



Article:

Lee F Schroeder, Michael A Bachman, Allison Idoni, Jennifer Gegenheimer-Holmes, Steven L Kronick, Riccardo Valdez, and Paul R Lephart.

Predicting Direct-Specimen SARS-CoV-2 Assay Performance Using Residual Patient Samples.

J Appl Lab Med 2022;3: 661–673. <https://doi.org/10.1093/jalm/jfab159>

Guest: Dr. Lee Schroeder is an Associate Professor of Pathology at the University of Michigan, where he is Director of Point-of-Care Testing, Associate Director of Chemical Pathology, and co-chair of the Laboratory Stewardship Sub-Committee.

Randye Kaye:

Hello and welcome to this edition of JALM Talk from *The Journal of Applied Laboratory Medicine*, a publication of the American Association for Clinical Chemistry. I'm your host, Randye Kaye.

SARS-CoV-2 diagnostic testing by PCR has been a critical tool throughout the COVID-19 pandemic. While many different PCR assays are now available, they differ in many ways, including in their pre-analytic sample handling. Many lab-based assays are approved for use with viral transport media.

Viral transport media stabilizes the specimen and allows for multiple samples to be used from one patient collection, which makes for a straightforward assay verification and method comparison studies. However, many point-of-care assays require direct or dry swabs, whereby the swabs are placed directly into assay reagents after the sample is collected from the patient. In order to verify these direct methods, typically two swabs must be collected from each patient, one for each method to be compared. This paired swab approach can be time-consuming and inconvenient. Further, it can be difficult to establish diagnostic sensitivities of these point-of-care methods because of variabilities in specimen type and population-specific viral loads.

In an article in the May 2022 issue of *JALM*, the authors described their development of a logistic regression model of SARS-CoV-2 assay performance that can use data from limit of detection and residual viral transport media studies to predict the performance of direct specimen assays in different patient populations. Today, we are joined by the first author of the article, Dr. Lee Schroeder. Dr. Schroeder is an Associate Professor of Pathology at the University of Michigan where he is Director of Point-of-Care Testing, Associate Director of Chemical Pathology and Co-Chair of the Laboratory Stewardship Subcommittee. Welcome, Dr. Schroeder. First question, what was your motivation for this study?

Lee Schroeder:

So, in short, it was our EDs [emergency departments] being backed up. It was late in 2020. We had been using DiaSorin as our rapid solution for COVID testing. It has a run time of about an hour and a half. So, we wouldn't get results out for two hours, three hours, and that just wasn't cutting it. So, there was an institutional decision that we had to find an alternative. And so, we started looking. You know, early on in the pandemic, there weren't many options. At this point, there were actually too many options.

So, we had a number of different assays we could look at: Xpert, ID NOW, and then a whole bunch of different antigen tests. And then, of course, with different populations to look at, and by that point, it was very clear that symptomatic and asymptomatic populations, different viral loads, and assays are going to have different sensitivities. And then on top of that, there was different specimen types, so nasal, nasopharyngeal, but also some assays use direct specimen testing and some use VTM. So, there's a lot of permutations and it really just wasn't an option to validate all this stuff ourselves, especially since a lot of the assays required direct specimens. You couldn't use residual VTM.

So, I went to the published studies and I was searching through the supplements, trying to find accuracy of the different antigen tests primarily by CT value, knowing that the higher the viral load, the lower CT value, the better they would perform, and then see what the accuracy was at different CT levels and then kind of mapping it in our heads to our populations that we're going to be looking at, to kind of guess sort of how they would do. And then, we realized we just needed to operationalize this as a regression model where we get the raw data and we fit curves that would estimate the ability to detect SARS at different viral loads or CT values.

And so, whenever I'm talking about curves or drop-off curves or regression models, that's what I'm talking about. It's this curve to know we know at CT of 10 or 15 with a high viral load, probably the assay is detecting 100%. But when you move over to 30, a cycle time of 30, a very low viral load, maybe it's only detecting 20% of those cases.

Now, actually, Gary Horowitz and his group at Tufts was ahead of us on this and they were doing a similar thing which then published in this journal, in *JALM*, and they did a very similar thing. They just broke the patients into four buckets of CT values, so low, little bit higher, and higher and so forth, and did a similar sort of assessment of the accuracy of two rapid tests. And so, understanding how those assays perform in these different buckets, then they could go look at their actual patient populations and extrapolate how they would perform overall.

So, we had this additional problem in that the assays we were looking at were all direct specimen assays.

And so, direct specimen assays, it's hard to get the data. So, we got a data problem. If you could use VTM, residual viral transport media, then you have all the samples you want and it's easy to generate your data set, you could fit the curves, see how the assays perform across a wide range of CT values, but that really wasn't a straightforward solution for us having direct specimens.

So, now we had previously published on the ID NOW accuracy. It was from April 2020, when we looked at it the first time around. And then, we were using it with a nasal swab and comparing that against nasopharyngeal VTM samples in Xpert and DiaSorin in our core lab. There, we found that the ID NOW didn't perform very well with the nasal specimen. They have 50% across-the-board sensitivity. However, two things. One, we also took the VTM samples and ran it on ID NOW, even though that actually was off label, just to see how it would perform. It actually performed better with the VTM samples from the nasopharyngeal swabs than it did with the direct specimens from a nasal. That was one clue that we really would need to focus on an NP swab. But the other one was that when we did the LOD evaluation of these different instruments, the LOD actually looked pretty good on ID NOW if you adjusted for the fact that it was going to get a direct specimen. So, you made our LOD dilutions, but then made an adjustment for the ID NOW to account for the fact that it was going to get like all the viral particles on that swab and not just a portion that gets diluted into the 3 mL VTM tube.

Those two things gave us the idea that what we could do is we could, in fact, use VTM samples on all these assays that are really direct specimen assays, just in the evaluation phase to generate a bunch of data, but then adjust for the fact that it's going to be a direct specimen test again that gives it a bump up, a benefit.

So, we made that adjustment after the fact. And then, with these curves, we could estimate a couple things. One, which assay is going to work? How well in different populations? And, two, we'd be able to provide some kind of consultation to our clinicians on that test accuracy in terms of missed cases per 1000. So, that was the background to this study.

Randye Kaye:

Okay. Thank you. So, that was a very detailed answer, which is great. I have a feeling you've already gone into the answers of some of my next questions. So, I'll just give you a chance to fill in any details here. For instance, let's talk about the different ways of measuring test performance. In

the article, it describes different methods, like analytic sensitivity curves, and predicted positive percent agreements and clinical sensitivity. So, anything else to say about why those different approaches and what the difference is between them?

Lee Schroeder:

Yeah. So, I have to say most of the projects I get involved in are pretty complicated. A lot of informatics and math usually. But, this was particularly complicated. And, we went back and forth a lot of times on what these different metrics should be called. It was really bordering on philosophical debate. But, in the end, we landed on what I think is the right foundation.

So, the first concept is analytic sensitivity curve. So, that's in Figure 1. This is those drop-off curves. And these are essentially limit of detection curves. But, instead, normally, in an LOD, you get dilutions of viral particles. You determine at what dilution the assay can detect 95% of samples and that's your LOD. That's basically what we're doing, except instead of using, first of all, viral concentrations as a metric, we're using CT values. And then we generate a bunch of tubes of different CT values to calculate what the accuracy would be at each of these levels, and we have this curve. That curve is really independent of anything else, or any other instrument or assay.

So, at first it was thought this should be some kind of PPA, percent positive agreement, because you're using a CT value from an instrument, comparing to this one, but really it's not, because it does not depend on which assay created the CT value for that specific tube. You can think of it as just almost a dilution set.

So, we decided to call that analytic sensitivity curves and in contrast, in Figures 2 and 3, once we took those analytic sensitivity curves and mapped them onto the distribution of CT values in our patient populations in the ED, symptomatic or asymptomatic. So, if, for all the patients that had a CT value of 20, then the assays are largely going to detect all of those cases, but as you move up to higher CT values, the assays are going to start missing them.

And so, once you add all those detected cases up and divide by total number of cases, then that is your percent positive agreement. We called it predicted because this is all prediction and it wasn't something we actually did, and the PPA does change when you use different assays as the reference and, in fact, that would be the case here as well.

So, we used the DiaSorin to generate all the ED data. So, you could imagine that the CT distribution might be different using the DiaSorin than it would have been had we been using our

core instrument for all the ED patients. There'd be much more low viral load detected. And if there's a lot more low viral load detected, then any assay is going to miss more of those compared to the distribution where there's mostly high viral load.

So, that's the PPA that we used to compare each of our different assay sensitivity curves to different patient populations. And while that number isn't really directly translatable to anything clinically useful, it's at least an index, a relative index to compare the different assays between each other to see which one is performing best.

One of the assays that we're looking at is the DiaSorin. So, immediately, you might think, well the PPA of the DiaSorin with itself should be 100%. But, that's not the case, and it's not the case because if an assay is detecting any samples near its LOD and all the assays, even the reference assays are detecting near the LOD, you're going to be missing some cases. So, if you re-run those samples all again on the same instrument, you would miss some of those hits and you would detect some of the ones that were missed the first time.

So, anyway, the PPA of an assay itself is not 100%, and that's what we were demonstrating, or that was part of what we would get out of Figures 2 and 3. Anyway, so those PPAs there are just some kind of relative measure between assays, but then, finally to get clinical sensitivity, which is what we really needed at the institutional level to make a decision on what we were going to use for testing, we are going to file that Infectious Disease Society of America benchmarks.

And so, that is to say, I'm going into details of that later, but basically, 20 missed cases per 1000 tested was acceptable. To get there, we had to not just use the CT distributions of positive patients in our health records, we had to add back the expected missed cases that we had missed the first time around in clinical practice. So, since we all know every instrument, even the core live instruments would be missing some cases, based on the analytic sensitivity curve, we can actually predict how many were missed based on what the distribution looks like and add those back, artificially enrich it. So, we actually have the true population, the true expected population. And then, once you applied our analytic sensitivity curves to those, now we actually have a clinical sensitivity and we could calculate the missed cases per 1000.

Randy Kaye:

Okay. Thank you. And just a note to the podcast listeners, those figures are available in the *JALM* article and I'll let you know the title of that once again at the end of this podcast. And we'll get to more of your predictions in a second, but just quickly, how did you translate the viral transport media data to direct specimen data?

Lee Schroeder: Okay. This is I hope pretty straightforward. So, with the ID NOW, one point, you could use VTM and I think the package insert was 200 microliters of VTM to be transferred in. So, we were transferring 200 microliters of VTM, but that was representing only 6.6% of the total VTM volume that the swab was in, which is 3 mL. So, you have the swab and you put it into the 3 mL VTM tube, diluting all those viral particles throughout the solution, but then you only take 200 microliters out. So, you're only getting 6.6% of the viral particles and putting that into ID NOW. For Sofia, it ended up being 1.6% because you only transferred 50 microliters.

So, what that means is that, in clinical practice, there would be 15 times more and 60 times more for those two assays, more viral particles in clinical practice than in these VTM studies. Now, since PCR approximately doubles gene targets each cycle, to get to 15 times ends up being 3.9 CTs and for 60 times it's 5.9. So, that's how we adjusted.

So, if we ran a VTM specimen that was characterized as having a cycle threshold of 25 on a reference instrument and ID NOW detected it, we would assume that in clinical practice, yeah, it could detect a cycle count of 25, but even if you detect a higher one, it's 25 plus 3.9 being 28.9 cycle threshold. That was the approach we used. And then we use that adjusted data to fit the logistic regression fall-off curve for the analytic sensitivity.

Randy Kaye: All right. Thank you. So, can you describe what the model allowed you to predict and how much confidence? So, what kind of confidence do the predictions have?

Lee Schroeder: We were pretty confident in this strategy. The findings were that these analytic sensitivity curves for the ID NOW using nasal specimens, because we actually had that data from a paired swab study we did earlier in the year. That one and the Sofia using nasopharyngeal specimens, those were pretty much overlapping, while the nasopharyngeal on the ID NOW and the nasopharyngeal in the DiaSorin, those two overlap, and the nasopharyngeal in the DiaSorin, that's what we had been using all along.

So, since the ID NOW using the NP sample overlapped pretty well with the DiaSorin, we were pretty confident that was going to work, but we still needed to get an idea of how many missed cases there would be. And so, again, those ideas as benchmarks is that in a symptomatic population, successful to have up to 10 to 20 missed cases per thousand tested while more than 60 per thousand would be unacceptable. So, it's a gray area there. So, we used 20 missed cases per thousand tested as our quality goal, essentially. And what we found is in our symptomatic population, both the DiaSorin and the ID

NOW using the NP swab were predicted to have less than 20 missed per thousand, far past the 20% prevalence. So, of course, the higher the prevalence, the more missed cases you'll have. They're probably going up to 40%, we actually didn't plot that far out. Whereas the ID NOW using a nasal specimen, or the nasopharyngeal specimen on antigen assay, they are across the threshold of 20 at around an 8% prevalence, which we commonly would see in peaks, even outside of peaks.

And then, in the asymptomatic population, both the DiaSorin and nasopharyngeal ID NOW were predicted to be less than 20 up to around 8%. And actually, we did hit 9%, I think, in the peak of the Omicron over Christmas, but, typically, it's much lower than that, while the ID NOW nasal and NP antigen crossed the threshold at about 4% prevalence in the asymptomatic population, which is pretty commonly seen.

So, we liked the idea of going with the ID NOW, but using nasopharyngeal specimen. We took these predictions to the highest level of leadership for decision-making. It's an institution. We decided to go forward with it. But, we did run a final paired swab study. You know, the ID NOW had been in the headlines for so long as being having low sensitivity and we ourselves published that, so our providers, we didn't think they were going to have much faith in it unless we ran it through and did the paired swab study with patients. And we did that and found that it detected every single sample that the DiaSorin detected in 100-person study and actually detected one additional patient. So, those were the findings and what gave us confidence.

Randye Kaye: Okay. Thank you, doctor. So, finally, how do you think this can be used going forward? What's the value? Does the study have value for other types of testing conditions?

Lee Schroeder: The short answer is yes, and an obvious extension would be to any infectious disease like influenza, strep throat, where there are rapid tests that require direct specimens, but also, in this case, there would be PCR assays. It's a directly analogous extension there. But, I think you could probably use this in a number of other tasks, as long as they are the quantitative result that is typically communicated as a dichotomized result, up or down, when the quantitative value is above some threshold, the decision limit, so it's dichotomized. And then, particularly, if there are lateral flow tasks that exist, for that analyte, that give just the dichotomized results. Obviously, you see this with COVID, but you could potentially see it with like drugs of abuse testing, pregnancy testing. And there are plenty of lateral flow tasks and we don't use them for TSH, D-dimer, and troponin, but they exist. So, I assume someone's using those. And, this approach could be helpful in predicting how

accurate these different assays would be in very disparate populations.

You know, package inserts often have performance characteristics, sensitivity, and specificity that are higher than what you end up seeing in published independent studies. And, I'm sure there's a number of reasons for this, but one of the reasons could be that the population that the assay was studied in has a very severe form of disease.

So, the textbook teaching is that sensitivity and specificity are characteristics of the tests and are independent of the population. Whereas predicted values are dependent on the population that we test probability, and so forth. But, that's only true if the only thing that varies between populations is prevalence. If the disease severity changes, then the sensitivity is clearly going to change, and we see that obviously in COVID, right? The symptomatic and asymptomatic populations have very different sensitivities because their viral loads are different.

Randye Kaye:

Right.

Lee Schroeder:

So, the same concept could be applied to any task where you could imagine there is some difference in the population expression of different antigens and so forth, and you would like to know in *your* population how is this actually going to work and not rely on the package insert or hopefully finding some published study that is representative.

Randye Kaye:

All right. Well, thank you so much. Thank you for joining us today.

Lee Schroeder:

Thank you very much. Great talking to you.

Randye Kaye:

That was Dr. Lee Schroeder from the University of Michigan describing the *JALM* article "Predicting Direct-Specimen SARS-CoV-2 Assay Performance Using Residual Patient Samples." Thanks for tuning in to this episode of JALM Talk. See you next time and don't forget to submit something for us to talk about.