AIMS SIG Update

The AIMS SIG team are continuing to focus on the use of R in the pharmaceutical industry. Many companies are now working towards infrastructure frameworks, to allow R to be used for non-regulatory and regulatory work. Collaboration with the R Validation Hub (https://www.pharmar.org/) and the R Consortium (https://www.r-consortium.org/) is key to ensuring we utilize company resources effectively, working towards a solution which meets regulatory needs.

To share some examples to date of how R is currently being used in the industry, AIMS will be showcasing case studies, ranging from R being implemented as part of a validated platform, to validation of single R package for single project use. If anyone would like to get more involved please reach out to psi.aims.r.validation@gmail.com.

R implementation of a Platform for Pharmaceutical Industry Use: Case #1

With the development of adaptive designs & complex methodologies, statisticians use R more frequently in pharmaceutical companies. Due to the open-source nature of the R language;

- most new graduates learnt R at school rather than SAS (as R free)
- the R community provides an important, active worldwide support network and so new methodologies are now quickly available & shared through R packages.

This first case study will detail a specific platform implementation of R in a pharmaceutical company.

In this company, at the beginning, users developed scripts with R (and RStudio) installed on their own laptop. It was a good start, however performances were not meeting expectations (limited laptop memory), and scripts/applications were hard to share between users (different R version and/or R packages installed). In addition, the lack of control due to these local installations was also not compatible with usage for regulatory purposes, as users could download any packages they wish.

Based on that, an important project was launched in 2018 to set up a new powerful R platform.

As the needs varied among users, from R “explorers” who want to test new statistical methodologies, to clinical statistical programmers, who needed to produce results in regulatory compliant environment, it was decided to build two separate platforms.

- An exploratory R platform, flexible, allowing users to have an up-to-date R version and R packages, used only for internal exploratory analyses
- A regulatory R platform, fully validated (platform, tools & packages) to support production of results for regulatory purposes.

Overview of both platforms

The main characteristics of the Exploratory R platform are:

- users could choose the R version they want (from 3.2 to 4.0),
- users can install the package(s) they needed directly from the source repository like CRAN, Bioconductor… on their R-Home repository,
- For R version 3.4 and later, more than 400 commonly used packages were installed on a centralized repository, avoiding local installation by each user,
- an RStudio interface (pro version allows multiple R versions to be run within the environment) for programming and RShiny Proxy for applications to be deployed for users,
- a High-Performance Cluster was connected to support greedy scripts (to have quicker results and to avoid overloading the platform),
- a new R version is deployed around twice a year (new major release) with the update of the list of packages centrally installed.

The main characteristics of the Regulatory R platform are:

- only one version available (3.5.2),
- around 60-70 R packages, externally validated, are installed in a central repository and available to all users,
- users do not have the availability to install their own packages in order to keep the platform fully controlled,
- an RStudio interface for programming and RShiny Proxy (to be able to grant access only to qualified users) for applications to be deployed,
- a High-Performance Cluster was connected to support greedy scripts (to have quicker results and to avoid overloading the platform),
- Shiny Applications developed internally or externally are deployed through containers (including all needed R packages), to avoid installation on the platform of R packages not useful in programing part (through RStudio),
- a training process has been set-up with several modules (overview, RStudio training, good practices, specific Team's process, …).

To ensure the validated status of the platform, several environments have been set up (test, validation & production) and complete documentation has been produced including:

- a System Validation Plan to detail the validation process & documents,
- User requirement, functional & technical specifications to detail user needs on both platforms
- Installation Qualification to detail technical steps of the installation process,
- Operation qualification to complete package qualification by user tests on RStudio (connection, folders access, multiple sessions, R coding, performances, sharing of R scripts …) and on Shiny (deployment of application, restricted access and performances),
- Performance qualification & validation reports.

In light of the time needed for the installation and all the validation documentation, it is planned to update the R version no more than every 2 years, but to have the flexibility to add, if needed, new packages, after external validation, on the current R platform.

The journey is not over yet: evolutions are still under evaluation (to accommodate new R and RStudio versions, and more autonomy on the deployment of Shiny Applications). However, both platforms are now frequently used with around 150 users on the Exploratory R & 30 for the Regulatory platform. This has increased the capabilities of the statisticians to use R to support both exploratory and regulatory work within the company.

We hope this case study and the future examples AIMS provides, will help to galvanize the adoption of R across the pharmaceutical industry!

Lyn TAYLOR & Yann ROBERT,
On behalf of PSI AIMS SIG
Declaimer: This case study was put together by AIMS SIG members, and by no means represents the views or opinions of the author's employers.