A Risk-based approach for assessing R package accuracy within a validated infrastructure

Juliane Manitz, EMD Serono; Andy Nicholls, GSK; Paulo Bargo, Janssen R&D; Doug Kelkhoff, Roche; Yilong Zhang, Merck & Co., Inc.; Lyn Taylor , Phastar; Joe Rickert, R Consortium; Marly Gotti, Biogen; Keaven M Andersen, Merck & Co. Inc.

What is the R validation Hub?

R Validation Hub is a collaboration to support the adoption of R within a biopharmaceutical regulatory setting

A RISK-BASED APPROACH FOR ASSESSING R PACKAGE ACCURACY WITHIN A VALIDATED INFRASTRUCTURE

Andy Nicholls, Statistics Director, Head of Statistical Data Sciences, GSK

Paulo R. Bargo, Director Scientific Computing, Statistics & Decision Sciences, Janssen R&D

John Sims, Director, Analytical Systems Architect & Data Science - Pfizer Vaccine Research

On behalf of the R Validation Hub, an R Consortium-funded ISC Working Group

January 23, 2020

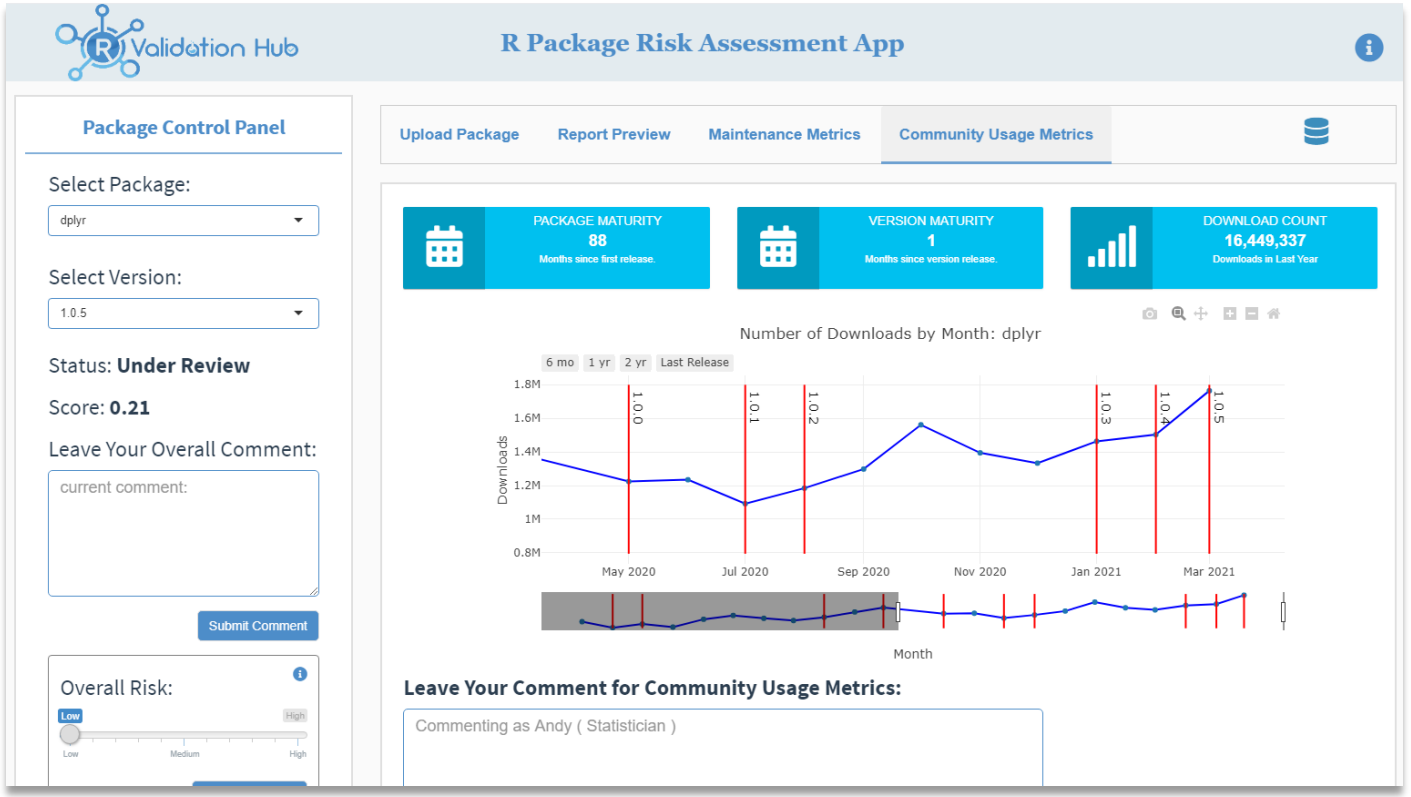
Download the PDF version of this white paper here.

1. Scope and Background

This white paper addresses concerns raised by statisticians, statistical programmers, informatics teams, executive leadership, quality assurance teams and others within the pharmaceutical industry about the use of R and selected R packages as a primary tool for statistical analysis for regulatory submission work. When discussing validation of software systems two areas should be considered:

1. Infrastructure validation
2. Software validation

Infrastructure includes the server, OS, necessary infrastructure software, etc... For example, a system may use a server running Redhat Enterprise Linux (RHEL) version 6 and several other infrastructure software pieces including proxy servers like Apache httpd. Documenting infrastructure (or the environment) is an essential part of the validation process and this validation could follow standard practices such as those proposed in GAMP5, particularly change control management. Discussions regarding infrastructure validation are not in scope of this paper.



~~How do I validate R?~~

Reliable Software

5.8 Integrity of Data and Computer Software Validity

The credibility of the numerical results of the analysis depends on the **quality and validity of the methods and software** (both internally and externally written) used both for data management (data entry, storage, verification, correction and retrieval) and also for processing the data statistically. Data management activities should therefore be based on thorough and effective standard operating procedures. **The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.**

The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) used both for data management (data entry, storage, verification, correction and retrieval) and also for processing the data statistically. Data management activities should therefore be based on thorough and effective standard operating procedures. The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

VI EVALUATION OF SAFETY AND TOLERABILITY

6.1 Scope of Evaluation

In all clinical trials evaluation of safety and tolerability (see Glossary) constitutes an

ICH E9: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf

Is my software reliable?

Reliable Software

Why do we trust a software output?

- *"I know the actual answer"*
- *"It's in the ballpark of what I might expect to see"*
- *"I've used the software before and it did what I expected"*
- *"Many others use the software and it does what they expect"*
- *"When I learnt statistics, I was taught using the software"*
- *"The software is used/cited in statistical literature"*
- *"I trust that the software owner develops it using best practice"*
- *"The software owner provides tests that I can use to verify that it is working"*

Intuition

Community Exposure

Developer SDLC

So why is it so difficult to use Open Source languages for GxP analyses?!

Challenge 1: The R Ecosystem

- **Core R** (Base+Recommended) - Low risk

- **Contributed** - Variable risk
  - Many different authors
  - Varying SDLCs
  - Varying levels of popularity
  - Potentially lots of unknowns

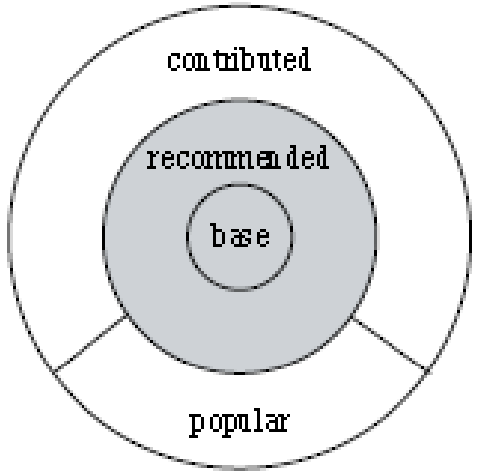


Image source: German, D.M. & Adams, Bram & Hassan, Ahmed E.. (2013). The Evolution of the R Software Ecosystem. Proceedings of the Euromicro Conference on Software Maintenance and Reengineering, CSMR. 243-252. 10.1109/CSMR.2013.33.

Summary

Definitely possible

*"The computer software used for data management and statistical analysis should be **reliable**, and documentation of **appropriate** software testing procedures should be available"*

ICH E9

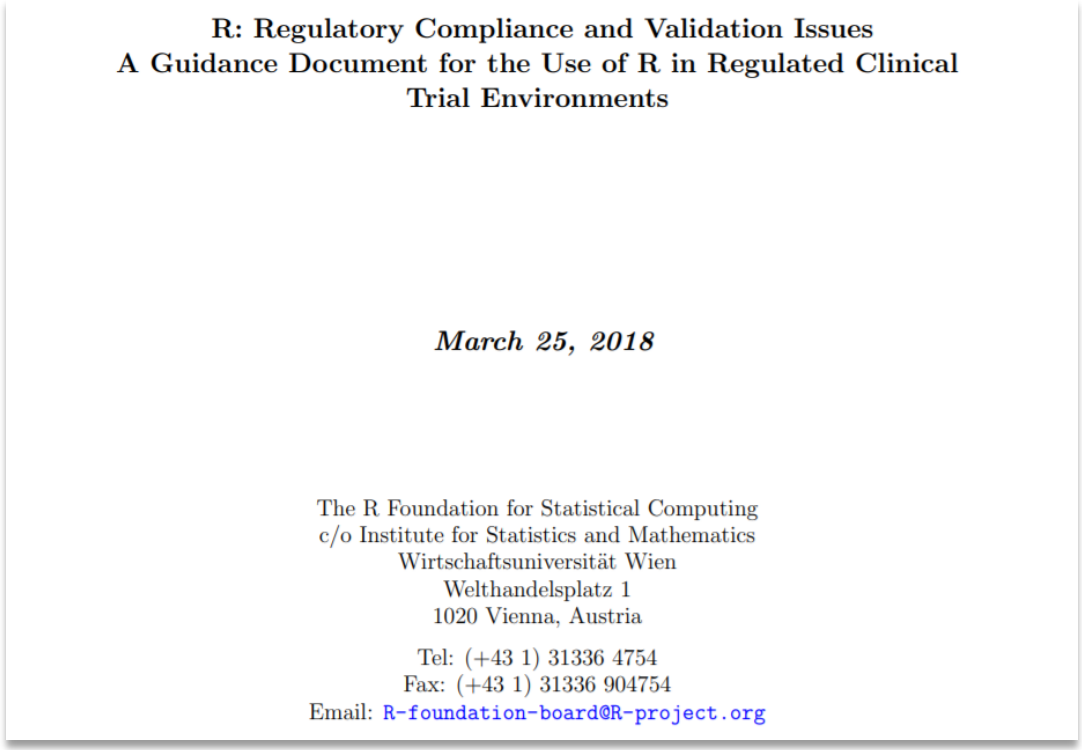
More of a challenge!

Thank You

Further Reading

- **R**
  - R Validation Hub
  - R: Regulatory Compliance and Validation Issues A Guidance Document for the Use of R in Regulated Clinical Trial Environments
  - tidyverse, tidymodels, r-lib, and gt R packages: Regulatory Compliance and Validation Issues
- **ICH**
  - E9
- **FDA**
  - FDA Statistical Software Clarifying Statement
  - 21 CFR Part 11
  - Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application
  - Glossary of Computer System Software Development Terminology
  - General Principles of Software Validation; Final Guidance for Industry and FDA Staff
- **EMA**
  - Notice to sponsors on validation and qualification of computerised systems used in clinical trials
  - Q&A: Good clinical practice (GCP)

I trust that the software owner develops it using best practice



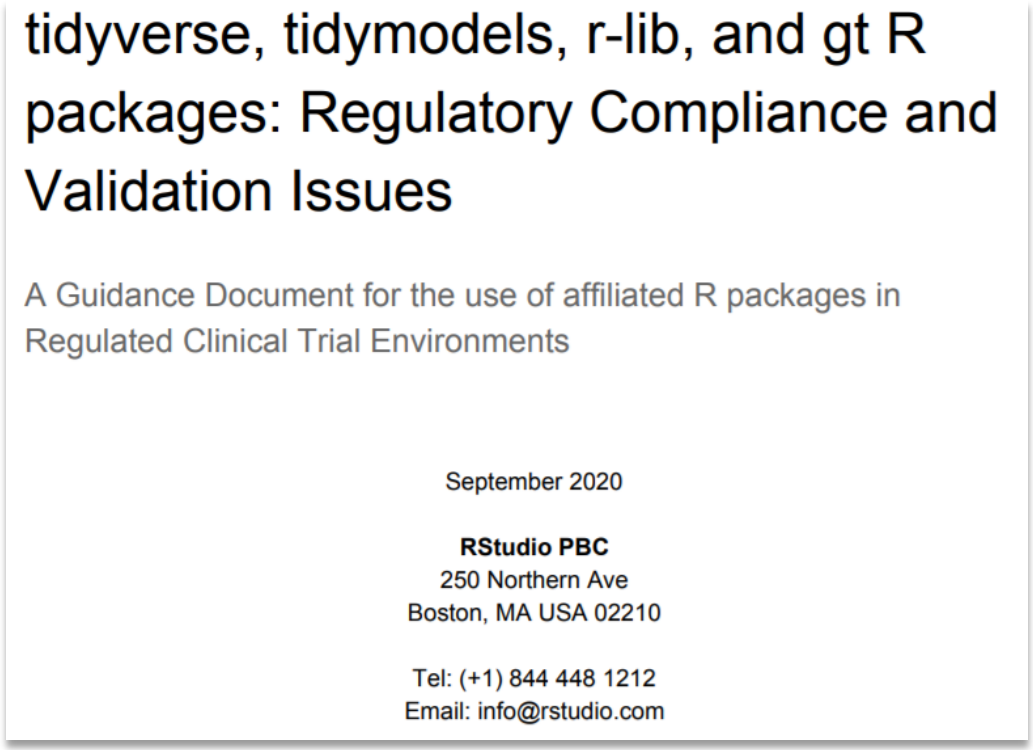
R: Regulatory Compliance and Validation Issues
A Guidance Document for the Use of R in Regulated Clinical Trial Environments

March 25, 2018

The R Foundation for Statistical Computing
c/o Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien
Welthandelsplatz 1
1020 Vienna, Austria

Tel: (+43 1) 31336 4754
Fax: (+43 1) 31336 904754
Email: R-foundation-board@R-project.org

https://www.r-project.org/doc/R-FDA.pdf



tidyverse, tidymodels, r-lib, and gt R packages: Regulatory Compliance and Validation Issues

A Guidance Document for the use of affiliated R packages in Regulated Clinical Trial Environments

September 2020

RStudio PBC
250 Northern Ave
Boston, MA USA 02210

Tel: (+1) 844 448 1212
Email: info@rstudio.com

https://resources.rstudio.com/assets/img/validation-tidy.pdf