PharmaSUG 2015 - Paper DS13

Implementing Various Baselines for ADaM BDS datasets

Songhui ZHU, A2Z Scientific Inc., East Brunswick, NJ

Abstract

Most ADaM datasets can be implemented as BDS data structure. In BDS datasets, if there is only one baseline per subject per parameter, variables ABLFL, BASE, and BASEC can be used to implement baselines and the implementation is pretty straightforward. However, when there are multiple baselines such as time-matched baselines and baselines for cross-over studies, the implementation can be tricky. In some cases, new records need be created. But in some other cases, no new records need to be created. In the meantime, BASETYPE and DTYPE need be implemented appropriately. In this paper, the author explains and provides examples showing how to implement baseline for a variety cases. In addition, this paper will also explain when new records need be created and when new records are not needed.

INTRODUCTION

ADaM defined several classes of data structures, including Subject Level Analysis Data (ADSL), Occurrence Data (OCCDS), and Basic Data Structure (BDS). Among these ADaM data structures, BDS data structure can be used implement majority of safety datasets such as ADLB, ADVS, and ADEG, and most efficacy datasets. Therefore, implementing ADaM-compliant BDS is essential to the CDISC compliance of the whole submission package. While implementing BDS, baseline implementation is almost always needed. In fact, implementation of baselines has caused a lot of confusions in industry.

Baseline is usually defined as the last non-missing value on or before the first treatment date. If the study is a parallel study and there is only one treatment phase, one can use ABLFL to flag the baseline records and populate baseline values using variable BASE to each record. The implementation is straightforward. However, for study designs that are not as simple as single phase parallel study, the implementation of baselines is not as simple as it appears. For example, for a crossover study, a subject may have multiple baselines. In crossover studies, some subjects can take the same treatment in different study phases, while some other subjects take different treatments in different study phases. In this paper, the author will explain why and how to implementation baseline in ADaM BDS datasets for different cases using examples.

IMPLEMENTING BASELINES IN BDS

While creating ADaM datasets, one may face various cases of baselines. A lot of studies are single phase and parallel for which the baseline is defined as the last non-missing value on or before the first dosing date. In this case, there is only one baseline per subject per parameter. This is the simplest case and the implementation is straightforward. But in other cases, for example, in a cross-over study, the baseline for a period could be defined as the last non-missing value on or before the first dosing date of each period. In this cases, there is one baseline per subject per parameter per period. In some other studies, baselines may be defined as time-matched. In these cases, there can be one baseline per subject per parameter per visit. In some studies, for instance in the extension phase, the baseline can be dependent on the arm. According to ADaM IG V1.0, new records need be created when there is more than one baseline. Do we always need create new records for all subjects and all parameters? When BASETYPE is needed? How to implement DTYPE accordingly? In this paper, implementation of ADaM-compliant BDS datasets for various cases of baseline definitions one may face will be discussed and illustrated using examples.

Case I: Single-phase parallel study

If the study is a parallel single phase study, the baseline is usually defined as the last non-missing value on or before first dosing date. The SDTM data are as shown in Table 1.

In this case, the implementation of baseline is in ADaM pretty straight-forward. One can use ABLFL to flag the baseline record and pop the baseline value to each record using variable BASE. So the ADaM data can be as shown in Table 2.

Table 1. SDTM data for single-phase parallel study

SUBJID	XOCAT	XOTEST	VISIT	XOSTRESN	XOSTRESU	XOBLFL	EPOCH
101-01	OD	Intraocular pressure	Visit 2	20	mmHg	Y	Screening
101-01	OD	Intraocular pressure	Visit 3	21	mmHg		Treatment
101-01	OD	Intraocular pressure	Visit 4	22	mmHg		Treatment
101-01	OD	Intraocular pressure	Visit 5	25	mmHg		Treatment

Table 2. ADaM data for single-phase parallel study

SUBJID	PARAM	AVISIT	AVAL	ABLFL	BASE	CHG	APHASE
101-01	Intraocular pressure (mmHg) - OD	Visit 2	20	Y	20	0	Screening
101-01	Intraocular pressure (mmHg) - OD	Visit 3	21		20	1	Treatment
101-01	Intraocular pressure (mmHg) - OD	Visit 4	22		20	2	Treatment
101-01	Intraocular pressure (mmHg) - OD	Visit 5	25		20	5	Treatment

While using this approach to implementing single baseline flag ABLFL, one of the widely-used approach is to define BASE, then merge into the dataset, and finally flag the baseline record. But this approach may results in multiple records flagged as baseline in some cases such as there are multiple records at the same time. To avoid this error, a modified approach can used. The approach is to define an intermediate variable first, for instance baseline candidate variable BASECAND, which identifies all non-missing values on or before the first dosing date, then sort the data by USUBJID PARAM BASECAND ADT, then flag the last BASECAND as ABLFL = Y. In this approach, there will never be multiple records flagged as baseline.

Case II: two-phase parallel studies

If a study is a two-phase study in which a subject can have two different treatments in different phases, assuming that there is no wash-out period between the two treatment phases. In this type of studies, the baseline is defined as the last non-missing value on or before first dosing date of each treatment phase. So, as a result, there are two different baselines per subject per parameter.

Table 3. SDTM data for two-phase parallel study

SUBJID	XOCAT	XOTEST	VISIT	XOSTRESN	XOSTRESU	XOBLFL	EPOCH
101-01	OD	Intraocular pressure	Visit 2	20	mmHg	Y	Screening
101-01	OD	Intraocular pressure	Visit 3	21	mmHg		Treatment A
101-01	OD	Intraocular pressure	Visit 4	22	mmHg		Treatment A
101-01	OD	Intraocular pressure	Visit 5	25	mmHg		Treatment A
101-01	OD	Intraocular pressure	Visit 6	26	mmHg		Treatment B
101-01	OD	Intraocular pressure	Visit 7	27	mmHg		Treatment B
101-01	OD	Intraocular pressure	Visit 8	28	mmHg		Treatment B

In this case, the implementation of baseline is not straight-forward. In the example shown in Table 3, for subject 101-01, the intraocular pressure value at Visit 2 (20 mmHg) is the baseline for all records. In the meantime, the intraocular

pressure value at Visit 5 (25 mmHg) is the baseline for all records the second treatment phase (Treatment B, highlighted in yellow). So for records at Visits 6, 7, and 8, there are two baseline values - one from Visit 2 and the other one from Visit 5. How to implement this? Intuitively, one may think about adding variables, ABLFL2, BASE2, CHG2, PCHG2, etc. Quite a few folks did implemented BDS this way. But this is not ADaM-compliant. According to ADaM IG v1.0 and draft v1.1 (**Rule 6**. **When there is more than one definition of baseline, each additional definition of baseline requires the creation of its own set of rows**.), new rows, not new columns need be created to implement the second baseline. Table 4 is one of the possible implementations. In Table 4, rows highlighted in green are newly created records.

In this example, there are two sets of records which can be distinguished by BASETYPE. The first set of records has BASETYPE = 'Screening' and the second set of records has BASETYPE = 'Period 01'. The first set of records uses the baseline from screening period, while the second set of records uses the baseline from Period 1 (Visit 5). BASETYPE is required when there are multiple baselines.

Table 4. ADaM data for two-phase parallel study

SUBJID	PARAM	AVISIT	AVAL	ABLFL	BASE	CHG	BASETYPE	APHASE
101-01	Intraocular pressure (mmHg) - OD	Visit 2	20	Y	20	0	Screening	Screening
101-01	Intraocular pressure (mmHg) - OD	Visit 3	21		20	1	Screening	Acute
101-01	Intraocular pressure (mmHg) - OD	Visit 4	22		20	2	Screening	Acute
101-01	Intraocular pressure (mmHg) - OD	Visit 5	25		20	5	Screening	Acute
101-01	Intraocular pressure (mmHg) - OD	Visit 6	26		20	6	Screening	Extension
101-01	Intraocular pressure (mmHg) - OD	Visit 7	27		20	7	Screening	Extension
101-01	Intraocular pressure (mmHg) - OD	Visit 8	28		20	8	Screening	Extension
101-01	Intraocular pressure (mmHg) - OD	Visit 5	25	Y	25	0	Period 01	Acute
101-01	Intraocular pressure (mmHg) - OD	Visit 6	26		25	1	Period 01	Extension
101-01	Intraocular pressure (mmHg) - OD	Visit 7	27		25	2	Period 01	Extension
101-01	Intraocular pressure (mmHg) - OD	Visit 8	28		25	3	Period 01	Extension

Discussions:

1) In this approach, no matter whether a subject received two different treatments in the two treatment phases or received the same treatment in two different treatment phases, there are two sets of records.

i) In the first set of records, the baseline is picked up from the last non-missing value on or before the first dosing date.

ii) In the second set of records, the baseline is picked up from the last non-missing value on or before the first dosing date for the second phase.

2) In some studies, the baseline for records in the second treatment phase can be dependent on whether a subject took different treatments in two treatment phases. If a subject took the same drug in both phases, the baseline for both phases are the same. If a subject took different drugs in the two phases, the baseline for the first phase is from the screening visit and the baseline for the second phase is the last non-missing value before the start of extension phase. For this analysis, the above implementation is not perfect for table programming. In table programs, one have to select the right sets of records for subjects who changed treatment or did not change treatment.

3) To make the table programming easier, one can modify the above approach as follows: for subjects who did not change treatment (such as Subject 101-01 in Table 5), just create the first set of records. For subjects who changed treatment (such as Subject 101-02 in Table 5), use the first baseline for acute phase analysis and use the second baseline for the extension phase analysis. So the enhanced ADaM data will be like the following Table 5.

Table 5. Enhanced ADaM data for two-phase parallel study

SUBJID	TRTP	PARAM	AVISIT	AVAL	ABLFL	BASE	CHG	BASETYPE	APHASE
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 2	20	Y	20	0	Screening	Screening
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 3	21		20	1	Screening	Acute
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 4	22		20	2	Screening	Acute
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 5	25		20	5	Screening	Acute
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 6	26		20	6	Screening	Extension
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 7	27		20	7	Screening	Extension
101-01	Drug A	Intraocular pressure (mmHg) - OD	Visit 8	28		20	8	Screening	Extension
101-02	Drug A	Intraocular pressure (mmHg) - OD	Visit 2	22	Y	22	0	Screening	Screening
101-02	Drug A	Intraocular pressure (mmHg) - OD	Visit 3	23		22	1	Screening	Acute
101-02	Drug A	Intraocular pressure (mmHg) - OD	Visit 4	24		22	2	Screening	Acute
101-02	Drug A	Intraocular pressure (mmHg) - OD	Visit 5	27		22	5	Screening	Acute
101-02	Drug B	Intraocular pressure (mmHg) - OD	Visit 5	27	Y	27	0	Acute	Acute
101-02	Drug B	Intraocular pressure (mmHg) - OD	Visit 6	29		27	2	Acute	Extension
101-02	Drug B	Intraocular pressure (mmHg) - OD	Visit 7	30		27	3	Acute	Extension
101-02	Drug B	Intraocular pressure (mmHg) - OD	Visit 8	28		27	1	Acute	Extension

Discussions about enhanced implementation:

1). For subjects (such as 101-01) who took the same drug in acute and extension phases (TRT01P = TRT02P), there is only one set of records.

2). For subjects (such as Subject 101-02) who took different drugs (drug switchers) in acute and extension phases, there are still two sets of records that can be distinguished by BASETYPE.

3) For drug switchers, each set of records has less records than in the previous implementation. Actually, only one new record (the baseline for the second set) has been created per subject per parameter.

4). It is worth noticing that for the newly created record (highlighted in green), Drug B is populated as TRTP, which can facilitate the table programming.

5). One may wonder whether creating only one new record violates Rule 6 in ADaM IG 1.0/1,1. The answer is no. This implementation is ADaM compliant. Rule 6 implies that new records should be created when there is more than one baseline **for the same record.** In the previous example, the record at VISIT = Visit 5 has two different baselines (and CHG), but no other records have multiple baselines. So only a duplicate record of VISIT = Visit 5 is created. In

general, if a subject has different baselines for different records, there is no need to create any extra records. New records need be created only when there are two or more baseline values for the same record.

Case III: Time-matched baseline

In some analyses, especially in Phase I studies, some parameters are measured in a series of time points at each of scheduled visits. In those studies, the baselines could be defined as the value at each time point at the baseline visit. So there is a baseline per subject per parameter per time point. In this case, the SDTM data will be like Table 6.

Table 6. SDTM data for time-matched baseline

SUBJID	XOCAT	XOTEST	VISIT	XOTPT	XOSTRESN	XOSTRESU
101-01	OD	Intraocular pressure	Baseline	8 AM	20	mmHg
101-01	OD	Intraocular pressure	Baseline	10 AM	21	mmHg
101-01	OD	Intraocular pressure	Baseline	12 PM	22	mmHg
101-01	OD	Intraocular pressure	Baseline	2 PM	25	mmHg
101-01	OD	Intraocular pressure	Baseline	4 PM	26	mmHg
101-01	OD	Intraocular pressure	Visit 3	8 AM	24	mmHg
101-01	OD	Intraocular pressure	Visit 3	10 AM	25	mmHg
101-01	OD	Intraocular pressure	Visit 3	12 PM	26	mmHg
101-01	OD	Intraocular pressure	Visit 3	2 PM	27	mmHg
101-01	OD	Intraocular pressure	Visit 3	4 PM	28	mmHg

In this case, there will be a baseline for each of the following time points: 8 am, 10 am, 12 pm, 2 pm, and 4 pm. But there is only one baseline for each record. Value at 8 am at VISIT = Baseline will be and will only be used as the baseline for all 8-am records. Similarly, 10-am value at VISIT = Baseline will be and will only be used as the baseline for 10-am records. No new records are to be created. Consider all these, one can implement ADaM data as shown in Table 7. In this example, there are five different baselines per subject per parameter, no extra records have been created.

Case IV: by-visit Baseline

For many trials, including ophthalmology studies, it is quite often that the difference of pre-dose and post-dose measurements such as intraocular pressures (IOP) need be analyzed. In these cases, we can define the pre-dose measurements at each visit as the baselines for the records measured at each visit. If we are only interested in the difference between pre- and post-dose values, then the implementation can be as shown in Table 8.

In this table, BASETYPE is needed since more than one record within one PARAM is flagged by ABLFL. According to ADaM IG V1.0, if there are multiple baseline records flagged for a given parameter within a subject then BASETYPE should be populated and contain different values for the baseline records within a subject. Please note that BASETYPE cannot be simply set as "By-visit Baseline".

Case V: Combination of Cases in Previous Cases

In some studies, one may be interested in different analyses using several different baseline definitions. For instance, in ophthalmology studies, the analyses of change in intraocular pressure from pre-dose at each visit and change in IOP from the IOP before the first dosing are both of special interest. In this case, we need use two baseline

definitions: the first baseline is the last no-missing value before the first dosing, and the second baseline is the time-matched baseline. The ADaM datasets should be implemented as the combination of Tables 2 and 7. Similarly, one can implement other combinations of the cases discussed above for various baseline definitions.

Table 7. ADaM data for time-matched baseline

SUBJID	PARAM	AVISIT	ATPT	ABLFL	AVAL	BASE	CHG	BASETYPE
101-01	Intraocular pressure (mmHg) - OD	Baseline	8 AM	Y	20	20	0	8 AM BL
101-01	Intraocular pressure (mmHg) - OD	Visit 3	8 AM		24	20	4	8 AM BL
101-01	Intraocular pressure (mmHg) - OD	Baseline	10 AM	Y	21	21	0	10 AM BL
101-01	Intraocular pressure (mmHg) - OD	Visit 3	10 AM		25	21	4	10 AM BL
101-01	Intraocular pressure (mmHg) - OD	Baseline	12 PM	Y	22	22	0	12 PM BL
101-01	Intraocular pressure (mmHg) - OD	Visit 3	12 PM		26	22	4	12 PM BL
101-01	Intraocular pressure (mmHg) - OD	Baseline	2 PM	Y	25	25	0	2 PM BL
101-01	Intraocular pressure (mmHg) - OD	Visit 3	2 PM		27	25	2	2 PM BL
101-01	Intraocular pressure (mmHg) - OD	Baseline	4 PM	Y	26	26	0	4 PM BL
101-01	Intraocular pressure (mmHg) - OD	Visit 3	4 PM		28	26	2	4 PM BL

Table 8. ADaM Implementation of by-visit baseline

SUBJID	PARAM	AVISIT	ATPT	ABLFL	AVAL	BASE	CHG	BASETYPE
101-01	IOP (mmHg) (OD)	Visit 2	Predose	Y	20	20	0	Baseline for Visit 2
101-01	IOP (mmHg) (OD)	Visit 2	Postdose		30	20	10	Baseline for Visit 2
101-01	IOP (mmHg) (OD)	Visit 3	Predose	Y	21	21	0	Baseline for Visit 3
101-01	IOP (mmHg) (OD)	Visit 3	Postdose		19	21	-2	Baseline for Visit 3
101-01	IOP (mmHg) (OD)	Visit 4	Predose	Y	22	22	0	Baseline for Visit 4
101-01	IOP (mmHg) (OD)	Visit 4	Postdose		25	22	3	Baseline for Visit 4

Discussions

1) When there is more than one baseline, one cannot add new variables such as BASE2 and CHG2. ADaM IG explains this clearly. In ADaM IG, there are 6 rules, among which the following two are related:

Rule 1. a parameter-invariant function of AVAL and BASE on the same row that does not involve a transformation of BASE should be added as a new column.

The derivation of the second baseline involves scanning all pre-treatment record and thus it cannot be added as a new column. Similarly, change from second baseline also involved more than one record and cannot be added a new column, either.

Rule 6. When there is more than one definition of baseline, each additional definition of baseline requires the creation of a set of its own rows.

The identification of which row is the baseline involves comparison of more than one row, Rule 6 is the direct inference of Rule 1.

2) When there is more than one baseline definition per subject per parameter, one does not always need create new rows for the same record.

Rule 6 implies that when there are two definitions of baselines, there must be two rows of the same record. However, this rule does not explain what means by "more than one baseline". The actual meaning of this rule is "When there is more than one baseline definition **for the same record (or AVAL)**, each additional definition of baseline requires the creation of a set of its own rows. This implies that no matter how many baseline definitions there are for a parameter, if there is only one baseline definition for a single record, **no** new rows need be created. For instance, for time-matched baseline and by-visit baseline, no new records are created.

3) DTYPE need be appropriately implemented when new records are created. ADaM has controlled terminology for DTYPE. The current existing CT terms include LOCF, WOCF, AVERAGE etc. There is no controlled terms for the baseline records. Neither ADaM IG v1.0 nor ADaM IG v1.1 has addressed this. However, CDISC ADaM team is working on this issue and more controlled terminologies will be added.

4) BASETYPE is needed whenever there is more than one baseline (ie, more than one record with ABLFL = Y) per subject per parameter. Having multiple BASETYPE does not imply creation of new sets of records. For instance, for the case of time-matched baseline, there are multiple values of BASETYPE, but no new records have been created.

5) Creating new records makes development and validation of ADaM datasets much harder. It is wise to avoid adding any new records if possible.

CONCLUSIONS

This paper presented the author's thoughts and practice on implementing CDISC-compliant ADaM datasets for various baselines. This paper also explained when new rows need be created and when new rows are not needed. Making good choices on these issues can make the ADaM datasets more compact and serve the analysis purpose well.

REFERENCES

[1] CDISC Analysis Data Model ,Version 2.1.
[2] CDISC ADaM Implementation Guide, Version 1.0.
[3] CDISC SDTM Implementation Guide, Version 3.1.2.
[4] CDISC ADaM Implementation Guide, Version 1.1 draft.
[5] The ADaM Structure for Occurrence Data (OCCDS), Version 1.0, Jan 13, 2015
[6] Nate Freimark, Susan Kenny, Jack Shostak, and John Troxell. Common Misunderstandings about ADaM Implementation. PharrnaSug 2012. San Francisco, CA. May 8 - 11, 2012.
[7] Songhui Zhu. Designing and Tuning ADaM Datasets. ParmaSUG 2013, Chicago, IL.
[8] Songhui Zhu. Applying ADaM Standards to Therapeutic Area Ophthalmology. ParmaSUG 2014, San Diego, CA.

CONTACT INFORMATION

Songhui Zhu, PhD
A2Z Scientific Inc.
East Brunswick, NJ 08816
songhui.zhu@a2zscientific.com
www.a2zscientific.com