R Validation Hub

1. Introduction

Over the past two years, the AIMS SIG has been focussing on the use of R software in the industry. This focus has covered both the usage and application of R, and the hot topic of validation. The topic of validation has been presented previously. Needless to say, it's a complex area, full of myths and opinions but with limited guidance!

Earlier this year AIMS set about trying to simplify the topic, with a goal to provide an "R Validation Hub" that would provide (pharmaceutical) companies with the tools and information they need to use R within a regulatory framework.

2. Our Journey

In May 2018 the AIMS SIG proposed the Validation Hub concept to the R Consortium, "a group organized under an open source governance and foundation model to support the worldwide community of users, maintainers and developers of R software". The Consortium's Infrastructure Steering Committee (ISC) approved our application for support in our efforts and granted us an initial $4,000 to get the project going.

In August, Andy Nicholls was invited to present on behalf of AIMS at the inaugural R in Pharma Conference. A validation workshop at the conference led to an expansion of our project team to include members from Pharmaceutical companies and regulatory bodies from around the world. At the time of writing the group now has representatives from:

Abbvie, Amgen, Biogen, Boehringer-Ingelheim, Eli Lilly, FDA, Genentech, GSK, IQVIA, Johnson & Johnson, Linux Foundation, Merck, Merck KGaA, Novartis, Pfizer, Phastar, PPD, PRA, RCPE, RHO world, Roche, RStudio, R-consortium, Sanofi, Servier and Syne Qua Non.

Following the conference, the expanded group has been working through the details of the Validation Hub.

3. Progress

As discussed in a previous AIMS article, the R Foundation released an update to their Guidance Document for the Use of R in Regulated Clinical Trial Environments in March 2018. The document, addresses the validation question in relation to the base and recommended R packages (those installed with R) and the general consensus within our group is that Base R and the "Recommended" packages (those packages that are installed with R as a single bundle) is fit for use in a regulatory context, although individual companies would still need to assess the distribution against their internal quality criteria.

Discussions have therefore centred around the many open source R packages. At the time of writing there are nearly 14,000 of these available via the Central R Archive Network (CRAN) alone. Although packages must pass a large number of technical checks before they make it onto CRAN, this is not a guarantee of quality. Packages do not have to contain examples or tests, and maintainers may not follow any development best practices. Further, since many are produced by individuals, it is simply not a realistic option for companies to audit the package maintainers for all the packages they wish to use.

4. The R Validation Hub

The group has been focussing on two aspects of validation that will be reflected in the Hub.

- Requirements and tests
- Risk assessment

Formal validation typically involves writing tests. These tests are written to test specific requirements, which are generally not available for R packages. The R Validation Hub will provide a mechanism for contributing both requirements and tests for R packages. By sharing our requirements and any additional tests that we write, everyone benefits. Especially if any of the tests find a bug in the package as this can be fed back to the author.

Another crucial area when dealing with open source tools is risk assessment. If an R package is being actively maintained, has had a wide exposure to the user community and bugs are clearly tracked and fixed, then one might consider it to be following good practice and we may determine it to be a low risk to our organisation. This is important as the amount of testing required will typically be guided by the perceived level of risk. For the Validation Hub, our goal is to determine a set of metrics that companies may use to assess risk. We aim to provide guidance for these metrics so that organisations can judge whether the risks are acceptable or whether additional testing of the packages is required.

5. Next Steps

The primary website, https://www.pharmar.org, went live on GitHub on 12th December 2018. Since then we have been working on content and focussing on the risk assessment criteria. Once that is available we will look to update the community further as we will be looking for people to contribute requirements and tests for packages.

Several members of our group are also involved in the recently approved Transcelerate initiative, Modernize Data Analytics for Clinical Development with R. It will be important to ensure that the two initiatives remain closely aligned to avoid any duplication of effort.

For further information on the project, please contact Andy Nicholls (andy.p.nicholls@gsk.com) or use our official psi.aims.r.validation@gmail.com account.