 'why test if the tests are unreliable? Why test if the test only provides a snapshot assessment which may change within the next hour?' We start by recalling the advice from our federal and state medicos. Given the limited availability of tests and processing labs, we are only supposed to request a test if we are symptomatic – fever, cough, headache, … . Don't burden the system if you are asymptomatic.

(A scary interlude here is to consider how the virus spreads through contact with pre-symptomatic infected people who don't 'socially distance' or stay at home. With Covid19, such a pre-symptomatic period may last well over a week, and some more fortunate infected people may never become symptomatic and recover. However, both cohorts can infect others during their pre-symptomatic interval. This is the prime causal factor that makes the infection rate to explode, and why distancing is the only sure way to stop/slow the spread. Consider this – if everyone stayed isolated for a month, Covid19 would be gone because people would either be uninfected, recovered, or dead. That's mostly how epidemics end.)

Onward. OK, so we are told that today's coronavirus tests have a detection rate between 60% and 75%. And we're also told that these tests have false alarm rates between 3% and 50%, a huge range as such diagnostics go. These are conditional probabilities, so we'll use some appropriately compact math symbols for what follows. Let V stand for infected with the virus, and $\neg V$ stand for not infected. (The minus sign with the little hangy-down denotes a logical NOT.) Also let TP stand for the test positive result, and $\neg TP$ stand for a test negative (NOT positive) result. Got that?

Now, the probability that X is true, given that Y is true, is expressed as $P(X|Y)$ – voiced as 'probability of X given Y', and is called a conditional probability of X being TRUE conditioned on Y = TRUE. $P(X)$ is just the (marginal) probability that X is TRUE (conditioned on everything known and unknown), like P(getting a three on the roll of a die) = 1/6.

So, why are suspect but still asymptomatic people being tested? What good does it do? Let's answer that with the help of the Rev Bayes (more on Bayes here). But first we'll express the above test parameters in their most hopeful form – detection rate $P(TP|V) = 0.75$, and the false-alarm rate $P(TP|\neg V) = 0.03$. From the Bayes theorem we have

$$P(V \mid TP) = \frac{L(TP \mid V) P(V)}{L(TP \mid V) P(V) + P(\neg V)}, \quad \text{where} \quad L(TP \mid V) = \frac{P(TP \mid V)}{P(TP \mid \neg V)}.$$

Where $L(\ )$ is called the likelihood ratio of obtaining TP, the test positive evidence, and $P(V)$ is called the prior probability of being infected before incorporating the results of a positive test. It should be clear that $P(\neg V) = 1 - P(V)$, the probability of not being infected is just the complement of being infected. Now assume that our target population is that of the entire US of 330M people, of whom, say, a 100K are already infected (a conservative doubling of the 24mar20 known infected cohort). The prior probability of a random walk-in person who is tested has $P(V) = 100\text{K}/300\text{M} = 0.0003$, pretty small. However, if a suspect individual who was exposed to potentially infected persons is to be tested, then we have to change his prior probability to reflect our knowledge that he may or not be infected. The probability that represents this kind of prior knowledge (or ignorance) is $P(V) = 0.50$, that is a 50-50 chance that he's infected.

The hopeful test reliability parameters from above give us a likelihood value of $L(TP|V) = 0.75/0.03 = 25$. Plugging this value of L and the prior probability $P(V)$ for a walk-in person into the Bayes formula, we get

$$P(V \mid TP) = \frac{L(TP \mid V) P(V)}{L(TP \mid V) P(V) + P(\neg V)} = \frac{25 \cdot 0.0003}{25 \cdot 0.0003 + (1 - 0.0003)} = 0.0074.$$

That's still a very low probability of having the virus given that the test came back positive. So we're not going to send this asymptomatic individual into our already over-burdened healthcare system. The existing test was able to increase his infection probability form 3 in 10,000 to 74 in 10,000, a marked increase to be sure, but still not enough to get excited about just yet. So we understand the policy of telling such people to stay home until symptoms appear (or not) and then report this to their physician.

But the situation is very different when we test a person who is suspected of being exposed and possibly infected. Plugging into the Bayes formula his prior probability of infection gives

$$P(V \mid TP) = \frac{L(TP \mid V) P(V)}{L(TP \mid V) P(V) + P(\neg V)} = \frac{25 \cdot 0.50}{25 \cdot 0.50 + (1 - 0.50)} = 0.9615.$$

This is a totally different picture. Now a positive test result indicates with high probability that the tested individual has been infected. He may still be told to go and stay home, or he may be admitted into the healthcare system for treatment. So we now see the value of testing such a suspect cohort of people to get them out of circulation and/or into early treatment.

Let's also look at what this test reveals when it reports ¬*TP*, a negative result obtained from testing a suspect individual. We hear a lot of such test results reported about prominent individuals ranging from President Trump on down. Here we want to find out the probability that the VIP is still infected given that he tested negative. In this case the Bayes formula for the posterior conditional probability incorporating the negative test evidence becomes

$$P(V \mid \neg TP) = \frac{L(\neg TP \mid V)P(V)}{L(\neg TP \mid V)P(V) + P(\neg V)}, \text{ where } L(\neg TP \mid V) = \frac{P(\neg TP \mid V)}{P(\neg TP \mid \neg V)} = \frac{1 - P(TP \mid V)}{1 - P(TP \mid \neg V)}$$

In this case, we use the indicated complementary probabilities that characterize our test's reliability to compute *L*, the likelihood ratio for a negative test result. This calculates to

$$L(\neg TP \mid V) = \frac{P(\neg TP \mid V)}{P(\neg TP \mid \neg V)} = \frac{1 - P(TP \mid V)}{1 - P(TP \mid \neg V)} = \frac{1 - 0.75}{1 - 0.03} = 0.2577.$$

Plugging this value into the above Bayes formula for a negative test result we get

$$P(V \mid \neg TP) = \frac{0.2577 \cdot 0.50}{0.2577 \cdot 0.50 + (1 - 0.50)} = 0.2049,$$

which tells us that there's still a one out of five chance of being infected even if they tested negative. So even though the prior probability of 0.5 was reduced by the negative test result to 0.20, having such people continue to mix with the uninfected is now a judgement call. Consider the number of suspect people who may subsequently have been infected by some very busy official continuing to meet with many others after having obtained a hopeful negative test result.

And now boys and girls, for your Lucky Strike Extra (never mind if you don't understand this piece of Americana) we're going to answer a related question that concerns negatively tested (remember, it's only a snapshot in the best case) VIPs having close contacts with tens of people every day as they do their job. The above result of 1 out of 5 still seems a bit high, at least it does to me. What if the President agreed and asked for a better test that would reduce the posterior probability down to something more like one out twenty or one out of a hundred? What would the test reliability parameters have to be then? A little thought tells us that we can determine whether a new test would qualify or not by calculating its likelihood ratio *L*, and then seeing if it satisfies the threshold determined by solving the Bayes equation for *L* and plugging in the specified prior *P*(*V*) and posterior P(*V*|¬*TP*) probabilities. Solving Bayes for *L* gives us

$$P(V \mid \neg TP) = \frac{L(\neg TP \mid V)P(V)}{L(\neg TP \mid V)P(V) + P(\neg V)} \xrightarrow{\text{Solving for } L} L(\neg TP \mid V) = \frac{[1 - P(V)]P(V \mid \neg TP)}{P(V)[1 - P(V \mid \neg TP)]} .$$

So let's the requirement for a new test is that when it reports a negative on a suspect person, the probability that the person is actually infected must be at most one out of twenty or that *P*(*V*|¬*TP*) = 0.05. Using this and *P*(*V*) = 0.50 to calculate the threshold value of *L* gives

$$L(\neg TP \mid V) \le \frac{\left[1-P(V)\right]P(V \mid \neg TP)}{P(V)\left[1-P(V \mid \neg TP)\right]} = \frac{(1-0.5)0.05}{0.5(1-0.05)} = 0.0526 \ge \frac{1-P(TP \mid V)}{1-P(TP \mid \neg V)} \ .$$

From this we can immediately calculate the required relationship between detection and false alarm probabilities for any candidate test that would satisfy the new reliability specification. The required relationship is

$$0.0526 \ge \frac{1-P(TP \mid V)}{1-P(TP \mid \neg V)} \longrightarrow P(TP \mid V) \ge 1-0.0526\left[1-P(TP \mid \neg V)\right] \text{ , or}$$

$$0.0526 \ge \frac{1-P(TP \mid V)}{1-P(TP \mid \neg V)} \longrightarrow P(TP \mid \neg V) \le 1-\frac{\left[1-P(TP \mid V)\right]}{0.0526} \ .$$

So if the candidate test had false alarm rate of, say, $P(TP|\neg V) = 0.10$, then it should have a detection rate P(TP|V) of at least 0.9526. And if the test had a detection rate of only 0.75, then its false alarm rate should not be above zero, since otherwise $P(TP|\neg V)$ would become the unallowed negative. (What does this say for the hopeful numbers we used for today's quoted test reliability?) Since all real tests have a positive false alarm rate, our specified false alarm rate of $P(V|\neg TP) = 0.05$ requires the test's detection rate to be $P(TP|V) \ge 0.9474$.

(Apologies to the purists offended by my casual treatment of significant digits in this expose.)