Application and Implementation of Methodologies in Statistics (AIMS) Special Interest Group Meeting

03 April 2018: 09:00 (BST)

Attendees:

Team Member	Present at meeting
Craig Mcilloney (PPD)	X
Lyn Taylor (PRA)	√
Chris Toffis (Syne Qua Non)	√
Andy Nicholls (GSK)	√
Yann Robert (Servier)	√
Helene Savel (Bordeaux University Hospital)	X
Sophie Canete (Bordeaux University Hospital)	X
Jules Hernandez-Sanchez (Roche)	X

Yearly Objects 2018

- Conduct a parallel session at conference
- Establish relationships with other companies who use the tools we are talking about
- Expand membership of the group (1 or 2 more)
- 1 article in each SPIN newsletter

Previous Action Items

Action Item	Assigned team member(s)	Deadline	Status
Mike Smith to send list of 50 top packages & assign volunteer to join group	Andy	Ongoing	Ongoing still want a Pfizer rep and can get input into packages at later date.
Distribute list of R packages used for standard statistical analyses.	Helene	Feb 6th	Closed – uses lm, glm and confint (functions rather than packages)
Continue work on PSI presentation Lyn to work with Chris on potential for Lab graphics demo (and if he can attend)	Craig/ Lyn/ Jules /Chris	June 4th	Ongoing – split into 2 items
Review Andy's Tidyverse article and Idea of a validation plan (above)	All	23rd March	Closed – Andy to send to Matt by 20th April. Lyn to provide link to website where data will be loaded
All to think about plans / agenda for a Brainstorming day with heads of Stats/prog departments	All	1st May	Open
Lyn to write a proposal for r-Consortium funding (due to submit by 1st April)	Lyn	16th March	Closed – Submitted 27th March decision by 1st May.

Application and Implementation of Methodologies in Statistics (AIMS) Special Interest Group Meeting

03 April 2018: 09:00 (BST)

Agenda/Discussion

Topic/Lead	Discussion/Decisions
Conference preparation for a parallel session	Conference proposal (Session is Tuesday 5th, 4-5:15): • 20 min Introduction to AIMS – Craig • 20 mins summary of articles to date (IDEs & Rstudio, R validation)– Jules • 25 min demonstration possibly including the Introduction to Rshiny article and show how RShiny can be used to produce safety (AE and Lab) reporting • 10 questions/conclusions Chris provided a demonstration of the R shiny Lab graphics App. Yann & Lyn provided feedback. (Update from Months to be Visit number, Add n to summary stats, update baseline for ALP to look same for both treatment groups so a narrative of how you may explore data can be given). Chris & Lyn to look into doing a video demonstration which will sit inside the powerpoint. Note: The conference will be held at the Beurs Van Berlage, Amsterdam, from 3rd to 6th June 2018. Lyn, Craig, Jules & Yann will attend. Chris may also be there (TBC). Continue work on presentations to be ready 1-2 weeks before.
Article writing progress/ All	Deadlines are: approx 20th of the month in January, April, July and October Current suggestions for future articles. 1) **Tidyverse – Andy [To send to matt by 20th April]** 2) Summary of PSI conference session [July] 3) R Markdown- part of Rstudio (mention that you get R Notebook within this) to write and publish HTML, PDF, DOC – Jules [October?] 4) R Shiny. –multiple articles to follow – Chris - 2nd installment on hold due to Chris leaving the group for a while 5) literate programming and Reproducible research Cloud computing – Wilmar is moving from AZ to contractor position and so will no longer have time to write articles - New volunteer needed if we want articles in these topics 6) Linking R with other software (ie. BUGS / SAS) MCMC – On hold since Dave left the group. 7) High performance in R vs SAS – Yann 8) How to submit a submission package to FDA including R – Any volunteers 9) Reporters – Andy 10) Sweave – Jules

Application and Implementation of Methodologies in Statistics (AIMS) Special Interest Group Meeting

03 April 2018: 09:00 (BST)

Topic/Lead	Discussion/Decisions
Collaboration for validation concept/ All	Not discussed today but Left on Agenda for future meetings: Brainstorming day (or 1/2 day) with heads of Stats/Prog to ask what they want and agree a principal of validation for open source code. We could announce this day at the conference to get buy in. Perhaps a format for the day would be: • Background on FDA guidance • Current work by R- consortium (feedback from R-Pharma conference (Rheinhart & Min?) • PSI AIMs thoughts and what we've done to date • Demonstration of the dplyr validation (see below) • Presentation by validation vendors (such as Mango) – (limit their attendance to this section only)? • Discussion – to include what companies want and how to handle version upgrades. Await for decision from r-consortium on collaboration for a web portal to store R validation, if approve our group would work on validation for dplyr as an example. Any would work on specification for this once the validation plan template is confirmed. 1. **Determine requirements/specifications** for the package - you can't validate something without documenting what it is that you're trying to use it for 2. **Assess the risks** – Mixture of quantifiable things (package lifetime, number of downloads, number of tests within package, does it have a news feed?) and non-quantifiable (reputation of author, quality of news feed) 3. **Test requirements** (i.e. produce evidence of mitigating risks) Note: For Statistical packages – all need to test methods Next steps • Continue to get list of key packages ? • Include stats packages? • Ideas for branstorming day?
Long standing items we might come back to/ All	• R in Pharma conference: http://rinpharma.com/ The first annual R/Pharma conference will be held Tuesday, August 7th and Wednesday, August 8th at Harvard University, Cambridge, Massachusetts, USA. Suggest we get feedback from Min Lee & Reinhold Koch after the conference. • r-Consortium funding to bring someone in to provide a website which contains pharma validation, https://www.r-consortium.org/announcement/2018/01/31/r-consortium-call-proposals-february-2018 Application was submitted 27th March 2018. Decision by 1st May. Application is attached at end of these minutes. • SPIN Competition? – Ask people to write How and Why they use R in the

Topic/Lead	Discussion/Decisions
	pharmaceutical industry: Find a company to sponsor? • R-foundation interaction • Roche's interest to use R for adaptive designs replacing Adplan/East etc.

Action Items

Action Item	Assigned team member(s)	Deadline	Status
Ongoing item: We want a Pfizer representative on the group. Continue to ask Mike Smith.	Andy	Ongoing	
Chris to do a video demo of the Lab App, and send to Lyn. Lyn to practice presenting it.	Chris/ Lyn	Ongoing	
Lyn to check PSI website to see where best to load the Tidyverse data. Update Chris's company	Lyn	9th April	
Continue work on PSI presentation	Craig/ Lyn/ Jules	1st May	
All to think about plans / agenda for a Brainstorming day with heads of Stats/prog departments	All	1st May	Open

Statisticians in the Pharmaceutical Industry (PSI) Application and Implementation of Methodologies in Statistics (AIMS) Special Interest Group (SIG) request for R Consortium collaboration to create an online repository for storage of regulatory standard validation material for R packages

1) The Problem

Pharmaceutical companies have to comply with strict regulations and laws set out by the regulatory authorities such as the Food and Drugs Administration (FDA), European Medicines Agency (EMA) and Pharmaceuticals and Medical Devices Agency (PMDA). Statisticians and data analysts must comply with these regulations for the analysis and reporting of clinical trials, including provision of validation documentation for any software used. The FDA (2014) defines validation as "Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes". The user is therefore required to validate the software installation (to confirm it is performing as the developer intended for given examples) and to validate the data manipulation or analysis performed is accurate and repeatable.

No statistical software can be purchased as "pre-validated" software in the regulatory context. Companies performing regulatory analysis have a responsibility to ensure software is installed correctly and that it behaves as expected. Given the dominance of SAS in the industry, many pharmaceutical companies have substantial evidence collected over a long period of time, detailing the testing & validating of SAS procedures. Therefore, rather than duplicating testing which has already been performed in the past by others, many companies make the decision to use SAS procedures based on their history of extensive previous use.

The desire to use R in the pharmaceutical industry is evident due to it being a very powerful object-orientated language, able to produce analyses and high quality graphics and being open source hence readily enabling bespoke extensions. It can be used for high performance computing and allows easy integration with other software. It can also be used for complex analyses which is especially important when these analyses are not available in SAS.

R is relatively new to the regulatory environment. R Users can easily download and use packages from CRAN which have had varying degrees of testing performed on them. This is problematic as companies either have to restrict the use of R packages, to only include those validated to an appropriate level or allow free use of R, but not for regulatory submission use. Currently each company, is working in isolation producing their own validation of a subset of packages or purchasing a validated R system from a software provider (such as Microsoft or Mango).

The Statisticians in the Pharmaceutical Industry (PSI) organization tasked a special interest group (SIG) made up of PSI members, called the Application and Implementation of Methodologies in Statistics (AIMS), to investigate the use of R software in the industry. The SIG is made up of members from GSK, Roche, Servier, PPD, PRA Health Sciences, Syne Qua Non, and Bordeaux University Hospital. More information about PSI and the work of the AIMS SIG can be found here: https://www.psiweb.org/about-us/sigs-special-interest-groups/aims

Initial discussions between the AIMS SIG members, and further representatives from Amgen and Roche, found a desire to freely share and collaborate on the documentation of R package validation for pharmaceutical industry use. The hope would be to reduce the duplication of effort and enable wider use of R packages for regulatory submission work.

Whilst the problem of testing and validation documentation was highlighted in the pharmaceutical industry, there would be many other R users such as those in Banking, Market Research and Academia who would benefit from our plan. Every user who handles data must have a desire for their work to be reliable, reproducible and accurate. Our plan (as detailed below) could be freely accessed and available to all R users.

2) The Plan

Overview: The documentation available for R packages currently widely varies. The AIMS SIG would like to collaborate with the R-Consortium and representatives from pharmaceutical companies on the setting up of an online repository /web portal, where validation which is of regulatory standard for R packages can be submitted and stored for free use. Companies (or individual R users) would still be liable to make their own assessment on whether the validation is suitable for their own use, however the online repository would serve as a portal for sharing existing regulatory standard validation documentation.

Level of Validation Required: FDA considers software validation to be "confirmation by examination and provision of objective evidence that software specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled." "Software verification and validation are difficult because a developer cannot test forever, and it is hard to know how much evidence is enough. In large measure, software validation is a matter of developing a "level of confidence" that the device meets all requirements and user expectations for the software automated functions and features of the device."

Initial Steps for 2018:

- The PSI AIMS SIG and the R-consortium work together on the creation of an online repository for R package validation in accordance with regulatory standards
- The PSI AIMS SIG and the R-consortium work together to define a set of "Validation Criteria" which would be used to inform contributors and R package authors what criteria we are using to assess their package, so they know what they have to do to write a good package and supply testing documentation. This will include using metrics from "R CMD check" and "covr" in combination with other meta-information and testing examples
- The PSI AIMS SIG, would validate a single R package (dplyr) following the methods described. This would then be used as a template (example) of the validation level expected. This validation is expected to be complete by 1st January 2019.

For each R package, the validator (person submitting to the online repository) should:

1. Determine requirements/specifications for the package – Package details including version and what it is that you're trying to use it for
2. Assess the risks – Mixture of quantifiable metadata (package lifetime, number of downloads, number of tests within package, does it have a news feed) and non-quantifiable (reputation of author, quality of news feed)
3. Test requirements - This section should include the tests performed as evidence of mitigating the risks. Either testing documentation itself or links to other validation evidence

The PSI AIMS SIG to put out a call to all pharmaceutical companies, advertising the online repository and requesting them in the future to contribute with their own validation documentation. PSI AIMS SIG to announce the plan for an online repository at the PSI Conference in Amsterdam (4th – 6th June). In addition, key opinion leaders / heads of pharmaceutical company statistics and programming departments will be invited to a "Validation of open source code" brainstorming day planned for late 2018, where the work of the PSI AIMS SIG to date will be presented, along with presentations on validation and discussions on validation of open source code in the pharmaceutical industry.

3) The Team

The PSI AIMS SIG consists of:
Craig Mcilloney (PPD)
Lyn Taylor (PRA Health Sciences)
Jules Hernández-Sánchez (Roche)
Andy Nicholls (GSK)
Chris Toffis (Syne Qua Non)
Yann Robert (Servier)
Helen Savel (Bordeaux University Hospital, Clinical Epidemiology Unit)
Sophie Canete (Bordeaux University Hospital, Clinical Epidemiology Unit)

Preliminary discussions regarding a collaboration for validation documentation have also taken place with:
Reinhart Koch (Roche)
Markus Elze (Roche)
Min Lee (Amgen)
Mark Stetz (Amgen)

4) Project Milestones

a) Design Phase
   i) Agreement of high level process
   ii) Technical solution design
b) Publication of "Validation Criteria"
c) Implementation of database
d) Finalization of front end, i.e. website
e) Proof of Concept – The loading of the first package, dplyr
f) System Go-Live

5) How Can The ISC Help

Provide the infrastructure for a knowledge-sharing platform of validation information. Contribute to the set up of the validation criteria that authors / testers would have to meet in order to load a package to the website.

6) Dissemination

- Details of how the validation code will be stored is one of the design deliverables of the project, which we would like to discuss with the R-Consortium.

- Regular updates would be made to the R Consortium Blog advertising the new validation web portal and encouraging it's use

- Regular updates on progress would be submitted to the quarterly PSI SPIN Newsletter and available on the following website: https://www.psiweb.org/about-us/sigs-special-interest-groups/aims

- PSI AIMS SIG to announce the plan for an online repository at the PSI Conference in Amsterdam (4th – 6th June).

- Key opinion leaders / heads of pharmaceutical company statistics and programming departments, will be invited to a "Validation of open source code" brainstorming day planned for late 2018, where the work of the PSI AIMS SIG to date will be presented, along with presentations on validation and discussions on validation of open source code in the pharmaceutical industry.