R Consortium's R-based Test Submission Package for FDA Evaluation

Ning Leng
Genentech, A Member of the Roche Group

Background: Regulatory Submissions to the Health Authorities

- To support the market approval of a treatment, sponsors will submit a comprehensive package of materials to health authorities to evaluate a product's efficacy and safety
    - Include analysis reports, datasets and software programs
- In addition to reviewing reports, HA reviewers also reproduce key analysis results
    - Review submitted data/code, independent programming
- Proprietary software has been widely used, due to reproducibility and quality considerations
    - In recent years, there are increasing interests in embracing open source software

The R Consortium
R submission Working Group:

A cross industry collaboration to improve open-source language usage in a regulatory setting

The R Consortium
R submission Working Group:

Our Mission

- Easier R-based clinical trial regulatory submissions today
    - by showing open examples of using current submission portals
- Easier R-based clinical trial regulatory submissions tomorrow
    - by collecting feedback and influencing future industry and agency decisions on system/process setup

Open to anyone who is interested in contributing!

https://rconsortium.github.io/submissions-wg/

Slack

Open-source language based submission to FDA: The Challenges

Reproducibility

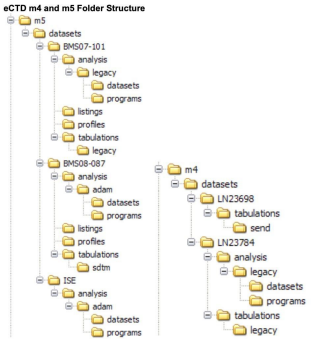
e.g. ensure reviewers access to the same package versions

Submission of R packages which are not publicly available

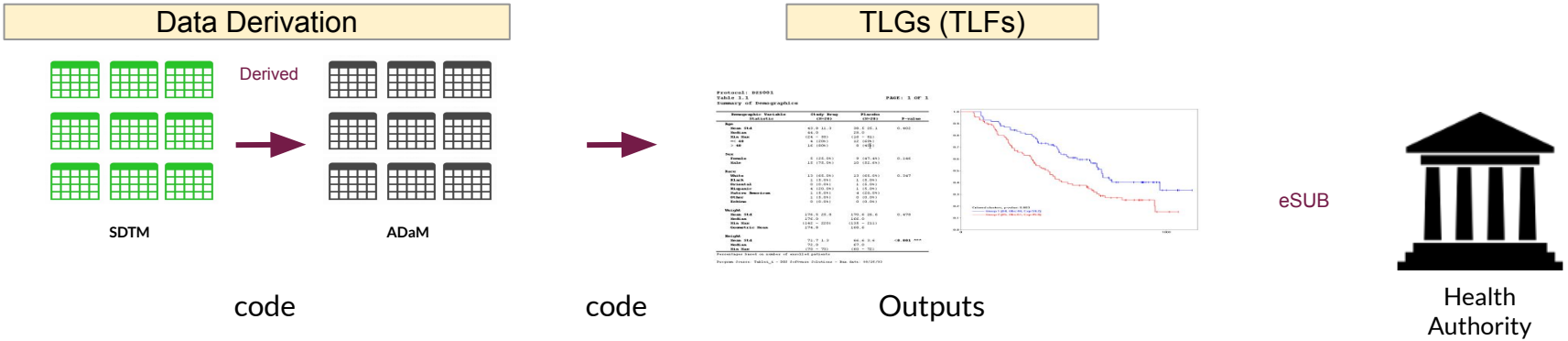
(FDA eCTD specifications, 2022, Source)

Fit into the well-established FDA required file structure (eCTD)

And ensure relevant information are easily findable by the reviewers



R Consortium Submissions Working Group: Pilot Overview



Pilot 1: TLG codes
Pilot 2: Static TLGs + shiny component

Pilot 3: TLG + ADaM codes
Pilot 4: deliver via webassembly/container technology
Pilot 5: alternative data format than xpt (json)

Submission Pilots to FDA: Current Status

Pilot 1 (finished)

FDA U.S. FOOD & DRUG ADMINISTRATION

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, AIS Hye Soo Cho -S
Supervisor	Maria Matilde Kam, AIS Maria Matilde S. Kam -S
Date(s):	March 10, 2022
Objectives of the submission	To test and support R-based clinical trial application submission
Location of datasets and programs	\\cdsesub3\evsprod\BLA111111\0002
Reviewed tables and figures	Table 14-2.01, Table 14-3.01, Table 14-3.02, Figure 14-1

Summary

- An FDA analyst was able to complete the following tasks:
  - Receive electronic submission package in eCTD format
  - Reconstruct and load the submitted proprietary R package
  - Install and load open-source packages used in this submission
  - Reproduce the analysis results
  - Share potential improvements to the submission deliverable and processes via a written communication
- FDA agrees that the initial phase of the R Pilot submission has been completed.
- For future reference, FDA suggest calculating 95% confidence intervals in a consistent manner.

Pilot 2 (finished)

FDA U.S. FOOD & DRUG ADMINISTRATION

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 and programs	BLA111111\0005
Reviewed tables and figures	Demographic Table, Kaplan Meier (KM) plot for time to first dermatologic event (TTDE), Primary Table, Efficacy Table, and Visit Completion Table in the Shiny app

Pilot 2 R Shiny Application Review

R Consortium's R Submission Working Group submitted the Pilot 2 R Shiny application (app) in November 2022. The objective of this pilot submission was to test whether a Shiny app created with the R-language could be successfully incorporated into a submission package and deployed to FDA reviewers. The app was built using the same source data sets and analyses contained in the Pilot 1 submission. This app is supplemental to the analysis programs and analyses submitted in Pilot 1. It is recommended that sponsors continue to follow the Study Data Technical Conformance Guide (SDTCG), and that any Shiny apps be provided as exploratory supplements rather than as replacements of required statistical programs, tables, listings, or figures. Note that the app is a supplementary material and does not replace any of R programs submitted in the Pilot 1. In future R based FDA submission, a Shiny app should be supplemental and should not replace any of analysis programs. Figures 1 - 7 show snapshots of each individual tab of the Pilot 2 Shiny App. Figure 1 illustrates the app information and provides a list of outputs with brief explanation. Figure 2 shows the relationship between each tab of Pilot 2 and previously submitted analysis from the Pilot 1 as well as a step-by-step example of how to use filters for the Kaplan

Pilot 3 (finished)

FDA U.S. FOOD & DRUG ADMINISTRATION

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 3 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: 2024.08.06 12:30:05 -04'00' Youn Kyeong Chang Youn Kyeong Chang -S Digitally signed by Youn Kyeong Chang -S Date: 2024.08.06 12:34:59 -04'00'
Secondary Reviewer	Paul Schuette Paul H. Schuette -S Digitally signed by Paul H. Schuette -S Date: 2024.08.08 11:29:59 -04'00'
Supervisor	Maria Matilde Kam Maria M. Kam -S Digitally signed by Maria M. Kam -S Date: 2024.08.08 12:14:05 -04'00'
Date(s):	August, 2024
Objective of the submission	Utilize R to produce Pilot 1 ADaM (Analysis Data Model) datasets from SDTM (Study Data Tabulation Model) datasets and generate Pilot 1 tables, listings, and figures (TLFs) using Pilot 3 R derived ADaM datasets
Location of datasets and programs	BLA111111\0006 and 0007
Reviewed tables and figures	Demographic Table, Primary Table, Efficacy Table, and Kaplan Meier (KM) Figure using R generated ADaM datasets

Pilot 3 Application Review

The R Consortium's R Submission Working Group submitted Pilot 3 (using R to derive Pilot 1 ADaM datasets from SDTM datasets) in August 2023. The objective of this pilot submission is to use R not only for analysis and visualization but also for data preparation in a regulatory submission to the FDA. The applicant used R to transform and manipulate SDTM datasets into ADaM datasets, and to produce the four analyses from Pilot 1. Note: in Pilot 1, the SDTM datasets were not included in the submission, and the applicant used the SAS derived ADaM datasets.

Pilot 3 was developed in R Posit Cloud using R version 4.2.3 on a Linux platform, and FDA reviewed it in RStudio Desktop using R version 4.2.3 on a Windows platform. The ADaM datasets as well as analysis results between Pilot 1 and Pilot 3 were expected to be identical, and no major discrepancies were observed. The analyses replicated by the FDA review team can be found in the Analysis Results Replication by FDA section below.

Pilot 4 and 5 in progress

Pilot 3 Project Scope

Data and analysis scope:

- SDTM Data from CDISC pilot 1 - publicly available, simulated data
  - Focus of this pilot was on ADaM and TLF generation sourcing the SDTM from CDISCPILOT01.
- 5 ADaMs
  - ADSL, ADAE, ADLBC, ADADAS, ADTTE
- 4 TLFs - 3 tables, 1 figure
  - This pilot will provide R scripts for ADaM and TLF generation

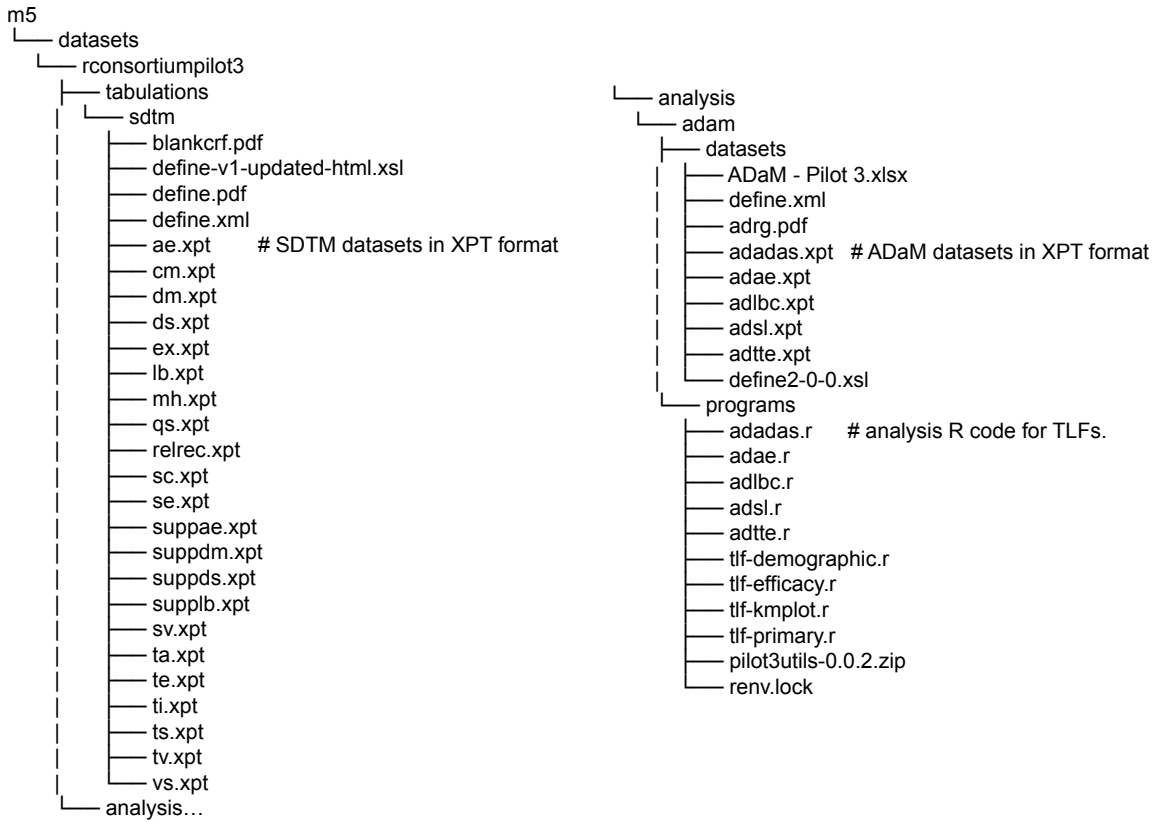
e-Sub package:

- m1

```
└── us
    ├── cover-letter.pdf  # Submission cover letter
    └── report-tlf-pilot3.pdf    # Submission TLFs
```

Pilot 3 Project Scope

e-Sub package:

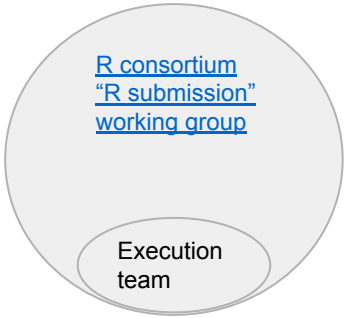


Pilot 3 Project Scope

Key evaluation aspects:

- For Submitter
  - Install and load proprietary R package {pilot3utils}, from a .zip file and run R scripts for ADaMs and TLFs from https://github.com/RConsortium/submissions-pilot3-adam-to-fda
  - Preparation of R-based submission materials (AD Reviewer's Guide, etc.)
- For FDA staff
  - Reproduce the ADaMs and the corresponding analysis result TLGs by installing the proprietary package, {pilot3utils} / retrieving open source packages with specific versions using {renv} and running the submitted R scripts for ADaMs and TLFs as instructed in the adrg.pdf.

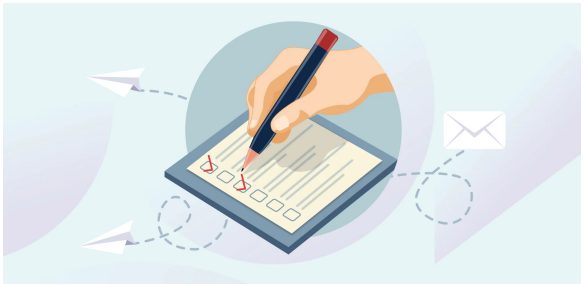
To provide feedback and/or review comments. **The Submission pilot 3 execution team.**



Joel Laxamana-Roche Author, maintainer Thomas Neitmann-Roche Author Phanikumar Tata-Syneos Author Robert Devine-J&J Author	Steven Haesendonckx-J&J Author Yutong Liu-Moderna Author Lei Zhao-Roche Author	Kangjie Zhang-Bayer Author Nicole Jones-Merck Author Benjamin Wang-Merck Author	Dadong Zhang-Illumina Author Benjamin Straub-GSK Author Declan Hodges-GSK Author

Working group remit: Beyond Submission Pilots

Conduct surveys and interviews on hot topics



Example blog: how is the R environment release managed across different healthcare organizations?

- Stability & reproducibility via version-controlled environments.
- Risk-based package assessments (testing, documentation, author reputation).
- Hybrid approaches (automation + manual review) for validation.
- Flexibility via tools like renv or rolling package updates within minor versions.

Source Control Validation blog - individual implementations

Roche

- Uses **container-based, bi-annual releases** (April & September) aligned with R major versions.
- Automated **risk assessment** of packages (test coverage, documentation, author reputation, etc.).
- Validated packages are published in a **continuously updated repository** per R minor version. Open-sourced **theValidator**.

Eli Lilly

- Updates **only after new R & Bioconductor major releases**; packages frozen post-deployment.
- Permits **only CRAN/Bioconductor packages** in the central library.
- Uses `renv` **for project-specific environments** to avoid central library disruptions.
- Balances **automation + risk assessment** for new package requests.

GSK

- Releases "**frozen R environments" every 6–12 months**, preferring stable R versions (e.g., 4.3.1).
- Assesses **internal & external packages** uniformly (author credibility, testing, documentation).
- Re-evaluates packages for **substantial updates** before allowing version changes.
- Ensures **reproducibility** via fixed package/R versions.

Pfizer

- **Annual R releases** (targeting stable versions like R-x.y.1), with **bi-annual package updates**.
- Uses **CRAN snapshots** for reproducibility while balancing stability & access to new packages.
- Validates and deploys **R containers every 6 months** (R version update yearly, packages mid-year).

Beyond Submission Pilots

Explore the role of AI/LLM to streamline R based submission process

Manually generated pilot 3 ADRG

Program Name	Output Name	Analysis Datasets & Variables	Selection Criteria
tlf-efficacy.r	tlf-efficacy-pilot3.rtf	ADSL.STUDYID ADSL.USUBJID ADSL.ITTFL ADLBC.TRTP ADLBC.TRTPN ADLBC.PARAMCD ADLBC.AVISITN ADLBC.BASE ADLBC.AVAL ADLBC.CHG	STUDYID== "CDISCPILOT01" Population: ADSL.ITTFL == "Y" & ADLBC.TRTPN in (0, 81) & ADLBC.PARAMCD == "GLUC" & ADLBC.AVISITN is not missing Treatment Groups: ADLBC.TRTPN Placebo Xanomeline High Dose
tlf-kmplot.r	tlf.kmplot-pilot3.pdf	ADSL.STUDYID ADSL.USUBJID ADSL.SAFFL ADSL.TRT01A ADTTE.STUDYID ADTTE.USUBJID ADTTE.PARAMCD ADTTE.AVAL ADTTE.CNSR	STUDYID== "CDISCPILOT01" Population: ADSL.SAFFL == "Y" Treatment Groups: ADSL.TRT01A Placebo Xanomeline Low Dose Xanomeline High Dose Parameters: ADTTE.PARAMCD == "TTDE"

LLM generated

script	output	Analysis Datasets & Variables	selection criteria
tlf-demographic.r	tlf-demographic-pilot3.out	ADSL.STUDYID; ADSL.ITTFL; ADSL.TRT01P; ADSL.AGEGR1; ADSL.RACE; ADSL.AGE; ADSL.HEIGHTBL; ADSL.WEIGHTBL; ADSL.BMIBL; ADSL.MMSETOT	ADSL.STUDYID == "CDISCPILOT01"; ADSL.ITTFL == "Y"
tlf-efficacy.r	tlf-efficacy-pilot3.rtf	ADSL.STUDYID; ADSL.USUBJID; ADSL.ITTFL; ADLB.STUDYID; ADLB.USUBJID; ADLB.TRTPN; ADLB.PARAMCD; ADLB.AVISITN; ADLB.CHG; ADLB.BASE; ADLB.TRTP; ADLB.AVAL	ADSL.ITTFL == "Y"; ADLBC.TRTPN %in% c(0, 81); ADLBC.PARAMCD == "GLUC"; !is.na(ADLBC.AVISITN); ADLBC.AVISITN == 20; !is.na(ADLBC.CHG); !is.na(ADLBC.BASE); ADLBC.AVISITN == 0
tlf-kmplot.r	tlf-kmplot-pilot3.pdf	ADSL.SAFFL; ADSL.STUDYID; ADSL.USUBJID; ADSL.TRT01A; ADTTE.PARAMCD; ADTTE.AVAL; ADTTE.CNSR; ADTTE.PARAM	ADSL.SAFFL == "Y"; ADSL.STUDYID == "CDISCPILOT01"; ADTTE.PARAMCD == "TTDE"; ADTTE.STUDYID == "CDISCPILOT01"
tlf-primary.r	tlf-primary-pilot3.rtf	ADADAS.EFFFL; ADADAS.ITTFL; ADADAS.PARAMCD; ADADAS.ANL01FL; ADADAS.TRTP; ADADAS.AVAL; ADADAS.AVISITN; ADADAS.CHG; ADADAS.USUBJID; ADADAS.TRTPN; ADSL.TRT01P	ADADAS.EFFFL == "Y"; ADADAS.ITTFL == "Y"; ADADAS.PARAMCD == "ACTOT"; ADADAS.ANL01FL == "Y"; ADSL.EFFFL == "Y" & ADSL.ITTFL == "Y"; ADADAS.AVISITN == 0; ADADAS.AVISITN == 24

R Consortium Collaboration with Global Groups

One More Step Forward: The R Consortium Submission Working Group's Presentation to Swissmedic on Regulatory Submission using R and Shiny

On January 30, 2024, the R Consortium Submission Working Group made a presentation to Swissmedic in Bern, Switzerland, with 10 attendees in person and 50 online.

Pharma RUG: The Rise of R in China's Pharmaceutical Industry

PharmaRUG, China organizer Joe Zhu, spoke with the R Consortium about the growing R community and the increasing use of R in the pharmaceutical industry in China. The group has…

R/Adoption Series:The Adoption Of R in Japan's Pharma Industry Confirmation

Look forward to the continued collaboration with JPMA!

Call for Collaboration

The best time to join the journey was 2 years ago.
The second best time is now.

R consortium **https://rconsortium.github.io/submissions-wg/**

oak **https://www.cdisc.org/oak**

pharmaverse admiral, NEST (as part of pharmaverse)
https://pharmaverse.org/

open stats ware **http://openstatsware.org**


Thank
You