Analytics in the Clinical Laboratory with R: An Expert Panel

AACC Webinar — August 20, 2020

Presented in collaboration with R/Medicine

Webinar Information

Transcript of live Q&A chat

Moderator: **Shannon Haymond, PhD**
Vice Chair for Computational Pathology
Lurie Children's Hospital of Chicago

Panelists: **Daniel Herman, MD PhD**
Assistant Professor, Pathology & Laboratory Medicine
University of Pennsylvania

Daniel Holmes, MD
Clinical Professor, Pathology and Laboratory Medicine
University of British Columbia

Stephan Kadauke, MD PhD
Assistant Professor of Clinical Pathology and Lab Medicine
Children's Hospital of Philadelphia

Patrick Mathias, MD PhD
Vice Chair of Clinical Operations, Laboratory Medicine and Pathology
University of Washington School of Medicine

Q: Is it possible to get mentors from this group to guide new learners?

Daniel Holmes: The best place to start is taking a short course with AACC, R/Medicine or MSACL. All of them have easy entry courses. There you will meet people who are part of the community and, of course, google your question. Someone will have answered it!

Q: If one knows nothing about R coding and would like to start learning, where would one go to look for training/courses?

Daniel Holmes: If you want to take a web-based course you could try Data Camp or a similar course. Otherwise I would take a face to face crash course. I also relied heavily on https://www.statmethods.net/ when I was starting.

Daniel Herman: AACC is actively working on how to extend beyond our existing in-person courses and figure out how create content and an active community that would help people learn and use R.

Shannon Haymond: A lot of people recommend starting with the *R for Data Science* book, which can be found here online and with exercises. We are going to get into lab medicine-specific resources in a later part of the webinar. https://r4ds.had.co.nz/

Q: Can you recommend a R-studio setup for optimum performance when creating real-time live view dashboards? Do you use R-studio Server Pro? If so, what is your architecture?

Daniel Holmes: Depends on how much computational power you need. If you need many users to be on simultaneously you can deploy on AWS or R-Studio server. However, if you just want one or two users, you can deploy locally on a single computer. You can also deploy for free on shinyapps.io.

[What about for an enterprise setup?]

Daniel Holmes: Dockerize and put on AWS or R-Studio Server are the answers that come to mind.

Patrick Mathias: We have a locally developed setup that integrates Git and Docker built by our data engineer. It was high overhead to set up but has worked well and as multi-language support. I've heard that RStudio Connect has worked really well for the team at CHOP.

Q: How do you handle version controls?

Daniel Herman: There are lots of strategies for this. I currently containerize apps. There are also some packages in R to help pin versions of libraries and config.

Q: How can you teach what you do not know?

Shannon Haymond: This can be a challenge. Particularly if the server or computer running R is not connected to the network that houses the LIS data extract.

Q: What issues did you run into trying to access the data from your LIS to use R?

Shannon Haymond: R has a lot of flexibility in the type and size of data it can read in. So, once you get your data out, you can definitely get it into R and parse it appropriately.

Daniel Holmes: This is a serious challenge in many environments. You will need to make friendly with the Lab Information System people. There are ways to get data by guerilla tactics if your LIS is accessed through a shell window. See the Expect language.

Q: Can you discuss what database(s) you draw from e.g. LIS, and how you tap into those DBs. Do you need significant institutional IT support to access the data?

Shannon Haymond: I have used LIS data extracts and also directly connected to AWS redshift database using R. The IT support needed may be related to networking and install of R.

Daniel Holmes: Many LIS will allow access using ODBC. R will do this.

Q: How do I get the rest of my team to get into R?

Daniel Holmes: You need to show off what it can do without making them feel defensive. I have found that there can be a fair bit of fear among traditional decision support people when you show them what R can do.

Q: Is it possible to verify and validate methods and verify reference intervals?

Shannon Haymond: This is a great use case. I know it is possible to do these types of analyses using various statistical packages in R (combined with the more basic functionality for data prep, analysis, and plotting, etc.)

Q: Can lab managers, scientists, or pathologists study this program, or it is the job of information technologists?

Daniel Holmes: Definitely. The point is to empower managers, scientists and pathologists to answer their own questions because they understand the clinical context.

Q: How do you access the raw data? Do you have to request it from IT? It can be difficult to get access to the data warehouse.

Shannon Haymond: Could you work with your IT or data analytics group to provide you with an abstraction layer of data or recurring data feed of some sort vs direct access? I agree that getting direct access to data warehouses can be difficult.

Daniel Holmes: If your LIS allows an ODBC connection you can get it that way. But ultimately, you are going to need to make your institution understand that you need free and easy access to data to answer the clinical and operational questions.

Q: Can R be used to "fill gaps" in LIS features? Are there specific LIS vendors who better support this kind of "user add-in" functionality?

Patrick Mathias: R is best suited to analytics or data analysis gaps. In general, the LIS feature that can help feed this is access to a reporting database or the ability to generate reports from the LIS.