b.next

Challenges of Interpreting Test Results and Implications for COVID-19

Diagnostic tests are a critical tool to contain epidemics, to support medical care, and for public health measures. Understanding when they are accurate and inaccurate is necessary for understanding which individuals have the virus, need isolation, and need their contacts traced.

Many diagnostic tests are reliable, though all are imperfect. And at a large scale, tiny errors in accuracy for single tests can aggregate into large errors if deployed without care. This is especially true when the rate of true infection in the tested population is expected to be low. For example, when testing for infection in a person who doesn't have symptoms or a history of exposure, or when testing for a history of infection when the overall prevalence of disease for a given population is low.

Accuracy and Errors in Diagnostic Tests

The accuracy of a test informs how to use it for diagnoses, understanding population level immunity, and an individual's health status. Saying a diagnostic test is "accurate" can mean several things; some diagnostic tests are accurate in some ways and inaccurate in others.

When medical or public health officials administer a diagnostic test to a person, they may seek answers to two key questions:

1. *Does the individual being tested have the virus?* To reliably answer this question, a diagnostic test needs a high **positive predictive value (PPV)**. If PPV is low, the test will produce many false positive results, possibly causing people to believe they have a virus they do not have.
2. *Can we rule out that the individual being tested has a virus?* To reliably answer this question, a diagnostic test needs a high **negative predictive value (NPV)**. If NPV is low, the test will produce many false negative results, possibly causing people with the virus to believe they are healthy (or that they have another virus).

It's ideal to have a diagnostic where both positive and negative test results are conclusive, meaning PPV and NPV are both high. Unfortunately, there is usually a tradeoff between being able to make strong positive and negative predictions.

PPV and NPV are closely related to two other accuracy measures: **sensitivity** and **specificity**. Sensitivity measures how well the test picked out the people in a study that had the virus. Specificity measures how well the test picked out the people in a study who did not have the virus. Sensitivity and specificity are metrics used by manufacturers to evaluate a diagnostic test when you know the status (positive or negative) of a sample. In contrast, these differ from PPV and NPV which measure how trustworthy a positive or negative result is when it is **NOT** known if a person has the virus.

PPV and specificity are closely related, as are NPV and sensitivity. If specificity is high, it will increase PPV. Similarly, if sensitivity is high, it will increase NPV. Sensitivity and specificity do not change with the rate of the virus in the population.

Low Disease Prevalence Can Cause Many False Positives

To complicate things further the rate of infection in the population impacts results as well. And in the case where a small percent of a population has the disease, PPV can be low, causing many false positive test results.

If, for example, specificity is 95%, then 5% of the *people who are healthy* have false positive results. If you have many more healthy people in the population being tested, then the number of false positive

tests rivals the number of true positive tests. In other words, when the virus is rare, false positives can become a problem.

To put this in concrete terms: Suppose you test 1,000 people, 20 of them (2%) have the virus causing COVID-19. (And, of course, you don't know who or how many people have the virus.) If the diagnostic test has 95% specificity, then 5% of the 980 people who don't have the virus will get positive test results. In other words, 49 healthy people may think they have the virus. If the test captured all 20 of the people with the virus, its PPV would be 20 / (20 + 50) or 28.6%. In other words, if a person gets a positive result, there's only a 28.6% chance they actually have the virus.

On the other hand, if 500 of the 1,000 people (50%) tested have the virus, 5% of the healthy people still get positive tests. But that's only 25 out of 500 people, a much smaller problem. This test would have a PPV of 500 / (25 + 500) or 95%. To get a 95% PPV e with only 2% of the population being sick, as above, the test's specificity would have to be about 99.9%, almost perfect.

This is not to say that you want to restrict testing so much that only a high percentage of tests are positive. That may minimize a problem of false positives while increasing the number of undetected cases. Instead, note that as testing gets more widespread in jurisdictions with a lower prevalence of the virus, false positives may become an issue requiring mitigation.

Diagnostic Testing in Normal Times: Testing for Influenza

Aren't most viruses rare? Wouldn't that mean that most diagnostic tests have a low PPV? Yes and no. Consider the case of diagnosing the flu. It's flu season and you feel ill, so you see a doctor. The doctor notes that your symptoms are consistent with the flu and orders a diagnostic test. If that test comes back positive, do you have the flu? Almost certainly.

The rate of influenza is not too low in the population of people who feel sick enough to see a doctor and have flu-like symptoms. And as the CDC notes in their guidance to clinicians[1], the infection rate can change with the season (prevalence of the flu in the wider population) and that affects the likelihood of false positives and negatives of this diagnostic test as described in the previous section.

Balancing Errors with Diagnostic Needs

A diagnostic test is reliable, in part, because it is not given to a random sample of the population. False positives distort our understanding of the true infection rate in a population, not necessarily because the diagnostic test is unreliable, but because tiny errors in misdiagnosis for any one healthy person get amplified across the many healthy people in the population when the true infection rate is low.

Additionally, some diagnostic tests have extremely high specificity. In those cases, false positive results are less of a problem when the disease is rare. Even so, healthcare professionals may do multiple independent tests before conclusively diagnosing a person with a rare disease.

If testing is done in a careful way, diagnostic tests can be quite reliable. *The key is to adopt a strategy that balances testing error with the need to diagnose many people.* Such a strategy factors in the type of test (e.g., PCR, antibody, antigen, etc.), tests' strengths and weaknesses, human factors when administering a test, and an understanding of the relationship between the rate of true infections and false detection rates. One size does not fit all, and tests need to be deployed and interpreted with care.

[1] https://www.cdc.gov/flu/professionals/diagnosis/clinician_guidance_ridt.htm