Flags for Facilitating Statistical Analysis Using CDISC Analysis Data Model

Chun Feng, Xiaopeng Li and Nancy Wang

Celerion, 621 Rose Street, Lincoln, NE 68502

ABSTRACT

CDISC Analysis Data Model (ADaM) is designed to be analysis-ready for statistical summary and analysis. Driven by analyses in Statistical Analysis Plan (SAP), correct analysis flags should be set up in ADaM data sets prior to analyses. The ADaM subject-level analysis data set (ADSL) contains subject-level population flags which indicate whether subjects are in efficacy, safety, pharmacokinetic, pharmacodynamic, food effect, and dose proportionality analyses. In Basic Data Structure (BDS) data sets, common record-level analysis flags include, for example, recheck flags, flags for exclusion, baseline flags, early termination identifiers, and treatment emergent flags. However, from our early phase clinical trial experience with ADaM, some analyses require additional customized analysis flags. For example, multiple baseline flags, average values as baseline flags, LOCF (last observation carried forward)/ WOCF (worst value carried forward) flags, extra treatment breakdown flags for adverse events summarization, and analysis flags for sub-group analysis. In the paper, we first categorize basic key analysis flags and summarize their functions in a checklist. In addition, we share some examples with customized analysis flags used in our practice. We also extend the topic to what to consider while setting up analysis-ready ADaM data. The paper will prepare ADaM data users (implementers and reviewers) to check analysis flags appropriate for intended analyses.

INTRODUCTION

Clinical Data Interchange Standards Consortium (CDISC) Analysis Data Model (ADaM) which supports review of data and replication of analyses is designed to be analysis-ready for summary and statistical analysis. "Analysis-ready" means that ADaM data sets should contain all variables needed for planned analyses and analyses can be replicated without manipulating data. In other words, analyses are "one PROC away" from ADaM data. To facilitate the intended analyses, correct flags should be set up in ADaM data sets.

The CDISC ADaM Version 2.1 and the ADaM Implementation Guide Version 1.0 (ADaM IG) provide fundamental principles of setting up analysis data sets. There are two ADaM standard data structures: the subject-level analysis data set (ADSL), and the Basic Data Structure (BDS). The ADSL data structure has one record per subject and contains subject-level population flags indicating whether subjects are in efficacy, safety, pharmacokinetic, pharmacodynamic, food effect, or dose proportionality analyses. In the Basic Data Structure (BDS) data sets, common record-level analysis flags include: recheck flags, flags for exclusion, baseline flags, early termination identifiers, and treatment-emergent flags.

From our early phase clinical study experience with ADaM, some analyses require additional customized analysis flags or additional records to fulfill analysis needs. In the paper, we first categorize some pre-defined analysis flags and summarize their functions. In addition, we share some examples with customized analysis flags used in our practice.

METHODS

PRE-DEFINED ANALYSIS FLAGS

Since ADaM IG provides details about the features and application of each type of flag, we will not discuss the basics any further. Instead, we will revisit the frequently used flags from the viewpoint of facilitating analyses. At the same time, we will compare these flags in terms of their function, which data structure they can exist in, and the type of situation or analysis they can assist with. The three aspects are referred to as function, source data, and application in Table 1.

Table 1. Summary of Commonly Used Analysis Flags in ADSL and BDS Data in Early Phase Clinical Studies

Analysis Flags	Variable Name	Function	Source Data	Application
Subject-Level Population Flag	FASFL, SAFFL, ITTFL PPROTFL, COMPLFL RANDFL, ENROLFL	Indicators of analysis populations	ADSL, BDS	Subset data to include intended population
Record-Level Population Flag	FASRFL, SAFRFL, ITTRFL, PPROTRFL, COMPLRFL	Indicators of inclusion of intended records	BDS	Exclude a value from analysis per definition in SAPs; Exclude records per protocol; Exclude follow-up records of early discontinued subjects; Exclude unintended recheck values; Exclude contaminated blood concentration; Exclude a parameter in a treatment arm etc.
Baseline Flag	ABLFL	Baseline indicator	BDS	Analysis with baseline
Criterion Flag	CRITxxFL	Indicators of a criterion is met	BDS	Flexibly used in many circumstances to indicate an analysis criterion is met
Analysis Flag	ANLxxFL	Indicators of records for intended analyses	BDS	Subset data to include records for planned analysis
Analysis Reference Range Indicator	ANRIND	Indicates if a value is normal	BDS	Indicate out-of-range values; Provide shift summary
Baseline Reference Range Indicator	BNRIND	Indicates if baseline value is normal	BDS	Pair with ANRIND for shift tables
Treatment Emergent Analysis Flag	TRTEMFL	Indentify treatment-emergent AEs	ADAE	Subset data to include only treatment-emergent AEs for summarization

---

Pre-defined analysis flags are compared in the following aspects:

Functions of Analysis Flags for Facilitating Analysis:
- Subject-level population flags are considered the primary source for sub-group analysis
- Record-level population flags indicate inclusion or exclusion of one or more records
- Criterion flags keep records that meet a pre-specified analysis criterion
- Baseline record flags identify baseline values
- Analysis flags are used in conjunction with other flags to select the intended records
- Analysis Reference Range Indicator (ANRIND) and Baseline Reference Range Indicator (BNRIND) are usually used together to form the change in value from baseline to a new time point
- Treatment Emergent Analysis Flags (TRTEMFL) identify treatment-emergent adverse events

Flexibility of Analysis Flags:
Some flags have only one pre-defined function while others can be used in multiple situations with more flexibility. Baseline record flags only define baselines. Treatment-emergent analysis flags only identify treatment-emergent records. ANRIND and BNRIND flags indicate if a record is within or out of normal range. Subject-level population flags are used for defining analysis population. Record-level flags are used more flexibly. Criterion and analysis flags are more flexibly and widely used while existing flags combined are not sufficient for analyses.

Frequency of Usage for Analysis Flags:
Subject-level population flags are used in almost every dataset. Each analysis data set requires at least one record-level population flag and it is required that ADSL contains all applicable subject-level population flags. However, no record-level population flags exist in the ADSL. The rest of flags can only exist in BDS. Treatment Emergent Analysis Flag is only for ADAE. Baseline records are used when change-from-baseline summaries are of interest. ANRIND and BNRIND are used when shift tables are requested in SAPs. To illustrate the frequencies of usage for each type of analysis flag in a data set, we list the data sets that are used often in early clinical studies and mark "√" where a type of flag can exist.

Table 2. Frequency of Analysis Flags in Early Phase Clinical Studies ADaM Data Sets

	Subject-Level Population Flag	Record-Level Population Flag	Analysis Flag	Criterion Flag	Baseline Flag	Analysis Reference Range Indicator	Baseline Reference Range Indicator	Treatment Emergent Analysis Flag
ADSL	√							
ADAE		√	√	√				√
ADLB	√	√	√	√	√	√	√	
ADEG	√	√	√	√	√	√	√	
ADVS	√	√	√	√	√	√	√	
ADPC	√	√	√	√				
ADPP	√	√	√	√				

CUSTOMIZED FLAGS

In addition to pre-defined population flags, customized population flags can be added as needed. Similar to pre-defined population flags, customized population flags can be subject-level or record-level.

Subject-level Customized Flags

To perform analyses with intended groups of subjects, more subject-level population flags can be added in the ADSL. In the example of ADSL data sets below, additional subject-level population flags are created (Table 3) to support Pharmacokinetic analysis (PKPOPFL), Pharmacodynamic analysis (PDPOPFL), Food Effect analysis (FEPOPFL), Bioequivalence analysis (BEPOPFL), and Dose-proportionality analysis (DOSPOP). The names of these customized subject-level population flags end with "FL" following the naming conventions of subject-level population flags.

Table 3. Customized Subject-Level Population Flags in ADSL

USUBJID	SAFFL	PKPOPFL	COMPLFL	RANDFL	PDPOPFL	BEPOPFL	FEPOPFL	DOSPOPFL
XX-XX-XX-101	Y	Y	Y	Y	Y	Y	Y	Y
XX-XX-XX-102	Y	Y	Y	Y	Y	Y	Y	Y
XX-XX-XX-103	Y	Y		Y				
XX-XX-XX-104	Y	Y	Y	Y	Y	Y	Y	Y
XX-XX-XX-105	Y	Y		Y				
XX-XX-XX-106	Y	Y	Y	Y	Y	Y	Y	Y
...	...	...	...	...	...	...	...	...

Record-level Customized Flags

In some studies, adding more criterion flags make planned analyses achievable. In the example below, a study with two periods is used to illustrate the above statement. As shown in Table 4, subjects are dosed with Treatment A in Period 1 and Treatment B in Period 2 as defined in the SAP. Looking closely at the study drug administration (Table 4), Treatment B in Period 2 breaks down to "B1" on Day 1 and "B1+A" on the remaining 6 days. Intuitively, the summary of AEs by treatments can be facilitated by TRTA and AERIOD (Table 5). However, as defined by the SAP in the study, the summary of interest is by "Period 1 Treatment A", "Period 2 Treatment B1", and "Period 2 Treatment A+B1."

Table 4. Study Drug Administration

Period 1	Period 2						
Day 1	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
A	B1	A+B1	A+B1	A+B1	A+B1	A+B1	A+B1

Extra criterion flags are needed in the data set to further break down treatments inside Period 2. As shown in Table 5, CRIT01FL, CRIT02FL, and CRIT03FL are added to represent three new treatments including: A, B1, and A+B1 to complete analysis defined in the SAP. To summarize AEs under Treatment B1, we can keep records whose CRIT02FL equals "Y".

Table 5. Customized Criterion Flags in ADAE

USUBJID	APERIOD	TRTA	ANYAE	AETERM	CRIT01	CRIT01FL	CRIT02	CRIT02FL	CRIT03	CRIT03FL
XX-XX-XX-101	1	A	Y	AE1	Treat A	Y	Treat B1	N	Treat A+B1	N
XX-XX-XX-101	1	A	Y	AE2	Treat A	Y	Treat B1	N	Treat A+B1	N
XX-XX-XX-101	2	B	Y	AE3	Treat A	N	Treat B1	Y	Treat A+B1	N
XX-XX-XX-101	2	B	Y	AE4	Treat A	N	Treat B1	Y	Treat A+B1	N
XX-XX-XX-101	2	B	Y	AE5	Treat A	N	Treat B1	N	Treat A+B1	Y
XX-XX-XX-101	1	A	N	None						
XX-XX-XX-102	2	B	N	None						
XX-XX-XX-103	1	A	Y	AE6	Treat A	Y	Treat B1	N	Treat A+B1	N
XX-XX-XX-103	1	A	Y	AE7	Treat A	Y	Treat B1	N	Treat A+B1	N
XX-XX-XX-103	2	B	Y	AE8	Treat A	N	Treat B1	Y	Treat A+B1	N
XX-XX-XX-103	2	B	Y	AE9	Treat A	N	Treat B1	N	Treat A+B1	Y
...	...	...	...	...	...	...	...	...	...	...

Analysis Flags for the Creation of Extra Records

There are times when creation of extra rows in combination of analysis flags can fulfill planned analysis. The situations include defined multiple baselines, average values as baseline, and derived records by methods of LOCF / WOCF. Multiple baselines can be handled by adding additional rows of records using other types of baselines. Variable "basetype" is required to specify the type of baseline by definition in the SAP.

As shown in Table 6, two types of baselines are defined in the SAP. One is a pre-dose non-missing record and the other one is a screening value. The first five observations use screening records as their baseline values, whereas, observations 6 to 10 use pre-dose non-missing values as baseline values. "BASETYPE" indicates the method of deriving the corresponding two baselines. In combination with the ABLFL flag, two types of change-from-baseline values are to be summarized with two different baselines.

Table 6. Two Baselines for One Parameter

Observation	AVISIT	AVAL	Base	ABLFL	CHG	BASETYPE
1	Screening	35.1	35.1	Y		Screening
2	Check-in	34.9	35.1		-0.2	Screening
3	Pre-dose	35.5	35.1		0.4	Screening
4	Week 1	32.8	35.1		-2.3	Screening
5	EndPoint	34.5	35.1		-0.6	Screening
6	Screening	35.1	35.5			Pre-dose
7	Check-in	34.9	35.5			Pre-dose
8	Pre-dose	35.5	35.5	Y		Pre-dose
9	Week 1	32.8	35.5		-2.7	Pre-dose
10	EndPoint	34.5	35.5		-1	Pre-dose

Table 7 is an example dataset for derived records. "LOCF" and "AVERAGE" are derivation methods of adding extra rows of records. The derivation type (DTYPE) documents the method of derivation. In combination with "DTYTE" and other analysis flags, planned analysis can be achieved as defined in the SAP. Observations 2 to 4 are triplicate measurements of a parameter in Table 7. Observation 5 is the baseline record whose DTYPE shows the derivation method is "AVERAGE". Observations 6 to 10 are post-dose measurements of the same parameter. However, the measurements are only collected from Days 1 to 5 (observations 6-9). Day 6 value is not collected but the analysis in the SAP defines the summary of change-from-baseline values for Days 1 to 6. The LOCF method is specified to be used for deriving the Day 6 record. The last line of the record which is observation 10 is derived by the method of LOCF. Combining "ABLFL" and "ANL01FL", the summary of change-from-baseline values of post-dose from Days 1 to 6 are able to be performed.

Table 7. Derived Values With Flags

Observation	ADY	ATPT	ATPTN	AVAL	DTYPE	ABLFL	BASE	ANL01FL	CHG
1	-2	2 DAYS BEFORE	-48	15			16		
2	1	45 MINUTES BEFORE	-0.75	18			16		
3	1	30 MINUTES BEFORE	-0.50	16			16		
4	1	15 MINUTES BEFORE	-0.25	14			16		
5	1			16	AVERAGE	Y	16	Y	
6	1	1 HOUR 15 MINUTES AFTER	1.25	13			16	Y	-3
7	1	3 HOURS 45 MINUTES AFTER	3.75	12			16	Y	-4
8	2	23 HOURS 45 MINUTES AFTER	23.75	11			16	Y	-5
9	5	3 DAYS 23 HOURS 45 MINUTES AFTER	95.75	14			16	Y	-2
10	6	4 DAYS 23 HOURS 45 MINUTES AFTER	119.75	14	LOCF		16	Y	-2

In summary, if pre-defined analysis flags in ADaM and ADaM IG are insufficient for planned analysis, additional population flags, criterion flags, and analysis flags can be customized to fulfill the analysis needs. For sub-group analysis, subject-level analysis flags are used in ADSL. If records that meet certain criteria are used for analysis, then criterion flags are used. Combining analysis flags and derived additional records provide convenience and flexibility to customize facilitation for analyses in the SAP.

CONCLUSION

Following CDISC ADaM IG, ADaM data structure is pre-specified. However, to meet the needs of statistical analysis, some flexibility is achieved through analysis flags. Analysis flags make it possible to review data and perform analysis on the same set of data. While setting up ADaM data, guidelines and required analyses are considered simultaneously.

The tables and examples detailed present items that can be considered before setting up analysis data sets. Our suggestions and examples are based on our early clinical research experience. The same ideas and principles can be extended to the summarization of later phase studies.

REFERENCES

1 Analysis Data Model (ADaM) version 2.1. Available at http://www.cdisc.org/adam.
2 Analysis Data Model (ADaM) Implementation Guide version 1.0. Available at http://www.cdisc.org/adam.
3 CDISC ADaM Data Structure for Adverse Event Analysis Version 1.0. Available at http://www.cdisc.org/stuff/contentmgr/files/0/5aee16f59e8d6bd2083dbb5c1639f224/misc/adam_ae_final_v1.pdf.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Chun Feng
Celerion, Inc.
621 Rose Street
Lincoln, NE 68502
402-437-4799
chun.feng@celerion.com

Xiaopeng Li
Celerion, Inc.
621 Rose Street
Lincoln, NE 68502
402-437-6260
xiaopeng.li@celerion.com

Nancy Wang
Celerion, Inc.
621 Rose Street
Lincoln, NE 68502
402-437-4850
nancy.wang@celerion.com

www.celerion.com