Perspective

False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications

Steven Woloshin, M.D., Neeraj Patel, B.A., and Aaron S. Kesselheim, M.D., J.D., M.P.H.

There is broad consensus that widespread SARS-CoV-2 testing is essential to safely reopening the United States. A big concern has been test availability, but test accuracy may prove a larger long-term problem.

While debate has focused on the accuracy of antibody tests, which identify prior infection, diagnostic testing, which identifies current infection, has received less attention. But inaccurate diagnostic tests undermine efforts at containment of the pandemic.

Diagnostic tests (typically involving a nasopharyngeal swab) can be inaccurate in two ways. A false positive result erroneously labels a person infected, with consequences including unnecessary quarantine and contact tracing. False negative results are more consequential, because infected persons — who might be asymptomatic — may not be isolated and can infect others.

Given the need to know how well diagnostic tests rule out infection, it's important to review assessment of test accuracy by the Food and Drug Administration (FDA) and clinical researchers, as well as interpretation of test results in a pandemic.

The FDA has granted Emergency Use Authorizations (EUAs) to commercial test manufacturers and issued guidance on test validation.[1] The agency requires measurement of analytic and clinical test performance. Analytic sensitivity indicates the likelihood that the test will be positive for material containing any virus strains and the minimum concentration the test can detect. Analytic specificity indicates the likelihood that the test will be negative for material containing pathogens other than the target virus.

Clinical evaluations, assessing performance of a test on patient specimens, vary among manufacturers. The FDA prefers the use of "natural clinical specimens" but has permitted the use of "contrived specimens" produced by adding viral RNA or inactivated virus to leftover clinical material. Ordinarily, test-performance studies entail having patients undergo an index test and a "reference standard" test determining their true state. Clinical sensitivity is the proportion of positive index tests in patients who in fact have the disease in question. Sensitivity, and its measurement, may vary with the clinical setting. For a sick person, the reference-standard test is likely to be a clinical diagnosis, ideally established by an independent adjudication panel whose members are unaware of the index-test re-

sults. For SARS-CoV-2, it is unclear whether the sensitivity of any FDA-authorized commercial test has been assessed in this way. Under the EUAs, the FDA does allow companies to demonstrate clinical test performance by establishing the new test's agreement with an authorized reverse-transcriptase–polymerase-chain-reaction (RT-PCR) test in known positive material from symptomatic people or contrived specimens. Use of either known positive or contrived samples may lead to overestimates of test sensitivity, since swabs may miss infected material in practice.[1]

Designing a reference standard for measuring the sensitivity of SARS-CoV-2 tests in asymptomatic people is an unsolved problem that needs urgent attention to increase confidence in test results for contact-tracing or screening purposes. Simply following people for the subsequent development of symptoms may be inadequate, since they may remain asymptomatic yet be infectious. Assessment of clinical sensitivity in asymptomatic people had not been reported for any commercial test as of June 1, 2020.

Two studies from Wuhan Province, China, arouse concern about false negative RT-PCR tests in patients with apparent Covid-19 illness. In a preprint, Yang et al. described 213 patients hospitalized with Covid-19, of whom 37 were critically ill.[2] They collected 205 throat swabs, 490 nasal swabs, and 142 sputum samples (median, 3 per patient) and used an RT-PCR test approved by the Chinese regulator. In days 1 through 7 after onset of illness, 11% of sputum, 27% of nasal, and 40% of throat samples were deemed falsely negative. Zhao et al. studied 173 hospitalized patients with

acute respiratory symptoms and a chest CT "typical" of Covid-19, or SARS-CoV-2 detected in at least one respiratory specimen. Antibody seroconversion was observed in 93%.[3] RT-PCR testing of respiratory samples taken on days 1 through 7 of hospitalization were SARS-CoV-2–positive in at least one sample from 67% of patients. Neither study reported using an independent panel, unaware of index-test results, to establish a final diagnosis of Covid-19 illness, which may have biased the researchers toward overestimating sensitivity.

In a preprint systematic review of five studies (not including the Yang and Zhao studies), involving 957 patients ("under suspicion of Covid-19" or with "confirmed cases"), false negatives ranged from 2 to 29%.[4] However, the certainty of the evidence was considered very low because of the heterogeneity of sensitivity estimates among the studies, lack of blinding to index-test results in establishing diagnoses, and failure to report key RT-PCR characteristics.[4] Taken as a whole, the evidence, while limited, raises concern about frequent false negative RT-PCR results.

If SARS-CoV-2 diagnostic tests were perfect, a positive test would mean that someone carries the virus and a negative test that they do not. With imperfect tests, a negative result means only that a person is less likely to be infected. To calculate how likely, one can use Bayes' theorem, which incorporates information about both the person and the accuracy of the test (recently reviewed[5]). For a negative test, there are two key inputs: pretest probability — an estimate, before testing, of the person's chance of being infected — and test sensitivity. Pretest probability might depend on

local Covid-19 prevalence, SARS-CoV-2 exposure history, and symptoms. Ideally, clinical sensitivity and specificity of each test would be measured in various clinically relevant real-life situations (e.g., varied specimen sources, timing, and illness severity).

Assume that an RT-PCR test was perfectly specific (always negative in people not infected with SARS-CoV-2) and that the pretest probability for someone who, say, was feeling sick after close contact with someone with Covid-19 was 20%. If the test sensitivity were 95% (95% of infected people test positive), the post-test probability of infection with a negative test would be 1%, which might be low enough to consider someone uninfected and may provide them assurance in visiting high-risk relatives. The post-test probability would remain below 5% even if the pretest probability were as high as 50%, a more reasonable estimate for someone with recent exposure and early symptoms in a "hot spot" area.

But sensitivity for many available tests appears to be substantially lower: the studies cited above suggest that 70% is probably a reasonable estimate. At this sensitivity level, with a pretest probability of 50%, the post-test probability with a negative test would be 23% — far too high to safely assume someone is uninfected.

The graph shows how the post-test probability of infection varies with the pretest probability for tests with low (70%) and high (95%) sensitivity. The horizontal line indicates a probability threshold below which it would be reasonable to act as if the person were uninfected (e.g., allowing the person to visit an elderly grandmother). Where this threshold should be set — here, 5% —
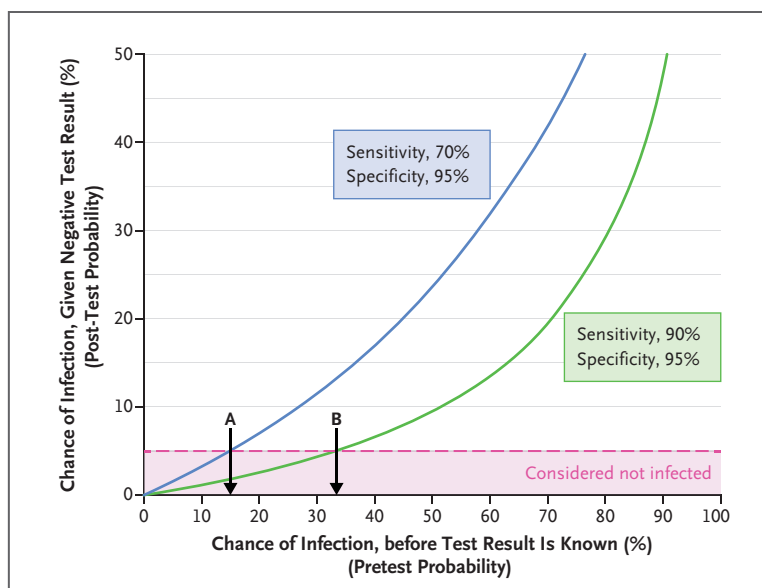
Chance of SARS-CoV-2 Infection, Given a Negative Test Result, According to Pretest Probability.

The blue line represents a test with sensitivity of 70% and specificity of 95%. The green line represents a test with sensitivity of 90% and specificity of 95%. The shading is the threshold for considering a person not to be infected (asserted to be 5%). Arrow A indicates that with the lower-sensitivity test, this threshold cannot be reached if the pretest probability exceeds about 15%. Arrow B indicates that for the higher-sensitivity test, the threshold can be reached up to a pretest probability of about 33%. An interactive version of this graph is available at NEJM.org.

is a value judgment and will vary with context (e.g., lower for people visiting a high-risk relative). The threshold highlights why very sensitive diagnostic tests are needed. With a negative result on the low-sensitivity test, the threshold is exceeded when the pretest probability exceeds 15%, but with a high-sensitivity test, one can have a pretest probability of up to 33% and still, assuming the 5% threshold, be considered safe to be in contact with others.

The graph also highlights why efforts to reduce pretest probability (e.g., by social distancing, possibly wearing masks) matter. If the pretest probability gets too high (above 50%, for example), testing loses its value because negative results cannot lower the probability of infection enough to reach the threshold.

We draw several conclusions.

First, diagnostic testing will help in safely opening the country, but only if the tests are highly sensitive and validated under realistic conditions against a clinically meaningful reference standard. Second, the FDA should ensure that manufacturers provide details of tests' clinical sensitivity and specificity at the time of market authorization; tests without such information will have less relevance to patient care.

Third, measuring test sensitivity in asymptomatic people is an urgent priority. It will also be important to develop methods (e.g., prediction rules) for estimating the pretest probability of infection (for asymptomatic and symptomatic people) to allow calculation of post-test probabilities after positive or negative results. Fourth, negative results even on a highly sensitive test cannot rule

out infection if the pretest probability is high, so clinicians should not trust unexpected negative results (i.e., assume a negative result is a "false negative" in a person with typical symptoms and known exposure). It's possible that performing several simultaneous or repeated tests could overcome an individual test's limited sensitivity; however, such strategies need validation.

Finally, thresholds for ruling out infection need to be developed for a variety of clinical situations. Since defining these thresholds is a value judgement, public input will be crucial.

Disclosure forms provided by the authors are available at NEJM.org.

From the Center for Medicine in the Media, Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH (S.W.); the Lisa Schwartz Program for Truth in Medicine, Norwich, VT (S.W.); the Program on Regulation, Therapeutics, and Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston (S.W., A.K.); and Yale University, New Haven, CT (N.P.).

1. U.S. Food and Drug Administration. Emergency Use Authorization (EUA) information, and list of all current EUAs (https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization).
2. Yang Y, Yang M, Shen C, et al. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. February 17, 2020 (https://www.medrxiv.org/content/10.1101/2020.02.11.20021493v2). preprint.
3. Zhao J, Yuan Q, Wang H, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin Infect Dis 2020 March 28 (Epub ahead of print).
4. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. April 21, 2020 (https://www.medrxiv.org/content/10.1101/2020.04.16.20066787v1). preprint.
5. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. BMJ 2020;369:m1808.

Copyright © 2020 Massachusetts Medical Society.