

Steps to verifying SARS-CoV-2 antibody assays and what's known about protective immunity

Sherrie Rice

August 2020—The CAP treats emergency use authorization assays similar to FDA-cleared assays and thus requires full verification. In a June 4 CAP webinar, Neil Anderson, MD, D(ABMM), assistant director of clinical microbiology, Washington University School of Medicine in St. Louis, walked through how to approach verification for SARS-CoV-2 assays.

Co-presenter Elitza Theel, PhD, D(ABMM), director of the infectious diseases serology laboratory at Mayo Clinic, reported what's known about protective immunity against SARS-CoV-2.



Dr. Anderson

For analytical interference, typical substances to investigate are hemoglobin, bilirubin, and triglycerides, and this should be performed at the limit of detection. “You may want to consider other exogenous inhibitors as well,” Dr. Anderson said, “maybe things you commonly see in samples submitted to your laboratory.” The laboratory can use data from the manufacturer in lieu of performing its own study.

Precision—the closeness of agreement between independent test measures—consists of intraassay (measurements collected under similar conditions) and interassay (under different conditions) precision. “Typical sources of imprecision need to be accounted for,” Dr. Anderson said, “and these include differences in timing of testing, temperature, mixing, pipetting, “pretty much anything you can introduce that might lead to a different result.” And it’s ideal to test concentrations at or near the limit of detection.

In his laboratory at Barnes Jewish Hospital, he and colleagues used a negative and a positive patient specimen and a positive QC specimen, and then compared the ratio between calibrator and signal intensity as seen in **Fig. 1**.



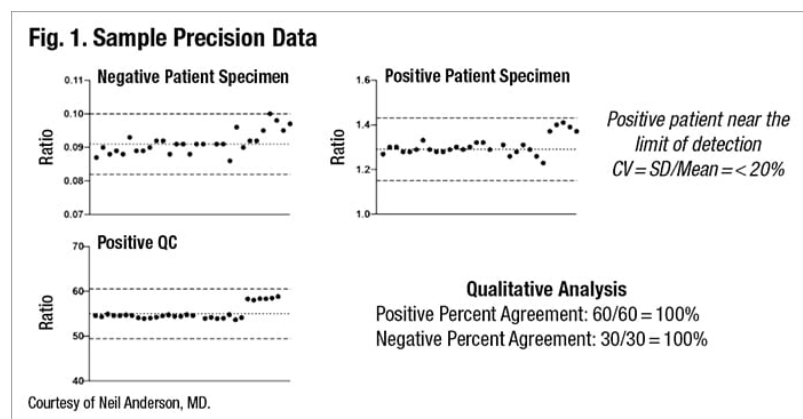
Dr. Theel

“What you can see is we have different results. We ran those single specimens using multiple replicates and were able to learn quite a bit. There is a spread of results,” he said. The lab wants to have an acceptable range for results when looking at this quantitatively, he said, and it’s often defined by CV (SD/mean), requiring that it remain below 20 percent.

Looking at the data qualitatively is also important. This allows calculation of positive percent agreement and negative percent agreement, “based on what the obtained results were expected to have been. Ideally this should be at or near 100 percent.”

Determination of reportable range typically doesn't apply to SARS-CoV-2 assays at this time because all are designated as qualitative, but it must be determined if a lab reports results quantitatively.

"You're probably going to spend the bulk of your assay verification in trying to determine accuracy, the extent to which a particular test is in agreement with a reference method or comparator," Dr. Anderson said. At this point, the ideal comparators are specimens from patients with known SARS-CoV-2 infection, established with molecular testing and collected at a known time post-symptom onset. "We all know molecular testing might be imperfect so there could be debate as to whether this is ideal, but that's what we're left with right now," he said. A secondary comparator would be specimens with known positive and negative antibody status tested using another validated or verified antibody test.



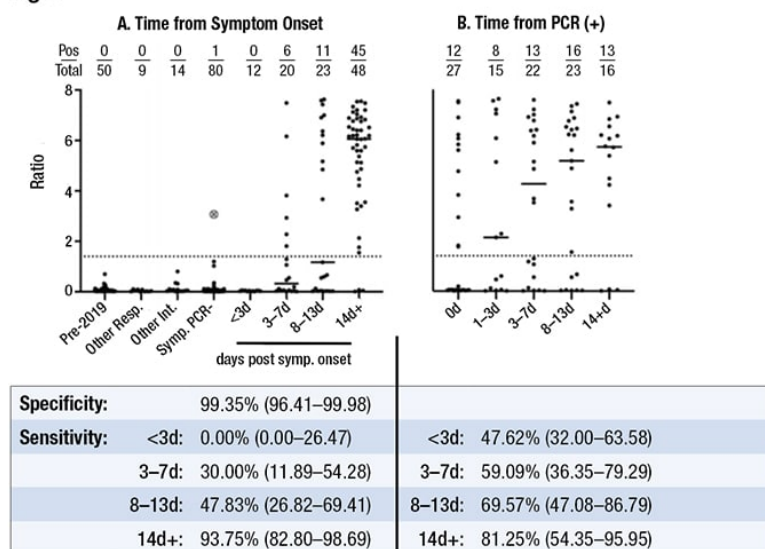
Comparing the reference method, if it's a gold standard, with the method being evaluated is used to determine sensitivity and specificity. If the reference method isn't a gold standard, it has to be phrased as positive or negative agreement. How the lab defines the reference method for accuracy studies affects the results. "So particularly if you're using serum from positive patients, you need to consider the timing of serum collection.

"For instance, if I'm using serum that had been collected from symptomatic patients near the time of symptom onset, my assay is going to look like it performs quite poorly because we know serology in general is going to have poor sensitivity early in disease. However, if I give my serologic assay a chance by looking at specimens that have been collected 14 days or later from symptom onset, using that type of analysis I'm going to show a better sensitivity."

A lab might want to investigate both pieces of data, he argues: sensitivity early and late in disease. That's how sensitivity was determined in his laboratory. They had 89 "positive" samples, based on PCR as a comparator. These were serum drawn at a variety of times post-symptom onset. They used remnant CBC samples from positive patients, with testing performed on the Abbott Architect.

"And what we found was an overall sensitivity of 56 percent. You might say that's pretty abysmal, but when you look at the data and analyze it at different time points, the story becomes a little clearer," Dr. Anderson said. "We did have low sensitivity at early onset of disease. However, once you get beyond 14 days, we had a 94 percent sensitivity"—sufficiently sensitive to put the assay into use.

Sensitivity varies based on how the data are analyzed, he said, pointing to his laboratory's data, which were analyzed in two ways (**Fig. 2**). When they looked at the time from symptom onset, they saw the expected low sensitivity early in disease. When the lab looked at time from PCR, the data looks different. "What's going to happen here is not everyone is going to necessarily get that PCR on their first day of symptomatology. So this is going to overestimate sensitivity early in disease." Keep that in mind, he advises, and keep in mind, too, that "a lot of manufacturers are defining their sensitivity that way likely because all they have is the PCR data and they may not be able to do that chart review to figure out exactly when the patient became symptomatic."

Fig. 2

Tang MS, Hock KG, Logsdon NM, et al. Clinical performance of two SARS-CoV-2 serologic assays. *Clin Chem*. 2020;hvaat120. doi: 10.1093/clinchem/hvaa120, by permission of Oxford University Press.

For an evaluation of specificity, formal and exhaustive cross-reactivity studies aren't needed for an EUA assay, but accuracy studies should take into account common cross-reacting targets. "What I mean by that is you should try to include samples from patients with documented seasonal coronavirus positivity, with disease processes similar to COVID-19, so other respiratory diseases, and with common conditions that can lead to cross-reacting antibodies such as lupus or infectious mononucleosis. If these lead to antibodies that are going to react, that's important information to know and communicate to your providers."

Dr. Anderson calls cross-reactivity with seasonal coronaviruses "a real phenomenon," as reported in the literature. "SARS-CoV-2 has a pretty high amino acid homology with SARS, less so with seasonal coronaviruses, though there is some homology there. So depending on the assay design, you may or may not see cross-reactivity." Some studies have shown high cross-reactivities; others have shown very little, he said. "The bottom line, though, is that it's theoretically possible, and to compound this, seroprevalence studies have suggested that a lot of us do have antibodies against seasonal coronaviruses. Sixty-five to 75 percent of young kids have antibodies to at least one seasonal coronavirus, and greater than 90 percent of adults older than 50 years of age have antibodies to all four coronaviruses."

On the FDA website is information about the seasonal coronaviruses included in a manufacturer's evaluation. It varies from no seasonal coronaviruses included to as many as 40, as of early June. "Most of them show no cross-reactivity, though I think I would argue that based on the amount of specimens tested even by some of the more thorough manufacturers, it doesn't really capture the entire risk," Dr. Anderson said.

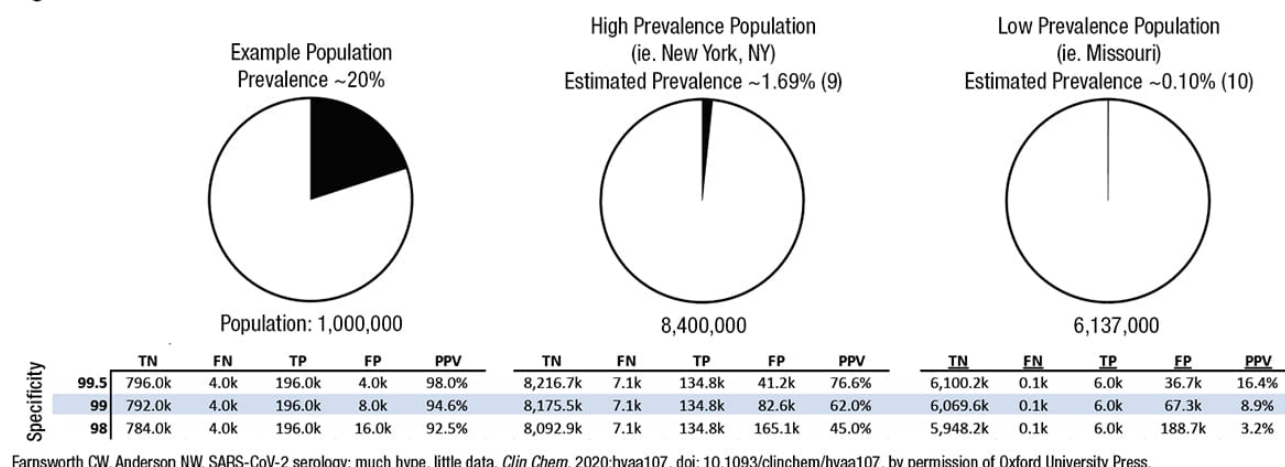
He shared an example of a specificity study performed in his laboratory (using remnant CBC specimens). He and colleagues studied 110 "negative" samples: 50 were pre-COVID-19 outbreak, nine were from patients with other respiratory illnesses, and 14 were from patients with viral infections or other possible interferents. They found an overall specificity of 100 percent.

Sensitivity and specificity thresholds are determined by the lab and there are two questions to consider, Dr. Anderson said. Are your providers going to want to test earlier than day 14 post-symptom onset? Guidance advises against this, he said, but "a lot of our providers may want to use it in this way and it may be hard to control, so you may consider a high sensitivity threshold early in the disease course."

Second, what patient population will be tested? Will it be used for symptomatic patients for diagnostic purposes, or for asymptomatic screening and surveillance? "You're going to approach those two tests very differently," he said. In **Fig. 3** are three different theoretical tests, with specificities in the lower left, all relatively high. They're used to

test the three example populations (the estimated prevalences of 20 percent, 1.69 percent, and 0.10 percent were based on molecular testing about mid-April).

Fig. 3



“We see that our positive predictive value and our prevalence of 20 percent is in the 90 percent range. However, it begins to drop as that prevalence drops, and when we get to the point where the prevalence is below one percent, the positive predictive value becomes quite abysmal,” meaning most of the positives will be false-positives. Thus, screening of asymptomatic populations must be performed using a high-specificity approach.

So a lab may wonder: How specific is our test? Dr. Anderson shared sample data (**Fig. 4**) on a test that has 100 percent specificity. “However,” he said, “we need to keep in mind the confidence intervals.” For the test in Fig. 4, “the 95 percent confidence interval goes as low as 83 percent. Your assay could be 100 percent specific or it could be 83 percent specific—you don’t really know.” Testing more specimens—in this case up to 200—will tighten the CI, in this case to a 98 percent specificity “you can be more sure of.”

“The bottom line here is if you are going to use an assay for population screening, you need to do a more rigorous verification to provide acceptable specificity, to be comfortable you’re doing the right thing.”

Fig. 4

What is the specificity of this assay?				What is the specificity of this assay?			
Gold Standard				Gold Standard			
Assay being evaluated		Positive	Negative	Assay being evaluated		Positive	Negative
	Positive	20	0		Positive	200	0
	Negative	0	20		Negative	0	200
Specificity: $TN/(TN + FP) \times 100 = 20/(20 + 0) = 100\%$ (95% CI 83.16–100%)				Specificity: $TN/(TN + FP) \times 100 = 200/(200 + 0) = 100\%$ (95% CI 98.17–100%)			
Assays to be used for population screening require more rigorous verification to prove acceptable specificity							

Courtesy of Neil Anderson, MD.

The CDC recommendations recognize this need for verified high-specificity assays for population-based screening. If a lab cannot achieve this, the CDC says, “it can avoid testing low pretest probability populations altogether,” Dr. Anderson said, or use a combination of assays in an algorithmic fashion. A PPV calculator is available for that purpose (www.fda.gov).

The Barnes Jewish Hospital clinical laboratory uses a frequently asked questions document to get SARS-CoV-2 test-related information to providers. “We found it useful to have a form of communication that’s centralized and can be updated frequently,” he said. Also used is a clinical decision support tool, in which providers are told at the

point of ordering what sensitivity to expect at different days post-symptom onset and what can cause a false-positive.

There are secondary benefits to such tools, Dr. Anderson said. “Depending on how they’re built, you can use them to monitor appropriateness of testing. What we’ve done at our hospital is have our providers answer a question about days post-symptom onset. We are then able to mine that data and figure out exactly what type of testing practices we have. This can be very important because it can give you insight into the effectiveness of education and where you need more education.”

Included in the educational material should be information on the interpretation of positive results. “That’s because there are so many misconceptions about what a positive means,” he said.

What a positive result doesn’t mean or doesn’t reveal is when the person was infected, whether he or she is shedding virus (live or remnant DNA), or whether the person is protected against reinfection, Dr. Theel said.

Binding antibodies are often produced at high levels but unable to independently prevent infection. Neutralizing antibodies are able to bind the virus and lead to loss of infectivity by blocking the virus from entering host cells, and this is largely accomplished independent of other immune system components.

The current commercially available assays do not distinguish neutralizing from non-neutralizing antibodies, Dr. Theel said, and testing for neutralizing antibodies is challenging because classically it requires plaque reduction neutralization testing using live virus, which for SARS-CoV-2 requires BSL-3 facilities for viral culture. “Increasingly, though, BSL-2-level methods are being developed for this purpose,” she said, using a variety of different viral constructs, including, for example, pseudotyped Vesicular Stomatitis Virus expressing the SARS-CoV-2 spike protein.

For common coronaviruses, studies performed decades ago showed that in volunteers infected with 229E, IgG levels peaked at about two weeks post-infection but then returned to baseline at about one year, Dr. Theel said. Re-challenge of these volunteers did not lead to symptomatic infection, although two-thirds of them still shed virus for a period of time. These studies also suggested that protective antibodies likely drop off to insignificant levels, leading to loss of protective immunity after about 18 to 24 months.

For SARS-CoV, Dr. Theel said, neutralizing antibodies peak at about four to five months post-infection and then decline over the next two to three years and become undetectable by six to seven years. For MERS-CoV, neutralizing antibodies remain detectable for at least three years. The studies didn’t go beyond that, she said.

“The one question these studies didn’t address or didn’t have the opportunity to address is what levels of neutralizing antibodies are clinically significant and correlate to protective immunity.”

For SARS-CoV-2, a few studies performed in rhesus macaques found that initial infection led to the development of binding and neutralizing antibodies against the virus, Dr. Theel said, and post-initial infection and recovery, re-challenge with SARS-CoV-2 at 30 to 35 days post-initial infection led to very low levels of detectable viral mRNA and no recoverable virus post day two (Chandrashekar A, et al. *Science*. 2020. doi:10.1126/science.abc4776). “These animals did not develop any clinically significant illness, suggesting that the presence of antibodies, and likely other components of the immune system, does lead to at least short-term immunity,” she said.

In another study, researchers looked at neutralizing antibodies in 175 recovered patients and found that while titers peaked at about 10 to 15 days post-symptom onset, those neutralizing antibody levels were variable across all individuals (Wu F, et al. medRxiv preprint. www.medrxiv.org/content/10.1101/2020.03.30.20047365v2). About six percent did not develop any neutralizing antibodies ($< 1:40$), and about 30 percent developed low-level neutralizing antibodies ($< 1:500$). “The unknowns that remain are what neutralizing antibody titer is clinically significant and potentially associated with protective immunity,” Dr. Theel said, “and then how long they persist.”

While knowing the neutralizing antibody level may play an important role in the future, tests to detect neutralizing antibodies will not be routinely performed in clinical laboratories, she said. “So one of the questions is, do the

currently available commercial tests in any way correlate with neutralizing antibody titers?” It’s a tricky question to answer, Dr. Theel said, because the EUA assays now are qualitative and few studies published to date explored this correlation. Among the few studies published, the methods are highly variable, making it hard to reach comparative conclusions. “But in at least three studies, the correlation between spike and nucleocapsid-based ELISAs with neutralizing titers does seem to occur” (To KKW, et al. *Lancet*. 2020;20[5]:565–574; Okba NMA, et al. *Emerg Infect Dis*. 2020;26[7]:1478–1488; Amanat F, et al. *Nat Med*. 2020. Epub ahead of print: doi.org/10.1038/s41591-020-0913-5). The findings are preliminary and the subset of samples was fairly small, “so we should interpret these results with caution,” she said.

A positive result suggests only recent or prior infection, though the positive predictive value is affected by the assay’s specificity and the anticipated prevalence in the community. Here is what Mayo Clinic reports with its positives: “SARS-CoV-2 IgG antibodies detected. Results suggest recent or prior infection with SARS-CoV-2. Correlation with epidemiologic risk factors and other clinical and laboratory findings is recommended. Serologic results should not be used to diagnose recent SARS-CoV-2 infection. Protective immunity cannot be inferred based on these results alone. False positive results for IgG antibodies may occur due to cross-reactivity from preexisting antibodies or other possible causes.”

“For your own comments,” Dr. Theel said, “review the manufacturer’s instructions for use to make sure there’s nothing specific they recommend including.”□

Sherrie Rice is CAP TODAY editor. The American Society for Microbiology EUA SARS-CoV-2 antibody tests verification protocols are at www.asm.org/Protocols/Verify-Emergency-Use-Authorization-EUA-SARS-CoV-2.