Analysis Data Reviewer's Guide

Sample Drug Company, Inc.

Study SAMPLE-9999

ADRG Template Version 2019-07-18

Analysis Data Reviewer's Guide

Contents

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 Acronyms

Acronym	Translation
ADaM	Analysis Dataset Model
ADRG	Analysis Data Reviewer's Guide
IG	Implementation Guide
NA	Not Applicable
SDTM	Study Data Tabulation Model
TAUG	Therapeutic Area User Guide

1.3 Study Data Standards and Dictionary Inventory

Standard or Dictionary	Versions Used
SDTM	v1.3/ IG 3.1.3
SDTM Controlled Terminology	2016-09-30
ADaM	v2.1/IG 1.0
ADaM Controlled Terminology	ADaM: 2016-09-30 Added 'REINDUCTION', 'CONSOLIDATION PERIOD 1', and 'CONSOLIDATION PERIOD 2' to EPOCH extensible codelist as the study design includes these periods.
Data Definitions	Define.xml v1.0
TAUG (if applicable)	NA
Medications Dictionary	WHO March 2007
Medical Events Dictionary	Initial: 9.0 Final: 10.0
Other standards (optional)	ADaM Data Structure for Adverse Event Analysis v1.0

Standard or Dictionary	Versions Used
	ADaM Basic Data Structure for Time-to-Event Analysis v1.0

1.4 Source Data Used for Analysis Dataset Creation

The ADaM datasets were derived from SDTM version 1.3. The datasets were derived from the final locked database.

In addition to the clinical database, the source data include a lookup file that was used to classify the cause of death. Details about this lookup table are in the Appendix.The lookup table spreadsheet was converted to a SAS transport file and included with the ADaM datasets.

2. Protocol Description

2.1 Protocol Number and Title

Protocol Number: SAMPLE-9999

Protocol Title: Phase 1b/2, Open-Label, Multicenter, Dose-Escalating Clinical Study of the Safety, Tolerability, and Pharmacokinetic and Pharmacodynamic Profiles of GoodDrug Injection in Patients with Relapsed or Refractory [Oncology Indication]

Protocol Versions: One

2.2 Protocol Design in Relation to ADaM Concepts

This is an open-label study of 3 dose regimens (schedules) of GoodDrug, with previously untreated [INDICATION]. Based on Protocol Amendment 3, the GoodDrug dose schedules were as follows:

- Schedule A (Sch A): weekly x 3 (Days 1, 8, 15) at 10 mg/m2
- Schedule B (Sch B): weekly x 2 (Days 1, 8) at 10 mg/m2
- Schedule C (Sch C): twice weekly x 1 (Days 1, 4) at 10 and 20 mg/m2

The original protocol included only Schedule A. The protocol was amended to remove the third dose (Day 15) of GoodDrug injection in each cycle by closing Schedule A to enrollment and opening Schedule B and Schedule C, a more dose-intense regimen. Patients could complete up to 4 cycles of treatment consisting of 1 induction, 1 reinduction, and 2 consolidation cycles.

Treatment group assignment was verified by Sponsor and stored in the SDTM DM domain. It was used in ADaM datasets to derive the treatment group (TRT01P, TRT01PN), dose (DSLVL), and schedule (SCHDL) variables. These contain a single value per subject and were included on all datasets. The variables TRTP/TRTPN are used on non-ADSL datasets for consistency with ADaM standards and contain the same values as TRT01P/TRT01PN.

The start date and time of each possible cycle was recorded in the ADSL dataset and carried over to all other datasets as a core variable. In datasets used for analyses by treatment period, the study period variable TRTTYP (Treatment Type) was derived using the date of the observation compared to the treatment cycle dates.

Population flag variables were used to designate all treated patients (ALLTRTFL), defined as patients who received at least one dose of study drug. In addition, a population variable was created for patients who have similar characteristics to the population of the pivotal phase 3 study of this drug (POOLFL: see SDTM guide for source information for this variable). Variable POOLFL was based on the information from Sponsor's Medical review. The 'pooled analysis set' subjects were used for selected subgroup analyses, which were generally grouped by the baseline disease status (DISEASBL).

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
STUDYID	Study Identifier
USUBJID	Unique Subject Identifier
SUBJID	Subject Identifier for the Study
SITEID	Study Site Identifier
AGE	Age
AGEU	Age Units
AGEGR1N	Age Group 1 (N)
SEX	Sex
RACE	Race
ETHNIC	Ethnicity
INFCNTDT	Informed Consent Date
DSLVL	Dose Level of GoodDrug
PHASE	Phase
SCHDL	Schedule
ENRLFL	Enrolled Population Flag
ALLTRTFL	All Treated Analysis Set Flag
POOLFL	Pooled Analysis Set Flag
DISEASBL	Baseline Disease Status
COUNTRY	Country
TRTSDT	Date of First Exposure to Treatment
TRTEDT	Date of Last Exposure to Treatment
TRT01P	Planned Treatment for Period 01
TRT01PN	Planned Treatment for Period 01 (N)

3.2 Treatment Variables

ARM versus TRTxxP

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

Yes

ACTARM versus TRTxxA

If TRTxxA is used, then are the values of ACTARM equivalent in meaning to values of TRTxxA?

Yes. Since there were no discrepancies between planned and actual treatment, TRTxxA variables were not used.

Use of ADaM Treatment Variables in Analysis

Are both planned and actual treatment variables used in analyses?

No. The TRT01P/TRT01PN variables were used for all analyses that were broken out by treatment group. TRTP/TRTPN variables were included in datasets other than ADSL for consistency with standards, but were not used in analysis since they were identical to TRT01P/TRT01PN.

Use of ADaM Treatment Grouping Variables in Analysis

Are both planned and actual treatment variables used in analyses?

No. The TR01PG1/TR01PG1N variables were used for all analyses that were broken out by treatment grouping.

3.3 Subject Issues that Require Special Analysis Rules

Issue #	Issue Description	Issue Explanation	Resolution
1	Partial death date for patient 9999.	Exact date of death is unknown but it was confirmed that the patient died in January 2009 per a partial date entry on the Death Report CRF.	A death date of 01Jan2009 was imputed in the ADSL dataset for this patient.
2	Patient 1111	Patient 1111 enrolled in SAMPLE-9998 as patient ID 2222 prior to being enrolled in SAMPLE-9999.	A synopsis of SAMPLE-9998 treatment was recorded in SAMPLE-9999 as prior medications. The death information was summarized in both studies.

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

Was windowing used in one or more analysis datasets?

No.

Were unscheduled visits used in any analyses?

No.

3.5 Imputation/Derivation Methods

If date imputation was performed, were there rules that were used in multiple analysis datasets?

Yes. For datasets that may contain partial dates, imputation rules are as follows: If day is missing and the year and month are the same as the year and month of the first date of study drug dosing, then the date is imputed as the first dosing date. Otherwise, if day is missing and month and year are present, day is imputed as the first day of the month.

The variable TRTTYP was used in BDS datasets to record the type of treatment (induction, reinduction, consolidation 1, consolidation 2, follow-up) the patient was receiving at the time of the data collection/assessment.

4. Analysis Data Creation and Processing Issues

4.1 Split Datasets

There are no datasets that required splitting due to size constraints.

4.2 Data Dependencies

Dataset ADTTE (Time To Event Analysis Data) is derived directly from ADSL. All other datasets get core variable values from ADSL. There are no other processing dependencies.

4.3 Intermediate Datasets

There are no intermediate datasets.

5. Analysis Dataset Descriptions

5.1 Overview

Are data for screen failures, including data for run-in screening (for example, SDTM values of ARMCD='SCRNFAIL', or 'NOTASSGN') included in ADaM datasets?

No. All subjects in the SDTM database were used in ADaM datasets. Screen failure subjects were not included in SDTM.

Are data taken from an ongoing study?

No. This is a final locked database.

Do the analysis datasets support all protocol- and statistical analysis plan-specified objectives?

Yes. All analyses were done using the ADaM datasets as input.

5.2 Analysis Datasets

Dataset – Dataset Label	CLASS	Efficacy	Safety	Baseline or other patient characteristics	PK/PD	Primary Objective	Structure
ADSL Analysis Data Subject Level	ADSL	X		X		X	One record per subject
ADTTE Time to Event Analysis Data	BDS	X				X	One record per parameter per date per subject
ADMH Medical History Analysis Data	ODS			X			One record per medical history event per date per subject
ADAE Adverse Event Analysis Data	ODS		X				One record per adverse event per onset date per subject
ADEX Exposure Analysis Data	OTHER		X				One record per treatment name per administration date/time per subject
ADCM Medications Analysis Data	ODS		X				One record per medication term per start date per subject

Dataset – Dataset Label	CLASS	Efficacy	Safety	Baseline or other patient characteristics	PK/PD	Primary Objective	Structure
ADLBHEMA Hematology Results Analysis Data	BDS		X				One record per laboratory parameter per visit per draw date per subject
ADLBCHEM Chemistry Results Analysis Data	BDS		X				One record per laboratory parameter per visit per subject
ADLBPGUA Urinalysis/Pregnancy Res. Analysis Data	BDS		X				One record per laboratory parameter per visit per subject
ADVS Vital Sign Analysis Data	BDS		X				One record per vital sign parameter per position per visit per subject
ADECOG ECOG Performance Status Analysis Data	BDS		X				One record per parameter per visit per date per subject
ADPE Physical Examination Analysis Data	BDS		X				One record per parameter per visit per date per subject

5.2.1 ADSL – Subject Level Analysis Dataset

The subject level dataset ADSL contains required ADaM variables for demographics (AGE), treatment groups, and population flags. In addition, it contains:

Baseline variables
- CYTOBL- Baseline Cytogenetics Category
- ECOGBL – Baseline ECOG Status

Death date and cause (coded and uncoded)
- DTHDT – Death Date
- DTHDECOD – Primary Cause of Death (Coded Term)
- DTHCAUS – Primary Cause of Death (Reported Term)

Date and study day variables relevant to conduct of study (start dates of treatment in each treatment period, end of treatment, calculated days associated with these dates)
- TRTSDT – Treatment Start Date

TRTEDT – Treatment End Date
REINDDT – Reinduction Start Date
CNSL1DT – Consolidation 1 Start Date
CNSL2DT – Consolidation 2 Start Date

Date and flag variables associated with treatment response.

The ADaM Controlled Terminology Codelist ANLPURP (Analysis Purpose) is extensible. The following was added for this study: GLOBAL PRODUCT SAFETY.

5.2.2 ADTTE – Time to Event Analysis Dataset

The time to event dataset was used to support the primary study endpoints. It followed standard ADaM conventions for a TTE file. The events of interest were overall survival (paramcd='OS') and progression free survival (paramcd='PFS').

5.2.3 ADLBCHEM – Chemistry Results Analysis Dataset, ADLBHEMA – Hematology Results Analysis Dataset

These laboratory datasets were used primarily to support shift table analysis based on toxicity grade. The CDC toxicity grades were calculated and stored in the source SDTM dataset in the variable LBTOXGR. In the ADaM datasets, the LBTOXGR values were used to populate the baseline and on-study toxicity grade variables. However, since some lab parameters have potential toxicities in both the high and low direction, the analysis dataset made a distinction on the basis of potential toxicity direction. The variable RESDIR contains the direction of the potential toxicity (e.g. high or low), and the variable TOXNAM contains a description of the potential toxicity (e.g., hyperglycemia or hypoglycemia). The dataset contains one record per subject, per lab test, per potential toxicity. For lab parameters where there is more than one potential toxicity, a separate record was generated for each direction. In each of these records, if either the baseline or the on study result had a nonzero toxicity grade, but in the opposite direction of the RESDIR value, then the toxicity grade was set to 0 for that record.

It is important to note that the RESDIR and TOXNAM are used as classification variables for summaries; they do not necessarily indicate that this toxicity occurred.

Example (glucose records):

SDTM Variables			
USUBJID	VISIT	LBTOXGR	LBNRIND
101	baseline	1	H
101	day 15	2	L

ADaM Variables						Notes
USUBJID	AVISIT	RESDIR	TOXNAM	ATOXGR	BTOXGR	

ADaM Variables						Notes
101	day 15	High	Hyperglycemia	0	1	The baseline grade in BTOXGR is taken from the tox grade on the baseline record. It is set to 1 because the grade is 1 and it is in the same direction as this RESDIR value. The analysis tox grade is 0 because the source record is abnormal, but in the opposite direction as this RESDIR value.
101	day 15	Low	Hypoglycemia	2	0	The baseline grade in BTOXGR is taken from the tox grade on the baseline record. It is set to 0 because the grade is 1 and it is in the opposite direction as this RESDIR value. The analysis tox grade is 2 because the source record is abnormal, and in the same direction as this RESDIR value.

6. Data Conformance Summary

6.1 Conformance Inputs

Specify the software name and version for the analysis datasets

Pinnacle 21 version 2.2.0

Specify the version of the validation rules (i.e. CDISC, FDA) for the analysis datasets

CDISC

Specify the software name and version for the define.xml

config-adam-1.0 xml

Specify the version of the validation rules (i.e. CDISC, FDA) for the define.xml

CDISC

6.2 Issues Summary

Pinnacle 21 Notices were evaluated for potential problems but are not listed here. The following is a summary of Error level messages. There were no Warning level messages.

Dataset(s)	Diagnostic Message	Severity	Count	Explanation
ADAE	Neither AVAL nor AVALC are present in dataset	Error	1	The dataset is a hierarchical occurrence structure, the message is not relevant to this structure.
ADAE	Required variable is not present	Error	2	Message refers to AVAL/AVALC. The dataset is a hierarchical occurrence structure, the message is not relevant to this structure.
ADLBCHEM, ADLBHEMA	Multiple baseline records exist for a unique USUBJID, PARAMCD, and BASETYPE	Error	234	This dataset was designed to generate a separate record for each potential toxicity direction (RESDIR='High' or 'Low'). For lab parameters that can potentially be toxic in either high or low direction, it is expected that two records would be generated. We reviewed the dataset and confirmed that there are no duplicates within a toxicity direction value.
ADLBPGUA	ABLFL is present but BASE is not present	Error	1	No baseline or change from baseline analysis was required for this file.

7. Submission of Programs

All SAS programs for analysis datasets and primary and secondary efficacy results are submitted. They were all created on a SAS platform using version 9.3. The internal reference date used to create dates in ADaM datasets is January 1, 1960.

7.1 ADaM Programs

Program Name	Output	Macro Used
adsl.txt	adsl	attrib
adae.txt	adae	attrib, partdate
adcm.txt	adcm	attrib, partdate
adco.txt	adco	attrib
adcssrs.txt	adcssrs	attrib
addv.txt	addv	attrib
adeg.txt	adeg	attrib
adephis.txt	adephis	attrib
adex.txt	adex	attrib
adexd.txt	adexd	attrib
adexs.txt	adexs	attrib
adges.txt	adges	attrib
adhads.txt	adhads	attrib
adie.txt.	adie.	attrib
adlb.txt	adlb	attrib
admdres.txt	admdres	attrib
admh.txt	admh	attrib, partdate
adpc.txt	adpc	attrib
adpe.txt	adpe	attrib
adqolie.txt	adqolie	attrib
adsps.txt	adsps	attrib
adsv.txt	adsv	attrib
adszd.txt	adszd	attrib
adszfr.txt	adszfr	attrib
adszp.txt	adszp	attrib
adtte.txt	adtte	attrib
advs.txt	advs	attrib

7.2 Analysis Output Programs

Program Name	Output Number	Title	Input
t_predopbo.txt	7.1.1	Percent Reduction Over Placebo for – 28-Day Adjusted POS Frequencycacy - ITT	ADSZP
t_predopbo.txt	7.1.2	Percent Reduction Over Placebo for 28-Day Adjusted POS Frequencycacy - PP	ADSZP
t_resp.txt	7.2.1	Fifty Percent Responder Outcome for POS Frequency – ITT	ADSZP
t_resp.txt	7.2.2	Fifty Percent Responder Outcome for POS Frequency – PP	ADSZP
t_szfr.txt	7.3.1	Seizure Freedom for All Seizure Types - ITT	ADSZFR
t_nthseiz.txt	7.4	Time to nth Partial Onset Seizure – ITT	ADTTE
t_50resp.txt	7.6	Fifty Percent Responder Outcome for POS Frequency By Monthly Periods – ITT	ADTTE

7.3 Macro Programs

Program Name	Purpose
attrib.txt	Automatically set variable attributes based on specifications
partdate.txt	Creates full analysis dates from partial start or stop dates based on imputation rules.

8. Appendix

A spreadsheet was used as a lookup table for specific derived variables:

Table "**SAMPLE-9999 Primary Cause of Death**" contains all the unique values of the variable DDSTRESC from the DD domain. It contains a sponsor-assigned coded death reason. The coded reason was used to populate the variable DTHDECOD (Primary Cause of Death (Coded Term)) in the ADSL dataset.

Table 1: SAMPLE-9999 Primary Cause of Death

DTHCAUS	DTHDECOD
Acute renal failure	Acute Renal Failure
Acute Respiratory Distress Syndrome	Acute Respiratory Distress Syndrome
ARDS	Acute Respiratory Distress Syndrome
Bacterial Sepsis	Sepsis
Cardiac arrest	Cardiac Arrest
Cardiac Arrest unknown etiology	Cardiac Arrest
Disease Progression	Disease Progression
GI Bleed	GI Bleed
Leukocytosis	Leukocytosis
Multi-organ failure	Multi-organ Failure
SAE: Nosocomial Pneumonia	Nosocomial Pneumonia
Sepsis	Sepsis
Septic Shock, Multi-system organ failure	Septic Shock
Sudden Cardiac Arrest	Cardiac Arrest
Unknown	Unknown