R "Validation" Hub Meeting

21 May 2019

Attendees:

Andy Nicholls (GSK) – Meeting Chair
Lyn Taylor (Phastar) – Meeting Secretary

Alexander Lock-Achilleos (GSK)
Alun Bedding (Roche)
Anthony Williams (Fred Hutchinson CRC)
Chris Toffis (Syne qua non) – PSI AIMS SIG
Claus Dethlefsen (Novonordisk)
Doug Kelkhoff (Roche)
John Sims (Pfizer)
Joseph Rickert (R-Studio)
Juliane Manitz (Merck Serono)
Kieran Martin (Roche)
Matthias Trampisch (Boehringer-Ingelheim)
Min Lee (Amgen) - TransCelerate
Nate Mockler (Biogen)
Nicollo Bassani (Quanticate)
Parker Simms (No affiliation) - Website design
Raj Malathker (J&J)
Rebecca Krouse (RHO Inc.)
Reinhold Koch (Roche)
Rinki Jajoo (Merck)
Satish Murthy (J&J)
Steve Noga (RHO Inc.)
Tilo Blenk (GSK)
Tomas Drgon (FDA)
Xiao Ni (Novartis)
Yilong Zhang (Merck)

Previous Action Items

Action Item	Assigned team member(s)	Deadline	Status
Communication, dissemination of information about the project to the wider group Possible Funding discussion if we need to write a package to pull metrics Will re-assess for the next Autumn application	Andy Nicholls/ Lyn Taylor/ Joe Rickert	Sept 2019	Ongoing
Review of the website content: Direct review back to Andy, via GitHub or message via slack as you prefer. - Andy confirmed that the website is Live with all latest updates on the live side. Please can the team review and feedback comments.	Kieran Martin, Min Lee, Juliane Manitz, John Simms, Keaven Andersen, Alex lock-Achileos	NA	Ongoing

- The team welcomed Parker who will help on website. Rinki Jajoo asked if parker will act as the gatekeeper and it was agreed this would work well for her to review before updates are made live. Juliane Manitz suggested that Parker bring things to the rest of the stream to review and approve any technical content. - Parker asked if there is any website issue tracking being used currently and Andy confirmed that there was none to date, but that it would be good for her to set something up. - Joe Rickert commented that we should use the R consortium website for pointing to the minutes on our website. Parker & Joe to work together on this.			
Risk assessment workstream – Continue work on objectives - Still discussion on how the package structure would look. Once this is finalized then they can implement the package development part. - Andy asked if there is a date for the package structure to be finalized. However there was no date set yet as team have been very busy. Andy requested that we set a date to ensure continual progress and align timelines with the TransCelerate & Rpharma conference work.	**Yilong Zhang,** Rebecca Krouse, Doug Kelkhoff, Matthias Trampisch, Eric Nantz	May 2019	Ongoing
Requirements/tests workstream set objectives + milestones - Keaven not on the call today, but he sent notes on where the stream are prior to the meeting and these are included below	**Keaven Anderson**, Nate Mockler, Tilo Blenk.	May 2019	Ongoing
Validation white paper workstream – appoint spokesperson and set objectives + milestones - Target to produce white paper by R in pharma conference.	Andy Nicholls, Paulo Bargo	May 2019	Ongoing
Write text for the overview website page and request review by website review team	**Parker Sims,** Min Lee, Kieran Martin, Juliane Manitz, John Simms, Kieven Andersen, Alex lock-Achileos to review	May 2019	Ongoing
Write blog post on gathering metrics and using these in a risk assessment	Doug Kelkhoff	TBC	Open
Investigate how to educate people in the structure of R packages. Yilong suggested including on the website an overview of the structure – so if someone	TBC	May 2019	Ongoing

has high level questions we can direct them to the information. Andy added this could be in the form of a FAQ page.			

Agenda

1. Review of actions / updates– Andy (40 mins)
    a. Update on white paper – Andy
    b. Update on metrics stream – Min
    c. Update on website – Andy
    d. Update on testing – Keaven
2. Executive committee – Andy – (10 mins)
3. AOB (5 mins)

Discussion

Updates on the streams progress was included under the actions above with the exception of the Requirements/tests workstream which was discussed in detail below.

Requirements/tests workstream

Keaven Anderson (Merck), Tilo Blenk (GSK), Nate Mockler (Biogen) and Eric Milliman (Biogen) had a 1-hour discussion on March 20 concerning testing as part of the R Validation Hub initiative.

Tilo discussed his work on validating an R installation. He wrote an extensive set of unit tests to be run at GSK at the time of an R installation. While this could not be considered an exhaustive validation, it was extensive enough to provide confidence that the installation is running correctly. Tests included comparison with published results (e.g., Stata documentation). It was also mentioned that database interfaces at a given company also need to be tested. Keaven had a note that the group agreed that Base R (including stat) and some RStudio software would be things we would want to consider OK without extensive testing. There is a need to work with the metrics group on a combination of unit testing and package maturity that might be required to include things from RStudio.

Also requiring consideration is production of a document that the FDA (or other regulatory body) could request that approves an updated installation. There is also a question of whether we could develop some test suites but concern regarding the size of effort needed.

Keaven also provided the below for consideration:

- I believe there are a couple of companies that will occasionally outsource validation of a package. It would certainly be of some interest if there were a funded project from multiple companies to form test suites for desired packages.
- I have been working (with considerable help from Metrum) on updating the gsDesign test suite from runit to testthat. However, the package still only has about 20% coverage and I hope to

improve this over the course of the year. This does take a dedicated effort and I wonder if Merck and other companies will have the ability to support a lot of this.

- I coordinated the initial meeting on testing, but have not scheduled another. The coordinating committee may want to discuss their vision of directions we might wish to pursue for testing and seek a strong leader and participation if we expect to get substantial results or even a good set of test cases.

Kieren Martin asked what's the make up of the R Validation Hub group and can anyone speak regarding qualification of software. Min confirmed experience in this IT area, however, re: qualification tests is it similar to package testing? Andy, suggested it would be 2 stages 1) Confirming the package does what it says it does – and more tests may need to be written here based on risk assessment and 2) have I installed it correctly? The requirements/tests workstream could go in the direction of what needs to be done once you qualify a package for risk. If comes out at high risk, do you 1) not use it or 2) write more tests for it?

Min asked what do you do when the assessment of risk is different in different companies – we cannot provide any suggestion but guidance only and each company has to make their own decision. Andy suggests that for qualification, there would be a set a set of base tests, that they have to do in addition to qualifying the risk. This is the sort of question that we would like to answer.. what additional tests are needed. Min is happy to provide input on this, and can work with Keaven & Tilo.

Tilo asked what is the best thing to start on, and Min suggested doing what's most tangible. Perhaps how would you qualify a base+recommended + tidyverse environment. Tilo confirmed that is what he's done at GSK and includes approximately 500 tests just to qualify this. Andy suggested that Tilo present at the next meeting to present the sort of qualification that could be done to qualify the environment. **ACTION:** Tilo to Plan to present on package qualification at the next session.

Rheinhold suggested base R test should be used as a base to qualify R when new R versions are installed. Even if there are tiny differences in the updated numbers, this doesn't mean you have to fail a test, instead perhaps just document the change in the results.

Andy suggests that qualification would be based on a subset of tests that we've specifically chosen and some detail about what would class as the fail of a test, minor change in numerical precision may be acceptable to continue. We should document how to qualify and provide a starting point of the tests, rather than everyone having to mine through the packages and pull out the tests. Reinhold agreed and added it should be user friendly & easy to implement.

Tomas Drgon (FDA) asked if there is anyone on the group that has a relationship with the R Foundation, and is there value in having someone join us. Joe Rickert confirmed that the R consortium has a relationship with the R foundation and as this project is funded by the ISC, then the more we consort with the R consortium then we can get the attention of the R foundation. Maybe we are a big enough stakeholder to see if the R foundation can help. Joe said that any changes to the core itself is a big job and so going through the consortium is the best way. One idea is that if we do intend to validate the CORE R, then we could ask for funding from the framework.

However, Andy suggested that we don't want to spend a lot of time to write tests for base R, since it is already validated internally by the R foundation and hence in terms of using it, we should have documented straight forward tests to ensure it's installed OK. Thomas Drgon had concerns about this

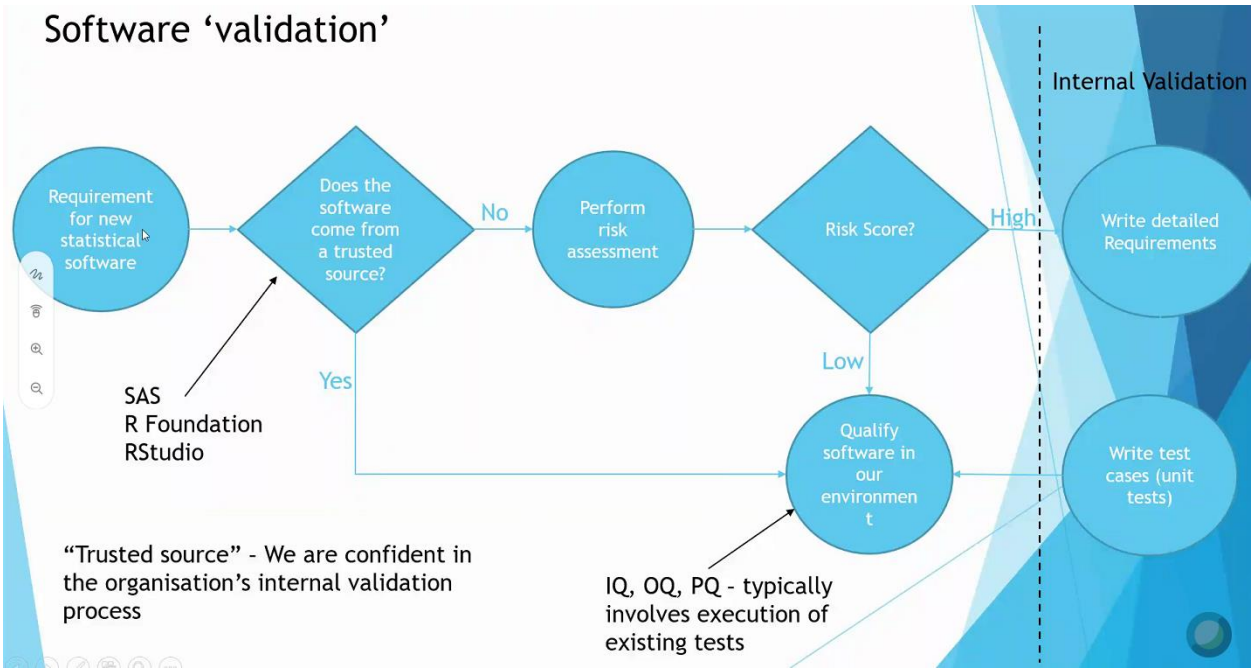
though if you are using it for regulatory submission then the validation issue becomes a legal issue. Andy questioned how is SAS qualified for regulatory work and suggested that if SAS is accepted (without the need for companies to write a large number of additional tests) because so many people use it, does the same apply to base R, which means that only basic qualification is required?

Joe Rickert suggested that this is a way we can levy the R consortium . The R pharma group has people representing the FDA, so we could utilize the co-operation between the groups. Reinhold informed that group that at Roche SAS installation is checked, and it's ensured that all tests pass in order to approve the installation. Roche try to do something similar to this with base R. This is the minimum that you should do.

As a follow up: Min (if possible), Reinhold, & Andy to follow up with Kevin/ Tilo workstream team to assess what we need moving forward.

Validation workflow

Andy presented the Idea of a validation workflow. The concept is that if you have a trusted resource (i.e. SAS, RStudio) then you only need to qualify it in our environment. If it's not trusted, then do a risk assessment. Using the risk score, you can determine if it passes validation or not. A low risk may require qualification only, High risk then you may decide to write detailed requirements for tests or may decide not to use the package.



John Sims responded that Pfizer have a lot of validated SAS systems, their SAS environment is on a high performance system and they are just using SAS base etc not the vertical products. They have a set of qualification documents, and follow a testing protocol. They leverage BT infrastructure for SAS. They also have a R environment. He hopes that this group can come up to writing test framework which would ideally generate the same results in SAS vs R to qualify the 2 systems.

Chris Toffis asked how does the High/Low rating get defined as it’s likely to be a company based setting. It was agreed that we could come up with numerical score, which Yilong’s team have discussed, however it’s likely to need Low, vs medium/high. Low may be OK to just do qualification, however both medium and high would need some tests but the level of testing may depend on the medium/ high rating.

We need to be careful in the wording of this on the website as we can only make recommendations but each company still has to make their own decision of what they need to do to be confident of their work.

John Sims asked how Andy was defining a “Trusted source”. Andy referred to the R foundations documentation regarding building R to best practices, as they have internal practices to validate the software themselves similar to the process SAS have for their software. Based on this documentation, we only need to qualify the software not do our own tests.

Doug Kelkoff asked why is the branch for the trusted sources needed. He thought we were trying to get a list of tests that we have to run for the package and would not just assume it works because it has low risk? He suggests we test it anyway and document that. However, Andy’s preference was not to implement an approach for R, that differs from what we do for SAS. For example, if SAS adds a new procedure / package / option in a procedure, then we don’t test that separately, but for R if someone adds more functions to the tidyverse, should we test or should we believe it’s built to tidyverse standards so we don’t have to individually test it. Doug responded that if we value the company then weight that in the risk such that a new package from a valued writer/author is less risk than a package from a lesser known author.

Claus agreed and informed the group that their attempts to reproduce work between two version of SAS has often been difficult and that if you can’t reproduce it then you have to keep both versions of SAS and document the differences.

It was then suggested to only update the R packages when you update to new R version, this could be implemented with a set of tests that you run to check everything is still OK. However, then we need to have a good surrogate and copy of the test to run. Perhaps this is something we can talk to Gabor Csardi about.

R "Validation" Hub Meeting

21 May 2019

Executive committee

Andy proposed that in order to have more structure to our work, we think that an executive committee would be useful to be assigned in order to steer the project. Members of this committee are proposed to be Andy (Chair), Lyn (For minutes) and also the leaders of each group shown below. All members were asked if they felt this was detailed enough or if anyone felt key to join.

The idea would be for the executive committee to meet more regularly to determine the direction of the project but that the regular meetings would continue with the wider group who would provide feedback from their progress and streams of work.

Joe asked if anyone was presenting on this project at USE-R. Neither Andy or Lyn are, but perhaps the exec committee can address that and see best direction to go forward. Andy will be at Pharma R and Lyn, Andy and Marcus Elze will be at PSI.

Actions

Action Item	Assigned team member(s)	Deadline	Status
Communication, dissemination of information about the project to the wider group Possible Funding discussion if we need to write a package to pull metrics Will re-assess for the next Autumn application	Andy Nicholls/ Lyn Taylor/ Joe Rickert	Sept 2019	Ongoing
Review of the website content: Direct review back to Andy, via GitHub or message via slack as you prefer. Any updates from Parker on the website and getting R consortium website to point to our minutes on our website.	Kieran Martin, Min Lee, Juliane Manitz, John Simms, Keaven Andersen, Alex lock-Achileos	NA	Ongoing
Risk assessment workstream – Continue work on objectives – set date for package structure to be agreed.	**Yilong Zhang,** Rebecca Krouse, Doug Kelkhoff, Matthias Trampisch, Eric Nantz	May 2019	Ongoing

Requirements/tests workstream set objectives + milestones - **ACTION:** TIlo to Plan to present on package qualification at the next session.	**Keaven Anderson**, Nate Mockler, **Tilo Blenk.**	May 2019	Ongoing
Validation white paper workstream – appoint spokesperson and set objectives + milestones Target to produce white paper by R in pharma conference.	Andy Nicholls, Paulo Bargo	May 2019	Ongoing
Write text for the overview website page and request review by website review team	Min Lee (and then Kieran Martin, Juliane Manitz, John Simms, Kieven Andersen, Alex lock-Achileos to review)	May 2019	Ongoing
Write blog post on gathering metrics and using these in a risk assessment	Doug Kelkhoff	TBC	Open
Investigate how to educate people in the structure of R packages. Yilong suggested including on the website an overview of the structure – so if someone has high level questions we can direct them to the information. Andy added this could be in the form of a FAQ page.	TBC	May 2019	Ongoing