Guide to Analysis Data

The analysis data of study CDISCPilot01 represent an update to the analysis data provided with the original SDTM/ADaM Pilot Project Submission. The analysis data in the original submission pre-date version 1.0 of the ADaM Implementation Guide (IG). The current data are compliant with version 1.0 of the ADaM IG.

This data guide is intended to illustrate an example of the type of information that might be incorporated with the analysis datasets for a study of this type.

This document provides an introduction to analysis datasets that benefit from additional explanation beyond what is available in the Data Definitions (define.xml) document.

Contents

Note on Input Data Sources

The analysis files for this study were derived from the submitted SDTM files. SDTM files were prepared from CRF data according to version 3.1.2 of the SDTM IG (with amendment 1). No non-CRF or non-SDTM data were used to create the ADaM data.

The datasets include only subjects who were enrolled in the study. Subjects who failed screening criteria were not included in analysis datasets.

Where to Find Key Data

Demographics and Populations

The ADSL (Subject Level Analysis Data) dataset contains all subject-level variables for demographics, subject characteristics, and population flags.

Safety

Key safety data are found in the datasets ADAE (Adverse Events Analysis Data), ADLBC (Laboratory Results Chemistry Analysis Data), ADLBH (Laboratory Results Hematology Analysis Data), ADLBHY (Laboratory Results Hy's Rule Analysis Data), and ADVS (Vital Signs Analysis Dataset). In addition, ADTTE is specifically for safety analyses of the time to the first dermatologic AE. Dermatologic AEs are considered an adverse event of special interest.

Efficacy

Study CDISCPilot01 has two primary endpoints, the Alzheimer's Disease Assessment Scale - Cognitive Subscale, total of 11 items [ADAS-Cog (11)] at Week 24 and the Video-referenced Clinician's Interview-based Impression of Change (CIBIC+) at Week 24. All ADAS-Cog data, including the first primary endpoint, can be found in the dataset ADQSADAS. All CIBIC+ data, including the second primary endpoint, can be found in the dataset ADQSCIBC.

Mean Revised Neuropsychiatric Inventory (NPI-X) data are considered secondary. These data can be found in the dataset ADQSNPIX.

Protocol Deviations

Protocol deviations were not reported in the source database and have not been incorporated into the SDTM data. As a result, such data do not exist among the analysis data.

Core Variables

The following core variables were merged from ADSL into all analysis datasets: STUDYID, SITEID, USUBJID, TRTSDT, TRTEDT, AGE, AGEGR1, AGEGR1N, SEX, RACE, and RACEN.

The table below summarizes the treatment variables and population flags that exist among different groups of analysis requirements:

Dataset(s)	Treatment Variables	Population Flags
ADSL	TRT01P, TRT01PN, TRT01A, TRT01AN	SAFFL, DSRAEFL, ITTFL, EFFFL, COMP24FL
ADAE	TRTA, TRTAN	SAFFL
ADLBH, ADLBC, ADLBHY	TRTP, TRTPN, TRTA, TRTAN	SAFFL,COMP24FL, DSRAEFL
ADVS	TRTP, TRTPN, TRTA, TRTAN	SAFFL
ADTTE	TRTP, TRTA, TRTAN	SAFFL
ADQSADAS, ADQSNPIX, ADQSCIBC	TRTP, TRTPN	ITTFL, EFFFL, COMP24FL

Treatment variable values are based on the numeric dose level (i.e., 0, 54, 81) to avoid the need for a separate variable that contains the treatment dose.

Overview of Analysis Datasets

Important information regarding specific analysis data sets can be found below. For full details of all analysis data sets, including key analysis results, please refer to the define.xml document.

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.

As with the SDTM AE data set, all MedDRA code variables (i.e. those variables that end in CD) have missing values and dummy terms have been applied to the MedDRA High Level Term (HLT) and High Level Group Term (HLGT). This is due to the proprietary nature of the MedDRA dictionary and the fact that the data with this project will be made available to the public. In a standard submission, these codes and terms should be non-missing and properly populated.

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.

ADAE also contains additional Occurrence Flags to facilitate traceability, reviewability, and easy of reporting between the analysis dataset and the unique counts in the summary tables. For treatment emergent adverse events refer to define.xml documentation for the following variables: AOCCFL, AOCCSFL, and AOCCPFL for summarization at the subject, System Organ Class, and Preferred Term levels, respectively. Similarly, refer to the define.xml documentation for AOCC02FL, AOCC03FL, and AOCC04FL for summarization of serious adverse events at the subject, System Organ Class, and Preferred Term levels.

The three deaths reported during the conduct of this study are captured in the Results in Death Flag (AESDTH='Y') and Outcome of Adverse Event (AEOUT='FATAL'). The Start Date of the Adverse Event in ADAE is imputed to the first of the month if the day is missing. The Study Day of Event Start (ASTDY) and the Treatment Emergent Analysis Flag (TRTEMFL) are derived based on this imputation and may differ from their corresponding SDTM AE/SUPPAE variables Study Day of Start of Adverse Event (AESTDY) and Treatment Emergent Flag (AETRTEM).

ADLBC – Laboratory Results Chemistry Analysis Data

ADLBH – Laboratory Results Hematology Analysis Data

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.

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.

ADQSADAS – ADAS-COG Data

ADQSADAS 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. If multiple visits fall into the same visit window, then the one closest to the target date is chosen for analysis. Records where ANL01FL='Y' are the ones that were used for analysis. The last observation carried forward (LOCF) algorithm only considered records used for analysis as candidates to carry forward. Records where DTYPE='LOCF' signify those where AVAL was imputed using the LOCF algorithm. Source data can be traced back to the SDTM.QS domain using USUBJID and QSSEQ. Details on how to derive the primary efficacy result based on ADAS-Cog data can be found in the analysis results metadata in the define.xml.

ADQSCIBC – CIBC Data

ADQSCIBC 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. If multiple visits fall into the same visit window, then the one closest to the target date is chosen for analysis. Records where ANL01FL='Y' are the ones that were used for analysis. The last observation carried forward (LOCF) algorithm only considered records used for analysis as candidates to carry forward.

Records where DTYPE='LOCF' signify those where AVAL was imputed using the LOCF algorithm. Source data can be traced back to the SDTM.QS domain using USUBJID and QSSEQ. Details on how to derive the primary efficacy result based on CIBIC+ data can be found in the analysis results metadata in the define.xml.

ADQSNPIX – NPI-X Item Analysis Data

ADQSNPIX 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). If multiple assessments fall into the same visit window, then the one closest to the target day is chosen for analysis. Records where analysis flag (ANL01FL) = 'Y' are the ones that were used for analysis. The last observation carried forward (LOCF) algorithm was not used for these data. Source data can be traced back to the SDTM.QS domain using USUBJID and QSSEQ. All the NPI-X parameters, except for the mean total score from Week 4 through Week 24 (NPTOTMN), are from SDTM.QS domain. The value of parameter, NPTOTMN, contains the mean total score for each patient who had any assessments from Week 4 through Week 24. The baseline value of the parameter, NPTOTMN, is the same as the baseline value of total score. The baseline value is a covariate in the analysis of covariance (ANCOVA) model.

CDISC Conformance

Best efforts have been made to ensure that all data are compliant with the relevant standards. This has been achieved primarily via manual reviews of the data. At the moment there is not a widely accepted programmatic way to verify ADaM data. As a result, formal conformance reports do not exist.

SAS Programs sas

SAS Version 9.2 was used for the generation of ADaM datasets and analysis displays. Since sources and derivations are described in the define.xml, the majority of programs used are not included with this sample submission.

Only a few programs which are considered as providing additional value for traceability are included in the submission. These are all referred to in the define.xml. Note that for the display

program submitted, at14-5-02.sas, the focus is on reproducing the numbers rather than the layout.

ADaM Datasets Creation Programs

The table below describes the ADaM datasets creation programs included in the submission:

Program	Description
adae.sas	Creation of Adverse Events Analysis Dataset ADAE

Analysis Display Creation Programs

The table below describes the display creation programs included in the submission:

Program	Description
at14-5-02.sas	Creation of Table 14-5.02: Incidence of Treatment Emergent Serious Adverse Events by Treatment Group