

Heading to base camp on our way up the open-source mountain...

Andy Nicholls (Phuse EU Connect)

November 2023



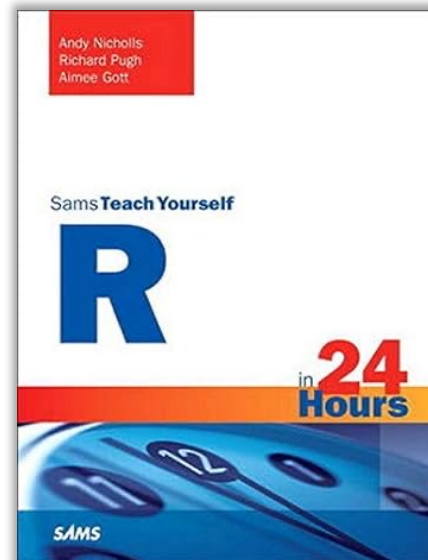
Me and my biases

I identify as a Data Scientist



Enabling
open-source
is my job!

I can also use these!



R user since ~2005

Making Commitments

GSK's commitments to open source in 2023

COMMITMENT 1: An Open-Source Backbone



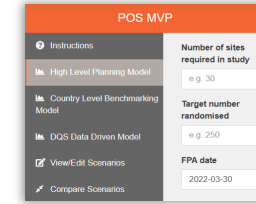
COMMITMENT 2: Parity by end 2025



Photo: <https://katelynmacmillan.com/>

GSK's Open Source Journey

Wow!



Period of the ad-hoc R course

80% staff trained in Tidyverse

AccelerateR

R Platform V1.0

R4GxP

Full study QC in R

Full prod'n study in R



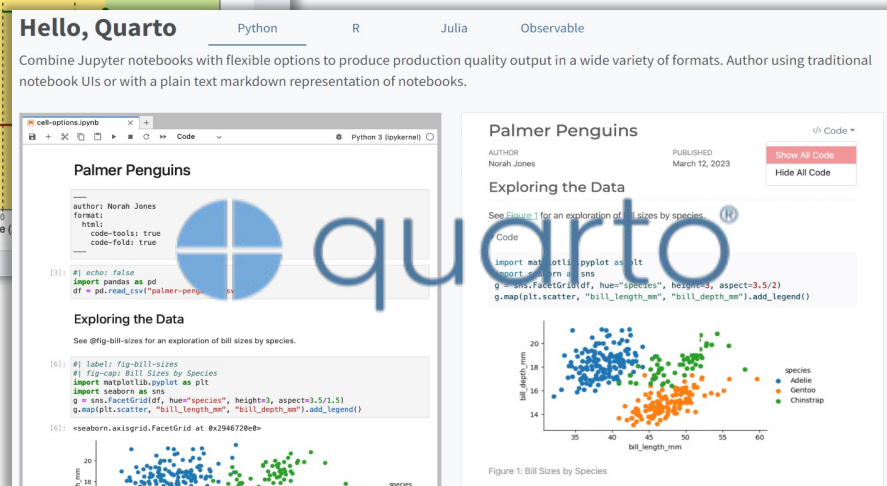
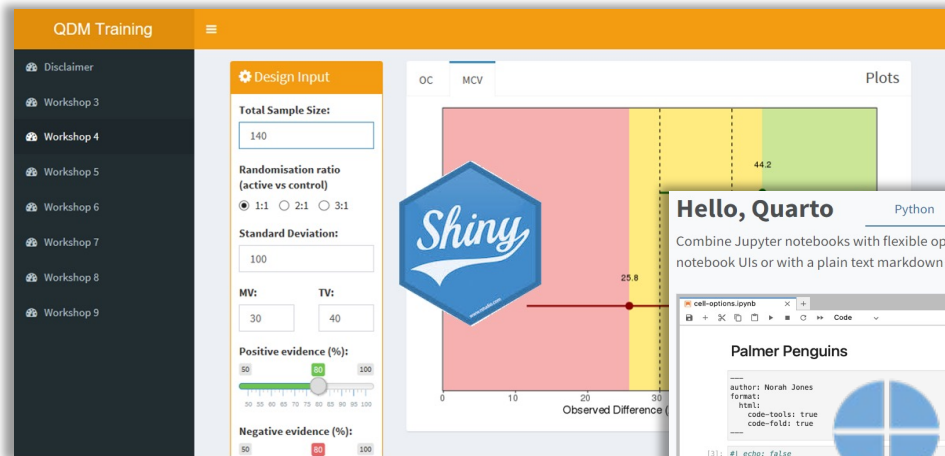
GSK

2017

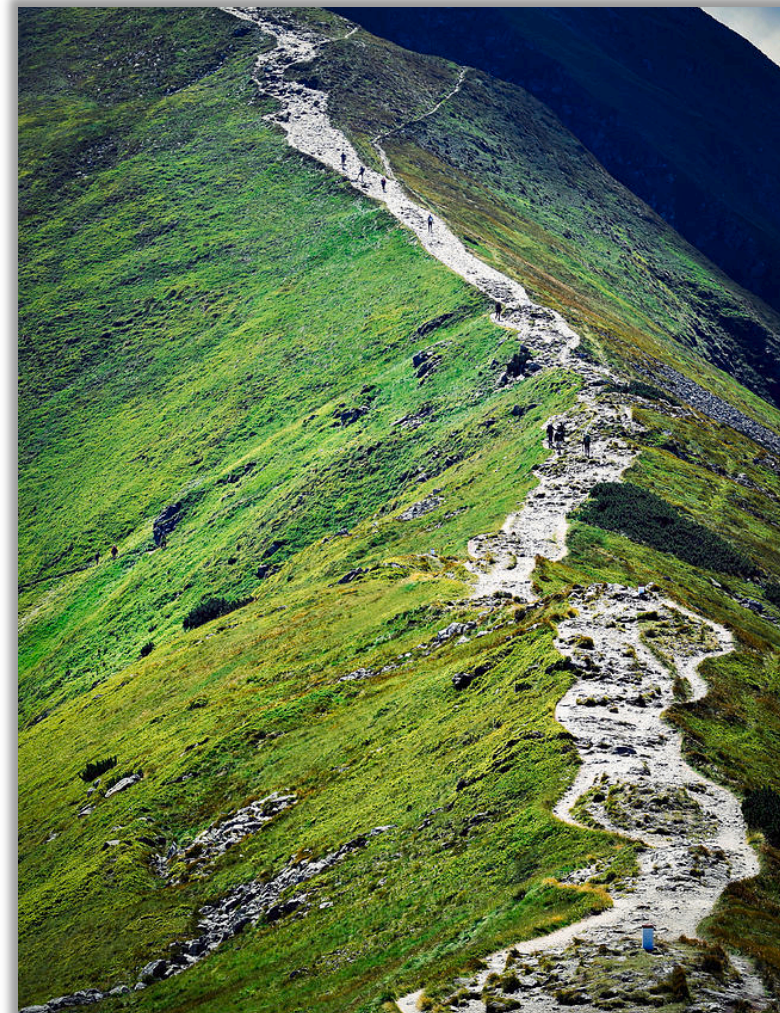
2024

Why are we climbing this mountain?

"I've used [Statistical Analysis Software] for 25 years..."



Libraries are available for C, C++, C#, Go, Java, JavaScript, Julia, MATLAB, Python, R, Ruby, and Rust. See [how to install](#) and get started.



Steep Mountain Trail Up Photograph by
Jozef Jankola - Fine Art America

Why are we climbing this mountain?

R and python are also pretty good at 'the day job' too!



Plenty of resources

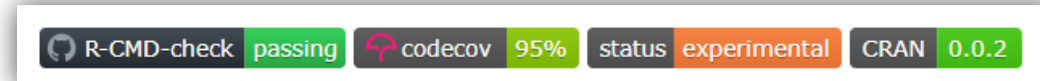


And well-established tools

Why are we climbing this mountain?

Package ecosystems are fantastic!

- The R/python package ecosystems are ideal for analytic engineering
 - Modular, build-and-extend model
 - Fully integrated help documentation with further user-guide capabilities via quarto (markdown)
 - Standardised test frameworks
- Pharma collaboration like we've never seen before!



<https://pharmaverse.org/>

The summit sounds nice
but ...



Making the journey in 2023

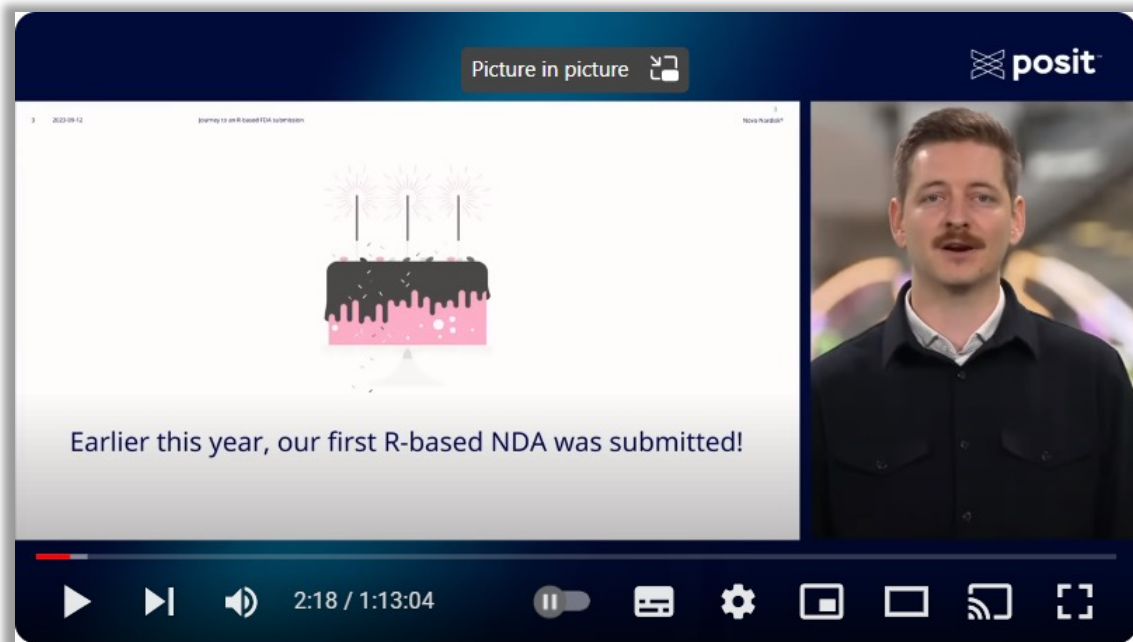




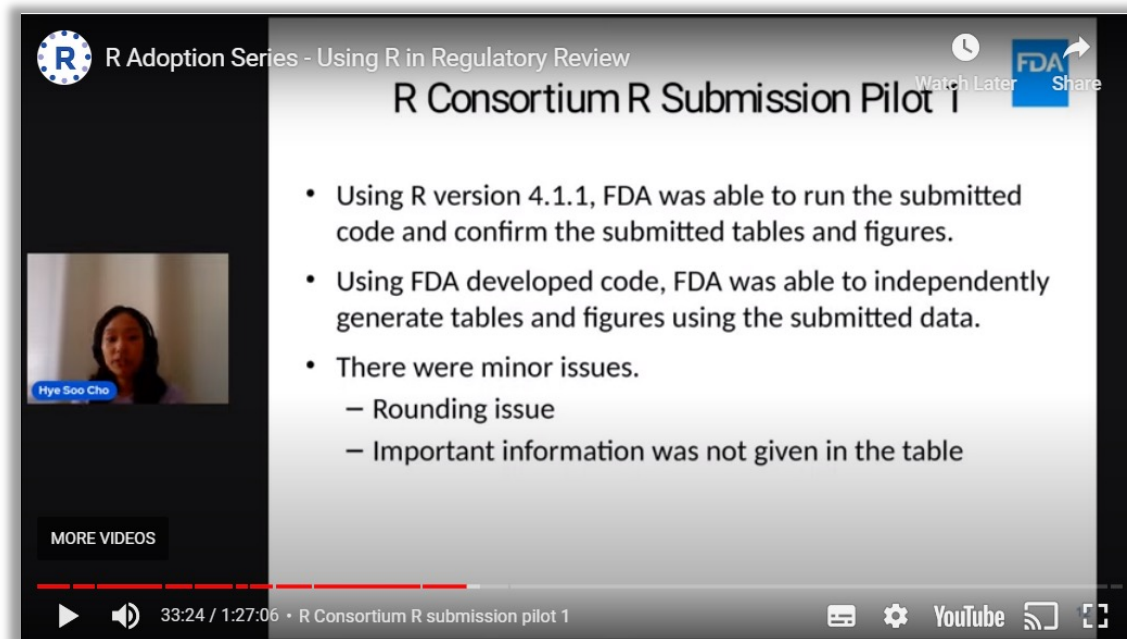
Learning from experience

R-based submissions

Novo Nordisk Submission *Posit Webinar*



Using R in Regulatory Review *R Consortium Webinar*



More from GSK soon...

Make use of available resources

Submissions working group recently piloted a Shiny submission



DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCES
OFFICE OF BIOSTATISTICS

STATISTICAL REVIEW AND EVALUATION

NDA/BLA #:	BLA 111111 (R pilot submission)
Applicant:	R Consortium's R Submission Working Group
Statistical Analyst	Hye Soo Cho Hye Soo Cho -S Digitally signed by Hye Soo Cho -S Date: 2023.09.25 17:33:08 -04'00'
Secondary Reviewer	Paul Schuette Paul H. Schuette -S Digitally signed by Paul H. Schuette -S Date: 2023.09.26 08:56:54 -04'00'
Supervisor	Maria Matilde Kam Maria M. Kam -S Digitally signed by Maria M. Kam -S Date: 2023.09.27 12:07:27 -04'00'
Date(s):	August, 2023
Objective of the submission	To test whether a Shiny application (app) created with the R-language can be successfully incorporated into a submission package and deployed to FDA reviewers.
Location of datasets	

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

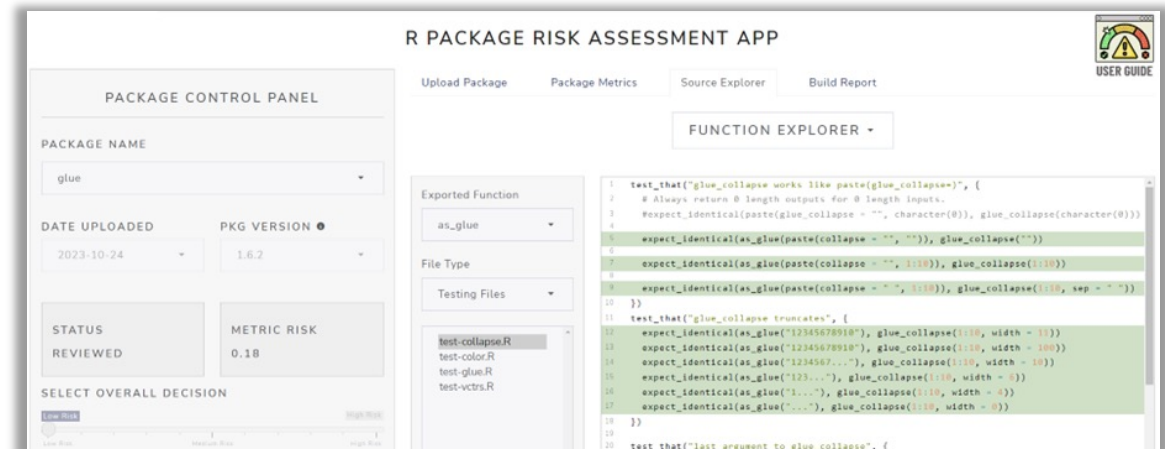
January 23, 2020

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

R Validation Hub:
White paper / risk assessment app



The screenshot shows the 'R PACKAGE RISK ASSESSMENT APP' interface. It includes a 'PACKAGE CONTROL PANEL' with fields for 'PACKAGE NAME' (glue), 'DATE UPLOADED' (2023-10-24), 'PKG VERSION' (1.6.2), 'STATUS' (REVIEWED), and 'METRIC RISK' (0.18). There is a 'FUNCTION EXPLORER' section showing a list of exported functions like 'as_glue', 'test-collapse.R', 'test-color.R', and 'test-vctrs.R'. On the right, there is a code editor displaying R test code for the 'glue' package, including tests for 'glue_collapse' and 'glue_truncates'.



[Home](#) [About](#) [Contributions](#) [Meeting Minutes](#) [Conferences](#) [White Paper](#)

CAMIS

Introduction

Several discrepancies have been discovered in statistical analysis results between different programming languages, even in fully qualified statistical computing environments. Subtle differences exist between the fundamental approaches implemented by each language, yielding differences in results which are each correct in their own right. The fact that these differences exist causes unease on the behalf of sponsor companies when submitting to a regulatory agency, as it is uncertain if the agency will view these differences as problematic. In its Statistical Software Clarifying Statement, the US Food and Drug Administration (FDA) states that it “FDA does not require use of any specific software for statistical analyses” and that “the computer software used for data management and statistical analysis should be reliable.” Observing differences across languages can reduce the analyst’s confidence in reliability and, by understanding the source of any discrepancies, one can reinstate confidence in reliability.

Motivation

The goal of this project is to demystify conflict when doing QC and to help ease the transitions to new languages and techniques with comparison and comprehensive explanations.

Methods		R	SAS	Comparison
Summary Statistics	Rounding	R	SAS	R vs SAS
	Summary statistics	R	SAS	R vs SAS
General Linear Models	Students t-test		SAS	

On this page

[Introduction](#)

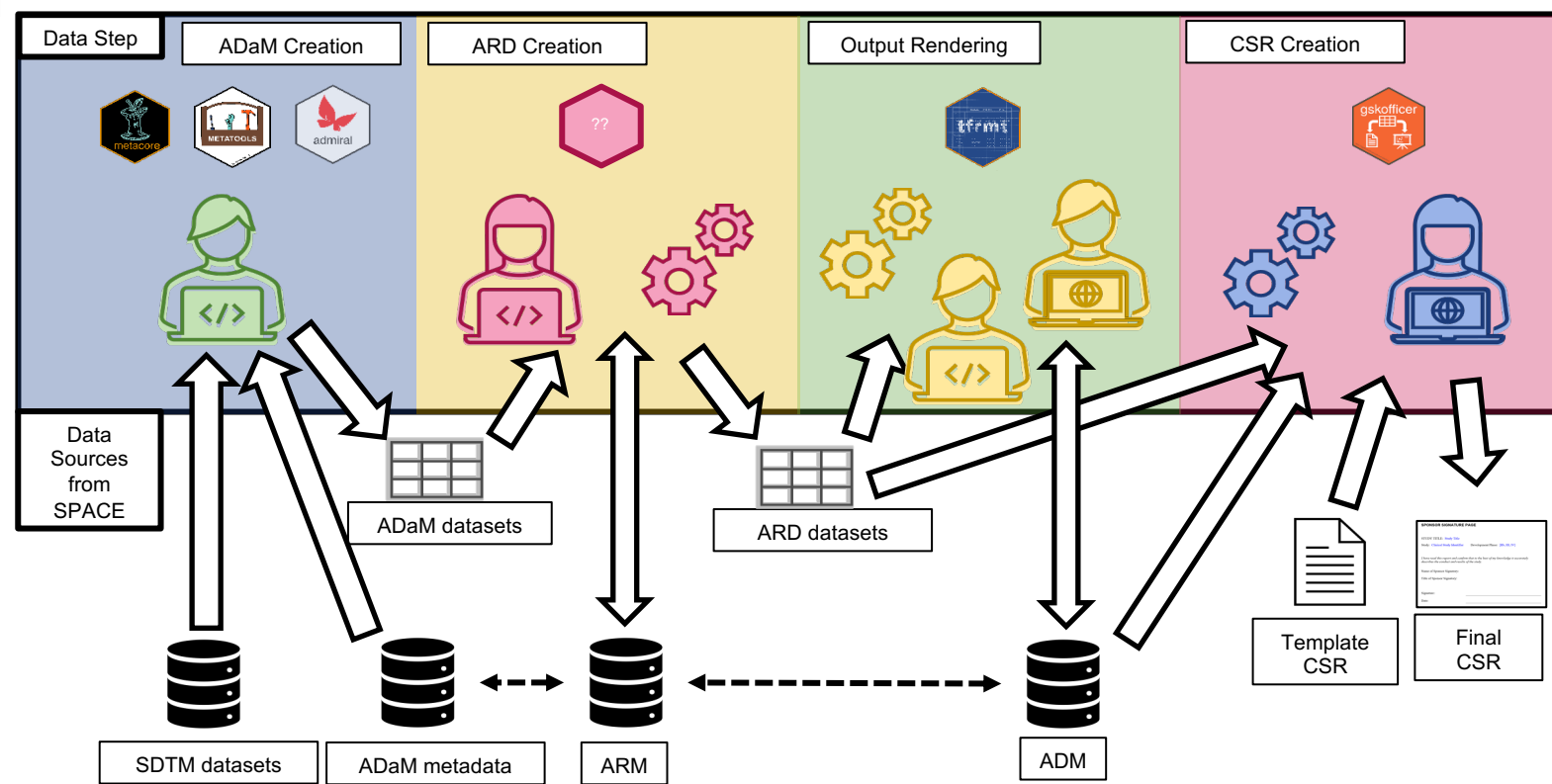
Motivation

Tooling is available

Implementing the Pharmaverse



- An end-to-end pipeline is available via the *nest* suite of tools
- ARD-centric pipeline one package away
- Companies already exploring CDISC 360 tooling
- LLM's accelerating learning



See <https://pharmaverse.org/> for more tools

Into the unknown...

Some internal questions / challenges discussed

- Wrong destination?
 - “It’ll never catch on”
 - Force majeure – LLMs?
- Challenges we expect to face in delivering the pipeline
 - Unsure how global regulators / HA’s will react
 - Our tools are not mature enough
 - Open-source skills are (currently) more expensive to hire
 - Existing staff not ready / not on board
 - FSP’s are not ready



Summary

COMMITMENT 1: An Open-Source Backbone



COMMITMENT 2: Parity by end 2025

