Analysis Data Reviewer's Guide

R Consortium

R Consortium R Submission Pilot 1

1. Introduction

1.1 Purpose

This document provides context for the analysis datasets and terminology that benefit from additional explanation beyond the Data Definition document (define.xml). In addition, this document provides a summary of ADaM conformance findings.

1.2 Study Data Standards and Dictionary Inventory

Standard or Dictionary	Versions Used
SDTM	SDTM v1.4/ SDTM IG v3.1.2
ADaM	ADaM v2.1/ ADaM IG v1.0
Controlled Terminology	SDTM CT as of 2011-12-09 ADaM CT as of 2011-07-22
Data Definitions	define.xml v2.0
Other standards (optional)	MedDRA v8.0

1.3 Source Data Used for Analysis Dataset Creation

The ADaMs we used to regenerate the outputs were the PHUSE CDISC Pilot replication ADaMs following ADaM IG v1.0. The ADaM dataset is publicly available at the PHUSE Github Repository (https://github.com/phuse-org/phuse-scripts/blob/master/data/adam/TDF_ADaM_v1.0.zip)

2. Protocol Description

2.1 Protocol Number and Title

Protocol Number: CDISCPilot1

Protocol Title: Safety and Efficacy of the Xanomeline Transdermal Therapeutic System (TTS) in Patients with Mild to Moderate Alzheimer's Disease

The reference documents can be found at https://bitbucket.cdisc.org/projects/CED/repos/sdtm-adam-pilot-project/browse/updated-pilot-submission-package/900172/

2.2 Protocol Design in Relation to ADaM Concepts

Objectives:

The objectives of the study were to evaluate the efficacy and safety of transdermal xanomeline, 50cm and 75cm, and placebo in subjects with mild to moderate Alzheimer's disease.

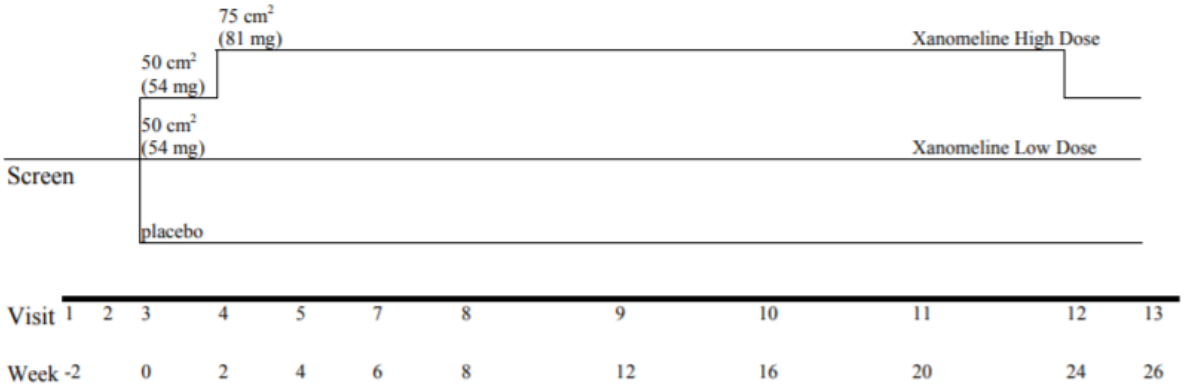
Methodology:

This was a prospective, randomized, multi-center, double-blind, placebo-controlled, parallel-group study. Subjects were randomized equally to placebo, xanomeline low dose, or xanomeline high dose. Subjects applied 2 patches daily and were followed for a total of 26 weeks.

Number of Subjects Planned:

300 subjects total (100 subjects in each of 3 groups)

Study schema:



3. Analysis Considerations Related to Multiple Analysis Datasets

3.1 Core Variables

Core variables are those that are represented across all/most analysis datasets.

Variable Name	Variable Description
USUBJID	Unique subject identifier
STUDYID	Study Identifier
SITEID	Study Site Identifier
TRTSDT	Date of First Exposure to Treatment
TRTEDT	Date of Last Exposure to Treatment
AGE	Age
AGEGR1	Pooled Age Group 1
AGEGR1N	Pooled Age Group 1 (N)
SEX	Sex
RACE	Race
RACEN	Race (N)

3.2 Treatment Variables

- Are the values of ARM equivalent in meaning to values of TRTxxP?

  Yes
- Are the values of TRTxxA equivalent in meaning to values of TRTxxP?

  Yes
- Are both planned and actual treatment variables used in analyses?

  Yes

3.3 Use of Visit Windowing, Unscheduled Visits, and Record Selection

- Was windowing used in one or more analysis datasets?

  Yes

- Were unscheduled visits used in any analyses?

  Yes

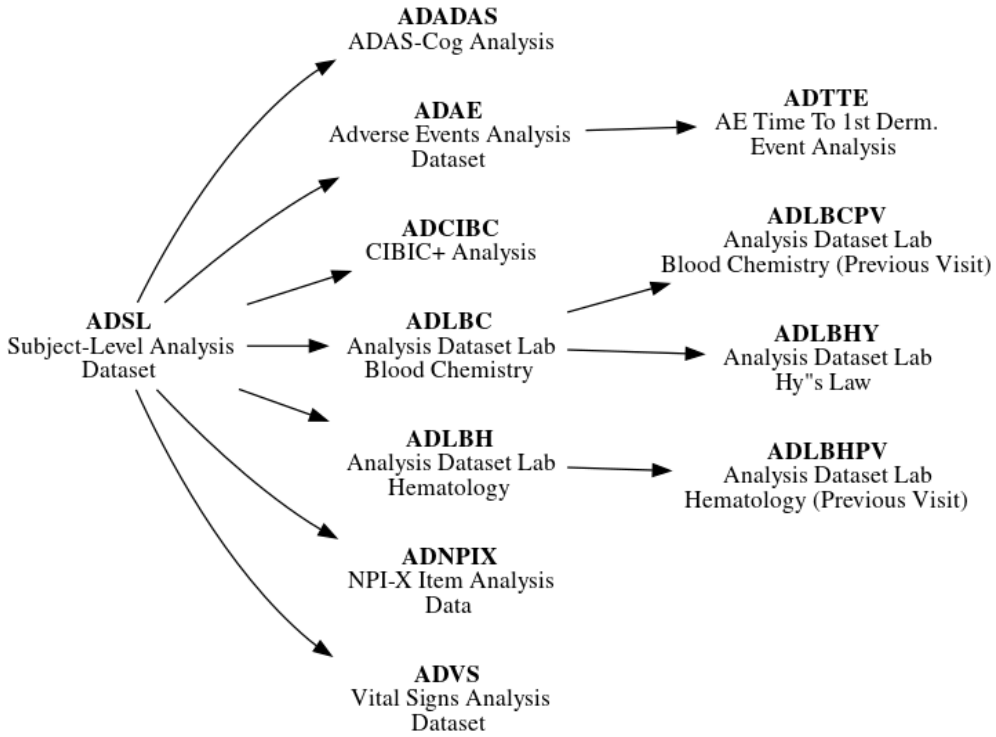
3.4 Imputation/Derivation Methods

Not applicable.

4. Analysis Data Creation and Processing Issues

Not applicable.

4.1 Data Dependencies



5. Analysis Dataset Descriptions

5.1 Overview

The analysis codes and outputs submitted in this pilot submission covers part of the efficacy and safety objectives of the initial protocol. More specifically, 4 analysis outputs are included, covering demographics analysis, primary efficacy endpoint analysis, and safety analysis.

5.2 Analysis Datasets

Dataset Dataset Label	Class	Efficacy	Safety	Baseline or other subject characteristics	Primary Objective	Structure
ADSL Subject Level Analysis Dataset	ADSL			X		One observation per subject
ADAE Adverse Events Analysis Dataset	ADAM OTHER		X			One record per subject per adverse event
ADTTE Time to Event Analysis Dataset	BASIC DATA STRUCTURE		X			One observation per subject per analysis parameter
ADLBC Analysis Dataset Lab Blood Chemistry	BASIC DATA STRUCTURE		X			one record per subject per parameter per analysis visit
ADLBCPV Analysis Dataset Lab Blood Chemistry (Previous Visit)	BASIC DATA STRUCTURE		X			one record per subject per parameter per analysis visit
ADLBH Analysis Dataset Lab Hematology	BASIC DATA STRUCTURE		X			one record per subject per parameter per analysis visit
ADLBHPV Analysis Dataset Lab Hematology (Previous Visit)	BASIC DATA STRUCTURE		X			one record per subject per parameter per analysis visit
ADLBHY Analysis Dataset Lab Hy's Law	BASIC DATA STRUCTURE		X			one record per subject per parameter per analysis visit
ADADAS ADAS-Cog Analysis	BASIC DATA STRUCTURE	X			X	One record per subject per parameter per analysis visit per analysis date
ADCIBC CIBIC+ Analysis	BASIC DATA STRUCTURE	X				one record per subject per parameter per analysis visit per analysis date
ADNPIX NPI-X Item Analysis Data	BASIC DATA STRUCTURE	X				one record per subject per parameter per analysis visit
ADVS Vital Signs Analysis Dataset	BASIC DATA STRUCTURE		X			one record per subject per parameter per analysis visit

5.2.1 ADSL – Subject Level Analysis Dataset

The subject level analysis dataset (ADSL) contains required variables for demographics, treatment groups, and population flags. In addition, it contains other baseline characteristics that were used in both safety and efficacy analyses. All patients in DM were included in ADSL.

The following are the key population flags are used in analyses for patients:

• SAFFL – Safety Population Flag (all patients having received any study treatment)

• ITTFL – Intent-to-Treat Population Flag (all randomized patients)

5.2.2 ADAE - Adverse Events Analysis Data

ADAE contains one record per reported event per subject. Subjects who did not report any Adverse Events are not represented in this dataset. The data reference for ADAE is 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.

5.2.3 ADTTE - Time to Event Analysis Dataset

ADTTE contains 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 used for the analysis of time to the first dermatological event is with PARAMCD of "TTDE".

5.2.4 ADLBHPV - Laboratory Results HematologyAnalysis Data(Previous Visit)

ADLBC and ADLBH contain one record per lab analysis parameter, per time point, per subject.

ADLBC contains lab chemistry parameters and ADLBH contains hematology parameters and these data are derived from the SDTM LB (Laboratory Tests) domain. Two sets of lab parameters exist in ADLBC/ADLBH. 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) is also presented.

The ADLBC and ADLBH datasets were split based on the values of the indicated variable. Note that this splitting was done to reduce the size of the resulting datasets and to demonstrate split datasets and not because of any guidance or other requirement to split these domains.

5.2.5 ADLBHY - Laboratory Results Hy’s Rule Analysis Data

ADLBHY contains one record per lab test code per sample, per subject for the Hy’s Law based analysis parameters. ADLBHY is derived from the ADLBC (Laboratory Results Chemistry Analysis Data) analysis dataset. It contains derived parameters based on Hy’s law.

5.2.6 ADADAS - ADAS-COG Data

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

5.2.7 ADCIBC - CIBC Data

ADCIBC contains analysis data from the from CIBIC+ questionnaire, one of the primary efficacy endpoints. It contains one record per subject per VISIT. Note that for all records, PARAM='CIBIC Score'. Visits are placed into analysis visits (represented by AVISIT and AVISITN) based on the date of the visit and the visit windows.

5.2.8 ADNPIX - NPI-X Item Analysis Data

ADNPIX contains one record per subject per parameter (NPI-X questionnaire item, total score, and mean total score from Week 4 through Week 24) per analysis visit (AVISIT). The analysis visits (represented by AVISIT and AVISITN) are derived from days between assessment date and randomization date and based on the visit windows that were specified in the statistical analysis plan (SAP).

6. Data Conformance Summary

6.1 Conformance Inputs

- Were the analysis datasets evaluated for conformance with CDISC ADaM Validation Checks?

  Yes, Version of CDISC ADaM Validation Checks and software used: Pinnacle 21 Enterprise version 4.1.1

- Were the ADaM datasets evaluated in relation to define.xml?

  Yes

- Was define.xml evaluated?

  Yes

6.2 Issues Summary

Rule ID	Dataset(s)	Diagnostic Message	Severity	Explanation
AD0258	ADAE	Record key from ADaM ADAE is not traceable to SDTM.AE (extra ADAE recs)	Error	There are derived records in ADAE, this has no impact on the analysis.

AD0018	ADLBC, ADLBCPV, ADLBH, ADLBHPV, ADVS, ADCIBC, ADLBNPIX	Variable label mismatch between dataset and ADaM standard	Error	The label for ANL01FL in these datasets are “Analysis Record Flag 01”, this is in conformance with ADaM IG 1.0, this is an issue in P21 checks, and has no impact on the analysis.
AD0320	ADSL	Non-standard dataset label	Error	The label for ADSL is “ADSL”, this has no impact on the analysis

7. Submission of Programs

Description

The sponsor has provided all programs for analysis results. They are all created on a Linux platform using R version 4.1.2.

7.1 ADaM Programs

Not Applicable. This pilot project only submits programs for analysis results.

7.2 Analysis Output Programs

Programs that produce analysis results are included in this package. The recommended steps to execute analysis results using R are described in the Appendix.

Program Name	Output Table Number	Title
tlf-demographic.txt	Table 14-2.01	Summary of Demographic and Baseline Characteristics
tlf-primary.txt	Table 14-3.01	Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 - LOCF
tlf-efficacy.txt	Table 14-3.02	ANCOVA of Change from Baseline at Week 20
tlf-kmplot.txt	Figure 14-1	KM plot for Time to First Dermatologic Event: Safety population

Proprietary R Analysis Package	Package version	Analysis Package Description
pilot1wrappers	0.1.1	A collection of R functions for this pilot project. Functions include wrappers for ANCOVA modeling and data/tableformatting.

Open-source R Analysis Package	Package version	Analysis Package Description
cowplot	1.1.1	Arrange figure
dplyr	1.0.7	Manipulate dataset.
emmeans	1.6.3	Calculate least square mean
ggplot2	3.3.5	Create figure
haven	2.4.3	Read in SAS dataset.
huxtable	5.4.0	Style data into presentation ready table
pharmaRTF	0.1.3	Write out a styled table to RTF format

pkglite	0.2.0	Prepare submission package
r2rtf	0.3.0	Create RTF table
rtables	0.3.8	Create and display complex tables with R
stringr	1.4.0	Manipulate string
Tplyr	0.4.1	Summarize and format clinical data for output
visR	0.2.0	Create figure

7.3 List of Outputs Programs

The following table contains a list of programs that generate outputs used in the R consortium R submission pilot 1. It shows the program file names, the related outputs, the input datasets and variables used, and any data restriction criteria that need to be applied.

Readable /Executable Code File Name	Output Name	Analysis Datasets & Variables	Selection Criteria
tlf-demographic.txt	tlf-demographic.out	ADSL.STUDYID ADSL.TRT01P ADSL.ITTFL ADSL.AGE ADSL.AGEGR1 ADSL.RACE ADSL.HEIGHTBL ADSL.WEIGHTBL ADSL.BMIBL ADSL.MMSETOT	STUDYID== “CDISCPILOT01” **Population**: ADSL.ITTFL == “Y” **Treatment Groups:** **ADSL.TRT01P** Placebo Xanomeline Low Dose Xanomeline High Dose
tlf-kmplot.txt	tlf.kmplot.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”**
tlf-efficacy.txt	tlf-efficacy.rtf	ADLB.TRTP ADLB.TRTPN ADLB.PARAMCD ADLB.AVISITN ADLB.BASE ADLB.AVAL ADLB.CHG	STUDYID== “CDISCPILOT01” **Population**: ADLBC.TRTPN in (0, 81) & ADLBC.PARAMCD == "GLUC" & ADLBC.AVISITN is not missing **Treatment Groups:** **ADLB.TRTPN** Placebo Xanomeline High Dose
tlf-primary.txt	tlf-primary.rtf	ADSL.TRT01P ADSL.USUBJID ADSL.EFFFL ADSL.ITTFL	**Population:** ADADAS.EFFFL == “Y” ADADAS.ITTFL == “Y” ADADAS.ANL01FL == “Y”

		ADADAS.TRTP ADADAS.TRTPCD ADADAS.EFFFL ADADAS.ITTFL ADADAS.PARAMCD ADADAS.ANL01FL ADADAS.AVISIT ADADAS.AVISITN ADADAS.AVAL ADADAS.CHG	**Treatment Groups: ADSL.TRTP** Placebo Xanomeline Low Dose Xanomeline High Dose **Parameters:** ADADAS.PARAMCD == "ACTOT"

8. Directory Structure

Study datasets and their supportive files are organized in accordance to Study Data Technical Conformance Guide.

module	1	Refers to the eCTD module in which clinical study data is being submitted.
datasets	2	Resides within the module folder as the top-level folder for clinical study data being submitted for m5.
cdiscpilot1	3	Study identifier or analysis type performed
analysis	4	Contains folders for analysis datasets and software programs; arrange in designated level 6 subfolders
adam	5	Contains subfolders for ADaM datasets and corresponding software programs
datasets	6	Contains ADaM datasets, analysis data reviewer's guide, analysis results metadata and define files

programs	6	Contains software programs for analysis datasets and key tables and figures

Appendix: Instruction to Execute Analysis Program in R

1. Install R

Download and install R 4.1.2 for Windows from
https://cran.r-project.org/bin/windows/base/old/4.1.2/R-4.1.2-win.exe.

2. Define Working Directory

Create a temporary working directory, For example, "C:\tempwork". Copy all submitted R programs into the temporary folder. All steps below should be executed in this working directory represented as "." in the example R code below.

3. Specify R package repository

The R packages are based on CRAN at 2021-08-31. To install the exact R package versions used in this project, run the code below to set the snapshot repository.

```
options(repos = "https://mran.microsoft.com/snapshot/2021-08-31")
```

4. Install open-source R packages

In the same R session, install the required packages by running the code below.

```
install.packages(c("haven", "dplyr", "emmeans", "pkglite" , "r2rtf", "rtables", "ggplot2",
                   "cowplot", "visR", "Tplyr", "pharmaRTF", "huxtable"))
```

5. Install Proprietary R packages

The proprietary R packages “pilot1wrappers” packed in the file r0pkg.txt. In the same R session, restore the package structures and install them by running the code below. Adjust the output path as needed to use a writable local directory.

```
pkglite::unpack("r0pkg.txt", output = ".", install = TRUE)
```

6. Update path to dataset and TLFs

INPUT path: to rerun the analysis programs, define the path variable

- Path for ADaM data: path$adam

OUTPUT path: to save the analysis results, define the path variable

- Path for output TLFs: path$output

All these paths require to be defined before executing the analysis output program. For example:

```
path = list(adam = "path/to/esub/analysis/adam/datasets",          # Modify path to the actual location
            outtable =  ".", outgraph = ".")                       # Output saved in current folder
```

7. Execute analysis program

To reproduce analysis results, one can rerun the following four programs:

- "tlf-demographic.r"
- "tlf-efficacy.r"
- "tlf-kmplot.r"
- "tlf-primary.r"