Paper 063-2022

Child Data Set: An Alternative Approach for Analysis of Occurrence and Occurrence of Special Interest

Lindsey Xie, Jinlin Wang, Lauren Xiao, Kite Pharma, Inc.

Richann Watson, DataRich Consulting

ABSTRACT

Due to the various data needed for safety occurrence analyses, the use of a child data set that contains all the data for a given data point aids in traceability and support of the analysis.

Adverse Events of Special Interest (AESI) represents adverse events (AEs) that are of particular interest in the study. These AESI could potentially have symptoms associated with them. AESI could be captured as clinical events (CEs) in the CE domain while the associated symptoms for each CE are captured as AEs in the AE domain.

The relationship between CEs and associated AE symptoms are an important part of the safety profile for a compound in clinical trials. These relationships are not always readily evident in the source data or in a typical AE analysis data set (ADAE). The use of a child data set can help demonstrate this relationship, which provides enhanced data traceability to help the review effort for both sponsor and regulatory agency reviewers.

This paper provides examples of using a child data set to preserve data with the relationship between CEs and associated AE symptoms in an ADAE child data set (ADAE<si>). In addition, the paper will show that ADAE and ADAE<si> serve as analysis-ready data sets for the summary of AE and AESI.

INTRODUCTION

What is AESI? Per the U.S. Food & Drug Administration, “An adverse event of special interest (serious or nonserious) is one of scientific and medical concern specific to the sponsor’s product or program, for which ongoing monitoring and rapid communication by the investigator to the sponsor can be appropriate. Such an event might warrant further investigation in order to characterize and understand it. Depending on the nature of the event, rapid communication by the trial sponsor to other parties (e.g., regulators) might also be warranted. (Based on CIOMS VI)” (Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 2020)

The purpose of collecting safety data from clinical trials is to monitor important safety signals, protect patients from unnecessary risks, and develop the safety profile of the drug contributing to its benefit-risk assessment. Adverse events are collected by an Adverse Event Case Report Form (CRF) page and reported, individually. However, a general reporting of adverse events sometimes may not be sufficient for the monitoring of the treatment’s safety as it may fail to reveal important risks. AESI are adverse events defined by the sponsor as being of special interest in a given clinical study. These are usually reported separately to support safety assessment of a new drug.

AESI can be identified in the following ways:

- using Medical Dictionary for Regulatory Activities (MedDRA) preferred terms (PTs) via
  - pre-defined Standardized MedDRA Queries (SMQs) that can be extracted using MedDRA PTs
  - Customized MedDRA Queries (CQs) specific to a study but are not pre-defined within MedDRA PTs
  - adjudication process with medical/clinical confirmation by providing a list of MedDRA PTs or derivation logic to select specific AEs
- capturing data on separate CRF pages

AESI of SMQs, CQs and adjudicated PT group usually are identified in the AE analysis data set. AESI collected by separate CRF pages are considered clinical event data. On clinical event CRF page, the linkage between clinical event and AE symptoms may be recorded so the relationship between the clinical event and AE symptoms can be reported as safety assessment of AESI for the clinical trial. Furthermore, AESI

captured as a clinical event that are related to group of AE symptoms may have its own measurement of severity, grading, or other items.

WALK THROUGH OF SDTM DATA

AE symptoms on AE CRF are captured in AE domain while AESI events from clinical event CRF are captured in CE domain. In this paper, Clinical Event 1 and Clinical Event 2 representing AESI events are used as examples for demonstration purposes. In addition, for easy illustration, all AE and clinical event data are from the same subject. Data for Clinical Event 1 and Clinical Event 2 are collected on separate CRF pages. SUPPCE stores the 'SYMPTOM' (AEs) log number(s) associated with that 'EVENT' and any additional information captured on the CRF that pertains to that 'EVENT' and does not fit in the parent domain.

Data Display 1 is the data example of SDTM AE domain containing data of AEs collected on CRFs.

Row	AESEQ	AESPID	AETERM	AESTDTC	AEENDTC	AETOXGR	AESER
1	20	1	ATRIAL FIBRILLATION WITH RVR	2016-05-19	2016-05-22	3	Y
2	2	2	COUGH (DRY, INTERMITTENT)	2016-05-11	2016-05-15	2	N
3	1	3	RIGHT MCP JOINT TENDERNESS	2016-05-04	2016-05-11	1	N
4	7	4	**FATIGUE**	2016-05-15	2016-06-02	1	N
5	13	9	DYSPHASIA	2016-05-17	2016-05-23	1	Y
6	14	11	SOMNOLENCE	2016-05-17	2016-05-23	2	N
7	12	12	HYPOXIA	2016-05-17	2016-05-24	3	Y
8	18	13	HYPOTENSION	2016-05-19	2016-05-20	3	N
9	37	17	MEMORY IMPAIRMENT	2016-05-24	2016-05-25	2	N
10	41	18	CONFUSION	2016-05-25	2016-06-03	2	N
11	9	26	FEVER	2016-05-16	2016-05-19	2	N
12	25	27	FEVER	2016-05-20	2016-05-21	1	N
13	16	50	ENCEPHALOPATHY	2016-05-17	2016-06-03	2	Y
14	51	51	WBC DECREASED	2016-06-14	2016-06-17	4	N

ROW	AEDECOD	AEBODSYS
1	Atrial fibrillation	Cardiac disorders
2	Cough	Respiratory, thoracic and mediastinal disorders
3	Arthralgia	Musculoskeletal and connective tissue disorders
4	Fatigue	General disorders and administration site conditions
5	Aphasia	Nervous system disorders
6	Somnolence	Nervous system disorders
7	Hypoxia	Respiratory, thoracic and mediastinal disorders
8	Hypotension	Vascular disorders
9	Memory impairment	Nervous system disorders
10	Confusional state	Psychiatric disorders
11	Pyrexia	General disorders and administration site conditions
12	Pyrexia	General disorders and administration site conditions
13	Encephalopathy	Nervous system disorders
14	White blood cell count decreased	Investigations

Data Display 1 SDTM AE domain data example

Data Display 2 and Data Display 3 are data examples of SDTM CE domain and SUPPCE supplement domain. From Data Display 2, we can see that two clinical events, Clinical Event 1 and Clinical Event 2, were observed. Each clinical event identified by CETERM can consist of multiple episodes with episode numbers represented by CE.CESPID and changed severities stored in CE.CETOXGR. Like AE or other occurrence data, clinical event start date (CE.CESTDTC) or end date (CE.CEENDTC) could be partially missing. CEENDTC in Row 5 of Data Display 2 has a partial missing date. An imputed date for "2016-06" can be created in a derived date variable in the analysis data set following the pre-defined imputation rule.

ROW	CESEQ	CECAT	CESPID	CETERM	CEDECOD	CESTDTC	CEENDTC	CETOXGR	CESER
1	1	CE1	1	CLINCIAL EVENT 1	Clinical Event 1	2016-05-16	2016-05-17	2	Y
2	2	CE1	2	CLINCIAL EVENT 1	Clinical Event 1	2016-05-18	2016-05-18	*1*	N
3	3	CE1	3	CLINCIAL EVENT 1	Clinical Event 1	2016-05-19	2016-05-24	*3*	
4	4	CE2	1	CLINCIAL EVENT 2	Clinical Event 2	2016-05-17	2016-05-23	1	N
5	5	CE2	2	CLINCIAL EVENT 2	Clinical Event 2	2016-05-24	2016-06	*2*	

Data Display 2 SDTM CE domain data example

In Data Display 3, SUPPCE.QNAM as 'CEAENO' refers to AE numbers captured on clinical event CRF page for AE symptoms that were related to each clinical event episode. SUPPCE.IDVARVAL is the clinical event episode identifier tracing back to CE.CESEQ. SUPPCE.QVAL contains a list of AE numbers, separated by a comma, that link the clinical event to its related AE symptoms captured on the AE CRF page. Some AE symptoms may be related to a single episode of a clinical event. For example, AE symptoms ATRIAL FIBRILLATION WITH RVR and HYPOTENSION (AESPID 1, 13) are related to the episode of Clinical Event 1 with CESPID=3. Some other AE symptoms can be ongoing through multiple episodes of a clinical event like the case for HYPOXIA (AESPID 12) and FEVER (AESPID 26) in Clinical Event 1. And some AE symptoms can be associated to more than one clinical event. FATIGUE (AESPID 4 in Data Display 1) is a perfect example for this case that went through all episodes for both Clinical Event 1 and Clinical Event 2. Besides having related AE numbers map to SUPPCE.QVAL where SUPPCE.QNAM is 'CEAENO', SUPPCE can also have additional data, such as Depressed Level of Consciousness, map to SUPPCE.QVAL where SUPPCE.QNAM is 'LVLCONS'. In this scenario, this data is collected on Clinical Event 2 CRF. However, it does not have a place within the CE domain, therefore it is stored in SUPPCE, as shown in Data Display 3.

IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
CESEQ	1	CEAENO	AE Log Lines Reported on the CE Episode	4,12,26
CESEQ	2	CEAENO	AE Log Lines Reported on the CE Episode	4,12,26
CESEQ	3	CEAENO	AE Log Lines Reported on the CE Episode	1,4,12,13,26, 27
CESEQ	4	CEAENO	AE Log Lines Reported on the CE Episode	4,9,11,50
CESEQ	4	LVLCONS	Depressed Level of Consciousness	Awakens to voice
CESEQ	5	CEAENO	AE Log Lines Reported on the CE Episode	4,17,18,50
CESEQ	5	LVLCONS	Depressed Level of Consciousness	Awakens spontaneously

Data Display 3 SDTM SUPPCE supplement domain data example

CESPID contains AESI clinical event log number for each clinical event. CESEQ is consecutive sequence number created for CE/SUPPCE domains based on the sorting by CECAT and CESTDTC. Data mapped into SUPPCE domain can be put back to the same episode record through CE and SUPPCE merging by CESEQ (Data Display 4).

ROW	CESEQ	CECAT	CESPID	CESTDTC	CEENDTC	CETOXGR	CESER	CEAENO	LVLCONS
1	1	CE1	1	2016-05-16	2016-05-17	2	Y	4,12,26	
2	2	CE1	2	2016-05-18	2016-05-18	*1*	N	4,12,26	
3	3	CE1	3	2016-05-19	2016-05-24	*2*		1,4,12,13,26,27	
4	4	CE2	1	2016-05-17	2016-05-23	1	N	4,9,11,50	Awakens to voice
5	5	CE2	2	2016-05-24	2016-06	*2*		4,17,18,50	Awakens spontaneously

Data Display 4 Data example of SDTM CE merged with SUPPCE

REPORTS TO BE PRODUCED

Now we want to describe the table shells for producing two types of safety reports, Treatment-Emergent Adverse Event (TEAE) and TE AESI summary.

Output 1 is what we typically see in our safety report. It is a standard AE summary table to summarize TEAE incidence with toxicity grade grouped by System Organ Class (SOC), PT. This table is to show risk signs of TEAE via each SOC and PT. For this table, we have the following items on the to-do list.

- obtain TEAE data for each subject
- identify the worst toxicity grade among all TEAEs for each subject
- identify the worst toxicity grade among all TEAEs of each SOC for each subject
- identify the worst toxicity grade among all TEAEs of each PT within a SOC for each subject

MedDRA SOC Term, n (%) Preferred Term, n (%)	Any Grade	Worst Grade 1	Worst Grade 2	Worst Grade 3	Worst Grade 4	Worst Grade 5
Subjects with Any Treatment-emergent Adverse Event	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
SOC Term 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 3	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
…..						

Output 1 Table Shell of Treatment-emergent Adverse Events by System Organ Class, Preferred Term and Worst Grade

Output 2 is the summary of TE Clinical Events, a AESI summary report with toxicity grade grouped by subjects who have a clinical event with associated TEAE symptoms by PT. This table is to show risk signs of TE clinical events with associated AE symptoms. These AE symptoms individually may have not indicated any risk sign when being reported separately. For example, FATIGUE from the example of Data Display 1 would have not shown a safety concern when it is reported in Output 1. However, when FATIGUE, HYPOXIA, HYPOTENSION, and FEVER (AESPID=4, 12, 13, 26, and 27) occurred within the same timeframe and contributed to Clinical Event 1 (CESEQ=3 and QNAM='CEAENO' in Data Display 3), then they could be a potential safety concern for Clinical Event 1, as seen in Output 2.

For Output 2, our to-do items are as summarized below.

- obtain TE clinical event data for each subject
- identify the worst toxicity grade among all episodes of a clinical event for each subject
- obtain associated symptoms, the TEAE data with PTs, that are linked to each individual clinical event for each subject
- identify the worst toxicity grade among all TEAE records of each related PT for each subject

Event, n (%)	Any Grade	Worst Grade 1	Worst Grade 2	Worst Grade 3	Worst Grade 4	Worst Grade 5
Subjects with Any Clinical Event	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Clinical Event by Preferred Terms						
PT Term 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
…..						

Output 2 Summary of AESI: Treatment-emergent Clinical Event by Preferred Term and Worst Grade

Notice that this table is similar to Output 1. However, a key difference is that in a regular AE Summary table (Output 1), a PT is associated with a single SOC and reported once, while in the AESI summary a PT may be associated with multiple events and would be summarized in each clinical event (Output 2). In our data example, FATIGUE would be reported in both Clinical Event 1 and Clinical Event 2.

Then how do we setup the analysis data set to support such a summary for AESI reporting? One approach that directly comes to mind is through data merge of ADAE and CE by each clinical event in the output programs. However, by doing so, the data process will be hard to trace back, which causes the issue with data traceability. In addition, the source data ADAE and CE are not analysis-ready data sets. A solution to preserve data traceability and to have ADaM data set analysis ready is to use a child data set for AESI reporting.

A child data set is a data set that combines various data from parent data sets. In our case, a child data set for AESI events would contain clinical event data and related AE symptom data as clustered data points to preserve the relationship between a clinical event and its associated AE symptoms. The goal of using the child data set for AESI analysis is to make the data set analysis ready. Before we go deeper into the concept of a child data set, let's refresh our knowledge with analysis-ready ADaM data sets.

ANALYSIS-READY ADAM DATA SET

There are five fundamental principles that analysis data sets must adhere to in order to be considered an ADaM data set.

- Clear and unambiguous communication
- Traceability
- Use of metadata
- Analysis ready
- Use of commonly available software

While most of these fundamental principles are straightforward, the idea of analysis-ready data sets may warrant further explanation. In the ADaM Implementation Guide (ADaMIG) it states:

> ADaM datasets should have a structure and content that allow statistical analyses to be performed with minimal programming. Such datasets are described as "analysis-ready." ADaM datasets contain the data needed for the review and re-creation of specific statistical analyses. It is not necessary to collate data into analysis-ready datasets solely to support data listings or other non-analytical displays. (CDISC Analysis Data Model Team, 2021)

Essentially, an analysis-ready data set means ready to produce a result. Outputs can have several results that are produced from a single data set or from multiple data sets. The key is that the result is generated by executing a procedure on the data set as-is with little to no manipulation.

But what is meant by 'little to no manipulation' or 'minimal programming'? A statement that can be used to subset the data set to obtain the appropriate records is considered no manipulation of the data or minimal programming. Sorting the data prior to using it in a procedure to produce the result could be classified as little manipulation and minimal programming. A simple joining of two data sets is also considered minimal programming. For example, joining ADSL with an SDTM domain in order to have all the necessary

treatment, covariates or subgrouping variables in the ADaM data set so that the variables are readily available when the procedure is executed.

If there are complex data manipulations involved such as altering values or removing records based on derivation rules that are not readily defined by variables in the ADaM data set, then the analysis data set is not analysis ready. In order to make it an analysis-ready data set, these complex data manipulations should be handled within the data set programming.

It is these complex data manipulations that lead to the approach taken to handle AESI.

UTILIZING CDISC OCCURRENCE DATA STRUCTURE IMPLEMENTATION GUIDE

When dealing with data that is used for counting, ADaM Structure for Occurrence Data (OCCDS) Implementation Guide v1.1 has pre-defined concepts that can help with preparing your data set to be analysis ready. Some of these concepts are 'U' prefix for stacked data, occurrence flags to identify specific records that should be used for counting and customized queries to illustrate the relationship of an AE to an AESI. Before discussing the 'U' prefix and the occurrence flags, we first need to identify the AEs associated with the clinical events.

CUSTOMIZED QUERIES FOR IDENTIFYING AESI

With analyses on AESI becoming more common, it is important to be able to easily identify which AEs are ones of special interest. These AESIs can be the event itself or symptoms of the event. To identify these AESI, Standardized MedDRA Queries (SMQ) or Customized Queries (CQ) can be used. SMQ are pre-defined groupings of MedDRA terms that help define a specific medical condition (International Conference on Harmonization, 2022). CQ are sponsor-defined groupings. In this paper, CQs are used to identify which AEs are associated with the event of interest.

OCCDS has only one variable convention defined for customized queries, CQzzNAM. It uses the zz index which must be 01 – 99, zero filled, and does not have to start at 01 nor does it have to be sequential. The label can also be modified as long as the portion that has been defined in the OCCDS is not modified. For example, CQ01NAM may represent CLINICAL EVENT 1. The label could be 'Customized Query 01 Name – CE1'.

Recall that there are two clinical events in Data Display 2 SDTM.CE and Data Display 3 SDTM.SUPPCE data examples. Per Data Display 3, AE symptoms associated with Clinical Event 1 are AEs with AE log numbers of 1, 4, 12, 13, 26, and 27 while AE symptoms associated with Clinical Event 2 are AEs with AE log numbers of 4, 9, 11, 17, 18, and 50. All these AE numbers trace back to ADAE, which allows for the population of ADAE.CQzzNAM. In ADAE data example (Data Display 5, see below), CQ01NAM is populated as "CLINICAL EVENT 1" for AEs with AESPID of 1, 4, 12, 13, 26, and 27 and CQ02NAM is populated as "CLINICAL EVENT 2" for AEs with AESPID of 4, 9, 11, 17, 18 and 50. Notice that both ADAE.CQ01NAM and ADAE.CQ02NAM are populated for FATIGUE as this AE is associated with both clinical events. Note that the variable APERIOD is usually derived in ADAE to support by analysis period summary. In Data Display 5, Row 1 – Row 13 are AEs that occurred in Analysis Period 1 and Row 14 is the AE that occurred in Analysis Period 2.

Row	AESEQ	AESPID	AETERM	AESTDTC	AEENDTC	APERIOD	AETOXGR	AESER
1	20	1	ATRIAL FIBRILLATION WITH RVR	2016-05-19	2016-05-22	1	3	Y
2	2	2	COUGH (DRY, INTERMITTENT)	2016-05-11	2016-05-15	1	2	N
3	1	3	RIGHT MCP JOINT TENDERNESS	2016-05-04	2016-05-11	1	1	N
4	7	4	**FATIGUE**	2016-05-15	2016-06-02	1	1	N
5	13	9	DYSPHASIA	2016-05-17	2016-05-23	1	1	Y
6	14	11	SOMNOLENCE	2016-05-17	2016-05-23	1	2	N
7	12	12	HYPOXIA	2016-05-17	2016-05-24	1	3	Y
8	18	13	HYPOTENSION	2016-05-19	2016-05-20	1	3	N
9	37	17	MEMORY IMPAIRMENT	2016-05-24	2016-05-25	1	2	N
10	41	18	CONFUSION	2016-05-25	2016-06-03	1	2	N
11	9	26	FEVER	2016-05-16	2016-05-19	1	2	N
12	25	27	FEVER	2016-05-20	2016-05-21	1	1	N
13	16	50	ENCEPHALOPATHY	2016-05-17	2016-06-03	1	2	Y
14	51	51	WBC DECREASED	2016-06-14	2016-06-17	2	4	N

Row	AEDECOD	AEBODSYS	CQ01NAM	CQ02NAM
1	Atrial fibrillation	Cardiac disorders	CLINCIAL EVENT 1	
2	Cough	Respiratory, thoracic and mediastinal disorders		
3	Arthralgia	Musculoskeletal and connective tissue disorders		
4	**Fatigue**	General disorders and administration site conditions	**CLINCIAL EVENT 1**	**CLINCIAL EVENT 2**
5	Aphasia	Nervous system disorders		CLINCIAL EVENT 2
6	Somnolence	Nervous system disorders		CLINCIAL EVENT 2
7	Hypoxia	Respiratory, thoracic and mediastinal disorders	CLINCIAL EVENT 1	
8	Hypotension	Vascular disorders	CLINCIAL EVENT 1	
9	Memory impairment	Nervous system disorders		CLINCIAL EVENT 2
10	Confusional state	Psychiatric disorders		CLINCIAL EVENT 2
11	Pyrexia	General disorders and administration site conditions	CLINCIAL EVENT 1	
12	Pyrexia	General disorders and administration site conditions	CLINCIAL EVENT 1	
13	Encephalopathy	Nervous system disorders		CLINCIAL EVENT 2
14	White blood cell count decreased	Investigations		

Data Display 5 Customized queries CQzzNAM in ADAE data example

Once the AEs associated with the clinical events are identified in the parent data set, ADAE, via CQzzNAM, we are ready for the next step. We can create a child data set from the two parent data sets, ADAE and CE, by using variables that have a 'U' prefix.

'U' PREFIX FOR STACKED DATA

With the release of the OCCDS v1.1 came the introduction of the 'U' prefix. When there are multiple data sources that have similar variable naming conventions, the SDTM variable names in the source data sets cannot be used for analysis purposes since the data would be spread across multiple variables. For example, combining data from ADAE and CE, the preferred term would be captured in AEDECOD for ADAE data and captured in CEDECOD for CE data. When the preferred term is in two different variables, the data set is not analysis ready. Therefore, the prefix 'U' could be used to replace the two-letter domain prefix. (CDISC Analysis Data Model Team, 2021)

'U' represents data that is unmodified from the data source. It is used to combine data from variables that have the same root across the multiple data sources. Thus, instead of using AEDECOD and CEDECOD, UDECOD would be used. Table 1 contains examples of other variables from ADaM.ADAE and SDTM.CE with the same root where the 'U' prefix could be used instead of the original SDTM variables.

Adverse Events (ADaM.ADAE)		Clinical Events (SDTM.CE)	
AETERM	Reported Term for the Adverse Event	CETERM	Reported Term for the Clinical Event
AEDECOD	Dictionary-Derived Term	CEDECOD	Dictionary-Derived Term
AEBODSYS	Body System or Organ Class	CEBODSYS	Body System or Organ Class
AEPRESP	Pre-Specified Adverse Event	CEPRESP	Clinical Event Pre-specified
AESTDTC	Start Date/Time of Adverse Event	CESTDTC	Start Date/Time of Clinical Event
AEENDTC	End Date/Time of Adverse Event	CEENDTC	End Date/Time of Clinical Event
AESPID	Adverse Event Number	CENUM	Clinical Event Number

ADaM.ADAE + SDTM.CE	
UTERM	Reported Term
UDECOD	Dictionary-Derived Term
UBODSYS	Body System or Organ Class
UPRESP	Pre-specified
USTDTC	Start Date/Time
UENDTC	End Date/Time
UNUM	Event Number

Table 1 ADaM.ADAE and SDTM.CE Variables with Same Root Stacked to Create 'U' Variables

In the scenario illustrated in this paper, AESIs are captured in SDTM CE domain, and all the symptoms related to the AESIs are captured in SDTM AE domain. Since we want to build in the linkage of AE symptoms to clinical events using CQzzNAM, we use ADAE and CE as parent data sets. We stack related AE symptom data from ADAE based on CQzzNAM values and clinical event data from SDTM.CE into a child data set ADAE<si> where the 'U' prefix allows for like variables to be combined into one. Note for simplicity we are using CE1 and CE2 to represent <si>.

Along with the pre-defined variable, CQzzNAM, for flagging AE symptoms associated with clinical events in ADAE (Data Display 5), the sponsor-defined variable, CQzzRCID, is also introduced in the child data set ADAE<si> where CQzzRCID is equal to the value of UNUM from CE data (SRCDOM='CE') which would trace back to SDTM.CE.CENUM.

In data example ADAECE1 (Data Display 6) and ADAECE2 (Data Display 7, see below), records with SRCDOM ='CE' are from SDTM.CE and records with SRCDOM as 'ADAE' are from ADAE where CQ01NAM is "CLINICAL EVENT 1" for Clinical Event 1 or CQ02NAM is "CLINICAL EVENT 2" for Clinical Event 2, respectively. In ADAECE1 and ADAECE2, CE data and ADAE data are also categorized by ACAT1 with "EVENT" for CE data and "SYMPTOM" for ADAE data. The combination of CQ01RCID/CQ02RCID and ACAT1 assembles the cluster of data points for each individual clinical event episode and associated AE symptoms.

Row	SRCDOM	SRCSEQ	UNUM	UTERM	USTDTC	UENDTC	ASTDT	AENDT
1	***CE***	***1***	***1***	***CLINICAL EVENT 1***	***2016-05-16***	***2016-05-17***	***16-May-16***	***17-May-16***
2	ADAE	7	4	**FATIGUE**	2016-05-15	2016-06-02	15-May-16	02-Jun-16
3	ADAE	12	12	HYPOXIA	2016-05-17	2016-05-24	17-May-16	24-May-16
4	ADAE	9	26	FEVER	2016-05-16	2016-05-19	16-May-16	19-May-16
5	***CE***	***2***	***2***	***CLINICAL EVENT 1***	***2016-05-18***	***2016-05-18***	***18-May-16***	***18-May-16***
6	ADAE	7	4	**FATIGUE**	2016-05-15	2016-06-02	15-May-16	02-Jun-16
7	ADAE	12	12	HYPOXIA	2016-05-17	2016-05-24	17-May-16	24-May-16
8	ADAE	9	26	FEVER	2016-05-16	2016-05-19	16-May-16	19-May-16
9	***CE***	***3***	***3***	***CLINICAL EVENT 1***	***2016-05-19***	***2016-05-24***	***19-May-16***	***24-May-16***
10	ADAE	20	1	ATRIAL FIBRILLATION WITH RVR	2016-05-19	2016-05-22	19-May-16	22-May-16
11	ADAE	7	4	**FATIGUE**	2016-05-15	2016-06-02	15-May-16	02-Jun-16
12	ADAE	12	12	HYPOXIA	2016-05-17	2016-05-24	17-May-16	24-May-16
13	ADAE	18	13	HYPOTENSION	2016-05-19	2016-05-20	19-May-16	20-May-16
14	ADAE	9	26	FEVER	2016-05-16	2016-05-19	16-May-16	19-May-16
15	ADAE	25	27	FEVER	2016-05-20	2016-05-21	20-May-16	21-May-16

Row	CQ01RCID	ACAT1	UDECOD	UBODSYS	UTOXGR	ATOXGR	USER	ASER
1	***1***	***EVENT***	***Clinical Event 1***		***2***	***2***	***Y***	***Y***
2	1	SYMPTOM	**Fatigue**	General disorders and administration site conditions	1	1	N	N
3	1	SYMPTOM	Hypoxia	Respiratory, thoracic and mediastinal disorders	3	3	Y	Y
4	1	SYMPTOM	Pyrexia	General disorders and administration site conditions	2	2	N	N
5	***2***	***EVENT***	***Clinical Event 1***		***1***	***1***	***N***	***N***
6	2	SYMPTOM	**Fatigue**	General disorders and administration site conditions	1	1	N	N
7	2	SYMPTOM	Hypoxia	Respiratory, thoracic and mediastinal disorders	3	3	Y	Y
8	2	SYMPTOM	Pyrexia	General disorders and administration site conditions	2	2	N	N
9	***3***	***EVENT***	***Clinical Event 1***		***3***	***3***		***Y***
10	3	SYMPTOM	Atrial fibrillation	Cardiac disorders	3	3	Y	Y
11	3	SYMPTOM	**Fatigue**	General disorders and administration site conditions	1	1	N	N
12	3	SYMPTOM	Hypoxia	Respiratory, thoracic and mediastinal disorders	3	3	Y	Y
13	3	SYMPTOM	Hypotension	Vascular disorders	3	3	N	N
14	3	SYMPTOM	Pyrexia	General disorders and administration site conditions	2	2	N	N
15	3	SYMPTOM	Pyrexia	General disorders and administration site conditions	1	1	N	N

Data Display 6 Data Example of Child dataset ADAECE1 for Clinical Event 1

A few other things worth mentioning.

- In Row 6 of Data Display 7, UENDTC has a partial missing date “2016-06”. Its imputed value “30-Jun-16” is captured in AENDT.
- In Row 9 of Data Display 6 and Row 6 of Data Display 7 (see below), USER is blank, and ASER is ‘Y’. In this case, whether the clinical event is serious was not answered on the CRF but was confirmed thorough medical review for analysis report (or may be imputed based on imputation rules).
- Row 1 and Row 6 of Data Display 7 show the additional data of Depressed Level of Consciousness that is collected on Clinical Event 2 CRF and captured in SUPPCE.QVAL where

SUPPCE.QNAM is 'LVLCONS'. The additional data is populated as a separate variable, ADAECE2.LVCONS, to support additional analysis as needed.

ROW	SRCDOM	SRCSEQ	UNUM	UTERM	USTDTC	UENDTC	ASTDT	AENDT
1	***CE***	***1***	***1***	***CLINICAL EVENT 2***	***2016-05-17***	***2016-05-23***	***17-May-16***	***23-May-16***
2	ADAE	7	4	**FATIGUE**	2016-05-15	2016-06-02	15-May-16	02-Jun-16
3	ADAE	13	9	DYSPHASIA	2016-05-17	2016-05-23	17-May-16	23-May-16
4	ADAE	14	11	SOMNOLENCE	2016-05-17	2016-05-23	17-May-16	23-May-16
5	ADAE	16	50	ENCEPHALOPATHY	2016-05-17	2016-06-03	17-May-16	03-Jun-16
6	***CE***	***2***	***2***	***CLINICAL EVENT 2***	***2016-05-24***	***2016-06***	***24-May-16***	***30-Jun-16***
7	ADAE	7	4	**FATIGUE**	2016-05-15	2016-06-02	15-May-16	02-Jun-16
8	ADAE	37	17	MEMORY IMPAIRMENT	2016-05-24	2016-05-25	24-May-16	25-May-16
9	ADAE	41	18	CONFUSION	2016-05-25	2016-06-03	25-May-16	03-Jun-16
10	ADAE	16	50	ENCEPHALOPATHY	2016-05-17	2016-06-03	17-May-16	03-Jun-16

ROW	CQ02RCID	ACAT1	UDECOD	UBODSYS
1	***1***	***EVENT***	***Clinical Event 2***	
2	1	SYMPTOM	**Fatigue**	General disorders and administration site conditions
3	1	SYMPTOM	Aphasia	Nervous system disorders
4	1	SYMPTOM	Somnolence	Nervous system disorders
5	1	SYMPTOM	Encephalopathy	Nervous system disorders
6	***2***	***EVENT***	***Clinical Event 2***	
7	2	SYMPTOM	**Fatigue**	General disorders and administration site conditions
8	2	SYMPTOM	Memory impairment	Nervous system disorders
9	2	SYMPTOM	Confusional state	Psychiatric disorders
10	2	SYMPTOM	Encephalopathy	Nervous system disorders

ROW	UTOXGR	ATOXGR	USER	ASER	LVLCONS
1	***1***	***1***	***N***	***N***	***Awakens to voice***
2	1	1	N	N	
3	1	1	Y	Y	
4	2	2	N	N	
5	2	2	Y	Y	
6	***2***	***2***		***Y***	***Awakens spontaneously***
7	1	1	N	N	
8	2	2	N	N	
9	2	2	N	N	
10	2	2	Y	Y	

Data Display 7 Data Example of Child dataset ADAECE2 for Clinical Event 2

OCCURRENCE FLAGS FOR SUMMARY OF AE AND AESI

With most events data, it is possible for a subject to have more than one record for a particular event. However, when counting the number of subjects for a specific preferred term, body system or maximum severity level, the subject should only be counted once. In order to make the data set analysis ready, the use of occurrence flags can help to easily identify a unique record per subject per indicated type (e.g., preferred term, body system, maximum severity).

OCCDS has several pre-defined occurrence flags for the more common uses. However, sponsor-defined occurrence flags can be created using the standard naming convention of AOCCzzFL which has the same rules as the CQzzNAM, where zz is an index from 01 – 99. Note that the index must be zero filled but it does not have to start at 01. If there are multiple sponsor-defined occurrence flags, they do not need to be sequential. This allows for organizations to setup company standards where a specific AOCCzzFL represents a particular type of occurrence flag across all studies. For example, AOCC11FL may represent maximum severity/intensity for serious AE per analysis period.

Labels for sponsor-defined occurrence flags can also be customized as long as the portion that is pre-defined in the OCCDS is kept and the label is 40 characters or less. For example, AOCC11FL could have a label of '1st Occurrence of WTX: ASER-PER'.

While occurrence flags could be created for every type of occurrence, in this illustration (Table 2) it is only used to identify AEs and serious AEs with a maximum severity/intensity per analysis period.

Variable	Label	Summary Block
AOCC01FL	1st Occurrence of WTX: PER	For "Any TEAE with the worst toxicity grade" summary per analysis period
AOCC11FL	1st Occurrence of WTX: ASER-PER	For "Any serious AE with the worst toxicity grade" summary per analysis period
AOCC02FL	1st Occurrence of WTX: SOC-PER	For "Any TEAE with the worst toxicity grade by SOC" summary per analysis period
AOCC12FL	1st Occurrence of WTX: ASER-SOC-PER	For "Any serious AE with the worst toxicity grade by SOC" summary per analysis period
AOCC03FL	1st Occurrence of WTX: PT-PER	For "Any TEAE with the worst toxicity grade by PT" summary per analysis period
AOCC13FL	1st Occurrence of WTX: ASER-PT-PER	For "Any serious AE with the worst toxicity grade by PT" summary per analysis period

Table 2 Occurrence Flags for Worst Toxicity in ADAE Data Set

In ADAE data example of Data Display 8 (see below), two records were flagged with AOCC01FL = 'Y', one for Row 1 record where AETOXGR is 3, the worst toxicity grade among all 13 records with APERIOD as 1, and another for Row 14 record, the only record with APERIOD as 2. Row 8 is also APERIOD 1 record with the same worst toxicity grade and the same onset date as Row 1 record. The rules for setting the first occurrence should be determined by the study needs. For this example, AE log number (AESPID) is used for sorting and Row 1 record has AESPID prior to the AESPID in Row 8. AOCC11FL is for serious AE records. As a result, Row 1 record is flagged with AOCC11FL= 'Y' among 4 serious AE records with AESPID be 1, 9, 12, and 50. The other four AOCCzzFL variables would be set according to the derivation rules.

ROW	AESPID	AESTDTC	AEENDTC	APERIOD	AETOXGR	AESER	AOCC01FL	AOCC11FL
1	1	2016-05-19	2016-05-22	1	3	Y		
2	2	2016-05-11	2016-05-15	1	2	N		
3	3	2016-05-04	2016-05-11	1	1	N		
4	4	2016-05-15	2016-06-02	1	1	N		
5	9	2016-05-17	2016-05-23	1	1	Y		
6	11	2016-05-17	2016-05-23	1	2	N		
7	12	2016-05-17	2016-05-24	1	**3**	Y	Y	Y
8	13	2016-05-19	2016-05-20	1	3	N		
9	17	2016-05-24	2016-05-25	1	2	N		
10	18	2016-05-25	2016-06-03	1	2	N		
11	26	2016-05-16	2016-05-19	1	2	N		
12	27	2016-05-20	2016-05-21	1	1	N		
13	50	2016-05-17	2016-06-03	1	2	Y		
14	51	2016-06-14	2016-06-17	2	4	N	Y	

ROW	AEBODSYS	AOCC02FL	AOCC12FL	AEDECOD	AOCC03FL	AOCC13FL
1	Cardiac disorders	Y	Y	Atrial fibrillation	Y	Y
2	Respiratory, thoracic and mediastinal disorders			Cough	Y	
3	Musculoskeletal and connective tissue disorders	Y		Arthralgia	Y	
4	General disorders and administration site conditions			Fatigue	Y	
5	Nervous system disorders			Aphasia	Y	Y
6	Nervous system disorders	Y		Somnolence	Y	
7	Respiratory, thoracic and mediastinal disorders	Y	Y	Hypoxia	Y	Y
8	Vascular disorders	Y		Hypotension	Y	
9	Nervous system disorders			Memory impairment	Y	
10	Psychiatric disorders	Y		Confusional state	Y	
11	General disorders and administration site conditions	Y		Pyrexia	Y	
12	General disorders and administration site conditions			Pyrexia		
13	Nervous system disorders		Y	Encephalopathy	Y	Y
14	Investigations	Y		White blood cell count decreased	Y	

Data Display 8 ADAE Data Example with AOCCzzFL for Worst Toxicity

The following annotated table shell (Output 3) indicates how ADAE.AOCCzzFL variables support a standard AE summary report. With ADAE data set and annotated table shell (Output 3), one can make the data set analysis-ready by implementing some occurrence flags (Table 2).

ADAE.ATOXGR

MedDRA SOC Term, n (%) Preferred Term, n (%)	Any Grade	Worst Grade 1	Worst Grade 2	Worst Grade 3	Worst Grade 4	Worst Grade 5
Subjects with Any Treatment-emergent Adverse Event in Period X ADAE.AOCC01FL = 'Y'	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
SOC Term 1 ADAE.AOCC02FL = 'Y' by ADAE.AESOC	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 2 (ADAE.AOCC03FL = 'Y' by ADAE.AEDECOD)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
PT Term 3	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
…..						

Output 3 Annotated Table Shell for Summary of Treatment-emergent Adverse Events by System Organ Class, Preferred Term and Worst Grade

AOCCzzFL variables can be utilized for ADAE<si> Table 3 the same way as they are for ADAE. Recall that in CE data example there are two clinical events, Clinical Event 1 and Clinical Event 2, When we create ADAE<si> child data set, we will have ADAECE1 for Clinical Event 1 analysis and ADAECE2 for Clinical Event 2 analysis. For demonstration purposes, we only use ADAECE1 as ADAE<si> data example. ADAECE1 data example has four AOCCzzFL variables for analysis category and analysis period level and for PT analysis category and analysis period level. These flags can be derived with the sorting key variables APERIOD, ACAT1, UDECOD, ATOXGR, and CQ01RCID (Data Display 9).

Variable	Label	Summary Block
AOCC01FL	1st Occurrence of WTX: ACAT1-PER	For "Subjects with any TE clinical event with the worst toxicity grade" summary by analysis category (EVENT/SYMPTOM) per analysis period
AOCC11FL	1st Occurrence of WTX: ACAT1-ASER-PER	For "Subjects with any serious clinical event with the worst toxicity grade" summary by analysis category (EVENT/SYMPTOM) per analysis period
AOCC02FL	1st Occurrence of WTX: ACAT1-PT-PER	For "Subjects with any TE clinical event with the worst toxicity grade" summary by analysis category (EVENT/SYMPTOM) and PT per analysis period
AOCC12FL	1st Occurrence of WTX: ACAT1-ASER-PT-PER	For "Subjects with any serious clinical event with the worst toxicity grade" summary by analysis category (EVENT/SYMPTOM) and PT per analysis period

Table 3 Occurrence Flags for Worst Toxicity in ADAE Child Data Set

Row	SRCDOM	SRCSEQ	UNUM	UTERM	ASTDT	AENDT	APERIOD
1	***CE***	***1***	***1***	***CLINICAL EVENT 1***	***16-May-16***	***17-May-16***	***1***
2	ADAE	7	4	**FATIGUE**	15-May-16	02-Jun-16	1
3	ADAE	12	12	HYPOXIA	17-May-16	24-May-16	1
4	ADAE	9	26	FEVER	16-May-16	19-May-16	1
5	***CE***	***2***	***2***	***CLINICAL EVENT 1***	***18-May-16***	***18-May-16***	1
6	ADAE	7	4	**FATIGUE**	15-May-16	02-Jun-16	1
7	ADAE	12	12	HYPOXIA	17-May-16	24-May-16	***1***
8	ADAE	9	26	FEVER	16-May-16	19-May-16	1
9	***CE***	***3***	***3***	***CLINICAL EVENT 1***	***19-May-16***	***24-May-16***	***1***
10	ADAE	20	1	ATRIAL FIBRILLATION WITH RVR	19-May-16	22-May-16	1

Row	SRCDOM	SRCSEQ	UNUM	UTERM	ASTDT	AENDT	APERIOD
11	ADAE	7	4	**FATIGUE**	15-May-16	02-Jun-16	***1***
12	ADAE	12	12	HYPOXIA	17-May-16	24-May-16	1
13	ADAE	18	13	HYPOTENSION	19-May-16	20-May-16	1
14	ADAE	9	26	FEVER	16-May-16	19-May-16	1
15	ADAE	25	27	FEVER	20-May-16	21-May-16	1

Row	CQ01RCID	ACAT1	UDECOD	UBODSYS
1	***1***	***EVENT***	***Clinical Event 1***	
2	1	SYMPTOM	**Fatigue**	General disorders and administration site conditions
3	1	SYMPTOM	Hypoxia	Respiratory, thoracic and mediastinal disorders
4	1	SYMPTOM	Pyrexia	General disorders and administration site conditions
5	***2***	***EVENT***	***Clinical Event 1***	
6	2	SYMPTOM	**Fatigue**	General disorders and administration site conditions
7	2	SYMPTOM	Hypoxia	Respiratory, thoracic and mediastinal disorders
8	2	SYMPTOM	Pyrexia	General disorders and administration site conditions
9	***3***	***EVENT***	***Clinical Event 1***	
10	3	SYMPTOM	Atrial fibrillation	Cardiac disorders
11	3	SYMPTOM	**Fatigue**	General disorders and administration site conditions
12	3	SYMPTOM	Hypoxia	Respiratory, thoracic and mediastinal disorders
13	3	SYMPTOM	Hypotension	Vascular disorders
14	3	SYMPTOM	Pyrexia	General disorders and administration site conditions
15	3	SYMPTOM	Pyrexia	General disorders and administration site conditions

Row	UTOXGR	ATOXGR	USER	ASER	AOCC01FL	AOCC11FL	AOCC02FL	AOCC12FL
1	***2***	***2***	***Y***	***Y***				
2	1	1	N	N			Y	
3	3	3	Y	Y	Y	Y	Y	Y
4	2	2	N	N			Y	
5	***1***	***1***	***N***	***N***				
6	1	1	N	N				
7	3	3	Y	Y				
8	2	2	N	N				
9	***3***	***3***		***Y***	***Y***	***Y***		
10	3	3	Y	Y			Y	Y
11	1	1	N	N				
12	3	3	Y	Y				
13	3	3	N	N			Y	
14	2	2	N	N				
15	1	1	N	N				

Data Display 9 ADAECE1 Data Example with AOCCzzFL for Worst Toxicity

Recall, in Row 9 of Data Display 9, USER was missing and ASER was confirmed through medical review for analysis report or can be imputed based on imputation rules. The annotated table shell (Output 4) gives a quick look of how easily an AESI clinical event summary can be created when ADAECE1.AOCCzzFLs are in play.

ADAECE1.ATOXGR

Event, n (%)	Any Grade	Worst Grade 1	Worst Grade 2	Worst Grade 3	Worst Grade 4	Worst Grade 5
Subjects with Any Clinical Event in Period X ADAECE1.AOCC01FL = 'Y' where ADAECE1.ACAT1 = 'EVENT'	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Clinical Event by Preferred Terms PT Term 1 } ADAECE1.AOCC02FL = 'Y' where PT Term 2 } ADAECE1.ACAT1 = 'SYMPTOM' ….. by ADAECE1.UDECOD	 xx (xx.x) xx (xx.x)	 xx (xx.x) xx (xx.x)	 xx (xx.x) xx (xx.x)	 xx (xx.x) xx (xx.x)	 xx (xx.x) xx (xx.x)	 xx (xx.x) xx (xx.x)

Output 4 Annotated Table Shell for Summary of AESI: Treatment-emergent Clinical Event by Preferred Term and Worst Grade

CONCLUSION

Monitoring the safety of subjects during a clinical trial is important. Sometimes it may not be sufficient to evaluate the risk by looking at AEs independently. In some cases, a combination of AEs may be an indication of a larger condition or syndrome. How these events are identified can vary (e.g., collected as a clinical event on the CRF or through medical/clinical confirmation) but they are typically captured in another domain. Being able to see the relationship between the AEs (symptoms) and the clinical event is not always obvious. By implementing the child data set approach, the symptoms can be linked to the clinical event that it is associated with. Furthermore, any additional information for the clinical event can also be captured in the child data set as well as standard OCCDS variables that will help to aid in the analyses needed for clinical events.

REFERENCES

CDISC Analysis Data Model Team. (2021, November). ADaM Structure for Occurrence Data (OCCDS) Implementation Guide. CDISC. Retrieved from https://www.cdisc.org/standards/foundational/adam

CDISC Analysis Data Model Team. (2021, November). Analysis Data Model Implementation Guide v1.3. CDISC. Retrieved from https://www.cdisc.org/standards/foundational/adam

Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. (2020, 04 09). *Guidance Document E2F Development of Safety Update Report*. Retrieved 04 2022, from U.S. Food & Drug Administration: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e2f-development-safety-update-report

International Conference on Harmonization. (2022, March). Introductory Guide for Standardised MedDRA Queries (SMQs) Version 25.0. ICH. Retrieved from https://admin.meddra.org/sites/default/files/guidance/file/SMQ_intguide_25_0_English.pdf

CONTACT INFORMATION

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

Lindsey Xie
Kite Pharma, Inc.
lxie@kitepharma.com

Jinlin Wang
Kite Pharma, Inc.
jwang20@kitepharma.com

Lauren Xiao
Kite Pharma, Inc.
lxiao@gilead.com

Richann Watson
DataRich Consulting
richann.watson@datarichconsulting.com