PharmaSUG 2015 - Paper DS06

Considerations in ADaM Occurrence Data: Handling Crossover Records for Non-Typical Analysis

Karl Miller, inVentiv Health, Lincoln, Nebraska
Richann Watson, Experis, Batavia, Ohio

ABSTRACT

With the release of the new ADaM Occurrence Data Model for public comment in the first quarter of 2014, the new model is clearly established to encompass adverse events as well as concomitant medications, along with other data into this standard occurrence analysis structure. Commonly used analysis for this type of occurrence structure data can be found utilizing subject counts by category, based on certain criteria (e.g. treatment, cohort, or study period). In most cases, the majority of the analysis data will be in a one-to-one relationship with the source SDTM record.

In this paper, the authors will discuss the creation of ADaM occurrence data for specific cases outside the common or typical analysis where analysis requires a record in SDTM data, which spans across multiple study treatments, periods or phases, to be replicated for inclusion in non-typical analysis with a record being analyzed under multiple study treatments, periods or phases. From the assignment and imputation of key timing variables (i.e. APERIOD, APHASE, ASTDT), through the appropriate derivation of indicator variables and occurrence flags (i.e. ANLzzFL, TRTEMFL, ONTRTFL and AOCCRFL, AOCCPFL, etc.) the authors guide you through the non-typical process in order to maintain efficiency along with ensuring the traceability in the generation of this analysis-ready data structure.

INTRODUCTION

In the planning and creation process of analysis data sets, it is best to work with the end goal (or deliverable) in mind. Doing so helps in maintaining efficiency through:

- Reducing the number of unnecessary variables being created for analysis
- Reducing unnecessary variables being carried over from the collected SDTM data
- Allowing the data sets to be focused on the specific information required to produce the analysis

Conforming to the ADaM standards for analysis data has shown overwhelming support in the process to create data sets that are analysis-ready, while being able to maintain data-point traceability back to the collected study data.

Of the available ADaM data structures, the focus of this paper is based around the occurrence data structure (OCCDS), initially released under the title “Analysis Data Model (ADaM) Data Structure for Adverse Event Analysis” back in 2012. The OCCDS is used for “occurrence analysis or the counting of subjects with a record or term, and often uses a structured hierarchy of dictionary coding categories.”

In terms of types of analysis, we reference two separate analyses: typical and non-typical; both being focused around the summary of subject counts per recorded event (i.e. medical history, concomitant medication, adverse event, etc.) in relation to a study specific treatment or period. For the purpose of this paper, the term ‘recorded event’ will be used to represent the medical history, concomitant medication, adverse event and any other type of occurrence data.

While the typical analysis is built on the onset or emergence of an event relative to the study treatment or period, based off the start date/time of the recorded event in the SDTM data, the non-typical analysis is built around the individual study periods or treatments that may be experienced throughout the duration of the same recorded event.

With the non-typical analysis there is a high potential for multiple treatments/periods occurring during the recorded event, and with this possibility, there exist several ways one could account for this in the analysis data set. Rather than trying to take the approach to create numerous flags or variables, or a combination thereof, we focus on the simple and acceptable approach to produce this non-typical analysis through generation of additional records in the corresponding ADaM data set from the source record in SDTM.

TYPICAL ANALYSIS

Common analysis for normal occurrence data from a clinical study or trial is comprised of the summary of a recorded event through the course of the study. For example, concomitant medications taken throughout the study or treatment-emergence of an adverse events experienced during the course of the study conduct are considered types of typical analysis. For both cases, the focus of analysis remains around the recorded event’s onset or start date/time in relation to the study treatment and/or period for the specific trial population of interest. Regardless of the type of recorded event, the summary table is typically structured with the summary reported event in the first column of the

table followed by the study treatment(s) or study period(s) as you can see in the sample output shell in Output 1 below. The selected layout of the summary table is dependent on the specific need of the analysis and is determined by the study statistician.

```
                                  <Treatment/Period X>     <Treatment/Period X>
Recorded Event                          n (%)                    n (%)
-------------------------------------------------------------------------------
< Recorded Event 1 >                  xx (xx.x%)               xx (xx.x%)

< Recorded Event 2 >                  xx (xx.x%)               xx (xx.x%)

< Recorded Event 3 >                  xx (xx.x%)               xx (xx.x%)

< etc. >

-------------------------------------------------------------------------------
Treatment/Period is assigned to the treatment/period in which the onset/start
date/time of the recorded event occurred.
```

Output 1. Output Table Shell of Summary Analysis (by Subject) per Onset Study Treatment/Period

Since the focus of the typical analysis is based on the onset or start date/time for the subject's recorded event, this creates a true one-to-one relationship between the analysis record and the record coming from the corresponding SDTM domain. This is illustrated with the examples shown in Figure 1 through the display of the collected records and their relationship in regards to the timeline for study conduct.

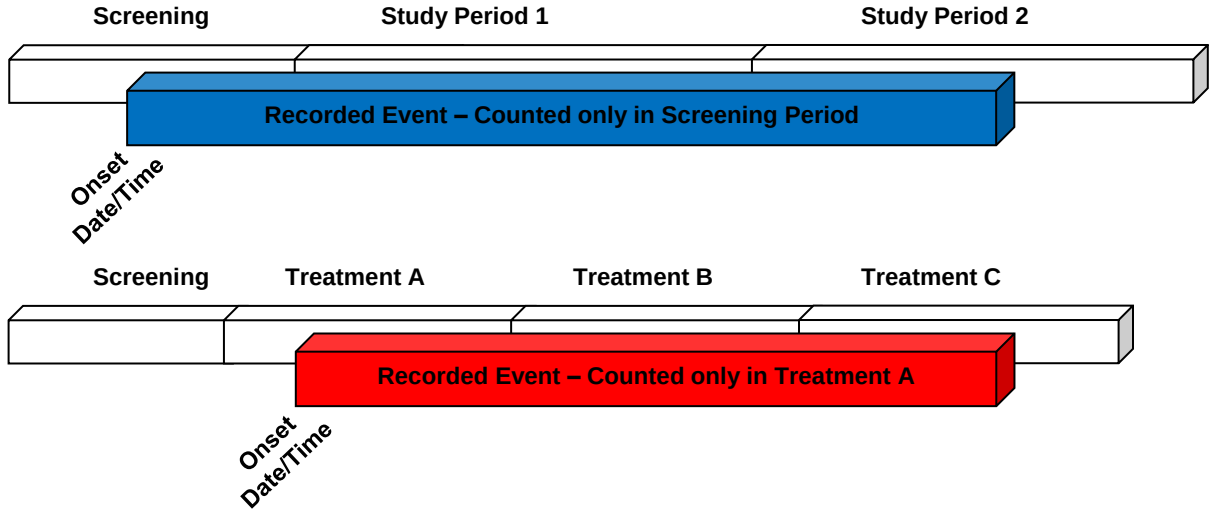

Figure 1. Examples of Recorded Events Spanning Multiple Study Treatments/Periods Counted at Onset

As you can see, based on the two examples above, the generation of the analysis data set is straight forward as it holds a one-to-one relationship with the collected SDTM data due to the focus on the recorded event's onset date/time only, and not on the number of treatments or periods that the recorded event spans. This can also been seen in the example's sample data tables found in Tables 1 and 2, for the first example, and in Tables 3 and 4 in respect to the second example.

CMSEQ	CMDECOD	CMSTDTC	CMENDTC	(cont.)
12	IBUPROFEN	2015-05-14	2015-09-30	…

Table 1. Sample SDTM Data Table for a Recorded Event - Concomitant Medication (CM)

CMSEQ	CMDECOD	CMSTDTC	CMENDTC	APERIOD	(cont.)
12	IBUPROFEN	2015-05-14	2015-09-30	Screening	…

Table 2. Sample ADaM Analysis Data Table for a Recorded Event - Concomitant Medication (ADCM)

USUBJID	AESEQ	AEDECOD	AESTDTC	AEENDTC	(cont.)
002	7	HEADACHE	2015-10-31	2016-01-03	…

Table 3. Sample SDTM Data Table for Recorded Event - Adverse Events (AE)

USUBJID	AESEQ	AEDECOD	AESTDTC	AEENDTC	TRTA	TRTEMFL	APERIOD	(cont.)
002	7	HEADACHE	2015-10-31	2016-01-03	A	Y	1	…

Table 4. Sample ADaM Analysis Data Table for Recorded Event - Adverse Events (ADAE)

NON-TYPICAL ANALYSIS

For the non-typical analysis, the analysis records are still generated from the same study collected data in the SDTM domain that is used for the typical analysis, however the focus for the non-typical analysis shifts from the onset/start date/time of the recorded event to the incidence of the study treatment(s)/period(s) that are encountered within the duration of the recorded event. For the subject recorded events which span multiple study treatment(s)/period(s) during study conduct, this non-typical analysis will require multiple analysis records in the ADaM data set which will be used in analysis and will hold a one-to-many relationship to the original SDTM record, as seen in the example recorded event below in Figure 2.

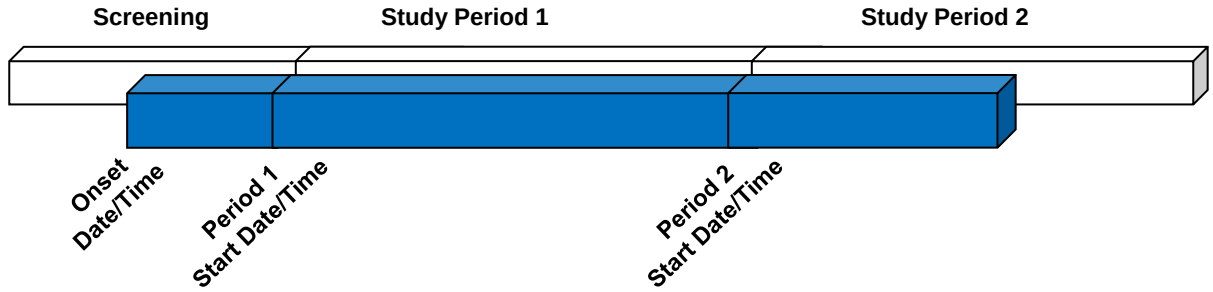

Figure 2. Example of the Division of a Recorded Event Spanning across Multiple Study Periods to be Counted in each Period for Analysis

Based on the example in Figure 2, you can see that the one collected record from the SDTM data would need to have three (3) corresponding records included in the ADaM data set to meet the need of the non-typical analysis, as described in the output shell displayed in Output 2.

```
                                    Screening        Period 1         Period 2
Recorded Event                         n (%)            n (%)            n (%)
-------------------------------------------------------------------------------
< Recorded Event 1 >                xx (xx.x%)       xx (xx.x%)       xx (xx.x%)

< Recorded Event 2 >                xx (xx.x%)       xx (xx.x%)       xx (xx.x%)

< Recorded Event 3 >                xx (xx.x%)       xx (xx.x%)       xx (xx.x%)

< etc. >

-------------------------------------------------------------------------------
Period is assigned to the study period in which the recorded event occurred or
continued to occur in until resolved or the end of study.
```

Output 2. Output Table Shell of Summary Analysis of Recorded Event Incidence per Study Treatment/Period

With the specific direction or scope of the analysis built around the incidence of the recorded event within each study period, the required analysis variable would be APERIOD, which would be assigned its value according to the period start date/time found in ADSL relative to the recorded event. To refrain from including unnecessary variables in the analysis data set, there is no need for the analysis start date and time variables (i.e. ASTDT, ASTTM, and/or ASTDTM), so they would be excluded from the data set as shown in the sample data found in Table 7. As per the ADaM IG, the data found in Table 7 is to be built from the SDTM data found in Table 5 along with the relevant subject-level information for period start dates/times that is located in the corresponding ADSL data set found in Table 6.

USUBJID	CMSEQ	CMDECOD	CMSTDTC	CMENDTC	(cont.)
001	12	IBUPROFEN	2015-05-14	2015-09-30	…

Table 5. Sample SDTM Data Table for Recorded Event - Concomitant Medication (CM)

USUBJID	AGE	ARM	ACTARM	AP01SDT	AP02SDT	(cont.)
001	35	A	A	2015-06-01	2015-08-01	…

Table 6. Sample ADSL Data Table for Subject

USUBJID	CMSEQ	CMDECOD	CMSTDTC	CMENDTC	APERIOD	(cont.)
001	12	IBUPROFEN	2015-05-14	2015-09-30	Screening	…
001	12	IBUPROFEN	2015-05-14	2015-09-30	Study Period 1	…
001	12	IBUPROFEN	2015-05-14	2015-09-30	Study Period 2	…

Table 7. Sample ADaM Analysis Data Table for Recorded Event - Concomitant Medication (ADCM)

Rather than implementing a complex creation of flag variables to mark the study periods in which the recorded event occurred, and keeping the analysis output and its needs in mind, the simple creation of additional rows in the ADaM data set is a straight-forward route to take, and retains appropriate record-level traceability while keeping the ADaM data set analysis-ready.

DATA STRUCTURE FOR BOTH TYPICAL AND NON-TYPICAL ANALYSIS

In specific cases, there may be a need to create an analysis data set that would be able to produce both the typical and non-typical analysis of the study data. In the case of having the need for both analyses, you would want to structure the ADaM data set in order to maintain the necessary analysis information for the typical analysis as well as the non-typical analysis, while at the same time being able to distinguish between records used for each analysis.

Based off of these separate analysis needs, one would need to create an analysis data set that would retain the relevant event information collected for the initial event captured in SDTM, as well as add records for any subsequent treatments/periods that might occur throughout the duration of the event. Figure 3 gives an accurate visual representation of the analysis need for a recorded event according to this scenario.

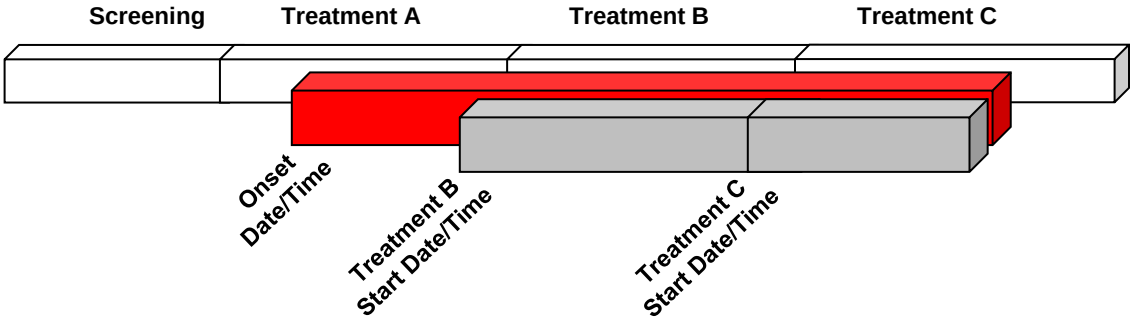

Figure 3. Example Division by Treatment of a Recorded Event Used for Typical and Non-Typical Analysis

For the example in Figure 3, the first bar (marked in red) represents the record used in the typical analysis, as well as for the first treatment in the non-typical analysis. The second bar (marked in gray) represents the additional records to be added for the subsequent treatments used in the non-typical analysis. Along with this division of the record in the analysis data set, there is an a need to keep the original start and stop date/time information from SDTM for an analysis record based on the additional requirement for the analysis of duration of the recorded events on study. For the duration (ADURN) used in this additional analysis, we would have to set the ASTDT/AENDT to the original AESTDTC/AEENDTC, imputing any part of the date/time according to the statistical analysis plan (SAP) as needed. This record can also serve the analysis need for the typical analysis as defined in the SAP.

Additionally, you would add two extra records for the subsequent treatments/periods that the recorded event occurred in for the non-typical analysis needs. Only two records are required, since the record which is used in the typical analysis can and will be used in the non-typical analysis. For all cases, the number of additional records needed in the ADaM data set is dependent on the study design and analysis needs, along with the total number of treatment(s)/period(s) encountered in the duration of the recorded event.

Handling of the analysis date(s)/time(s) is dependent on the analysis needs as there is the potential for many different cases; however, in this example there is no additional need for assigning or deriving any analysis date/time for the additional records.

Another challenge of generating an ADaM data set that can be used for multiple analyses is the correct assignment of records for each appropriate analysis. One variable that is designed to be used for this purpose is the analysis record flag (ANLzzFL). For our example, there are two (2) analysis needs:

1. Data records to be used in the typical analysis which can be used to retain any additional information from the SDTM record for other appropriate analysis needs.
2. Data records to be used in the non-typical analysis that indicate the treatment(s)/period(s) in which the recorded event spanned.

By designating the analysis record flag to mark the records used for the typical analysis (e.g. ANL01FL='Y') we can easily highlight records only to be used for that specific analysis programming, while still being able to keep the non-typical analysis records in the data set. What may not be evident is why there is a need for only one analysis record flag in the data set, since there are two types of analysis. For a non-typical analysis, all of the records in the data set will be used; therefore, it is not necessary to set a flag for each record (i.e. ANL02FL='Y' for all records). So with the use of just one single analysis flag variable, we can separate records for the two different analyses without any additional data steps required in the table programming.

USUBJID	AESEQ	AEDECOD	AESTDTC	AEENDTC	(cont.)
002	7	HEADACHE	2015-10-31	2016-01-03	…

Table 8. Sample SDTM Data Table for Recorded Event - Adverse Events (AE)

USUBJID	AGE	ARM	ACTARM	TR01SDT	TR02SDT	TR03SDT	(cont.)
002	32	ABC	ABC	2015-10-10	2015-11-11	2015-12-12	…

Table 9. Sample ADSL Data Table for Subject

	USUBJID	AESEQ	AEDECOD	AESTDTC	AEENDTC	ASTDT	AENDT	ADURN	(cont.)
1	002	7	HEADACHE	2015-10-31	2016-01-03	31OCT2015	03JAN2016	63	…
2	002	7	HEADACHE	2015-10-31	2016-01-03				…
3	002	7	HEADACHE	2015-10-31	2016-01-03				…

	TRTA	TRTEMFL	APERIOD	ANL01FL	(cont.)
1	A	Y	1	Y	…
2	B		2		…
3	C		3		…

(Note: Shaded columns are for display purposes to show corresponding rows within the analysis data set)

Table 10. Sample ADaM Analysis Data Table for Recorded Event - Adverse Events (ADAE)

CONCLUSION

When you are tasked with generation of any ADaM data set, it is always best to start with the end in mind and work your way back to how the data set best fits the needs of the analysis. In the creation of the data set(s), it is also best to keep things as simple as possible.

One of the most over-looked aspects of keeping things simple, which is stated directly in the ADaM IG, is the case where generating additional records from one SDTM record in the ADaM data set meets the requirement for analysis. In our experience and opinion, adding records, along with the use of the analysis record flags, has been found to be clearer and more beneficial than coming up with a complicated set of flag variables in order to re-use one record multiple times, for multiple treatments and/or study periods, in the programming of TFL analyses.

An added benefit of implementing the use of multiple records with the analysis-level flag is the efficiency gained, not only in the data set programming, but in the TFL programming as well, since there will be no need to do additional data manipulation in the TFL program. This not only allows the data set to be CDISC compliant by providing the necessary traceability, it also has the additional key component of being analysis-ready. Although it may not be a one-to-one relationship, the record level traceability is made available through the continued use of the initial SDTM sequence variable (--SEQ), and the SDTM record date and time (--DTC, --STDTC, --ENDTC), along with any additional SDTM variables that are retained in the ADaM data set that will be beneficial to the analysis.

Last, but definitely not least of the efficiencies gained, there is the additional cost/benefit of being able to program and maintain one single data set rather than all the time and effort that goes in to programming and maintaining multiple data sets to generate the analysis.

Simply put, where one might need to use a single SDTM record for multiple analyses, adding additional records along with the use of the analysis record flags maximizes the analysis-ready capabilities for your data set while maintaining the appropriate level of traceability between your SDTM and ADaM records for submission.

ACKNOWLEDGMENTS

We would be remiss without thanking Nancy Brucken and Paul Slagle for letting us harass them and pick their brains for their experience in applied CDISC standards.

CONTACT INFORMATION

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

Karl Miller
inVentiv Health
karl.miller@inventivhealth.com

Richann Watson
Experis
richann.watson@experis.com