AIMS at the PSI 2018 Conference and Our Plans for R Validation Documentation Storage

Authors: Jules Hernandez-Sanchez (Roche), Lyn Taylor (PRAHS), Craig McIlloney (PPD), Yann Robert (Servier)

The Special Interest Group (SIG) called Applications and Implementation of Methodologies in Statistics (AIMS) presented at the 2018 PSI Conference ([www.psiweb.org/psi-2018/psi-conference-2018](http://www.psiweb.org/psi-2018/psi-conference-2018)) in Amsterdam. The title was “The Future is heRe”. Craig introduced AIMS, Jules summarized some of the articles that AIMS published in SPIN and Lyn showed several examples of R Shiny applications including an adverse event tabulation app and laboratory data graphics app, both created by another AIMS member Chris Toffis. All AIMS Articles can be found on <https://www.psiweb.org/sigs-special-interest-groups/aims>. The three AIMS articles discussed in detail during the conference were: 1) Want to use R? Here are some tips before you start (Lyn Taylor), 2) Introduction to IDEs and RStudio (Chris Toffis), and 3) Introduction to tidyverse (Andy Nicholls). AIMS were also pleased to announce at the conference, their future plans for a web based portal, to store metadata/tests/ documentation relating to R validation. The portal will be freely accessible to all, allowing individuals to load up and share information, which will support the use R for late phase studies and submission filing.

AIMS is currently focused on R and specifically R validation: the main hurdle for widespread use of R in late phase trials. AIMS was created from the ashes of a previous SIG called the Statistical Computing Committee. The charter for AIMS was drawn in 2015 which was approved by both the board of directors at PSI and EFSPI ([www.efspi.org](http://www.efspi.org)). In addition to the authors of this article, the other members of AIMS are: Chris Toffis (Syne Qua Non), Helen Savel (Bordeaux University Hospital), Sophie Canete (Bordeaux University Hospital), Andy Nicholls (GSK) and Markus Elze (Roche). Additional members will be very welcome, especially those with an interest in R validation in our industry. The AIMS vision is encapsulated in the following paragraph from the charter: “*To support PSI Committees and PSI/EFSPI Special Interest Groups (SIGs) with the technological application and implementation of statistics. To develop understanding of new analytical tools and approaches to share with PSI & EFSPI members via appropriate forums. To ensure PSI & EFSPI members are supported with understanding the requirements for the implementation of industry data standards.*” The Top priorities identified for AIMS were: 1) Estimands, 2) Simulation and modelling, 3) Computationally Intensive Methods, and 4) R validation. In a meeting in June 2016, statistical leaders recommended the latter point be the highest priority of AIMS. In order to help with R validation initiatives, AIMS contacted the R Consortium who granted $4000 to develop a webpage to be the single portal for the storage of metadata, tests and validation type documentation. AIMS is now actively seeking to engage key stakeholders and encourage support and contributions to the storage of this information. To help to make people aware of the project, some members of AIMS will attend and present at the R in Pharma conference between 15 and 16 of August 2018 at the University of Harvard (USA). At the conference, all matters related to the use of R in the pharmaceutical industry will be discussed, including R validation issues.

There are many positive things about R, the sheer number of statistical packages developed by a global scientific community (both from industry and public sectors), more than 14,000 on 1st June 2018, makes it possible to tackle almost any analysis within R. R is free. It is maintained by a group of volunteers, the R Foundation, that includes some of the best statisticians in the world. The R Foundation issued a new version (on the 25th march 2018) of an article in titled: Regulatory compliance and validation issues -A guidance document for the use of R in regulated clinical trial environments. That guidance assures us that the base installation of R (base and recommended packages), with which the most common statistical needs are fulfilled, are kept in as high standards as is possible to keep any statistical software, including SAS. The Food and Drugs Administration (FDA) issued a document in 2002 titled: General principles of software validation; Final guidance for industry and FDA staff. This document is quite general and most of it dedicated to software used in medical devices, like CT scans. And although some of the principles and recommendations can be applied to any software, it lacks details of how to validate statistical code. AIMS defends the idea that no software can be purchased pre-validated in a regulatory context. In most cases, validation has happened through a historical process of use and debugging of software. This is certainly the case of SAS and, to a lesser extent, the case of R (R was created in 1996, SAS is older and with widespread use in pharmaceutical industry). Most importantly, FDA does not recommend the use of any statistical software, it just demands that the user should be capable of demonstrating that the software has been validated, i.e. it does what it says in the tin. A few tips were put forward by AIMS in order to help deciding whether some R packages could be safely used in a regulatory context, i.e. filing. Those tips include researching the author (is the author well known and reputed?), or the extent of software use (has the package reached a stable version after several improving iterations?). Nevertheless, more specific guidelines that could be easily applied to all packages are necessary.

**The next steps:** TheAIMS SIG is now working on designing a framework which will specify a set of requirements, including metadata and examples of tests, which together would form evidence of the quality of an R package.   Initially we will use dplyr as an example, and will make this “evidence of validation” available to the wider community.   Whether the evidence provided is sufficient will be the decision of the end user, but it can be a starting point for further testing or may be sufficient in itself depending on the user’s attitude to risk.  After review by our peers and with agreement that the framework is sufficient for use in the regulatory environment, we will be calling on all R users to submit similar evidence of validation for other packages.  By sharing this evidence, we hope to reduce the amount of re-work being done by multiple companies eager to use R, but fearful of doing so in a regulatory environment without documentation of validation.