Paper DH12

Validation and Open-Sourcing: Can We Tame the Two Headed Beast?

Ben Barnaby-Pass, PHASTAR, UK

ABSTRACT

Whenever we speak about open sourcing it always comes with the rebuttal of 'How do you do validation?'. While the validation process completely differs from SAS®, the process in R® can be as 'simple' as SAS too.

Of course, we externally source packages to help make R function, tidyverse and Shiny® are just two examples, these packages must have a certain level of integrity to them. We are unable to run a package that hasn't had an official release, and we are unable to use these packages in a continuous update state, this prolongs the validation process.

To get past this issue there's the opportunity to freeze the packages at a certain version, however, it puts the pressure on companies to keep kicking the can down the road. Even from this abstract you can see the troubles that open source has on Validation.

INTRODUCTION

Validation is a cornerstone of the pharmaceutical landscape, however, we are always moving and trying to progress the landscape for good. This is where open sourcing comes into play. There is a mass movement of open sourcing and working on coding languages that can do this. RStudio® and Python® are leading the way, and for good reason.

R is a mathematical language which allows the user to create graphs and use data efficiently. Of course, this is a massive help in the work we do. However, R is a very baseline language that doesn't seem to wrap around the complexity we need within the pharmaceutical world. This means that the more experienced R coders can create in-depth packages for people to use to be able to make the intricate databases and tables that we need to be able to provide to clients.

This process means we need to create new ways and macros to be able to work with the new processes that are being created. However, this isn't as easy as copy and pasting the processes that we have done within our given company's coding language. Coding languages differ vastly, and this means that we must adapt.

Adapting means leaving behind the known and going into the unknown. For an industry that needs solidity, that can be a struggle. We have processes to adhere to certain governing bodies while also making sure our client confidentiality is upheld. This means that there's always some pushback from new ideas as there are many hoops to jump through. Validation is one of the main issues that the pharmaceutical landscape hasn't settled on, especially within open-sourced products.

Open-Sourcing

While open sourcing is quite new and shiny, there have been quite a few pushes to make this the main way to code and operate within companies. There are a few reasons to go to open sourcing and their languages.

- Security – We can review code, and anyone is able to review this code. Ensure there is a better understanding of what the code might be doing and what it should be doing (this falls into the validation side of things which I will talk about later).
- Cost effective– The languages which allow open sourcing generally are more cost efficient than others. This allows the budgets for companies to lower.
- Flexibility – We can create UIs within RStudio, using Shiny. These allow us to make a user-friendly area where people can easily pick the program up and be able to work on studies at a high and consistent level.
- Transparency – This allows for continuous review of code, allowing for a tight-knit level of program which should be of very high quality.

These are only a few of the advantages of using open-source software. The sudden urge to go to open source isn't all just advantages. As I have said, there are also drawbacks. The biggest is the fact that we will have to move all these processes over to the language that we choose. This is a harder task than at first glance. If you choose to swap to RStudio®, the language is pretty baseline and takes a lot of manipulation to be able to conform to the needs for a pharmaceutical company. Luckily enough, there

are packages out there already been made that companies can use. For instance, the tidyverse package, this is a massive package which makes it a lot easier for users to be able to wrangle data as we wish.

However, we know that getting external packages must be scrutinized to a massive standard. This is where the validation steps come into it. Validation is needed to ensure that we are doing what we are meant to be doing and showing certain tables as they are meant to be shown. This means that any inclusion of a package (be it internal or external) must be checked over to ensure that the package has what we are hoping to be using within the code. This process can take a very long time and has proved to be an issue when starting an open-sourcing package in the past.

While the process can take a while, it is also worth noting that with the use of external packages, it is wise to keep note of the version number of each. As you need to make sure that the package has had an official release and been able to pass through the designated checks that a package must go through to be able to have an official release. This will help keep the integrity of the study and ensure that your work keeps the high quality that is needed for pharmaceutical companies.

VALIDATION

Validation is the process which we use to make sure that the programs we make are doing what we intend and there are no issues within the program. We can do validation by:

- Double Programming – Independent code created by two coders where they compare their findings and make sure that they match while having no errors or warnings present.
- Code Review & Spot Checks – Code gets assessed to identify issues to increase code and output quality. And spot checks are done to ensure that the code is producing an accurate output.
- Creation of Test Cases – Test criteria and test data can be created and designed to produce and expected result. Then the created output is tested based on the given test cases to see if there are any errors.

After these steps are done, we include a stage 3 QCer to back up the findings that is done by the stage 2 programmers. This is done by a different, third individual. They are responsible for:

- Reviewing the stage 2 QC conducted – Ensuring there were no errors missed.
- Review the program output – Ensuring that the output is how the shells are constructed, while making sure headers and footnotes are processed correctly.
- Document their evidence and actions – This is so the stage 2 programmers can correct their mistakes if there are any.

These are the processes that we know within SAS and have been using these processes for years without any change. However, with the move over to open sourcing there have been some issues trying to get these steps passed. This obviously is a massive issue, as it makes verifying tables harder that their quality and accuracy are what are needed to pass the validation steps.

The problems arise with the use of outside packages. Outside packages are great in the fact that they can create easier functions to work with within RStudio. However, these packages need to be validated so that they are needed within the tables that we are creating. This is a lengthy process as some packages within RStudio are quite large and we may not have to use the full package within the program which creates some issues on the use of packages.

This isn't the only issue with using outside packages. Packages can be created and not have an official release. This means that these packages may not have gone through the necessary steps for quality control which is troubling because this may lower the quality and accuracy of our programs which means that validation may not pass. This is the biggest question mark around validation and there hasn't been a solid process in which pharmaceutical companies are able to iron this issue out. Of course, there are work arounds (which I will go into later), however these issues will always be present if we use outside packages.

FUTURE OUTLOOK

With open sourcing on the rise, these are the processes that we need to make sure are 100% set in stone before being able to move studies over to these coding languages. Validation is just one of the many processes which we need to have official review of before we can conduct this in real studies.

The process of validation shouldn't move too far away from the process that we do right now. There are a few things that need to change. The language that we work within has a different way of working so this means we will have to change the processes there. This includes moving away from the use of macros and changing to functions. RStudio generally is built using a lot of functions as this is the most efficient way to be able to work with RStudio, and for good reason as this will help with time and resourcing when we work on studies.

Another process that needs to change is the fact that we need to find new ways to be able to work on the stage 2 QC, this means that we need to find new ways to compare tables together. Comparing tables is the pillar to validation and we can't fully validate any table unless we are able to check that the table is doing what we want, and the code is error free.

To iron out these issues there are a few ways to be able to validate open-source languages. One way is to freeze the external packages that are used, this means that we use a version of the package and ensure that the validation process is done for these packages. If we can freeze and validate the package, we will be able to use this version until the package becomes obsolete **OR** if the package updates to include code that we will need to make the process more efficient. Freezing packages seems like the optimum way (currently) to be able to use open-source software to work on pharmaceutical studies. The only way I can personally see the process for this changing is if we are able to find a way to validate the external packages at a faster rate, this doesn't seem likely as the validation part of the process is a very important and already optimized to take the least amount of time possible.

Another process that we will be able to adapt is the process in which we validate. This isn't 100% linked to the use of validation, however with the use of automation tools we will be able to speed up the process of creating all 3 parts of a pharmaceutical study. If we can make an open-source automation tool for each of the processes this will help save time and keep consistency between tables. This allows for validation to take less time as we are using the same shells for programs which means we know that the table is doing as we want and is likely error free. Not only would it speed up the validation process, it also would improve study timelines overall as we are able to create tables at a faster rate. Creating a better environment for stage 3 QCers and creating a general case that the tables created would have consistency between each of their versions.

CONCLUSION

We are moving in the right direction when it comes to open sourcing. For instance, the industry has been able to make moves within the open-source world like Pharmaverse, they are doing a brilliant job at creating packages that we are able to use! And shows the work that the industry has put in is 100% worth the effort that we have put in already. This is just one step of the process that has been able to be made and has helped outline the direction that we should be going in. However, I believe we are a few steps away from being able to create the spine of the process. There are still obvious validation issues, however, there are processes which are being penciled in that we can use while we find a concrete way to be able to use this process within open-source programs.

We are moving over to an autonomous world where we can create shell programs and use these to create tables. With this transition over to this kind of work this will help create an easier way for validation to be used. If we can work on these in tandem, we will be able to create a full portfolio of tables in which we have the shells have a general validation check done over rather than the full table every time, which will help improve the time of studies while also making sure that the consistency of tables is always present between studies.

Overall, as a final thought. I believe we are on the verge of a breakout. We have the spine of the work in place, we just need to iron out some of the smaller issues. I just ask the question: Are these issues as small as we really think?

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Ben Barnaby-Pass
Company: Phastar
Address: WFH, UK
Email: ben.pass@phastar.com