

Implementation of ADaM Basic Data Structure for Cross-over Studies

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

Abstract

ADaM standards defined several classes of data structures, including Subject Level Analysis Data (ADSL), Occurrence Data (OCCDS), Basic Data Structure (BDS), and ADaM Other. These classes of datasets usually are all utilized in a complete study. But BDS is the most widely-used class of ADaM datasets and suitable for most safety and efficacy datasets such as lab data analysis, accountability analysis, exposure analysis, time-to-event analysis, and questionnaire analysis. To seasoned ADaM programmers, the implementation of ADaM-compliant study package is relatively easy when the study is a parallel design. However, the implementation of ADaM datasets can be challenging when the trial is a cross-over study. The major challenging issues include the selection of appropriate variables (in ADSL, BDS, and OCCDS datasets), the creation of extra sets of records in BDS or OCCDS datasets when necessary, the identification of baseline records, the definition of change from baseline due to the existence of multiple baselines, the definition of period/phase, the definition of record-level planned or actual treatment for post-baseline records and/or baseline records. This paper will illustrate how to deal with these issues using examples of cross-over studies.

INTRODUCTION

ADaM defined several classes of data structures, including Subject Level Analysis Data (ADSL), Occurrence Data (OCCDS), Basic Data Structure (BDS), and ADaM Other. Among these ADaM data structures, BDS data structure can be used to 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. However, implementing BDS datasets, especially when the study design is cross-over, is challenging.

Compared to parallel studies, a cross over study has at least two phases or periods and thus two possible baseline records for each parameter. Due to the existence of multiple phases or periods, a lot of issues arise. For example, first of all, one need add a lot subject-level variables such TR02SDT, TR02EDT, TRT02P, TRT02PN, TRT02A, TRT02AN, etc. Also, one can add more variables in BDS datasets for cross-over study design. For example, one can add record-level variables APERSDT, APEREDT, ASPRSDT, ASPREDT, etc. Secondly, one need define multiple baselines using baseline flag ABLFL, BASE, and BASEC. Thirdly, one need define multiple changes from baselines for a single record. One will face the issue about how to implement multiple baselines and changes from baselines, how to define periods and phases, and how to implement treatment variables TRTA and TRTP for post-baseline and/or baseline records. This paper will use example to illustrate how to handle these issues.

IMPLEMENTING ADaM FOR CROSS OVER STUDIES

In a parallel study 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 a cross-over study, there are two periods and each subject takes two different drugs in sequence. As a result, the implementation of ADaM is much complex. For instance, the baseline for a period could be defined as the last non-missing value on or before the first dosing date of each period. Therefore, there is one baseline per subject per parameter per period. To handle all the issues resulted from one extra period compared to a parallel study, one has to add appropriate variables in ADSL (such as TR02SDT, TR02EDT, TRT02P, TRT02PN, TRT02A, TRT02AN, etc), variables in BDS datasets (such as APERSDT, APEREDT, ASPRSDT, ASPREDT, BASETYPE, DTYPE, etc), and other classes of ADaM datasets. Especially, according to ADaM IG V1.1, new records need be created when there is more than one baseline. In this paper, examples will be given to illustrate how to handle this new issues one may face when implementing ADaM datasets for a cross-over study.

I: Adding new subject-level variables for a cross-over study

For a cross-over study, the whole study duration can be divided into two phases or periods. In some cases, a period can be further divided into multiple sub-periods. Depending on the study design of the cross-over study, one first needs to choose appropriate subject-level variables in ADSL.

Compared to a parallel study, a cross-over needs many more variables in ADSL. Among these variables for a cross-over study, the most important variables are those for defining phase or period and treatment in each phase or period.

Implementing Various Baselines for ADaM BDS datasets, continued

According to ADaM IG 1.1, quite a few variables that are not needed for parallel studies may have to be added for cross-over studies. For example, it is very beneficial for traceability to have some period, sub-period, and phase timing variables such as APxxSDT, APxxSTM, APxxADTM, APxxEDT, APxxETM, APxxEDTM, and APxxEDT defined in ADSL, especially when the period end date/time is different from treatment end date/time. For instance, if a subject only takes one dose of study drug in a single period, the treatment start date/time is identical to treatment end date/time. In this case, APxxEDT is usually different from TRxxEDT. Adding APxxEDT and keep it in all BDS datasets will make the BDS implementation much easier. Table 1 lists some subject-level variables defined in ADaM IG 1.1 that are usually needed for cross-over studies.

Table 1, ADSL variable useful for cross-over studies

Category	Variables
1. Treatment Variables	TRTxxP, TRTxxPN, TRTxxA, TRTxxAN, TRxxPGy, PRxxPGyN, TRxxAGy, TRxxAGyN, TSEQPGy (Planned Pooled Treatment Sequence y), TSEQPGyN, TSEQAGy (Actual Pooled Treatment Sequence y), TSEQAGyN, DOSExxP (Planned Dose), DOSExxA, DOSExxU,...
2. Timing Variables	TRxxSDT, TRxxSTM, TRxxSDTM, TRxxEDT, TRxxETM, TRxxEDTM,...
3. Period, sub-period, and phase timing variables	APxxSDT, APxxSTM, APxxADTM APxxEDT, APxxETM, APxxEDTM, PxxSwDT, PxxSwTM, PxxSwSDM, PxxSwEDT, PxxSwETM, PxxSwEDM, PHwSDT, PHwSTM, PHwSDTM, PHwEDT, PHwETM, PHwEDTM, ...
4. Trial Experience Variables	EOTxxSTT, EOTxxDT (End of Trt Decision Date in Period xx), EOPxxSTT, DCTxxRS (Reason for Discout of Treat in Period xx), DCTxxRSP (Reason Spec for Disc of Trt in Period xx), DCPxxRS (Reason for Discout from Period xx), DCPxxRSP (Reason Spec for Discout from Study), TRCMP (Compliance with Treatment (%)), TRCMPGRy (Compliance with Treatment (%) Grouping y), TRxxDURD, TRxxDURM (Treatment Duration in Period zz (Mon)), TRxxDURY, RFICyDT, ENRLyDT, RANDyDT, LSTSVDT (Last Subject Visit (SV) Date), LSTALVDT (Last Known Alive Date), ...

Assuming that in a two-period cross-over study, each subject takes either Drug A or Drug B for four weeks in the first period. Then each subject takes either Drug B or Drug A for four weeks in the second period. The duration of the first period is from the first dosing date in the first period to the day prior to the first dosing date in the second period. The end date of the second period is assumed to be the last treatment date in the second period + 14 days. After that second treatment period, each subject should have a month of safety follow-up phase. In this case, the subject-level variables can be as shown in Table 2.

Discussions of ADSL for cross-over study:

It is noteworthy that in this example, each subject has AP01EDT ≠ TR01EDT and AP02EDT ≠ TR02EDT. For example, subject 101-01 has TR01EDT = 2015-10-28, but AP01EDT = 2015-10-31, which is TR02SDT - 1. Similarly, TR02EDT = 2015-11-29, But AP02EDT = 2015-12-13. Defining the start and end dates of each period explicitly makes the definitions of treatment emergent adverse event and concomitant medications much easier. In addition, this approach also improves data-point traceability in BDS datasets including efficacy and safety analysis datasets and OCCDS datasets.

Implementing Various Baselines for ADaM BDS datasets, continued

Table 2. Example of subject-level variables for a two-period cross-over study

USUBJID	TR01SDT	TR01EDT	TR02SDT	TR02EDT	TRT01P	TRT02P	TRT01A	TRT02A	DOSE01P	DOSE02P
101-01	2015-10-01	2015-10-28	2015-11-01	2015-11-29	Drug A	Drug B	Drug A	Drug B	100	200
101-02	2015-04-04	2015-04-18			Drug B	Drug A	Drug B		200	100
101-03	2015-05-06	2015-06-03	2015-06-05	2015-06-10	Drug B	Drug A	Drug B	Drug A	200	100

DOSE01A	DOS02A	DOSEU	TR01DURD	TR02DURD	AP01SDT	AP01EDT	AP02SDT	AP02EDT
100	200	mg	28	28	2015-10-01	2015-10-31	2015-11-01	2015-12-13
200		mg	15		2015-04-04	2015-05-02		
200	100	mg	28	6	2015-05-06	2015-06-04	2015-06-05	2015-06-24

PH1SDT	PH1EDT	PH2SDT	PH2EDT	PH3SDT	PH3EDT	EOS01STT	EOS02STT	DCP01RS	DCP02RS
2015-10-01	2015-10-31	2015-11-01	2015-12-13	2015-12-14	2016-01-13	Completed	Completed		
2015-04-04	2015-05-02			2015-05-03	2015-06-01	Discontinued		Adverse Event	
2015-05-06	2015-06-04	2015-06-05	2015-06-24	2015-06-25	2015-07-24	Completed	Discontinued		Lack of Efficacy

In order to capture the follow-up phase start and end dates, variables PHySDT and PHyEDT are introduced. One may wonder why not use AP03SDT and AP03EDT to represent the start and end dates of the follow-up period. The reason of not using variables APxxSDT and APxxEDT is that xx need be associated with treatment variables TRTxxP and TRTxxA. But there is no treatment associated with the follow-up phase. Similarly, if a subject need to define the start and end dates of a screening phase, one can also use PHySDT and PHyEDT.

Also, EOS01STT and EOS02STT capture the information of the end of period status. For instance, subject 101-01 completed both periods (EOS01STT = Completed and EOS02STT = Completed). Subject 101-02 discontinued in period 1 which can be seen since EOS01STT = Completed. Subject 101-03 completed the first period and discontinued in the second period (EOS01STT = Completed and EOS02STT = Discontinued).

The reason for each subject's discontinuation is captured in variables DCP01RS and DCP02RS. For instance, Subject 101-02 discontinued from the study in period 1 due to adverse event. Subject 101-03 discontinued from period 2 due to lack of efficacy.

The treatment duration of each period is captured by variables TR01DURD and TR02DURD. In this example, for subject 101-01, treatment durations in Periods 1 and 2 are 28 and 28 days, respectively. For subject 101-003, the treatment durations in Periods 1 and 2 are 28 and 6 days, respectively.

The planned dose information is captured in variables DOSE01P and DOSE02P and actual dose information is captured in variables DOSE01A and DOSE02A.

II: Adding new BDS variables for a cross-over study

Similarly, some variables also need be added in BDS datasets for cross-over studies on top of the variables needed for parallel studies. Table 3 lists some selected variables defined in ADaM IG 1.1 that are likely needed for cross-over studies.

For example, APERIOD/APERIODC can be used to define periods. Variable APERSDT can be added to store the start date associated with the period (APERIOD). It is worth pointing out that APERSDT is different from subject level APxxSDT in ADSL. TRTP and TRTA are the planned and actual treatment associated with the record, respectively. In general, TRTP/TRTA should be the same as TRTxxP/TRTxxA for records in period xx. But TRTxxP/TRTxxA are subject-level variable TRTxxP and TRTxxA and should be from ADSL. DOSEP is the planned dose level associated with the AVAL of the record, and DOSCUMP is the cumulative planned treatment dose up to the time point associated with AVAL.

Implementing Various Baselines for ADaM BDS datasets, continued

Table 3. BDS variables for cross-over studies

Category of variables	Variables
1. Record-level Treatment Variables	TRTP, TRTPN, TRTA, TRTAN, TRTPGy, TRTPGyN, TRTAGy, TRTAGyN, DOSEP, DOSECUMP (Cumulative Planned Treatment Dose) , DOSEA, DOSECUMA, DOSEU
2. Timing Variables for BDS	ADT, ATM, ADTM, ADY, ASTDT, ASTTM, ASTDTM, ASTDY, AENDT, AENTM, AENDTM, AENDY, AVISIT, AVISITN, ATPT, ATPTN, ATPTREF, APERIOD, APERIODC, ASPER, ASPERC, APHASE, APHASEN, APERSDT, APERSTM, APERSDTM, APEREDT, APERETM, APEREDTM, ASPRSDT, ASPRSTM, APSRSDTM, ASPREDT, ASPRET, ASPREDTM, PHSDT, PHSTM, PHSDTM, PHEDT, PHETM, PHEDTM.

The example below is from a CNS study and assume that the efficacy measurement MADRS total scores is in SDTM QS data as shown in Table 4.

The first subject has measurements at screening visit and all eight visits in two treatment periods. In addition, this subject took drug A 100 mg daily dose from 2015-10-01 until 2015-10-28 and then took Drug B 200 mg daily dose from 2015-11-01 to 2015-11-29. The information in EX is omitted.

Table 4 SDTM QS domain MADRS total score

USUBJID	QSTESTCD	VISIT	VISITNUM	QSSTRESN	QSBFL	QSDTC	EPOCH	QSSEQ
101-01	MADRSTOT	Screening	1	40		2015-09-30	Screening	1
101-01	MADRSTOT	Visit 2	2	39	Y	2015-10-01	Period 1	2
101-01	MADRSTOT	Visit 3	3	38		2015-10-08	Period 1	3
101-01	MADRSTOT	Visit 4	4	37		2015-10-15	Period 1	4
101-01	MADRSTOT	Visit 5	5	36		2015-10-22	Period 1	5
101-01	MADRSTOT	Visit 6	6	35		2015-10-29	Period 1	6
101-01	MADRSTOT	Visit 7	7	34		2015-11-01	Period 2	7
101-01	MADRSTOT	Visit 8	8	33		2015-11-08	Period 2	8
101-01	MADRSTOT	Visit 9	9	32		2015-11-15	Period 2	9
101-01	MADRSTOT	Visit 10	10	31		2015-11-22	Period 2	10
101-01	MADRSTOT	Visit 11	11	30		2015-12-01	Period 2	11

ADaM dataset for the efficacy measurement ADMADRS can be implemented as shown in Table 5.

Discussions about this implementation:

Compared to SDTM data, three new records are added - two endpoint records (one for each period, highlighted in green) and one baseline record for Period 02 (highlighted in yellow).

Variables APERIOD and APERIODC are introduced. For records with AP01SDT <= ADT <= AP01EDT or the baseline record for the first treatment period, APERIODC = Period 1. For records with AP02SDT <= ADT <= AP02EDT and the records for the second treatment period, APERIODC = Period 2. For other records such as the

Implementing Various Baselines for ADaM BDS datasets, continued

measurement at the screening visit, APERIODC is set to blank. Note that ADaM IG states that APEROD/APERIODC are associated with TRTxxP. For the screening visit, there is no treatment associated, and thus, APERIOD and APERIODC are set to missing.

It is worth notice that APERIOD is set to 1 for the added baseline record (ADT = 2015-10-01 and AVISITN = 19) for Period 01. This record is from pre-treatment period and is before AP01SDT (2015-10-01). Similarly, APERIOD is set to 2 for the added baseline record (ADT = 2015-10-29 and AVISITN = 69) for Period 02, even though ADT for this record is before the start date of Period 02 (AP02SDT = 2015-11-01). This implementation can facilitate table programming.

Table 5. Example of BDS datasets for a two-period cross-over study

SUBJID	PARAMCD	AVISIT	AVISITN	VISIT	VISITNUM	ADT	ABLFL	APERIODC	TRTP
101-01	MADRSTOT			Screening	1	2015-09-30			
101-01	MADRSTOT	Baseline 01	19	Visit 2	2	2015-10-01	Y	Period 01	Drug A
101-01	MADRSTOT	Visit 30	30	Visit 3	3	2015-10-08		Period 01	Drug A
101-01	MADRSTOT	Visit 40	40	Visit 4	4	2015-10-15		Period 01	Drug A
101-01	MADRSTOT	Visit 50	50	Visit 5	5	2015-10-22		Period 01	Drug A
101-01	MADRSTOT	Visit 60	60	Visit 6	6	2015-10-29		Period 01	Drug A
101-01	MADRSTOT	Endpoint 01	65	Visit 6	6	2015-10-29		Period 01	Drug A
101-01	MADRSTOT	Baseline 02	69	Visit 6	6	2015-10-29	Y	Period 02	Drug A
101-01	MADRSTOT	Visit 70	70	Visit 7	7	2015-11-01		Period 02	Drug B
101-01	MADRSTOT	Visit 80	80	Visit 8	8	2015-11-08		Period 02	Drug B
101-01	MADRSTOT	Visit 90	90	Visit 9	9	2015-11-15		Period 02	Drug B
101-01	MADRSTOT	Visit 100	100	Visit 10	10	2015-11-22		Period 02	Drug B
101-01	MADRSTOT	Visit 100	110	Visit 11	11	2015-12-01		Period 02	Drug B
101-01	MADRSTOT	Endpoint 02	115	Visit 11	11	2015-11-22		Period 02	Drug B

TRTP = Drug A for records with APERIOD = 1, and TRTP = Drug B for records with APERIOD = 2. TRTP is the planned treatment associated with the AVAL on date ADT. In this implementation, TRTA and TRTP for the baseline records are set to TRTA and TRTP, respectively in the associated periods.

It noteworthy that TRTP and TRTA are record-level variables, which means that in a BDS dataset, TRTP and TRTA can be different from record to record. In contrast, variables TRTxxA and TRTxxA are subject-level variable and should be invariant of records.

APERSDT and APEREDT are also record-level variables. For instance, APERSDT = 2015-10-01 and 2015-11-01 for records in Period 01 and Period 02, respectively.

Two baseline records are both flagged using variable ABLFL = Y. Please note that there is only one baseline flag variable (ABLFL) allowed in a BDS dataset no matter how many different baselines there are per subject per parameter. Similarly, there is only one change from baseline variable (CHG) allowed. There is only one baseline value variable (BASE) allowed.

Implementing Various Baselines for ADaM BDS datasets, continued

For the analysis of the last record in each period, a new record with DTYPE = Endpoint is added for each period. The record with AVISITN = 65 is the last-record for Period 01. The record with AVISITN = 115 is added for Period 02. The last records cannot be identified by adding variables. In ADaM IG V1.1 or V1.0, there are 6 rules that decide whether a row or a column should be added. Rule 3 states that **function of one or more rows within the same parameter for the purpose of creating an analysis timepoint should be added as a new row for the same parameter.**

Table 5 (Cont.)

APERSDT	APEREDT	DOSE P	DOSCUM P	DOSEU	AVAL	BASE	CHG	ABLF L	DTYPE	QSSE Q	BASETY PE
					40					1	
2015-10-01	2015-10-31	0	0	mg	39	39	0	Y		2	Run-in
2015-10-01	2015-10-31	100	700	mg	38	39	-1			3	Run-in
2015-10-01	2015-10-31	100	1400	mg	37	39	-2			4	Run-in
2015-10-01	2015-10-31	100	2100	mg	36	39	-3			5	Run-in
2015-10-01	2015-10-31	100	2800	mg	35	39	-4			6	Run-in
2015-10-01	2015-10-31	100	2800	mg	35	39	-4		Endpoint	6	Run-in
2015-11-01	2015-12-13	0	0	mg	35	35	0	Y		6	Period 01
2015-11-01	2015-12-13	0	0	mg	34	35	1			7	Period 01
2015-11-01	2015-12-13	200	1400	mg	33	35	-2			8	Period 01
2015-11-01	2015-12-13	200	2800	mg	32	35	-3			9	Period 01
2015-11-01	2015-12-13	200	4200	mg	31	35	-4			10	Period 01
2015-11-01	2015-12-13	200	5600	mg	30	35	-5			11	Period 01
2015-11-01	2015-12-13	200	5600	mg	30	35	-5		Endpoint	11	Period 01

III: Implementing baselines for a cross-over study

As briefly discussed before, for cross-over studies, there are multiple baselines per subject per parameter. Depending on the baseline definitions, there can be very different implementations. But keep in mind, there is only one ABLFL, one BASE, one CHG, and one PCHG.

Consider still a two-phase study in which a subject can have two different treatments in two different periods and MADRS total scores are shown in Table 4. In this type of studies, there is no wash-out period between the two treatment periods and thus the baseline is defined as the last non-missing value on or before first dosing date of each treatment period. So, as a result, there are two different baselines per subject per parameter.

In this case, the implementation of baseline is not straight-forward. In the example shown in Table 4, assuming that for subject 101-01, the MADRS total score at Visit 2 (ADT = 2015-10-01) is the baseline for all records and that in the meantime, the MADRS total score at Visit 6 (ADT = 2015-10-29) is the baseline for all records in the second treatment period (Drug B). So for records at Visits 7, 8, 9, 10, and 11, there are two baseline values - one from Visit 2 and the other one from Visit 6. To implement two different baselines for a record, one cannot think about adding variables, ABLFL2, BASE2, CHG2, PCHG2, etc. Unfortunately, according to ADaM IG v1.0 and 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 6 is one of the possible implementations. In Table 6, 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 6. BDS Baseline for cross-over study

SUBJID	PARAMCD	AVISIT	VISIT	AVAL	ABLF L	BASE	CHG	QSSE Q	BASETYP E	APERIOD C
101-01	MADRSTOT	Screening	Screening	40		39		1	Period 01	Run-in
101-01	MADRSTOT	Baseline 01	Visit 2	39	Y	39	0	2	Period 01	Run-in
101-01	MADRSTOT	Visit 30	Visit 3	38		39	-1	3	Period 01	Run-in
...
101-01	MADRSTOT	Visit 60	Visit 6	35		39	-4	6	Period 01	Run-in
...
101-01	MADRSTOT	Visit 110	Visit 11	30		39	-9	11	Period 01	Run-in
101-01	MADRSTOT	Baseline 02	Visit 6	35	Y	35	0	6	Period 02	Period 01
101-01	MADRSTOT	Visit 70	Visit 7	34		35	-1	7	Period 02	Period 01
...
101-01	MADRSTOT	Visit 100	Visit 10	31		35	-4	10	Period 02	Period 01
101-01	MADRSTOT	Visit 110	Visit 11	30		35	-5	11	Period 02	Period 01

Discussions:

In this approach, no matter whether a subject received two different treatments in the two periods or received the same treatment in two different treatment periods, there are two sets of records. For the first set of records, the baseline is the last non-missing value on or before the first dosing date. For the second set of records, the baseline is the last non-missing value on or before the first dosing date for the second phase. This approach provides all possible changes from baselines. But it also leaves the choice of which set of records to use to table programming, which sometimes is not straightforward for table programming.

In some studies, the baseline for records in the second treatment period can be dependent on whether a subject took different treatments in two treatment periods. If a subject took the same drug in both periods, the baseline for both periods are the same. If a subject took different drugs in the two periods, the baseline for the first period is from the screening visit and the baseline for the second period is the last non-missing value before the start of the second period. 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.

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 6), just create the first set of records. For subjects who changed treatment (such as Subject 101-02 in Table 6), use the first baseline for the first period analysis and use the second baseline for the second period analysis. So the enhanced ADaM data will be like the following Table 7.

Table 7. Enhanced ADaM data for Cross-over study

SUBJID	TRTP	PARAMCD	AVISIT	VISIT	AVAL	ABLFL	BASE	CHG	BASETYPE	APERIODC
101-01	Drug A	MADRSTOT	Screening	Screening	40		40		Run-in	Period 01
101-01	Drug A	MADRSTOT	Baseline 01	Visit 2	39	Y	39	0	Run-in	Period 01
101-01	Drug A	MADRSTOT	Visit 30	Visit 3	38		39	-1	Run-in	Period 01
...
101-01	Drug A	MADRSTOT	Visit 60	Visit 6	35		39	-4	Run-in	Period 01
101-01	Drug A	MADRSTOT	Visit 70	Visit 7	34		39	-5	Run-in	Period 02
...
101-01	Drug A	MADRSTOT	Visit 110	Visit 11	30		39	-9	Run-in	Period 02
101-02	Drug A	MADRSTOT	Screening	Screening	50		51		Run-in	Period 01
101-02	Drug A	MADRSTOT	Baseline 01	Visit 2	51	Y	51	0	Run-in	Period 01
...
101-02	Drug A	MADRSTOT	Visit 50		48		51	-5	Run-in	Period 01
101-02	Drug A	MADRSTOT	Visit 60		45		51	-7	Run-in	Period 01
101-02	Drug B	MADRSTOT	Baseline 02		45	Y	45	0	Period 01	Period 02
101-02	Drug B	MADRSTOT	Visit 70		47		45	2	Period 01	Period 02
...
101-02	Drug B	MADRSTOT	Visit 110		42		45	-3	Period 01	Period 02

Discussions about enhanced implementation:

For subjects (such as 101-01) who took the same drug in both periods (TRT01P = TRT02P), there is only one set of records.

But for subjects (such as Subject 101-02) who took different drugs (drug switchers) in the two treatment periods, there are still two sets of records that can be distinguished by BASETYPE. 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.

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.

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 6 has two different baselines (and CHG), but no other records have multiple baselines. So only a duplicate record of VISIT = Visit 6 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.

Discussions about BDS Baselines

In cross-over studies, it is often that there are multiple baselines per subject per parameter. The most intuitive choice is to define multiple baseline variables such as BASE1, BASE2, CHG1, CHG2, etc. However, this is not ADaM-compliant. 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 records on or before the first dosing date of Period 02 and thus the second baseline cannot be added as a new column. Similarly, change from second baseline also involves more than one record and cannot be added as 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.

However, when there is more than one baseline definition per subject per parameter, this rule does not imply that one always has to 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.

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, ENDPOINT, etc. There is no controlled terms for the baseline records.

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 the creation of new sets of records. For instance, for the case of time-matched baseline, there are multiple values of BASETYPE due to multiple baselines' existence, but no new records have been created. But the discussion of the implementation of other cases of multiple baselines is out of the scope of this paper.

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

IV: Occurrence datasets for cross-over study

CDISC released ADaM Structure of Occurrence Data (OCCDS) V1.0 and retired ADaM Data Structure for Adverse Event Analysis V1.0 on Feb. 12, 2016. OCCDS class includes ADAE, ADCM, and ADMH. The most significant changes from ADAE to OCCDS is that it allows a record in SDTM data to be replicated in ADaM datasets ADCM, ADMH, and ADAE. This change increased the flexibility of ADaM data structure to be applied to various study designs including cross-over studies.

For a cross-over study, the sponsor has the freedom to decide the data structure of ADAE, ADCM, and ADMH. For example, the sponsor can implement ADAE one record per adverse event in SDTM AE domain or multiple records per adverse event in SDTM AE domain. Similarly, the sponsor can also implement ADCM and ADMH the same way. For instance, for one record in CM, the sponsor can create two records of the same CMDECOD if this concomitant medication spans over two periods.

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 SDTM Implementation Guide, Version 3.1.2.
- [3] CDISC ADaM Implementation Guide, Version 1.1.
- [4] The ADaM Structure for Occurrence Data (OCCDS), Version 1.0, Feb. 12, 2016
- [5] Nate Freimark, Susan Kenny, Jack Shostak, and John Troxell. Common Misunderstandings about ADaM Implementation. PharmaSug 2012. San Francisco, CA. May 8 - 11, 2012.
- [6] Songhui Zhu. Designing and Tuning ADaM Datasets. ParmaSUG 2013, Chicago, IL.
- [7] Songhui Zhu. Applying ADaM Standards to Therapeutic Area Ophthalmology. ParmaSUG 2014, San Diego, CA.
- [8] Songhui Zhu. Implementing Various Baselines for ADaM BDS datasets, ParmaSUG 2015, Orlando, FL.

CONTACT INFORMATION

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

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.