R consortium

DEC 02, 2022

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

Re: R Consortium R Submission Pilot 3 ADaM

Dear Sir/Madam:

This letter serves as a proposal to the R Consortium R submission Pilot 3 (an ADaM extension to Pilot 1) focusing on a working example of R submission in eCTD format to the pharmaceutical industry in compliance with the FDA Electronic Submissions Gateway requirements. Based on the Pilot 1 submission package, Pilot 3 will plan to focus on the generation of ADaM datasets that were used as source for the tables and figures generated in Pilot 1. The Pilot 3 ADaM datasets will be generated using R. Upon submission to the FDA, we propose to the FDA Staff to review and reproduce the submitted R codes. More specifically, we would expect the FDA Staff to

- Receive an electronic submission package in eCTD format
- Reconstruct and load the submitted proprietary R package (i.e. "pilot3wrappers")
- Install and load open source packages used in this submission (i.e. including, but not limited to admiral, metacore, metatools, tidyverse, xportr)
- Reproduce the ADaM datasets and analysis results using the submitted R code and packages
- Confirm that the submitted ADaM dataset .xpt files can be opened in the FDA's installation of SAS
- Share potential improvements to the submission deliverables and processes via written communication

All data, code, material and communications from this pilot will be shared publicly.

On behalf of the R Consortium R Submission Working Group, we hope the R submission Pilot 3 can establish a working example to guide the industry for future submissions using the R language.

Analysis Dataset Description

The following provides detailed information for each analysis dataset we plan to include in the Pilot 3 submission, which were used to generate the outputs in Pilot 1. These ADaM datasets are ADSL, ADAE, ADTTE, ADADAS, and ADLBC.

ADSL - Subject Level Analysis Dataset
The subject level analysis dataset (ADSL) should contain required variables for demographics, treatment groups, and population flags. In addition, it will contain other baseline characteristics that were used in both safety and efficacy analyses. All patients in DM will be included in ADSL.

The following are the key population flags that will be used in the analyses for patients:
- SAFFL – Safety Population Flag (all patients having received any study treatment)
- ITTFL – Intent-to-Treat Population Flag (all randomized patients)

ADAE - Adverse Events Analysis Data

ADAE will contain one record per reported event per subject. Subjects who did not report any Adverse Events will not be represented in this dataset. The data reference for ADAE will be the SDTM

- AE (Adverse Events) domain and there is a 1-1 correspondence between records in the source and this analysis dataset. These records can be linked uniquely by STUDYID, USUBJID, and AESEQ.

- Events of particular interest (dermatologic) are captured in the customized query variable (CQ01NAM) in this dataset. Since ADAE is a source for ADTTE, the first chronological occurrence based on the start dates (and sequence numbers) of the treatment emergent dermatological events are flagged (AOCC01FL) to facilitate traceability between these two analysis datasets.

ADTTE - Time to Event Analysis Dataset

ADTTE will contain one observation per parameter per subject. ADTTE is specifically for safety analyses of the time to the first dermatologic adverse event. Dermatologic AEs are considered an adverse event of special interest. The key parameter that will be used for the analysis of time to the first dermatological event is with PARAMCD of "TTDE".

ADLBC - Laboratory Results Chemistry Analysis Data

ADLBC will contain one record per lab analysis parameter, per time point, per subject for lab chemistry parameters. These data will be derived from the SDTM LB (Laboratory Tests) domain. Two sets of lab parameters exist in ADLBC. One set contains the standardized lab value from the LB domain and the second set contains change from previous visit relative to normal range values. In some of the summaries the derived end-of-treatment visit (AVISITN=99) will also presented.

ADADAS - ADAS-COG Data

ADADAS will contain analysis data from the ADAS-Cog questionnaire, one of the primary efficacy endpoints. It will contain one record per subject per parameter (ADAS-Cog questionnaire item) per VISIT. Visits will be placed into analysis visits (represented by AVISIT and AVISITN) based on the date of the visit and the visit windows.

Submission of Programs

Description
The sponsor will provide all programs for analysis results. They will be created on a Linux platform using R version 4.2.1.

List of Analysis Dataset Programs
The following table contains the list of programs that we will prepare to generate the analysis datasets in Pilot 3. It shows the program file name, the analysis dataset name and the label of the analysis dataset.

Program Name	Analysis Dataset Name	Analysis Dataset Label
adsl.R	adsl.xpt	Subject Level Analysis Dataset
adas.R	adadas.xpt	ADAS-Cog Analysis Dataset
adlbc.R	adlbc.xpt	Lab Blood Chemistry Analysis Dataset
adae.R	adae.xpt	Adverse Events Analysis Dataset
adtte.R	adtte.xpt	AE Time to 1st Derm. Event Analysis

Analysis Output Programs

This pilot project may also submit potentially updated programs which generated the Tables and Figures from the Pilot 1 submission.

Pilot 3 Proposed Installation and Usage for FDA Staff

The proposed plan to install and execute the analysis dataset R programs, should follow all of the procedures below. Ensure that you note the location of where you will download the Pilot 3 eCTD submission files. For demonstration purposes, the procedures below will assume the transfer will be saved to this location: `C:\pilot3`.

In addition, we would propose to create a new directory, which should hold the unpacked Pilot 3 ADaM programs and files. For demonstration purposes, the procedures below will assume the new directory is this location: `C:\pilot3-files`.

Proposed Installation of R and Optional Software

We would like to propose to download and install R 4.2.1 for Windows from https://cran.r-project.org/bin/windows/base/old/4.2.1/R-4.2.1-win.exe . While optional, it is also recommended to prepare to view this Pilot 3 within the RStudio IDE. You can download RStudio for Windows by visiting https: //www.rstudio.com/products/rstudio/download/#download.

Proposed Installation of R Packages

A minimum set of R packages may be required to ensure the Pilot 3 analysis dataset files and programs are successfully unpacked and the custom package environment used for the planned R programs will be replicated correctly. The first packages to install are the `remotes` and `pkglite` packages:

```
install.packages(c("remotes", "pkglite"))

# install version 0.16.0 of the renv
package:
remotes::install_version("renv",
version = "0.16.0")
```

Ĺ Note

The console may display a warning message about Rtools being required to build R packages. However the packages required for this submission will not require custom compilation involving other languages like C++, hence the Rtools utility will not be required.

Extract Application Bundle

It is proposed to use the `pkglite` package to unpack the custom R package `r3pkg.txt` within the Pilot 3 eCTD submission transfer. This file is located in the following relative path within the eCTD transfer directory:

```
m5\datasets\rconsortiumpilot3\analysis\adam\programs\r3pkg.txt
```

Enter the following command in the R console to extract the Pilot 3 programs and files to the destination directory.

```
pkglite::unpack(
input =
"C:/pilot3/m5/datasets/rconsortiumpilot3/analysis/adam/programs
/r3pkg.txt",
output = "C:/pilot3-files"
)
```

The console should display messages of unpacking and writing files to the destination directory. Note that the procedure will create a sub-directory called `pilot3wrappers` in the destination directory. Take note of that particular directory path on your system, as you will use this in the remaining procedures. In this example, the directory is located in the following path:

```
C:\pilot3-files\pilot3wrappers
```

Initialize R Package Environment for ADaM generation

The dependencies for the ADaM generation will be managed by the `renv` R package management system. To bootstrap the customized R package library used to generate the analysis datasets, you will launch a new R session in the directory where you should have unpacked the R programs and source files in the previous step. You will use either of the following procedures depending on your R computing environment:

RStudio

Create a new RStudio Project within the `pilot3wrappers` directory:
1. Select `File -> New Project`
2. In the Create Project dialog box, choose **Existing Directory**
3. In the Create Project from Existing Directory dialog box, click the **Browse** button and navigate to the `pilot3wrappers` directory.
4. Once the location has been confirmed, click the **Create Project** button.

RStudio should refresh the window and automatically and install the `renv` package into the project directory. To complete the process of restoring the pilot 3 R packages, you will run the following command in the R console:

```
renv::restore(prompt = FALSE)
```

The package installation procedure may take a few minutes or longer depending on internet bandwidth.

Console

Launch a new R session in the `pilot3wrappers` directory of the unpacked application directory. By default, the R Gui interface on Windows will launch a new R session in your default Windows home directory (typically the **Documents** folder). Perform the following steps to ensure R is launched in the proper directory.

> Ĺ Note
>
> The procedure below assumes R 4.2.1 has been installed in a default location. If you are unsure of the full path to the R GUI executable on your system, you can find the location on your system by performing the following steps:
> 1. Open the Windows Start Menu and expand to show all applications.
> 2. Navigate to the R entry and expand the section such that all R program entries are visible.
> 3. Right-click the `R x64 4.2.1` entry and select `More -> Open file location`.
> 4. A new folder window will open with the shortcut `R x64 4.2.1` highlighted. Right-click this entry and select **Properties**
> 5. In the Properties window, copy the path specified in the **Target** text field. The portion of the text in quotations gives the full path to the `Rgui.exe` location on your system.

1. Open the Windows PowerShell program by searching for Windows PowerShell in the Windows Start menu.
2. Change the current directory to the `pilot3wrappers` directory by running the following command (substitute the `pilot3-files` location for your appropriate directory as needed):

```
Set-Location -Path "C:\pilot3-files\pilot3wrappers"
```

3. Launch the Windows R GUI in this session by running the following command:

```
C:\"Program Files"\R\R-4.2.1\bin\x64\Rgui.exe
```

The R GUI should launch and automatically install the `renv` package into the project directory. To complete the process of restoring the pilot R packages, run the following command in the R console:

```
renv::restore(prompt = FALSE)
```

The package installation procedure may take a few minutes or longer depending on internet bandwidth.

Planned Source SDTM from the CDISC SDTM/ADaM Pilot Project and Analysis Datasets generated from Pilot 3

Upon completion of proposed package installation, you should be able to import the source SDTM taken from the CDISC SDTM/ADaM Pilot Project and the ADaM data sets, both contained in the eCTD transfer. Only the ADaM data will be generated using R in this Pilot sourcing the CDISC SDTM Pilot Project data. The data files are located in the following relative path within the eCTD transfer directory:

```
m5\datasets\rconsortiumpilot3\analysis\sdtm\datasets
m5\datasets\rconsortiumpilot3\analysis\adam\datasets
```

You will run the following command in the R console (substitute the `pilot3` location for your appropriate directory as needed):

```
Pilot3wrappers::set_data_path("C:/pilot3/m5/datasets/rconsortiump
ilot3/analysis/sdtm/datasets")
```

Executing the R code to reproduce the Analysis Datasets

After importing the SDTM datasets, you should now be able to run the R code, which were developed to reproduce the Analysis Datasets for further review. The R programs will be located in the following relative path within the eCTD transfer directory:

```
m5\datasets\rconsortiumpilot3\analysis\adam\programs
```

To reproduce the analysis datasets generated in Pilot 3, rerun the following programs sourcing the SDTM datasets imported:

Program Name	Analysis Dataset Name	Analysis Dataset Label
adsl.R	adsl.xpt	Subject Level Analysis Dataset
adas.R	adadas.xpt	ADAS-Cog Analysis Dataset
adlbc.R	adlbc.xpt	Lab Blood Chemistry Analysis Dataset
adae.R	adae.xpt	Adverse Events Analysis Dataset
adtte.R	adtte.xpt	AE Time to 1st Derm. Event Analysis

Once the reproduction of the analysis datasets in R have been completed, you should then be able to use these data as source analysis datasets to reproduce the analysis outputs from the Pilot 1 programs. Please refer to the Pilot 1 submission here: https://github.com/RConsortium/submissions-pilot1-to-fda to run the Pilot 1 programs sourcing the analysis datasets generated from the Pilot 3 programs. Should there be any further potential updates needed to the Tables and Figures programs from Pilot 1, we will describe and include these updates in this Pilot 3 submission.