



Article:

Mahdi Mobini, Nancy Matic, J Grace Van Der Gugten, Gordon Ritchie, Christopher F. Lowe, and Daniel T. Holmes.

End to End Data Automation for Pooled Sample SARS-CoV-2 Using R and Other Open-Source Tools

J Appl Lab Med 2023;8(1): 41–52. <https://doi.org/10.1093/jalm/jfac109>

Guest: Dr. Daniel Holmes is a Clinical Professor of Pathology and Laboratory Medicine at the University of British Columbia and Head and Medical Director of the Department of Pathology and Laboratory Medicine at St. Paul's Hospital in Vancouver.

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. The COVID-19 pandemic generated renewed interest in sample pooling for SARS-CoV-2 viral testing in an effort to save on labor and supply costs and to manage supply chain shortages. With this approach, clinical samples are mixed in pools and each pool is tested collectively as one sample. If the pool's test result is negative, it is assumed that all samples within that pool are negative and their testing is complete. If the pool's test result is positive, samples within the pool are subsequently retested individually to determine which sample or samples are positive. Sample pooling may be an effective strategy in times of low disease prevalence. However, the benefits are lost in times of high disease prevalence where re-running samples from positive pools is required frequently.

An article in the 2023 special issue of *JALM*, on data science and the clinical laboratory, describes one lab's approach of using entirely open-source software tools to develop a modular software application stack to manage the pre-analytical, analytical, and post-analytical processes for pooled testing for SARS-CoV-2. To today, we're joined by the senior author of the article, Dr. Daniel Holmes. Dr. Holmes is a Clinical Professor of Pathology and Laboratory Medicine at the University of British Columbia. He is Head and Medical Director of the Department of Pathology and Laboratory Medicine at St. Paul's Hospital in Vancouver. His many interests include data science and its application to data automation, visualization, and clinical utilization. Dr. Holmes, welcome. As your lab responded to the COVID-19 pandemic, what was the primary driver for this project?

Daniel Holmes:

Well, listeners may be aware that in the summer of 2020, some of the big diagnostic companies were running short on their reagent supply, their proprietary reagent supply for their instruments, and so, we were put in a position where we weren't able to get reagent for our instruments at least not our high-throughput automated instruments. And so, we

needed to find a way to one, use our less automated methods and two, to stretch the reagent that we used for those methods. And so, working on a data automation project that could simultaneously take care of the tracking and reporting of pooled COVID-19 samples would permit us to increase our throughput by about three-fold and get rid of any manual transcription events that were occurring in our lab. So, the goal was to streamline the operation from all fronts, both in the consumption of reagent and in the improvement of the data throughput.

Randye Kaye: Thank you. So, your lab took an approach of developing your own software using open-source tools. Was there a commercial product that could have been used for this purpose and if so what limitations led you to develop your own product?

Daniel Holmes: Well, there was a pipeline commercial product available to do some of what we wanted to do. One of the vendors did have data automation software developed for their platform, but it did not handle COVID pooling, or any kind of sample pooling, and it was not yet Health Canada cleared. So, there was no way that we could adopt that software and have it deployed in the timeline that we wanted to. We started working on this project on about August the 15, 2020 and we had a target go live date of September the 20th. So, we were really forced into a position where we had to develop our own software and I had done similar kinds of data automation projects for our mass spectrometry lab, and so I was pretty sure about the process by which we could get this achieved in the timeline that we had.

Randye Kaye: All right. So, I'm thinking that maybe some other labs might consider a similar approach and they might want to know about money. Did you need a particular budget for this project?

Daniel Holmes: We actually didn't really have a budget when we started because it was done in such a precipitous fashion. So, the funny thing is we walked over to anatomical pathology and scavenged an old computer and then at lunch time, we walked down to Staples, my friend Mahdi and I, and we purchased a solid state drive to speed up the computer and then we ordered some RAM online and that afternoon, we began programming. Now, that isn't to say that we didn't have any budget at all. Subsequent to that Mahdi's time was obviously remunerated and we did get a small budget for purchasing a serious computer to do the data processing.

Randye Kaye: All right. Thank you, but you kind of started it and just with your fingers crossed it sounds like.

Daniel Holmes: Yeah. I mean I knew that they were going to pay him one way or another. The other thing we did get budget for was that we got a new liquid handler. That wasn't budgeted. I mean it was not budget that came to us. It was budget that was used to buy liquid handlers for everybody, but the important part is that we got Hamilton Liquid Handler that had a large deck and a barcode scanner and that permitted us to automate a lot of the liquid handling and data transfer between the liquid handler and the devices.

Randy Kaye: Okay. So thank you. What technical challenges did you find along the way and if so, how did you overcome them?

Daniel Holmes: So, there's a few technical challenges. Let's just start with ones that are based on the laws of physics that you can't control very much. A lot of the COVID samples came to our lab with mucus adhered to the nasopharyngeal swab. That nasopharyngeal swab stays in, at least in our setting, it stayed in the collection tube until it arrived at our facility. What that meant is that the original primary collection tube frequently had this mucus in it and when we put that onto the liquid handler, the liquid handler would detect that as a clot and jam up, or it might pull the strand of mucus across the deck and potentially contaminate many other specimens, which is not something that you want to have happen. So, unfortunately, we had to do manual pour-off of those specimens into smaller containers, avoiding the mucus strands. And I wished we could have automated that. If I were doing it all again, I would figure out a way to do that and have ideas about how it could be achieved because I know that others have done so.

The second thing is that, although there are many complex and optimized mathematical strategies for COVID pooling, these are rather opaque to technologists and in healthcare, people like to know how things work, particularly in a lab-developed task. They really want to know how things work. So, if there's a technical challenge, they can intervene. So, we did not opt for some of the more complex pooling strategies that people like Chris Builder have written about in literature. We wanted to stick with a simple Dorfman pooling strategy because everybody would understand the principles behind it. That meant that we didn't have the most fully optimized strategy, but we made the programming of the liquid handler much simpler and there would be no recursive kinds of liquid handling that would need to occur in order to get to the end stage where we could finally do all the reporting.

So that was another piece of the puzzle. I'm just back to the point about clots or whether pieces of mucus in the sample, sometimes they would plug our liquid handler tip up and then we would have to remove that particular specimen from the

pooling list. So, there were all kinds of little technical challenges like that. Another example was that the extraction device couldn't handle more than 25 characters in its specimen identifier, so we couldn't just concatenate all of our tube IDs together. You know, there were a lot of communication challenges behind firewalls that the company had put in place and network-attached storage that the company wanted us to use rather than connecting directly to their devices. You know, if I talk much more, we'll be too down in the weeds, but suffice it to say that everything that seems easy at the beginning, ends up being harder than you think and you just have to take one challenge at a time and make that today's problem and sort it out and by and by you get through all of these little technical challenges until you have a fully end-to-end solution which you can finally test.

Randy Kaye: Just like life. So they're not only challenges but there are also changes, like there were a lot of changes in protocols for SARS-CoV-2 testing. So, how did your technical staff adapt to those changes?

Daniel Holmes: The change that we had to deal with pretty quickly was that in September of 2020 in our jurisdiction, the positivity rate was less than 1% and that's highly amenable to throughput performance improvements by using COVID pooling, or sample pooling I should say. But the problem was that as the positivity rate climbs, the value proposition associated with sample pooling diminishes quickly. So by November, our positivity rate was such that we were having to do a lot of positive pool resolution as they say. It was becoming a frustration for the technologists. So, my colleague said to me just casually, "Hey, you know, you can tell whether a sample's going to be likely positive based on its ZIP code," which we call postal code in Canada.

And I said "Oh, that's a very interesting idea" and then I thought to myself, well why don't... We have that information. We have the historical results. Why don't we just see if the sample comes from a positive postal code and if it is, then we would divert it from the pooling process.

And so at 3:00 in the afternoon on, it was about November the 4th, I said to Mahdi, "Hey, like we need to make a little app that we can swipe the sample barcode, it will look at the old postal codes and we'll see whether this sample is likely to be positive" and we start working on that at 3:00 in the afternoon. By 2:00 in the morning, Mahdi had completed this app and we deployed it the next morning and it significantly decreased the positivity of samples going into the COVID pooling, and then we were able to run the process for quite a bit longer until the positivity rates became overwhelming. Subsequent to that, we did work on a machine learning model to add other features, other than postal code, to determine

whether a sample could be pooled, but the positivity rate never came down to a point where pooling was a good strategy anyway, but the good thing was that the automation workflow was not just for pooled samples. It was for all kinds of samples and so we can continue to use the software that we had written for singleton analysis all the way through the pandemic.

Randye Kaye: And still using it, I would imagine.

Daniel Holmes: Yes. So, we are still using and the techs were so happy with it, they wanted us to add the other tests that we do up in that virology lab. So, we do CMV, NBK, and other viruses, and so we used the same software to do all of the testing on the more manual platforms we use for extraction and thermocycling.

Randye Kaye: And finally, you already answered this one instance but if you could do it all over again, is there anything else that you would do differently? What advice would you give to other labs considering a similar approach?

Daniel Holmes: I think I would be, I mean this maybe sounds a bit trite, I would be prepared for something like this. I would not build lab-developed test processes that don't scale. I would think of scaling up right from the beginning. So, even if you're only doing a small batch of such and such a test every week, I would automate everything I could right from the beginning under the assumption that things would/might have to scale, particularly if it's the sort of thing that could scale, whether there was a centralization process or whether it was associated with an infectious disease where the demand would naturally increase. So I think I would build that into every lab-developed test process, automate it as much as I can. The other thing I would do, as I alluded, I think I would try to automate the pre-analytical a little bit more.

The slowest part of our testing process was not the analytical portion. It was logging the specimen in and decanting the specimen to avoid that strand of mucus. So, I would work on whatever automation strategy I could, whether it was connecting the various lab information systems or working on robotic process automation process for handling the paper requisitions, doing optical character recognition, taking the burden off the manual processes because the lab part is mostly automated right from the beginning even before we added our software to it. But the sample handling process is highly manual and if I were doing it again, I would put more energy into the pre-analytical process right from the beginning.

Randye Kaye: All right. Thank you so much. Thank you for joining me today.

Daniel Holmes: Thanks very much Randye.

Randye Kaye: That was Dr. Daniel Holmes describing the *JALM* article, "End to End Data Automation for Pooled Sample SARS-CoV-2 Using R and Other Open-Source Tools." This article is part of the January 2023 special issue of *JALM* on data science and the clinical laboratory. 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.