Validating R for Pharma

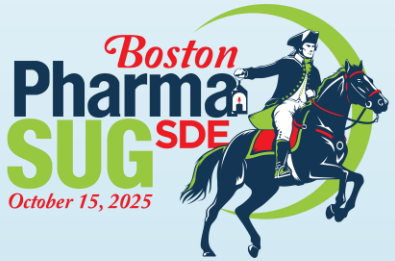
Streamlining the Validation of Open-Source R Packages within Highly Regulated Pharmaceutical Work

Anuja Das

Anuja Das is a Technology Project Specialist within the Data Visualizations and Application Development group at Biogen and a member of the Communications Workstream within the R Validation Hub. She enjoys bringing innovation, automation, and new ways of working to the existing clinical submissions process and has led projects that deploy applications created in multiple languages that enable better data review and output generation.

Jaxon Abercrombie

Jaxon Abercrombie is a Data Scientist at Genentech with a core focus on analyzing clinical trial data and supporting regulatory submissions, particularly for GLP-1/GIP receptor agonist therapies. Beyond his core work, Jaxon is an advocate for open-source tools, loves creating visualizations, and has contributed to the R Validation Hub Communications Workstream since 2023.



Agenda

1. What is the R Validation Hub?
2. The White Paper
3. {riskmetric}
4. {riskassessment}
5. {val.*} Packages
6. Regulatory R Repository
7. Communications

But First...

What is the R Validation Hub?

Some companies are:

- *Abbvie*
- *Biogen*
- *Lilly*
- *Merck*
- *Parexel*
- *Posit*
- *Roche*
- *Sanofi*

And many more…

Affiliates: R consortium

- Nested under the **R Consortium**
- Cross industry collaboration of over **fifty biotech and pharmaceutical companies**
- **Initial aim:** Consolidate definitions, regulations, and best practices for qualification and validation of the open-source language
- **Current State:** Grown to include a variety of workstreams that produce resources and tools for open-source use aligning with the expectations of regulatory agencies

The White Paper

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

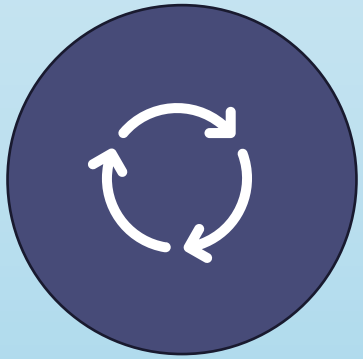
On behalf of the **R Validation Hub**, an R Consortium-funded ISC Working Group

January 23, 2020

A Risk-Based Approach for Assessing R Package Accuracy within a Validated Infrastructure

Focuses on systems validation and adequate testing and documentation in keeping with the FDA's "Statistical Software Clarifying Statement"

Accuracy

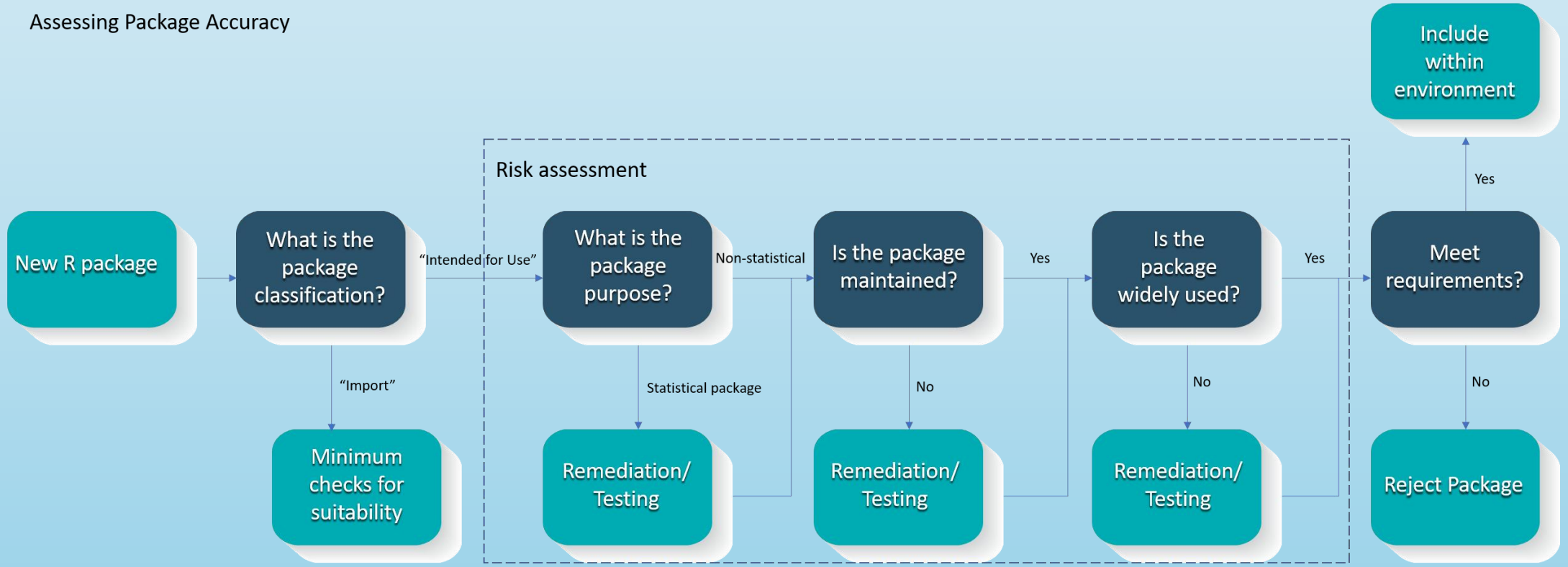


Reproducibility

Traceability

"For Intended Use"

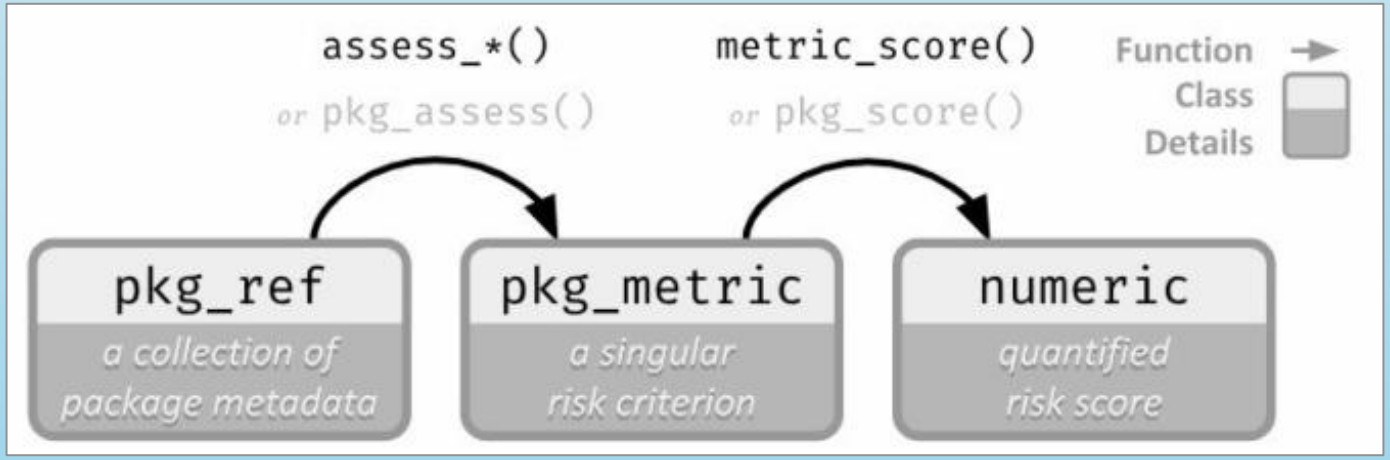
Proposed Risk-Based Approach to Assess Accuracy

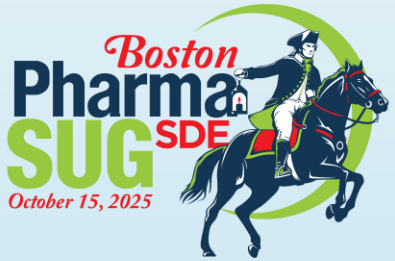


{riskmetric}

How It Works

Perceived Risk → Quantified Risk





How Quality can be Quantified

- A license
- Source code available for browsing
- An easy to contact maintainer
- Up-to-date news regarding new features and bugs
- A place to report bugs
- Evidence that new bugs are being addressed
- Package documentation & user guides/ perhaps a website
- Complete Function documentation
- Adequate test coverage
- Community usage
- *and many more!*

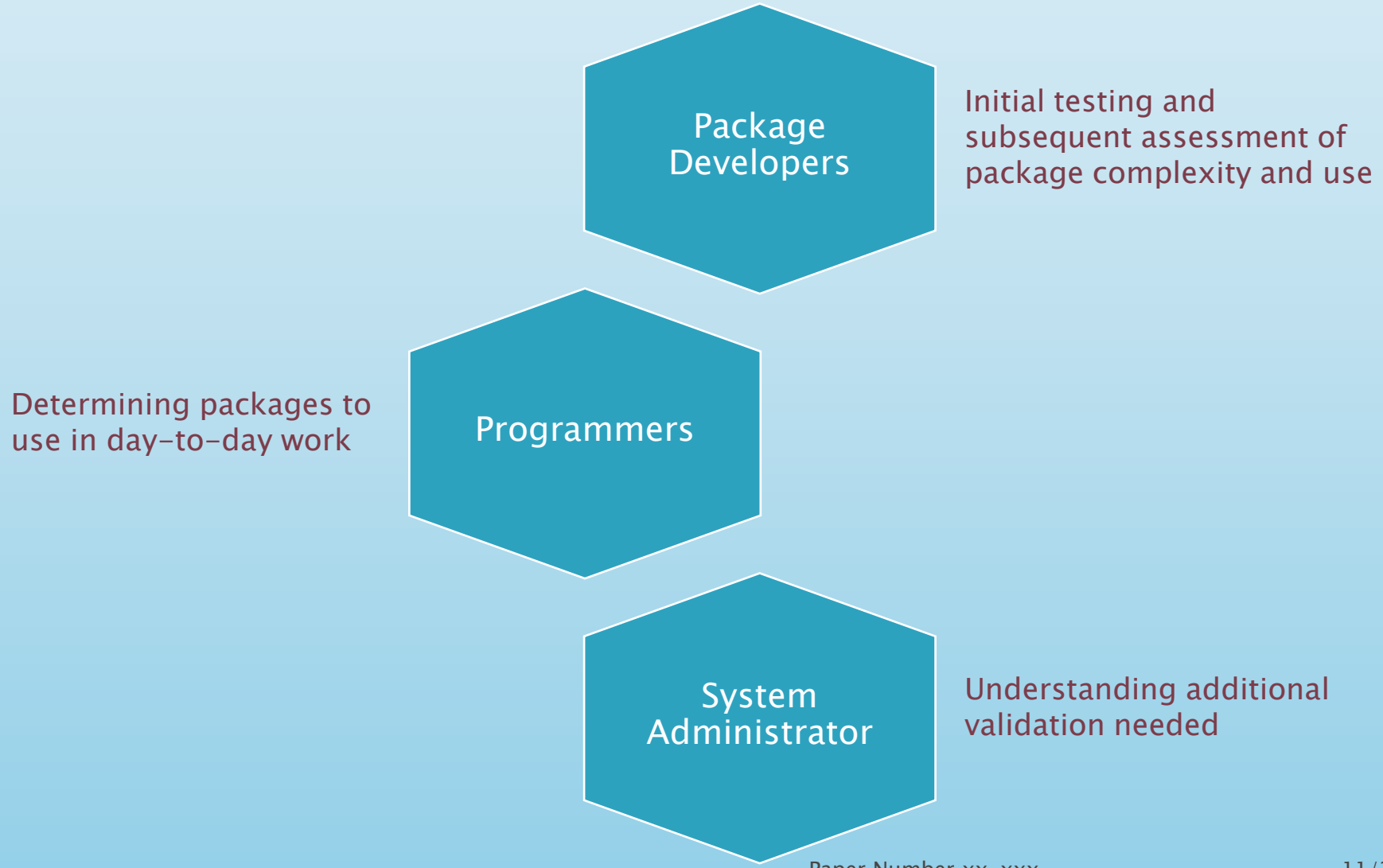
Unit Testing Metrics

Documentation Metrics

Community Engagement

Maintenance & Reusability

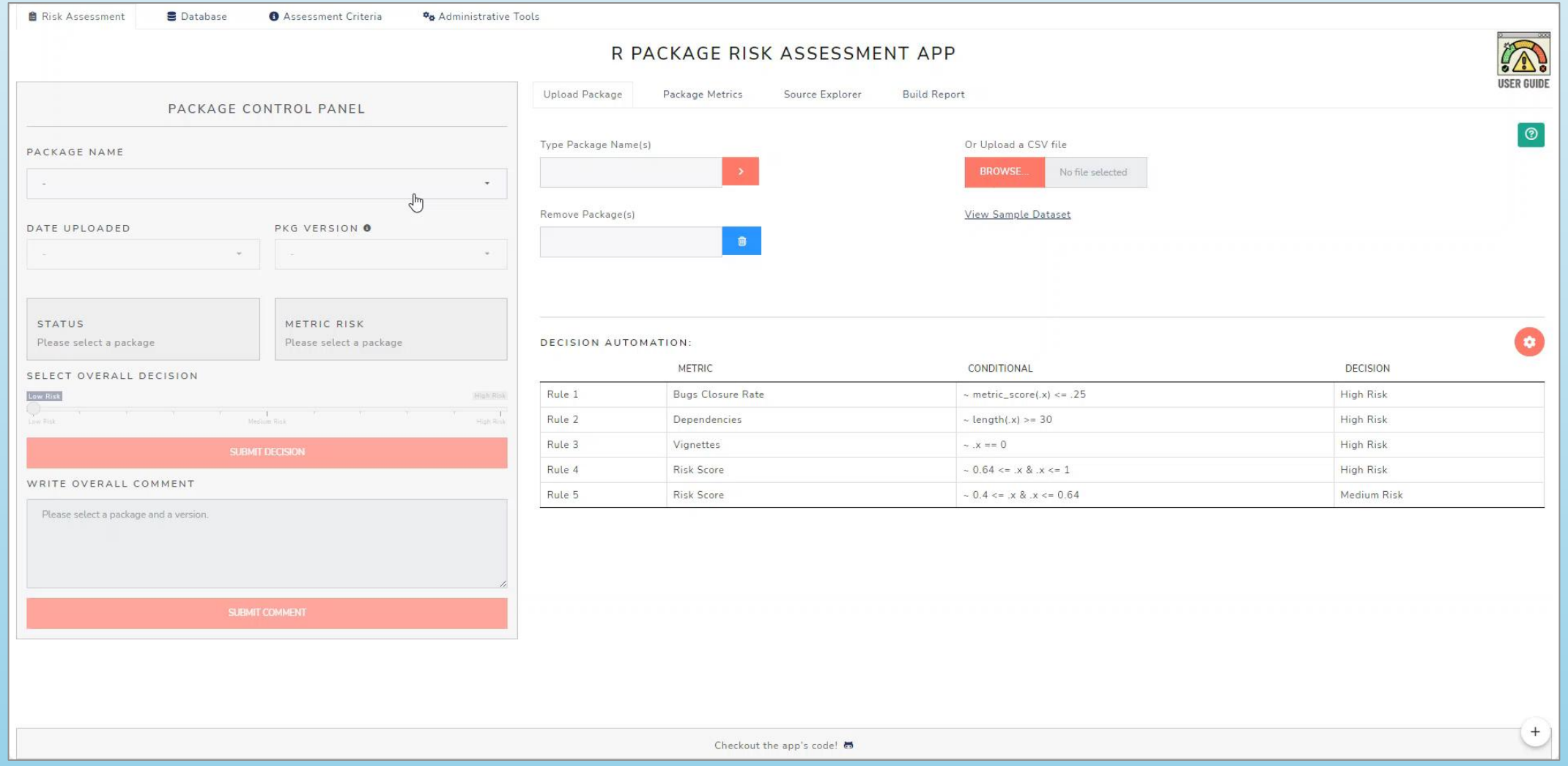
Who can use {riskmetric}?



{riskassessment}

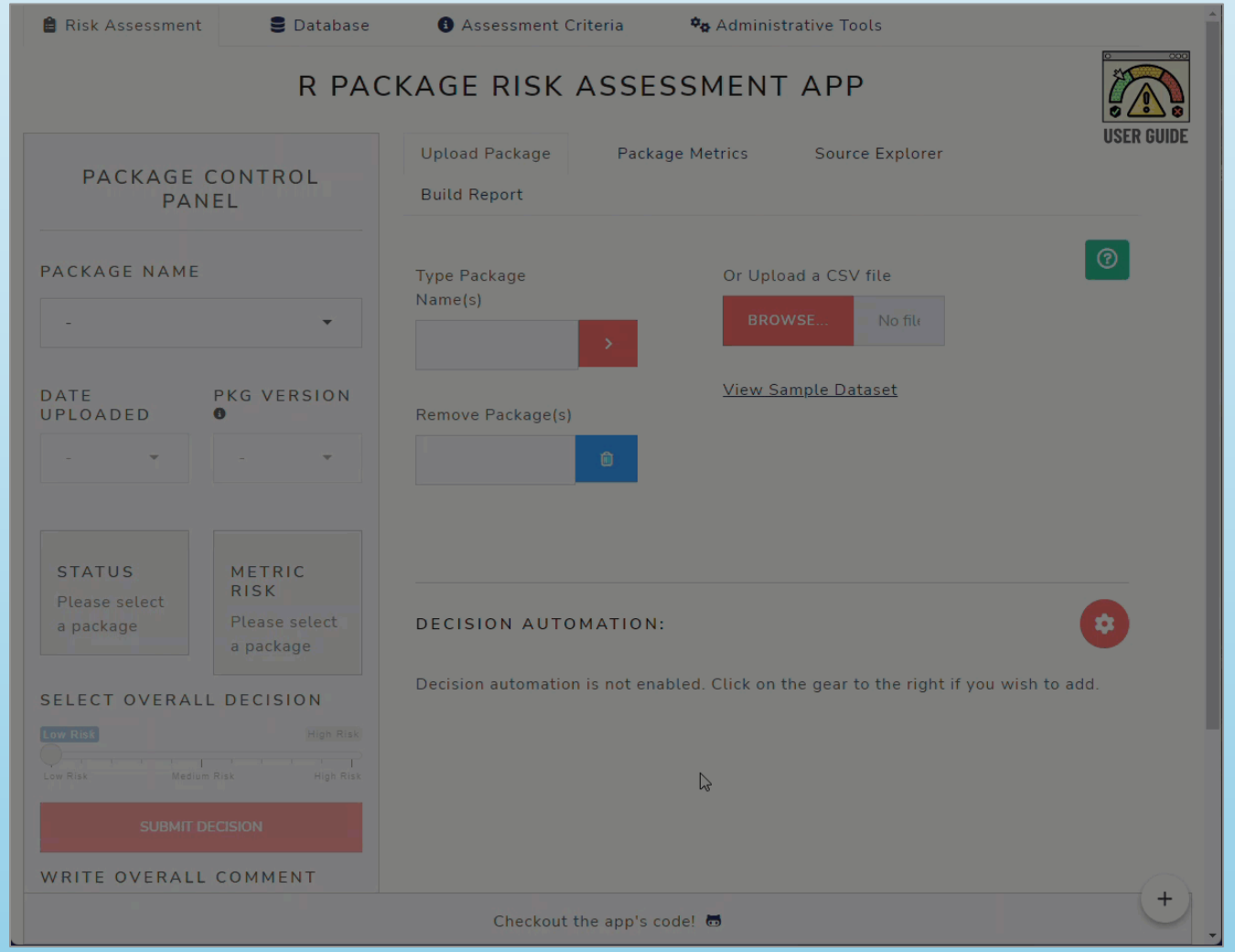
Voted 'Best App' at Shiny Conference 2023

How It Works

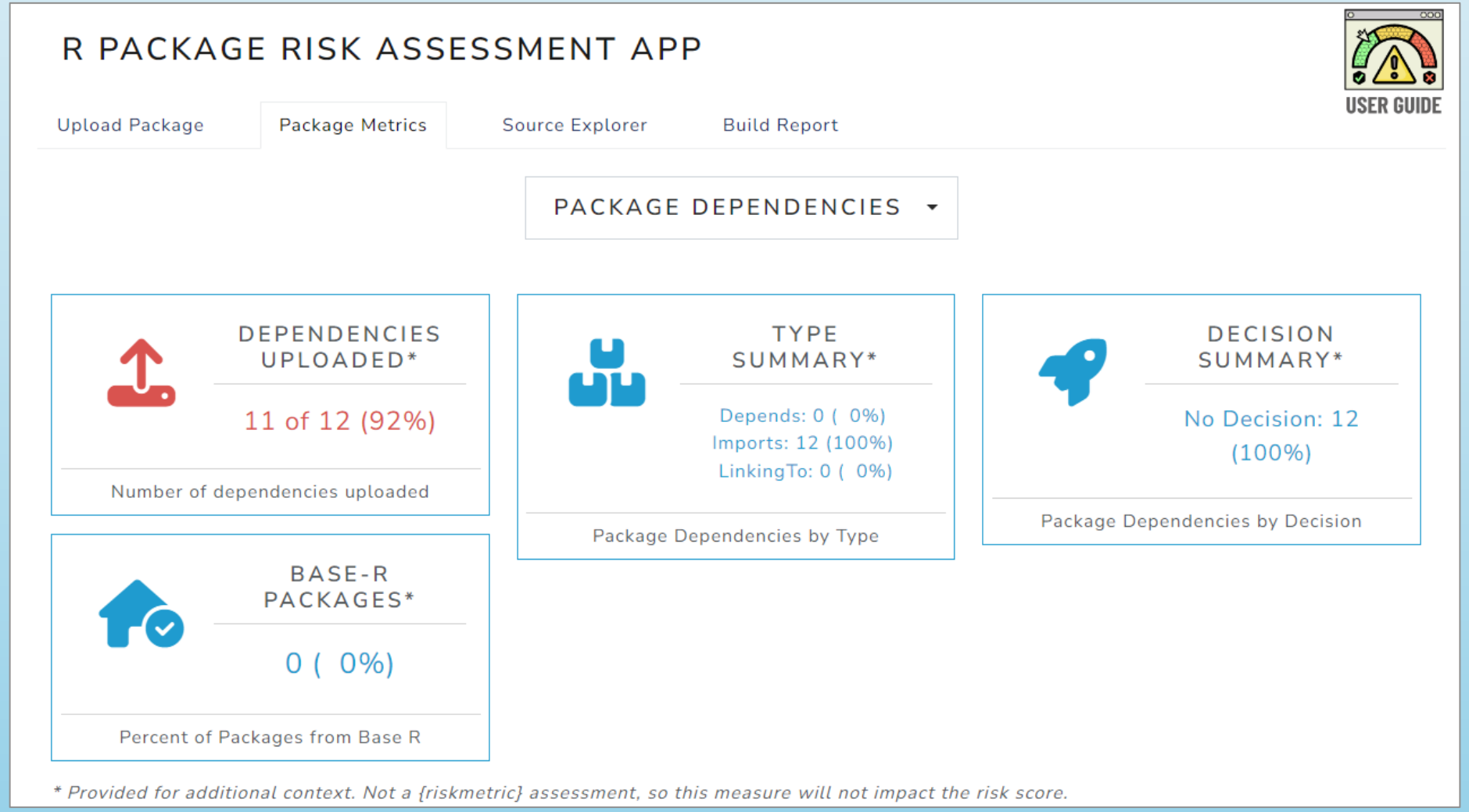


	METRIC	CONDITIONAL	DECISION
Rule 1	Bugs Closure Rate	~ metric_score(.x) <= .25	High Risk
Rule 2	Dependencies	~ length(.x) >= 30	High Risk
Rule 3	Vignettes	~ .x == 0	High Risk
Rule 4	Risk Score	~ 0.64 <= .x & .x <= 1	High Risk
Rule 5	Risk Score	~ 0.4 <= .x & .x <= 0.64	Medium Risk

Decision Automation – Integration with {riskassessment}



Exploring Dependencies



{val.*} Packages

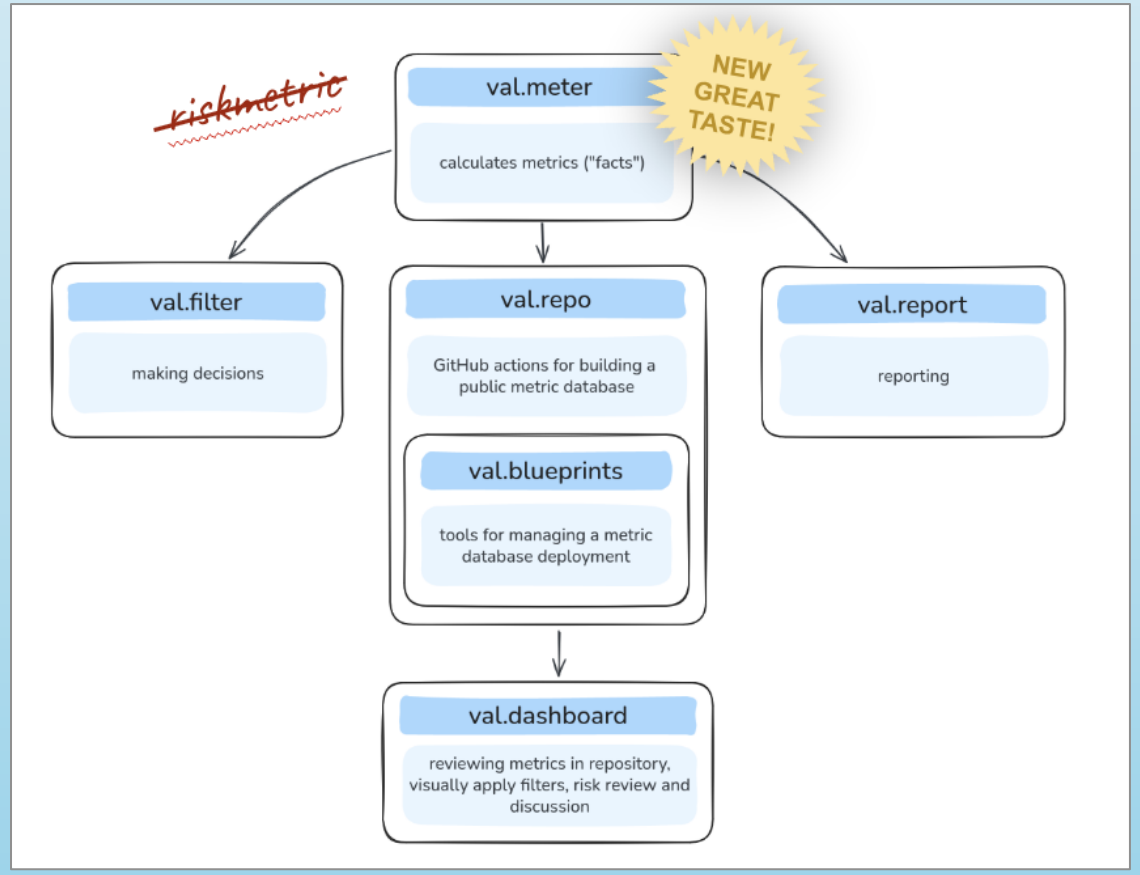
Our newest endeavor for package quantification in support of the regulatory use of R!

★ Currently exist

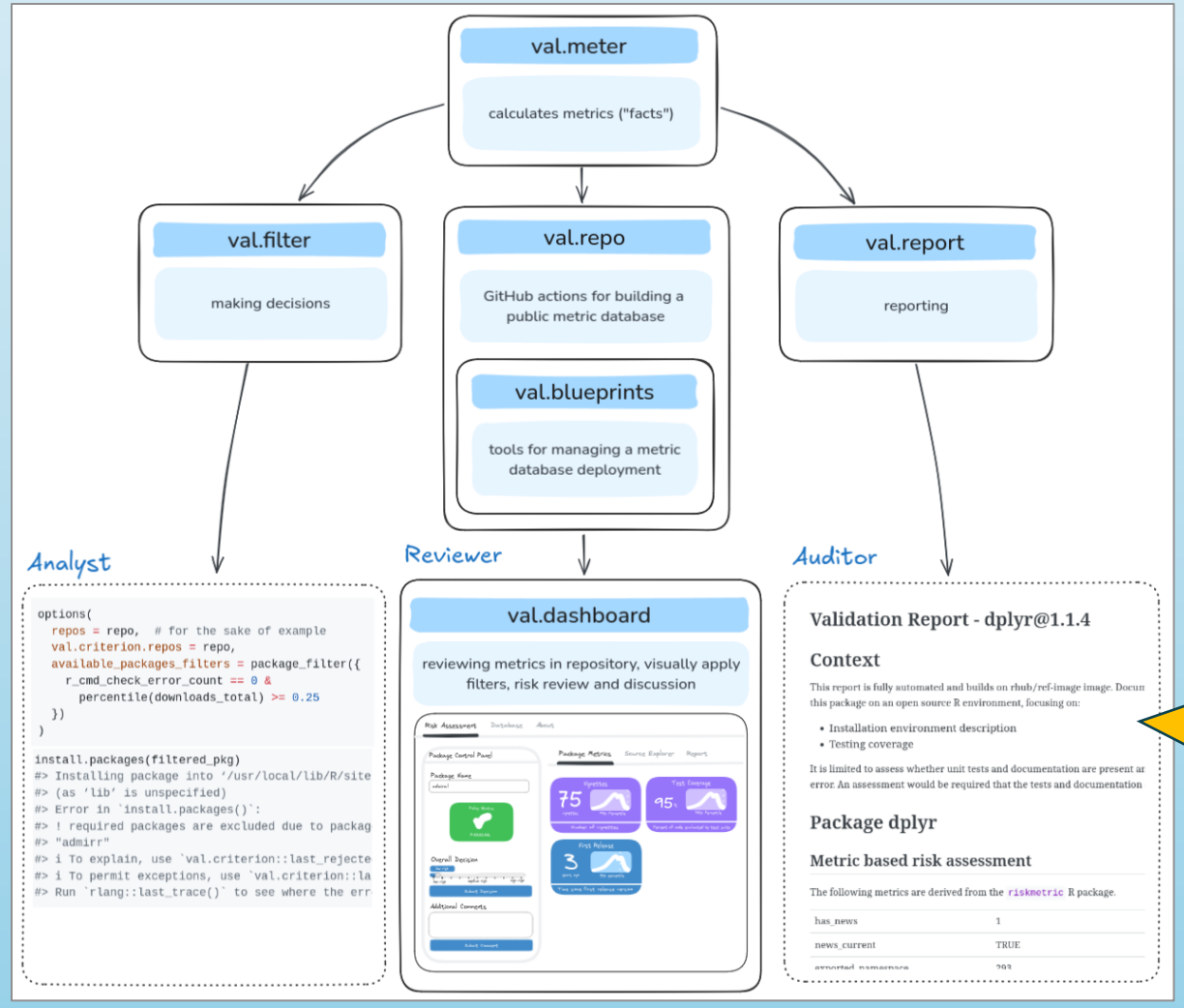
Evolution into {val.*} Packages

- A Dev Day hosted in April 2025 inspired **an evolution from {riskmetric}**; some decisions:
  - Focus on PACKAGES file
  - Commitment to “no guessing” strategy and focus on “facts” not “opinions” (eg 1M downloads, not “enough downloads”)
  - Adopt S7 and avoid S3-style double dispatch
  - Data to be co-located

- Plan: split up the mature {riskmetric} solution into a family of packages



Who is {val.*} Intended for?



Analysts, reviewers, and auditors alike!

Regulatory R Repository

About the Initiative

- **What It Is / Our Vision**
  - An open-source repository (CRAN-like) of risk assessments available to all, providing quantifiable package qualities
  - A future part of an R Validation Hub Submissions working group pilot (2026); health authority experimentation

- **Goals**
  - Empower **consistent**, **reproducible**, and **transparent** validation
  - **Lower the barrier** to entry for start-ups
  - Empower **consistency among contract research organizations (CROs)**
  - **Improve the speed of regulatory review** for established enterprises

- **Recent strides**
  - Progress with {val.*} packages to bring us closer to public-facing dashboard
  - Companies providing talent for development

Check Out More on Our Site!

REGULATORY METRICS REPOSITORY

An open solution to regulatory quality

Every regulatory submission requires evidence that the software used meets quality standards appropriate for the type of work. The *R Validation Hub* proposed standards for this process in our 2020 White Paper.

Now we're translating those standards into an open-source repository of risk assessments, available to all. We hope to empower validation that is **consistent**, **reproducible** and **transparent**, lowering the barrier to entry for start-ups, empowering consistency among contract research organizations (CROs), and improving the speed of regulatory review for established enterprises.

Intimidated by the climb to validation?
Don't forge your own path. Let us give you a lift!

Validation Hub

Why invest in the Regulatory R Repository

As our industry looks for ways to lower costs and shorten regulatory review processes, there has been a move to investigate and employ the R statistical programming language for broader use in clinical research and drug submissions. While quickly accomplishing the goal of lowering internal software and hardware costs, there are still some hurdles to overcome in this transition to broader use of open source across the pharmaceutical industry.

As companies move to deliver more R-based submissions an industry-wide effort is needed to standardize package validation for submissions to our government agency partners. A consistent and publicly available repository of risk-assessed packages will allow our agency partners to develop familiarity and trust, allowing for submissions to be processed faster and with a high level of confidence. To cement cost savings and speed up agency review of R-based submission packages, we are proposing the development and maintenance of the Regulatory R Repository.

The Regulatory R Repository will provide a collection of assessed R packages utilizing **the R Validation Hub's riskmetric framework**. The framework allows the user to quantify package quality and provide transparent, risk-based decision-making for regulatory use cases. This repository will eliminate the need for pharmaceutical companies, large and small, to spend time and money investigating packages and performing their own evaluations. Instead, the repository will allow companies to select packages based upon their own risk profile from a public repository of risk-assessed R packages that meet regulatory requirements for pharmaceutical submissions.

How to help build and maintain the Regulatory R Repository

We need funding partners to help us raise $250,000 USD to build the first version of the R Regulatory Repository. A grant or donation through the **R Consortium**, where our working group is maintained, will allow us to hire contractors to do the testing and publishing of our first set of packages.

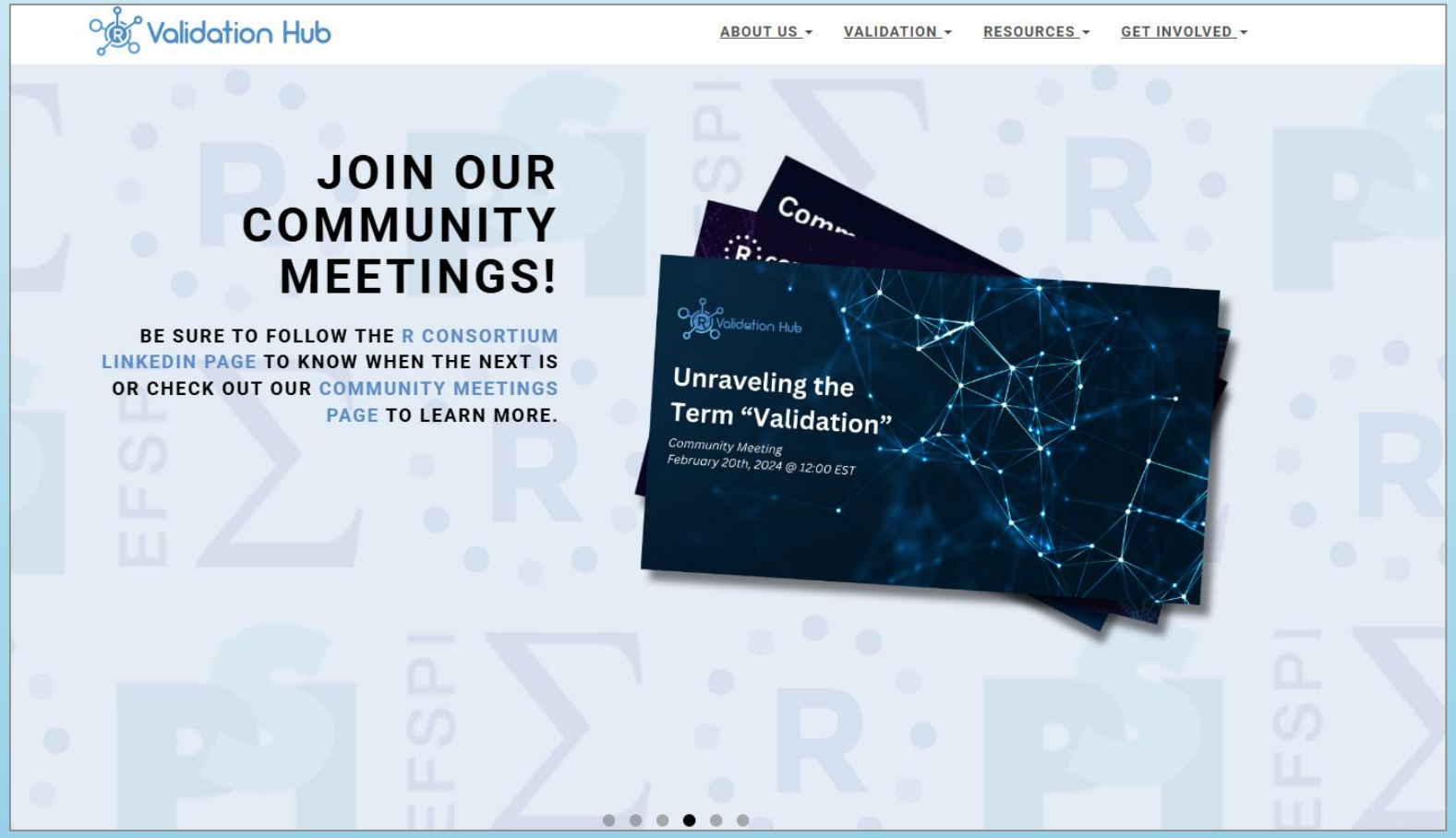
If you or someone you know in your organization would like to help fund this work, please visit our project landing page, review our materials, and opt-in to our mailing list as either a potential sponsor, project champion, or project contributor.

https://pharmar.org/repository/

Communications

Our Platforms

Website (pharmar.org)



Community Meetings

Past topics have included…

1. Unraveling the term 'Validation'
2. Shiny App Validation in Regulatory Submissions
3. Analyzing Change in Assessed Risk Across Package Releases

Get Involved!!

The R Validation Hub is always looking to expand its network. These are a few ways to join in!

Be a part of our Working Groups

Join our mailing list and attend Community Meetings

Tell us your ideas

Submit a Case Study

Name: Anuja Das
Organization: Biogen
E-mail: Anuja.das@biogen.com

Web: pharmar.org