

Johnson & Johnson's Hybrid R Submission Strategy: Successful story of the R submission to the FDA

Alicia Humphreys, Linshan Yuan

PHUSE SDE, 23 July 2025



Highlights

Johnson & Johnson Hybrid R/SAS strategy.

- R adaptive Pilot trial selection and regulatory submission.
- Validation of packages within the Johnson & Johnson's statistical computational environment.
- Hybrid R/SAS strategy

Two successful Submission applications to FDA.

- One Submission approved with designed timeline.
- One Submission ongoing with positive agency response.

Solutions and Challenges

- Real-world solution adapted from the R Consortium pilot projects

Contents

Pilot trial selection and regulatory submission.

Validation of packages within the Johnson & Johnson's statistical computational environment (with link provided)

Organization of R scripts

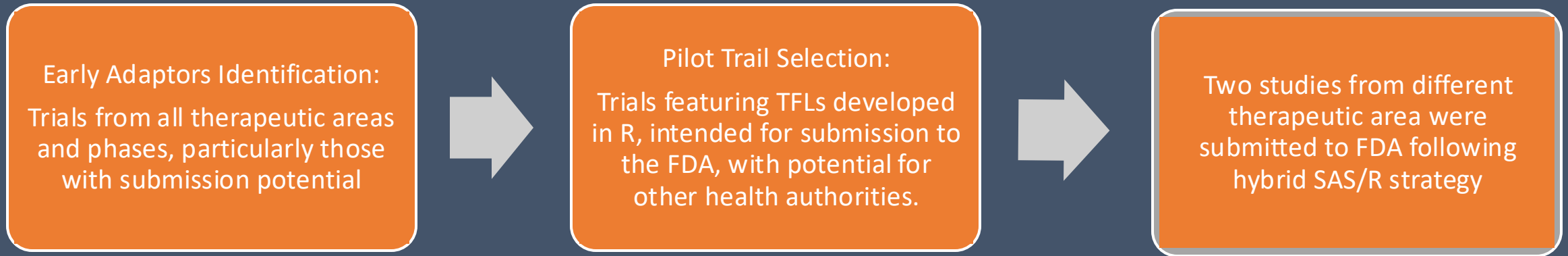
Verification of TFL results

ADRG streamline based on document from R Consortium pilot project

Conclusion and Challenges

R Consortium pilot: <https://github.com/RConsortium/submissions-pilot3-adam-to-fda>

Pilot trial selection & Hybrid SAS/R strategy



Hybrid SAS/R strategy

- ADaM datasets are created using SAS: Keeping current standards/knowledges reusability.
- TFLs are generated by R: more flexible and extensible visualization framework benefiting from a large, active and constantly evolving open-source community.
- Seamlessly moving workflow integration towards multiple languages while maintaining compatibility with existing tools that were built before with one language centric.
- Take advantage of the strengths offered by various programming languages.

Validation of packages within the Johnson & Johnson's statistical computational environment

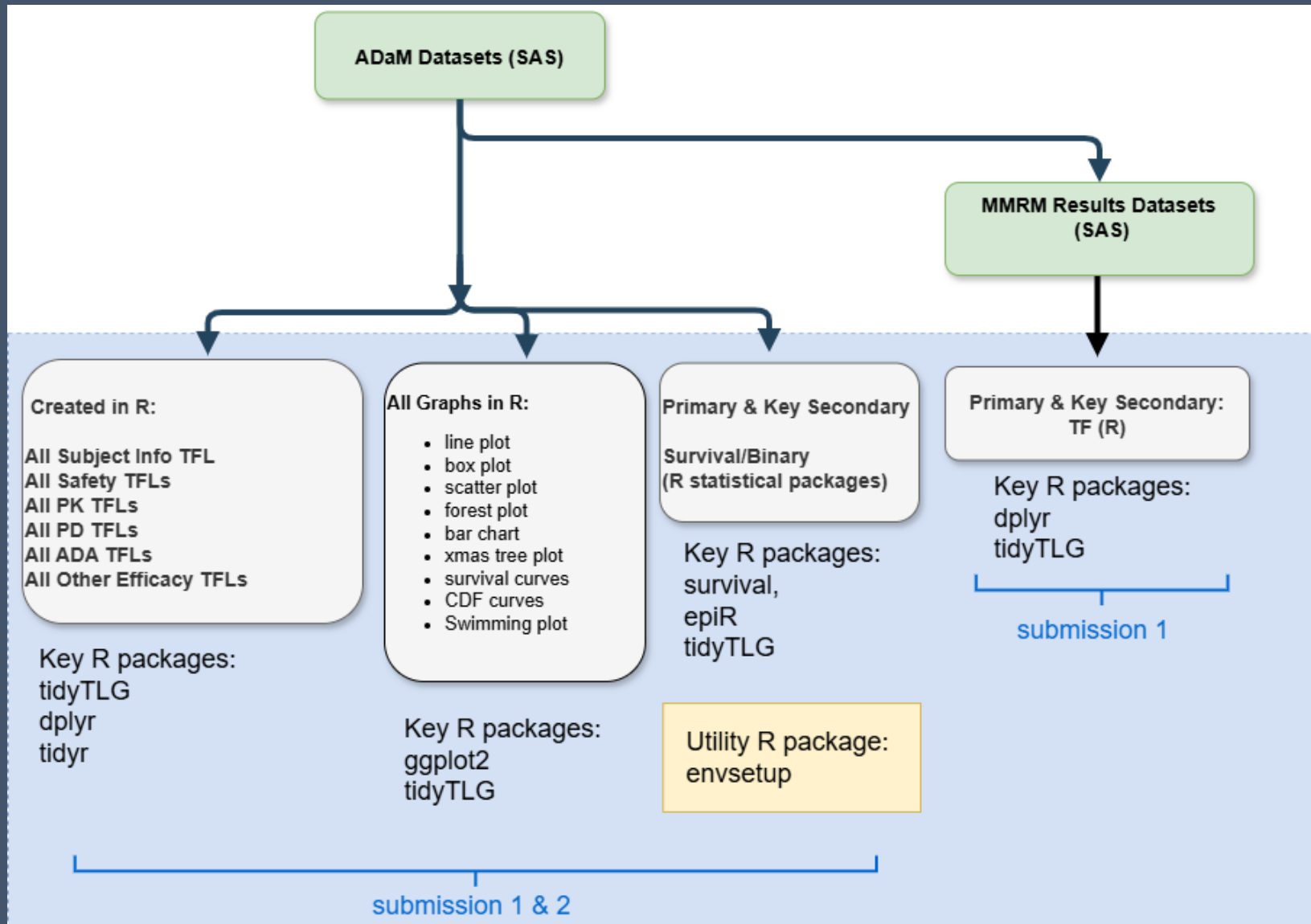
Using R in a GxP environment:

<https://www.youtube.com/watch?v=lWXqfuaxNL8>

Johnson & Johnson's Open Source Journey in R

<https://posit.co/webinar/johnson-johnsons-open-source-journey-in-r>

Organization of Hybrid R/SAS strategy



- All ADaM datasets are generated using SAS.
- All Tables/figures/Listings (TFLs) are created in R.
- All R packages are open-source .
- All R packages are validated following JNJ GxP.

Example of R script for analysis output

Follow the clear analysis logic to develop the R code

Applying column meta data →

Adding titles/footnotes →

```
#-----libraries-----  
|  
library(dplyr)  
library(haven)  
library(tidytlg)  
library(stringr)
```

```
#---update: specify the population flag-----  
  
popfl<-"RANDFL"  
adslfilter<- expr(popfl == 'Y' & TRT01PN %in% c(1,2))  
##define the ID of TLGs  
tblid<-"TSIDEM01"
```

```
#-----process data-----  
  
adsl<-read_sas(read_path(a_in,"adsl.sas7bdat"))%>%  
  mutate_if( is.character, ~ na_if(.x, c(''))) %>%  
  filter(eval(adslfilter))
```

```
#-----generate results-----  
|  
## AGE group  
t2<-adsl%>%  
  univar(colvar = "colnbr",  
        rowvar = "AGE",  
        row_header = "Age, years",  
        decimal = 0,  
        statlist = statlist(c("N", "MEANS", "MEDIAN", "RANGE")))
```

```
#-----bind together-----  
tbl<-bind_table(t1,t2,t2.1,t3,t4,t5,t6,t7,t8,t8.1,  
               column_metadata_file = read_path(dpspath, "column_metadata.xlsx"),  
               tbltype = "CL_DBTOT")
```

```
#-----output results-----  
|  
gentlg(huxme = tbl,  
      file = tblid,  
      orientation = "portrait",  
      opath=opath[[environ]],  
      title_file = read_path(dpspath, "titles.xlsx"))
```

Verification of R TFLs results: multiple languages

General summary table verification:

- Utilizing SAS standard macros/programs
- Keep qc results in rtf file
- SAS rtf and R rtf file comparison by extracting contents using R or SAS utilities functions.
- Allow multiple files Comparison in batch or through application.

Critical statistical method analysis verification:

- Utilizing SAS corresponding statistical procedures
- Review discrepancies with team whether they may affect the conclusion, taking appropriate actions.

Graph (figure) verification:

- Utilizing the ggplot2 built-in functions such as 'ggplot_build' to extract values from the figure created.
- Present the extracted values to R log file useful to cross check with the values in associated table.



At current stage, utilizing our established stable in-house SAS standards alongside the extensive functions available in R to accelerate the overall verification progress

ADRG for Hybrid SAS/R: Section 7

ADRG from Pilot 3 as template

7. Submission of Programs

7.1. Description

The sponsor has provided all programs for analysis results. They are all created on a Linux platform using R version 4.2.3.

7.2. ADaM Programs

The following table contains the list of programs that generate the analysis datasets in

7.3. Analysis Output Programs

The following table contains a list of programs that generate outputs used in the R

Program Name	Output Name	Analysis Datasets & Variables	Selection Criteria
tlf-demographic.r	tlf-demographic-pilot3.out	ADSL.STUDYID ADSL.TRT01P	STUDYID=="CDISCPIL0T01"

7.4. Proprietary R Packages

Proprietary R Package	Package version	Analysis Package Description
Pilot3	0.0.1	The objective of this utility package is to support the R

7.5. Open-source R Analysis Packages

Open-source R Analysis Package	Package version	Analysis Package Description
admiral		A toolbox for programming Clinical Data Interchange Standards Consortium (CDISC) compliant Analysis Data

ADRG in submission: adding description of major R functions

7. Submission of Programs

7.1 ADaM Programs

7.2 Statistical Analysis Results Datasets (RD Programs) & Associated Interim Programs

7.3 Analysis Output Programs

7.4 Macro Programs

7.5 Packages Used

7.6 Common R Functions Used

7.3 Analysis Output Programs

Program name	Output	Inputs	Macros and functions used
tsidem01b.r	Table TSIDEM01B	adsl, column_metadata.xlsx, titles.xlsx	base::c, base::factor, base::ifelse, base::is.na, base::lapply, base::paste0, base::rbind, base::source, dplyr::ensym, dplyr::filter, dplyr::mutate.

7.6 Common R Functions Used

For functions from open-source packages in CRAN or github, the full descriptions and usage can be found in more detail.

Function Name	Package name	Function Description
read_path	envsetup	Check each environment for the file and return the path to the first.

ADRG for Hybrid SAS/R: Appendix

The appendix is added to facilitate the R environment setup, including installation of required R packages. It provides guidance on how tables, figures, and listings developed in R can be executed to reproduce the analysis report in RTF format independently of the sponsor's computational system.

ADRG from Pilot 3 as template

Appendix

Appendix 1: Pilot 3 Installation and Usage

1. Installation of R and R Studio
2. Create a new R Studio project within the pilot3-files directory
3. Installation of R Packages
4. Set the paths to run the analysis programs
5. Execute analysis program

Appendix 2 : Common Issues

Common Issues

1. Package issue tracking
2. Knowledge sharing

Appendix 3 : QC Findings

ADRG in submission: update to reflect our system configuration

- Appendix 1: Set-up R environment and Analysis
1. Installation of R and RStudio
 2. Define Working Directory
 3. Set up the Working Environment
 4. Create a new RStudio Project to
 5. Reproduce the Working Environment
 6. Installation of Required Packages
 7. Create a renv-lock File
 8. Set the paths for analysis
 9. Execute the analysis programs

R Packages Installation & instruction for Window System

The overall R environment setup and package installations are provided to align with effective practices for the Windows system, aiming to maximize user accessibility. Simple solution using a configuration file is implemented whenever possible.

- Step 2,3,4 and 5: Provide instructions to create R project and setup the folder structure same as in the sponsor computing system.
- Step 6: All package installations are conducted through a single file, 'packages.txt', which contains all the necessary R packages mirrored from our container in the computing system embedded in Linux. The 'packages.txt' file is included in the submission package.

```
renv::init(bare = TRUE, repos = "https://packagemanager.posit.co/cran/2022-09-15")
pkgs <- read.csv(file.path(getwd(), "packages.txt"))
install.packages(pkgs$Package, repos = "https://packagemanager.posit.co/cran/2022-09-15")
```

- Step 7: Create a renv-lock file. This file capture packages versions to ensure the stability of the R environment recently configured on the reviewer's end.

```
renv::snapshot(type = "all", repos = "https://packagemanager.posit.co/cran/2022-09-15")
```

- Step 8: Source data libraries and paths utilized in R TFL scripts are configured via a file named 'envsetup.txt', which includes the folder path information.

m5 file structure in eCTD

FDA study data technical conformance guide (March 2025)

```
\---analysis
  +---adam
    +---programs
      adsl-sas.txt
      adae-sas.txt
      adlb-sas.txt
      tsidem01-r.txt
      tsfae01-r.txt
      toc-sasprograms.pdf
      toc-sasmacros.pdf
      toc-rprograms.pdf
      toc-rfunctions.pdf
    \---datasets
      adae.xpt
      adlb.xpt
      adsl.xpt
      adrg.pdf
      define.pdf
      define2-0-0.xml
      define.xml
  \---misc
```

4.1.2.10 Software Programs

Sponsors should provide the source code used to create all ADaM datasets, tables, and figures associated with primary and secondary efficacy analyses. Sponsors should submit **source code in single byte ASCII text format**. Files with MS Windows executable extensions (.cmd, .com, and .exe) should NOT be submitted. For a list of acceptable file extensions, refer to the document entitled *Specifications for File Format Types Using eCTD Specifications*.³²

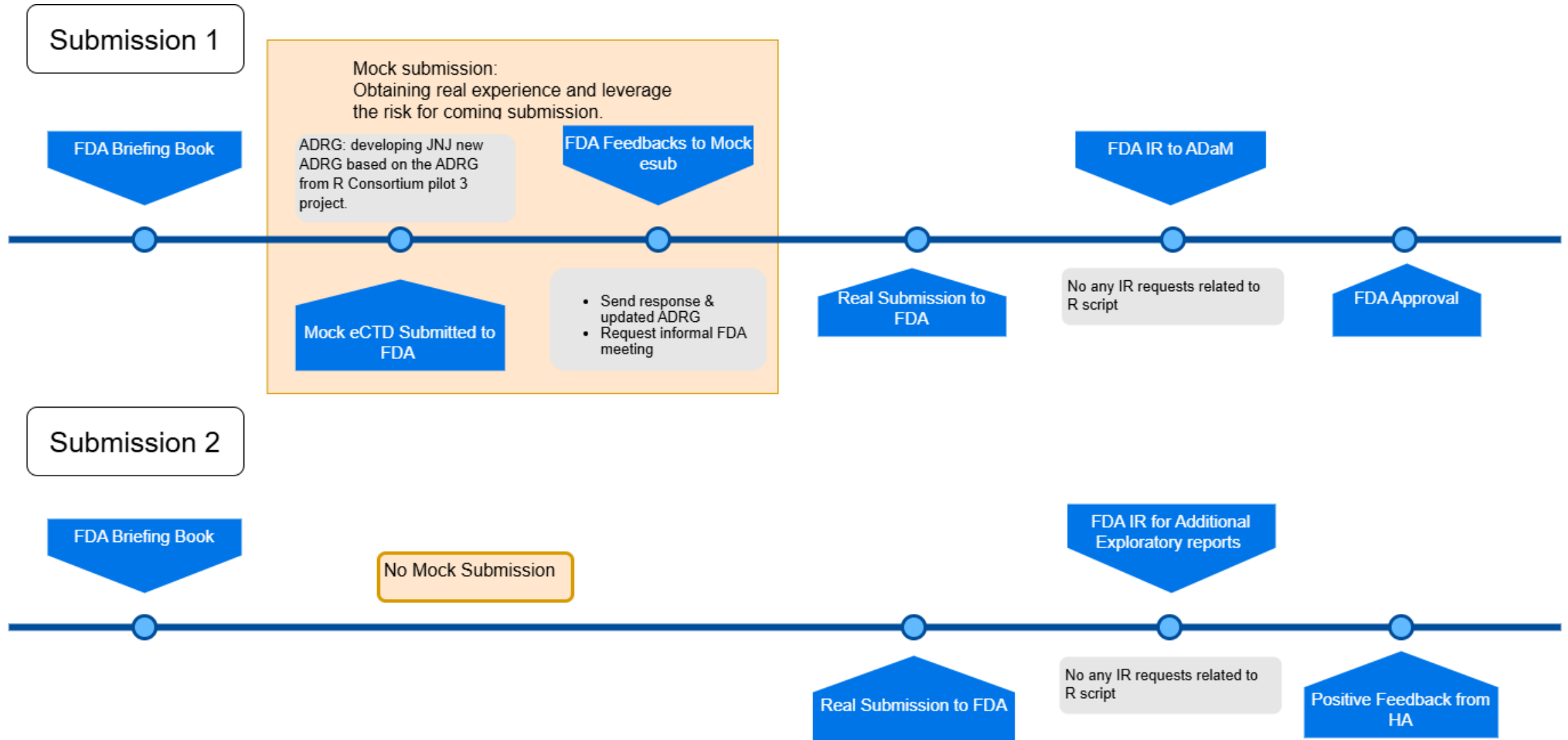
Furthermore, sponsors should submit the source code used to generate additional information included in Section 14 CLINICAL STUDIES of the Prescribing Information,³³ if applicable. The specific software utilized (version and operating system) should be specified in the ADRG.

- To facilitate hybrid SAS/R submissions, R scripts are converted similarly to SAS program files while maintaining the original naming convention. The file extension is appended with a suffix indicating the original programming language (SAS or R) before the .txt extension. For example, adsl.sas becomes adsl-sas.txt, and tsidem01.r becomes tsidem01-r.txt.
- Separated toc pdf files are created for R TFL scripts: toc-rprograms.pdf, toc-rfunctions.pdf

Interaction with regulatory agencies

	Submission 1	Submission 2
Briefing book	• ADaM datasets in xpt • SAS & R scripts in txt format • ADRG: providing R packages/version information • Mock eCTD requested and accepted	• ADaM datasets in xpt • SAS & R scripts in txt format • ADRG: providing R packages/version information
Mock eCTD	• Dummy ADaM datasets in xpt • 5 R table/figure scripts in txt format • ADRG: new JNJ ADRG template with R packages/version information and recreation of R environment. • Mock eCTD validated and submitted through FDA gateway	
Real submission	• All ADaM datasets in xpt • All MMRM results datasets in xpt • All ADaM SAS macros/programs in txt • Over 200 R TFLs scripts in txt • ADRG with study related R sections • Validated eCTD submitted to the FDA	• All ADaM datasets in xpt • All ADaM SAS macros/programs in txt • Over 300 R TFLs scripts in txt • ADRG with study related R sections • Validated eCTD submitted to the FDA

Milestones in Two Submissions



Conclusions

- ❑ Adhere to open-source R packages to avoid extra work for the submission, simplify the R setup process for reviewers with more transparent.
- ❑ R packages used to generate the analysis results are validated following GxP.
- ❑ Enhancing Reliability of R Analysis Results and building trust to reviewer:
 - Detailed explanations of program logic
 - R session information saved in R log file
 - Results being validated
- ❑ ADaM with the traceability of one step away to TFL: minimized of customized functions developed to handle additional derivations or data manipulations.
- ❑ ADRG incorporate clear R packages information including R environment set up guidance by taking Consortium FDA pilot projects and other R submission cases as reference.
- ❑ Using single file which contains all the required R packages and included in the submission package. It is to simplify the installation process for the reviewer. The renv.lock file is required to create after installation serves to stabilize and lock the project environment.
- ❑ Meet the requirements of eCTD file structure and file formats following the latest FDA Study Data technical conformance guide and passed validation.

Challenges for R Submission

❑ Challenges/Findings, from R Consortium & internal testing/mock submissions

- Replicating the sponsor's R environment.
- Different operating systems require different settings and generate different warnings
→in the 'renv' set up, different options might need to be selected to make it work
- Different R versions don't perform similarly,
- Not ideal to switch R versions in Rstudio for every submission
→'renv' package is part of the solution to reproduce the R environment but creates challenges when transferred between different operating systems.

❑ Recommendations/Exploration

- Add the question for R based submission in the pre-submission briefing book,
- Use CRAN or a curated repo for sourcing packages (github is allowed),
- Use standard packages and minimize dependencies on sponsor developed packages,
- The required R packages are installed using a single package list file. The renv.lock file is created upon completion of the R package installation to lock down the R environment .
- Provide thorough documentation and detailed comments in ADRG.

Reference

R packages validation process in Johnson&Johnson

❑ Using R in a GxP environment: <https://www.youtube.com/watch?v=lWXqfuaxNL8>

Johnson&Johnson's Open Source Journey in R

❑ <https://posit.co/webinar/johnson-johnsons-open-source-journey-in-r/>

R Consortium Pilot Projects

❑ R Consortium Pilot 1, <https://github.com/RConsortium/submissions-pilot1-to-fda>

❑ R Consortium Pilot 2, <https://github.com/RConsortium/submissions-pilot2-to-fda>

❑ R Consortium Pilot 3, <https://github.com/RConsortium/submissions-pilot3-adam-to-fda>

❑ R Consortium Pilot 4, <https://github.com/RConsortium/submissions-pilot4-container>

❑ R Consortium Pilot 5, <https://github.com/RConsortium/submissions-pilot5-datasetjson>

The opinions in this presentation are our own
and do not necessarily reflect the views and
policies of Johnson&Johnson