

## **Validating R for Pharma Streamlining the Validation of Open-Source R Packages within Highly Regulated Pharmaceutical Work**

Anuja Das, Biogen;  
Juliane Manitz, EMD Serono;  
Doug Kelkhoff, Roche;  
Jaxon Abercrombie, Genentech;  
Antal Martinecz, Certera

### **ABSTRACT**

Healthcare regulators require documentation showing that the quality of the programming packages used for submissions was adequately assessed. This can pose a challenge when using R or other open-source software due to the varying development practices and quality expectations. This paper will highlight the R Validation Hub's risk assessment framework for R packages that has been utilized by various companies across the pharmaceutical industry. We showcase the products that working groups have developed: The {riskmetric} R package which evaluates the risk of an R package using a specified set of metrics and validation criteria. The {riskassessment} app which provides a dashboard to the {riskmetric} package with a Shiny app front end. The repositories workstream which develops a technical framework to maintain a 'repository' of R packages with accompanying evidence of their quality and the assessment criteria. Lastly, we will discuss objectives of the newly instituted Communications Workstream that seeks to bring individuals who are interested in R Validation together through a variety of forums to foster exchange across the industry. All of our work is designed to facilitate the use of R within a highly regulated space and ease the burden of using R packages within a validated environment.

### **INTRODUCTION**

#### **THE R VALIDATION HUB**

The R Validation Hub sits within the R Consortium and is a cross-industry collaboration to support the adoption of R within a biopharmaceutical regulatory setting through appropriate tools and resources that leverage the open source, collaborative nature of the language. It is comprised of a core set of workstream leads and team members along with participants from over fifty biotech and pharmaceutical companies who are brought together through regular community meetings, case studies, and other collaborations.

The R Validation Hub was founded in 2018 and has since worked on developing several resources that are utilized industry-wide. While the initial impetus was for the group to consolidate definitions, regulations, and best practices for qualification and validation of the open-source language, the initial success has now given rise to a variety of workstreams that work on white papers, frameworks and prototypes, applications, and forums for open discussion. We will discuss several of these completed and ongoing efforts throughout this paper.

### **THE R VALIDATION HUB WHITE PAPER**

#### **A REFERENCE FOR A RISK-BASED APPROACH TO R VALIDATION**

The white paper titled *A Risk-Based Approach For Assessing R Package Accuracy Within A Validated Infrastructure* was published in January, 2020. Authors of the paper represented several companies throughout the industry; however it is crucial to note that the framework presented reflects the thoughts of the R Validation Hub working group and not the companies themselves. Since its publication, several

pharmaceutical companies have referenced this framework within their teams as they turn to R as a tool for regulatory submission.

The paper focuses on systems validation, not infrastructure validation, and highlights the need to test for accuracy, reproducibility, and traceability of R packages for intended use as part of the validation process, especially within the clinical submission space. This work highlights R's suitability for regulatory deliverables as described in the FDA "Statistical Software Clarifying Statement."

## BASE R CORE PACKAGES VS CONTRIBUTED OPEN-SOURCE PACKAGES

The R Validation Hub makes a distinction between base R and recommended R packages contributed to the R Foundation and other contributed open-source packages when assessing the risk of each.

The R Foundation develops the base and recommended packages and ensures a stringent software development life-cycle (SDLC) process is followed for maintenance. Additionally, the R user community continually tests these packages and provides feedback as needed. The various practices put in place to ensure validity of this set of packages automatically moves them towards a lower risk classification.

Contributed packages can be developed by anyone who uses R, and they only go through a series of high-level technical checks prior to being added to the CRAN repository. A risk assessment is still necessary for validation of such packages.

## THE RISK ASSESSMENT FRAMEWORK

The recommended risk-based approach assesses Intended-For-Use packages based on Purpose, Maintenance Good Practice, Community Usage, and Testing. An overview of the proposed process is in the figure below:

Figure 1: Proposed Process for Risk Based Approach

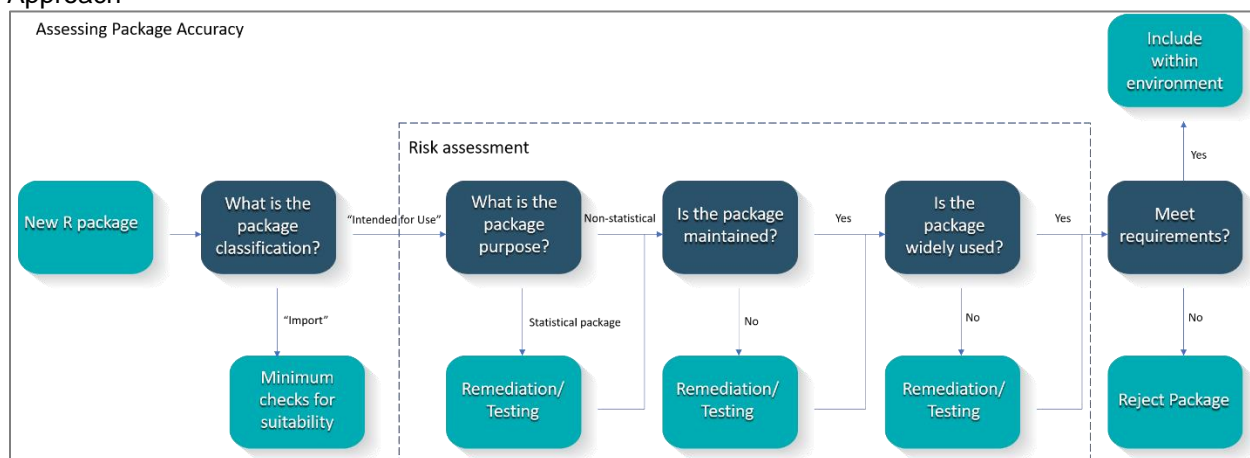


Figure 1: Proposed Process for Risk Based Approach

In practice, this framework can be used to classify packages into overall categories of riskiness. It is important to remember that following initial risk assessment and validation, long-term maintenance practices must be established.

The {riskmetric} package and {riskassessment} app are tangible products developed by the R Validation Hub that can help to implement this framework.

## R PACKAGE {RISKMETRIC}

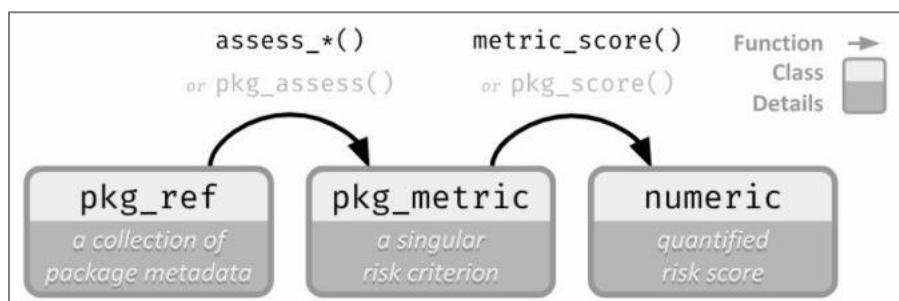
The {riskmetric} package encourages the adoption of R in a regulated environment by providing a tangible metric by which to evaluate and quantify package risk. Moving from perceived risk to quantified risk allows companies to evaluate the requirements more easily for validation, thereby leading to a more efficient timeline and method for the adoption of packages into their infrastructure.

The package is helpful for various end users, from programmers to system administrators. It is intended to provide a foundation for companies to build their own risk assessment framework.

## HOW IT WORKS

{riskmetric} works in three primary steps to assess the risk of an R package:

- Finding a source of package information to compile package metadata
- Assessing a package based on outlined validation criteria
- Assigning a numeric risk score based on the results of risk criteria



**Figure 2: Primary Steps to Assess R Package Risk**

The results of running a package through {riskmetric} are assembled in a dataset of validation criteria along with the resulting risk score.

## FUTURE WORK

The {riskmetric} is being continuously improved based on feedback, scenario planning, and complexities including, but not limited to, dependency footprints and package interoperability. Additionally, the push to make {riskmetric} more usable for non-technical or novice R users to increase the potential pool of end users resulted in the creation of the {riskassessment} application.

## THE {RISKASSESSMENT} APPLICATION

The {riskassessment} app is a Shiny frontend that extends the principles of {riskmetric} to be more accessible. The development of the Shiny app essentially negates the need for technical coding prowess to reap the benefits of the risk assessment framework. Additionally, the app allows users to log assessments, generate reports, and store communication for assessed packages in one single database.

## HOW IT WORKS

The {riskassessment} app has several pre-defined roles and requires authentication. You can use the app as an admin, lead, reviewer, or viewer. Upon deployment, users are presented with the main screen where they can select packages and versions to assess. Once a package is selected, an overall risk classification is assigned, and users have the ability to check the results of each criterion using the "Package Metrics" tab. In the figure below, the {markdown} package was assessed and given a risk score of 0.19 with a Low-Risk classification.

**R PACKAGE RISK ASSESSMENT APP**

Upload Package | Package Metrics | Source Explorer | Build Report

**PACKAGE CONTROL PANEL**

PACKAGE NAME: markdown

DATE UPLOADED: 2023-10-24 | PKG VERSION: 1.11

STATUS: REVIEWED | METRIC RISK: 0.19

SELECT OVERALL DECISION: [Dropdown]

SUBMIT DECISION

WRITE OVERALL COMMENT: [Text Area]

SUBMIT COMMENT

**MAINTENANCE METRICS**

- VIGNETTES**: 0 ✓, 5 (Number of vignettes)
- NEWS FILE**: 0 ✓, 1 (Number of NEWS files)
- NEWS CURRENT**: 0 ✓, Yes (NEWS contains current version)
- REPORT BUGS**: 0 ✓, Yes (URL to report bugs exists)
- WEBSITE**: 0 ✓, https://github.com/rstudio/m... (Package public website)
- MAINTAINER**: 0 ✓, Yihui Xie <xie at yihui.name> (Package maintainers)
- SOURCE CONTROL**: 0 ✓, https://github.com/rstudio/m... (Package source control url)
- DOCUMENTATION**: 0 ⚠, Not found (% of documented objects)
- BUGS CLOSURE RATE**: 0 ✓, 100% (% of the last 30 bugs closed)
- LICENSE**: NA, MIT + file
- DEPENDENCIES**: 0.38, 3
- TEST COVERAGE**: 0 ⚠

**Figure 3: Screenshot from Demo Risk Assessment App**

To provide the overall context for both the risk assessment framework used and the resulting metrics, the application includes options to review assessment criteria, add comments, and build a report for the assessed package and resultant score. There is a demo application that is open to all and allows users to familiarize themselves with the interface.

## RECOGNITION AND FUTURE WORK

During the 2<sup>nd</sup> Shiny Conference in March 2023, the app was awarded the Best App by popular vote among 20+ other app submissions. The {riskassessment} application continues to evolve based on user feedback and metrics are adjusted according to {riskmetric} development.

## REGULATORY R PACKAGE REPOSITORY

The Regulatory Repository working group is in its early stages and was initiated to simplify the burden of assembling and storing R packages with accompanying evidence and documentation of quality testing and validation assessments. The intention is to create a cross-industry repository to magnify the benefits of true open-source collaborations and ultimately reduce the risk of using R packages for regulatory submissions. The pilot will involve the development of a centralized and public repository along with its accompanying validation pipeline and reference container image.

## FUTURE WORK

The pilot is currently ongoing with a slated readout in Q3 2024. The pilot will include a base reference image, basic automated validated pipeline, and a basic repository to store and manage validated R packages.

Outputs from the workstream could also incorporate best practices that help teams prepare for audits when using open-source packages as part of their regulatory work, and that help build industry consensus on avoiding uncertainty in health authority response to submissions.

## THE COMMUNICATIONS WORKSTREAM

The newest working group within the R Validation Hub is the Communications Workstream which seeks to engage and leverage the essence of all open-source work – the users. Engagement with our entire gamut of stakeholders, from data scientists within the industry to casual R users, ensures that the R Validation Hub is working on the right scope of work that brings the highest value to the open-source community within the pharmaceutical space. The workstream also seeks to foster exchange between individuals who are interested in open-source software validation from across industry in a variety of forums.

## ONGOING AND FUTURE WORK

The Communications Workstream uses four primary communication pathways: [the website](#), regular community meetings, industry case studies, as well as social media and company outreach. Different types of outreach help to find commonalities within the pharmaceutical industry, and to identify areas that could lead to the creation of future working groups

The working group is working to broaden the impact of the Validation Hub through a series of goals.

- A website refresh that allows increased interaction with all the resources developed by the R Validation Hub including the white paper blog posts, and updates on ongoing and future work.
- Updates to previous case studies to showcase examples of the R Validation Framework in action.
- The Community Meetings which cover a variety of topics that range from open discussions on future work, a focus on specific company case studies, and breakouts to talk about what validation means (the answer varies depending on who you talk to!). These have been a great way to ensure that R Validation Hub projects have been on the pulse of the most pressing issues at hand and are also a great way to generate feedback on ongoing efforts and use cases.
- Improving outreach and establishing connections with company-internal R User Groups.

## CONCLUSION

In conclusion, the efforts of the R Validation Hub have significantly contributed to streamlining the validation of open-source R packages within the highly regulated pharmaceutical industry. The risk assessment framework, along with tangible products like the {riskmetric} package and {riskassessment} app, has provided a structured approach to assess package risk, enabling efficient adoption into validated environments. The ongoing workstreams, including the Repository and Communications Workstream, showcase a commitment to continuous improvement and industry collaboration. These endeavors collectively aim to ease the burden of using R in regulated spaces, fostering a more accessible and validated environment for leveraging open-source tools in pharmaceutical work. We are always looking for new volunteers and members. Please reach out to us if you want to get involved or have any questions about our work.

## REFERENCES

Nicholls, A., Bargo, P. & Sims, J. on behalf of the R Validation Hub. 2020. “A Risk-Based Approach for Assessing R Package Accuracy within a Validated Infrastructure.” Available at [A Risk-based Approach for Assessing R package Accuracy within a Validated Infrastructure \(pharmar.org\)](#)

U.S. Food and Drug Administration. 2015. “Statistical Software Clarifying Statement.” Available at [Statistical Software Clarifying Statement \(fda.report\)](#)

“The R Foundation.” Available at [R: The R Foundation \(r-project.org\)](#)

The R Validation Hub. “Demo R Package Risk Assessment App.” Available at [https://bit.ly/raa\\_demo](https://bit.ly/raa_demo)

## ACKNOWLEDGEMENTS

Our thanks to the R Validation Hub Executive Committee and Working Group Leads – Aaron Clark, Eric Milliman, Coline Zeballos.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Anuja Das  
[Anuja.das@biogen.com](mailto:Anuja.das@biogen.com)

Juliane Manitz  
[juliane.manitz@emdserono.com](mailto:juliane.manitz@emdserono.com)

Doug Kelkhoff  
[doug.kelkhoff@roche.com](mailto:doug.kelkhoff@roche.com)

Jaxon Abercrombie  
[abercrombie.jaxon@gene.com](mailto:abercrombie.jaxon@gene.com)

Antal Martinecz  
[antal.martinecz@gmail.com](mailto:antal.martinecz@gmail.com)