AIMS Update

The AIMS SIG continue 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 and the R Consortium is key to ensuring we utilize company resources effectively, working towards a solution which meets regulatory needs.
In this edition of SPIN, we share another example of how R is being used in the industry. For those who missed Case study #1, please feel free to look back at the Spring 2021 SPIN edition or you will also find the article on the AIMS SIG website under "Useful References" https://www.psiweb.org/sigs-special-interest-groups/aims. If anyone would like to get more involved with AIMS, please reach out to psi.aims.r.validation@gmail.com.
During Case study #2, for the purposes of anonymization, the 'Company' will be the ones that carried out any analysis on behalf of the 'Client'. The use of 'platform' is used to encompass the R installation and of RStudio.

Analysis requested by client which can only be carried out in R Case #2

Background

A Client contracted our Company for one of their products to develop a statistical analysis plan, TFL shells and to provide programming for reporting. The Client-side statistician had provided the statistical input into the protocol (the Company had no direct involvement in developing this document). In the statistical analysis section of the protocol, it stated that the analysis would use a methodology referenced in a specific paper[1] and that the details would be documented in the analysis plan.
The paper in question referred to a Bayesian analysis which used a novel approach for exchangeability across groups, treatments, and periods. The paper explicitly referenced R (and provided some code) to reproduce the approach. Unfortunately, there was no 'proc' in SAS that could easily replicate the analysis, at least, not without having to reverse engineer the approach in R to replicate a function/macro in SAS.
The Company had no official R infrastructure in place to carry out any sort of analysis using R.

Qualification

The Company felt they had two choices – either provide a fully validated platform of R to be used across the entire company or do a 'thin' validation for the purposes of this project. Due to time and resourcing limitations, they went for the latter.
Following discussions with the Head of Statistics, and the Head of Quality Assurance (QA), the following approach was decided upon:

1. 'Qualify' R to be used solely for this project but also only for this solitary analysis; the rest of outputs for the project would be produced using SAS, following the Company's standard operating procedures.
2. Identify and document which packages would be used for analysis and hence which would need the following QA documentation:
    - Qualification plan

        - What would be needed, why it was needed and what would be the intended use of the 'platform'.
    - Installation log
        - A detailed description of the installation process of the platform (to cover the two entities of R and RStudio) using screenshots to demonstrate the installations were successful.
    - User requirement specification documentation
        - A list of which functions (and packages) are needed to perform the analysis.
    - Qualification testing
        - A set of tests to demonstrate that R, RStudio and any additional packages to support the analysis, perform as to be expected.
    - Qualification report
        - A document to summarize the above that everything was 'successful' in terms of installation and performance (as intended). This document also highlighted if there were any differences in outcome that would be acceptable in context. As the analysis performed using Bayes follows a non-deterministic algorithm, it was decided to use a fixed 'seed' for testing purposes. Similarly, for performing the analysis, the seed was shared with the Client to allow them to independently check the output we generated (code was shared along with source ADaM)

All the above would need to be completed before using the code for the analysis. The Client-side statistician had written R code to perform the analysis. The code would need to be adapted to read in the analysis dataset for reporting purposes. The paper focused on binary endpoints; analysis required by the Client used continuous endpoints – an additional layer of complexity for conducting the analysis.

Challenges

Initial intention was to make use of the R *haven* package. This is a package that supports using SAS7BDAT files. However, at the time of analysis, whilst I was able to read in SAS datasets (the source ADaM for the analysis), post-analysis, I was not able to write out to SAS7BDAT (to generate the table) - the plan was to read the required derived dataset, conduct the analysis in R and then write out the results to a SAS dataset which would then be read into SAS to run some basic level code to produce the output. Instead, an in-built R function was used to write to CSV file instead and this read into SAS. Not perfect, but for practical purposes this was sufficient. There are alternatives to *haven* depending on user requirements, such as packages supporting the more 'open' XPT file format; SAS7BDAT files are more proprietary now.

The Company's QC process involved double programming table outputs (especially stats analysis outputs). As part of QC process for this output. It would not be feasible to ask a colleague to double program using SAS, equally, it would be a challenge to independently program this analysis. The steps taken instead were 3-fold:

1. Code on the main side using R and output the results to a CSV file. This CSV file would then be read into a 'stub' SAS program which would simply output the dataset into the desired layout, per shell.
2. An experienced colleague (knowledgeable in Bayes and R) would then perform a code review of the R code.

3. Lastly, they would run the R code independently to check the R console to see if there were any warnings/errors generated. This last step was necessary as SAS is fantastic at formally producing an output log to demonstrate that it has performed the steps that were programmed.

This was then passed over to the Client for them to review and confirm that they were happy with the analysis performed (as coded) and that the results made sense.

Initially, this analysis was going to be used to decide whether the study would 'select' one version of the treatment over another. However, due to the pandemic the analysis' importance in decision making was downgraded and used to merely reinforce the next stage of the study.

Outcome

The output was delivered during several runs and the results were accepted without comment. This case study demonstrates that it is possible to use R for project work, not just as an exploratory tool, but in a formal way to help support drug development. Many of the AIMS members foresee a future where R will be used side-by-side with SAS and other future technologies (perhaps Python) in a regulatory environment with a robust QA-process in the background, providing confidence in the platform, and therefore, the analysis. This experience will hopefully foster more confidence that R can be used (without jumping through too many hoops).

Christina FILLMORE, Lyn TAYLOR & Chris TOFFIS,
On behalf of PSI AIMS SIG

[1] Neuenschwander B, Wandel S, Roychoudhury S, Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. Pharmaceut. Statist. 2015, 15 123–134