R Consortium R Submission Working Group Pilot 3

Health Authority Request

Response to FDA Information Request

April 12, 2024

Food and Drug Administration Center for Drug Evaluation and Research 5901-B Ammendale Road Beltsville, MD 20705-1266

Re: Response to FDA's Statistical Review and Evaluation Summary of the R Consortium R submission Pilot 3 eCTD submission

Dear Sir/Madam:

On February 2, 2024, the R Consortium R submission Working Group received FDA's Statistical Review and Evaluation Summary of the R Consortium R submission Pilot 3 submission. The working group would like to thank FDA staff for the comprehensive review and thoughtful recommendations. We are thrilled to learn that the FDA staff was able to reproduce the ADaM (Analysis Data Model) datasets as well as the analysis outputs using the R programs and the proprietary R package submitted through the eCTD portal.

The purpose of this letter is to provide the working group's responses to FDA's Statistical Review and Evaluation Summary of the R Consortium R submission Pilot 3. The following information is provided in response to the findings from FDA Staff's review :

Agency Comment 1

The agency found issues switching between R versions. The FDA would like the wg [working group] to explore the impact of different versions of R, RStudio and RTools, as well as Linux vs. Windows

The R Consortium R Submission Working Group Response

It is recommended by the Working group to download and install the R and R Studio versions specified in the ADRG. Switching to other versions, not specified, could most likely be the cause of issues when running the R ADaM and analysis output programs. Certain versions of the R packages used in these analysis programs only correspond to the specified R version.

Regarding RTools, referencing the installation instructions on its website,

"Rtools42 is only needed for installation of R packages from source or building R from source. R can be installed from the R binary installer and by default will install binary versions of CRAN packages, which does not require Rtools42.

Moreover, online build services are available to check and build R packages for Windows, for which again one does not need to install Rtools42 locally. The Winbuilder check service uses identical setup as the CRAN incomming packages checks and has already all CRAN and Bioconductor packages pre-installed."[1]

As discussed in the R Consortium R Submission working group, *"It appears that the most practical path forward as to what preemptive measures sponsors can take to match submission environments with the FDA test environment is to work with the FDA to lock down the environment as best as can be done just prior to submission."*[2]

For Linux vs Windows, many others in the R community have asked similar questions. *"R programming language is well-suited for both Windows and Linux operating systems. It is an open-source programming language, so it can be used on a variety of platforms. The choice between Windows and Linux depends on your specific needs, familiarity with the operating systems, and the tools and packages you plan to use with R. Both platforms have their own advantages and it ultimately comes down to personal preference and the specific requirements of your project."*[3]

For the Pilot 3 project team, due to cross-industry collaboration, Linux was the best option for us to work in allowing us work in a stable R environment.

1. https://cran.r-project.org/bin/windows/Rtools/rtools42/rtools.html

2. https://rconsortium.github.io/submissions-wg/minutes/2024-02-02/

3. https://www.quora.com/Is-the-R-programming-language-better-suited-for-windows-or-Linux

Agency Comment 2

Differences between generated Pilot 2 ADaM and CDISC data sets : attributes, types (e.g. integer vs. double), and NA vs. NULL

The R Consortium R Submission Working Group Response

The Pilot 3 project team explored all differences between the original CDISC datasets generated in SAS vs the Pilot 3 datasets generated in R. More detailed findings are documented in the ADRG in the section 'QC findings'. To address the agency's concerns, we explored these specific discrepancies that are present that could impact the analysis results.

The team used three different packages:

```
library(arsenal)
library(diffdf)
library(waldo)

path <- list(sdtm = "~/submissions-pilot3-adam/submission/sdtm",
             adam = "~/submissions-pilot3-adam/submission/adam",
             cdisc = "~/submissions-pilot3-adam/adam")
```

ADSL

```
adsl <- read_xpt(file.path(path$adam, "adsl.xpt"))
adsl_cdisc <- read_xpt(file.path(path$cdisc, "adsl.xpt"))

arsenal::comparedf(adsl, adsl_cdisc)
diffdf::diffdf(adsl, adsl_cdisc, keys = c("STUDYID", "USUBJID"))
waldo::compare(adsl, x_arg = "old", adsl_cdisc, y_arg = "new")


`attr(old, 'label')` is a character vector ('Subject-Level Analysis Dataset')
`attr(new, 'label')` is absent
```

ADTTE

```
adtte <- read_xpt(file.path(path$adam, "adtte.xpt"))
adtte_cdisc <- read_xpt(file.path(path$cdisc, "adtte.xpt"))

arsenal::comparedf(adtte, adtte_cdisc)
diffdf::diffdf(adtte, adtte_cdisc, keys = c("STUDYID", "USUBJID"))
waldo::compare(adtte, x_arg = "old", adtte_cdisc, y_arg = "new")
```

```
=================================================
 VARIABLE   ATTR_NAME   VALUES.BASE   VALUES.COMP
-------------------------------------------------
     AGE    format.sas     NULL           3
  AGEGR1    format.sas     NULL          $5
 AGEGR1N    format.sas     NULL           3
EVNTDESC    format.sas     NULL          $25
   PARAM    format.sas     NULL          $32
 PARAMCD    format.sas     NULL          $4
    RACE    format.sas     NULL          $32
   RACEN    format.sas     NULL           3
   SAFFL    format.sas     NULL          $1
     SEX    format.sas     NULL          $1
-------------------------------------------------
```

ADAE

```
adae <- read_xpt(file.path(path$adam, "adae.xpt"))
adae_cdisc <- read_xpt(file.path(path$cdisc, "adae.xpt"))

arsenal::comparedf(adae, adae_cdisc)
diffdf::diffdf(adae, adae_cdisc, keys = c("STUDYID", "USUBJID", "AESEQ"))
waldo::compare(adae, x_arg = "old", adae_cdisc, y_arg = "new")
```

VARIABLE	ATTR_NAME	VALUES.BASE	VALUES.COMP
ADURN	label	Analysis Duration (N)	AE Duration (N)
ADURU	label	Analysis Duration Units	AE Duration Units
AOCCFL	label	1st Occurrence within Subject ...	1st Occurrence of Any AE Flag

ADLBC

```
adlbc <- read_xpt(file.path(path$adam, "adlbc.xpt"))
adlbc_cdisc <- read_xpt(file.path(path$cdisc, "adlbc.xpt"))

arsenal::comparedf(adlbc, adlbc_cdisc)
diffdf::diffdf(adlbc, adlbc_cdisc, keys = c("STUDYID", "USUBJID", "LBSEQ",
"PARAMCD", "AVISIT"))
waldo::compare(adlbc, x_arg = "old", adlbc_cdisc, y_arg = "new")
```

VARIABLE	CLASS.BASE	CLASS.COMP
ADT	Date	numeric
TRTEDT	Date	numeric
TRTSDT	Date	numeric

ADADAS

```
adadas <- read_xpt(file.path(path$adam, "adadas.xpt"))
adadas_cdisc <- read_xpt(file.path(path$cdisc, "adadas.xpt"))

arsenal::comparedf(adadas, adadas_cdisc)
diffdf::diffdf(adadas, adadas_cdisc, keys = c("STUDYID", "USUBJID", "QSSEQ",
"PARAMCD", "AVISIT"))
waldo::compare(adadas, x_arg = "old", adadas_cdisc, y_arg = "new")
```

```
=========================================================================================
VARIABLE  ATTR_NAME            VALUES.BASE                        VALUES.COMP
-----------------------------------------------------------------------------------------
ANL01FL    label              Analysis Flag 01                   Analysis Record Flag 01
ITTFL      label     Intent-To-Treat Population Flag  Intent-to-Treat Population Flag
-----------------------------------------------------------------------------------------
```

Any observed difference in variables is documented in QC findings, including NA versus NULL which we did not encounter. Any observed difference in attribute will not affect the analysis.

In conclusion, the team was unable to identify any differences that could have an impact on the analysis.

Agency Comment 3

In the eSub Package, the proprietary Pilot 3 package had the incorrect package name {pilot3} instead of {pilot3utils}. For one of the file names, it was renamed from Pilot 3.xlsx to adam-pilot-3.xlsx.

The R Consortium R Submission Working Group Response

The Pilot 3 package and the .xlsx files is now correctly named to {pilot3utils}[1] and adam-pilot-3.xlsx, respectively. These are referenced as so in the Pilot 3 ADaM and analysis output programs as well as the ADRG.

1. https://github.com/RConsortium/submissions-pilot3-utilities/blob/main/DESCRIPT ION

Agency Comment 4

File path - Relative vs Full path name. The location of a directory was specified using relative file path, where the Agency recommends specifying and using the full path.

The R Consortium R Submission Working Group Response

This has been updated in the ADRG with further detailed notes on how to specify the full file path upon execution of the R programs.

Agency Comment 5

Different primary output in Pilot 1 and Pilot 3

- There are discrepancies between Pilot 3 ADaM and CDISC datasets

- Discrepancies are noted in QC findings in ADRG

The R Consortium R Submission Working Group Response

Upon the Pilot 3 team's investigation, they noted that there are 818 available records in the QS domain. The CDISC ADADAS only brought in 799 records and imputed the rest, whereas Pilot 3 brought in all available 818 records into ADADAS and then imputed. When the Pilot 3 team adjusted by subsetting to ANL01FL='Y' records first before doing the LOCF imputation the results matched Pilot 1 and the discrepancy was resolved.[1]

1. https://github.com/RConsortium/submissions-pilot3-adam/pull/146

Kind regards,

The R Consortium R Submission Pilot 3 Project Team