PharmaSUG 2025 - Paper DS-065

Which ADaM Data Structure Is Most Appropriate? Gray Areas in BDS and OCCDS.

Veronica Gonzalez [formerly Renauldo], Biogen, Inc.

ABSTRACT

Choosing the ADaM dataset structure, based on the analysis requirements, is usually straightforward. Statistical analyses involving change from baseline, like those from Lab or Vital Signs, can only be done in a ADaM Basic Dataset Structure (BDS) dataset. Analyses that involve counts/frequencies of hierarchical dictionary data, like Adverse Events, Concomitant Medications, and Medical History, can only be completed in ADaM Occurrence Data Structures (OCCDS) dataset.

However, there are times that the ADaM structure needed to support the statistical analysis is not well defined and can be completed using either BDS or OCCDS. This paper outlines an instance where both BDS and OCCDS ADaM datasets could be appropriate to achieve the statistical analysis. The strengths and weaknesses of each data structure will be scrutinized, utilizing one in depth example, to illustrate that an analysis structure choice is not always straightforward and that either could be used to execute the analysis requirements.

INTRODUCTION

The study protocol describes what data is to be collected, and at what frequency, along with the objectives/endpoints of the trial. The schedule of activities lays out all of the data the study is required to collect at each specific visit. A case report form (CRF) dictates how the protocol requirements are executed through collected data. CRF collected data, or raw data, is not submitted to regulatory agencies. Instead, the raw data is mapped into SDTM making SDTM “formatted observed data.”

The analysis of SDTM data occurs in ADaM datasets. The ADaM structure that is most appropriate is based on the statistical analysis. In instances where multiple ADaM data structures are able to execute the statistical analysis, which should you choose? This gray area will be explored through one example where OCCDS may be appropriate, but also BDS may be appropriate. Both ADaM structures achieve the required analysis however both have strengths and weaknesses.

The example used for this paper revolves around historical data for a particular disease of interest, disease *x*. Historical information such as if the patient has experienced any flares, have comorbidity diseases #1 or #2, as well as details on the flares are collected. The analysis of this data revolves around one table summarizing the collected information across all participants.

DATA COLLECTIONS AND SDTM

Data from the CRF designed below is collected to better assess the history of the disease *x* of interest for the study. The first six questions are answered only once per subject while the log lines section, also known as *add-entry*, can be many records per subject. The investigator may choose to add just one organ/system if the subject only had one flare historically or there may be 10+ log lines (records) if the subject had experienced several historical flares. The number of entries in that section is dependent on the subject’s recall of historical flares experienced.

The data collected on this CRF goes into two different SDTM domains. The medical history (MH) domain would contain information about the occurrence of disease *x*, organ/system areas associated with disease flares, and occurrence of any comorbidities. The findings about (FA) domain would contain all the additional information collected that pertains to the different records in the MH domain. These two domains would be linked via –LNKGRP, and RELREC would be required.

Form: Disease *x* History		**SDTM Mapping**
Did the subject have disease *x*?	◦Yes ◦No	MHOCCUR where MHTERM = “DISEASE X” and MHLNKGRP = “DIS01”
[If yes]		
History of disease *x* flares?	◦Yes ◦No	MHOCCUR where MHTERM = “DISEASE X FLARES”
Current Course of Disease	◦Chronic Active ◦Relapsing-Remitting	FAORRES where FATESTCD = “STATUS”, FAOBJ = “DISEASE X”, and FALNKGRP = “DIS01”
History of comorbidity disease #1?	◦Yes ◦No	MHOCCUR where MHTERM = “COMORBIDITY #1” and MHLNKGRP = “COM01”
[If yes]		
Class of comorbidity disease #1	◦Class I ◦Class II ◦Class III	FAORRES where FATESTCD = “CLASS”, FAOBJ = “COMORBIDITY #1”, and FALNKGRP = “COM01”
History of comorbidity disease #2?	◦Yes ◦No	MHOCCUR where MHTERM = “COMORBIDITY #2”
[Log Lines]		
Organ/System involved with flare	◦Musculoskeletal ◦Renal ◦<7 other locations>	MHOCCUR = “Y” where MHTERM = “DISEASE X FLARES: <selected organ system>”
Flare Date		MHSTDTC where MHTERM = “DISEASE X FLARES: <selected organ system>”
Resolved?	◦Yes ◦No	If yes, MHENRTPT = “BEFORE” where MHTERM = “DISEASE X FLARES: <selected organ system>”. If no, MHENRTPT = “ONGOING” where MHTERM = “DISEASE X FLARES: <selected organ system>”.
Severity of flare	◦Mild ◦Moderate ◦Severe	MHSEV where MHTERM = “DISEASE X FLARES: <selected organ system>”

MH AND FA DOMAIN EXAMPLE DATA

To illustrate how the two ADaM displays will differ the example data provided below will be utilized. This is factitious data of 3 different subjects in this trial. Below are a subset of the MH and FA domains per subject for data pertaining to the *Disease x History* CRF. Only relevant variables from each domain are shown to reflect the pieces collected in the example CRF.

- Subject 1001 answered all questions on the CRF and noted two historical flares in the same organ/system with differing severities.
- Subject 1002 had disease x but did not have history of flares nor either comorbidity.
- Subject 1003 did not have disease x which is why this subject only has one record in MH and no FA records.
- For subject 1002 MHSEQ = 3 and subject 1003 MHSEQ = 1, MHLNKGRP is null because there is no FA data for these records that require linkages.

SDTM.MH

USUBJID	MHSEQ	MHLNKGRP	MHTERM	MHPRESP	MHOCCUR	MHSEV	MHSTDTC	MHENRTPT
1001	1	DIS01	DISEASE X	Y	Y			
1001	2		DISEASE X FLARES	Y	Y			
1001	3	COM01	COMORBIDITY #1	Y	Y			
1001	4		COMORBIDITY #2	Y	Y			
1001	5		DISEASE X FLARES: RENAL			MILD	2024-01-01	ONGOING
1001	6		DISEASE X FLARES: RENAL			SEVERE	2020-10-10	BEFORE
1002	1	DIS01	DISEASE X	Y	Y			
1002	2		DISEASE X FLARES	Y	N			
1002	3		COMORBIDITY #1	Y	N			
1002	4		COMORBIDITY #2	Y	N			
1003	1		DISEASE X	Y	N			

SDTM.FA

USUBJID	FASEQ	FALNKGRP	FAOBJ	FATESTCD	FATEST	FAORRES
1001	1	DIS01	DISEASE X	STATUS	Status	CHRONIC ACTIVE
1001	2	COM01	COMORBIDITY #1	CLASS	Classification	CLASS II
1002	1	DIS01	DISEASE X	STATUS	Status	RELAPSING-REMITTING

ANALYSIS REQUIREMENTS

A summary table is required for this analysis. The table summarizes the count and percentages of each response per collected item per treatment group. The organ/system collected data can have multiple records per subject. The following selection criteria will be utilized to ensure that only one record per organ/system is selected per subject.

- The record with maximum severity will be selected.
- If multiple records of maximum severity exist, then the record that is not resolved will be selected.
- If multiple ongoing maximum severity records exist, then the record with the latest date will be selected.

DISEASE X HISTORY – SAFETY ANALYSIS SET

	Placebo (N=xx)	**Active Arm (N=xx)**
Diagnosed with Disease *x*		
Yes	nn (pp.p)	nn (pp.p)
No	nn (pp.p)	nn (pp.p)
History of Flares		
Yes	nn (pp.p)	nn (pp.p)
No	nn (pp.p)	nn (pp.p)
Current Course of Disease		
Chronic Active	nn (pp.p)	nn (pp.p)
Relapsing-Remitting	nn (pp.p)	nn (pp.p)
History of Comorbidity #1		
Yes	XX (pp.p)	nn/XX (pp.p)
Class I	nn/XX (pp.p)	nn/XX (pp.p)
Class II	nn/XX (pp.p)	nn/XX (pp.p)
Class III	nn/XX (pp.p)	nn/XX (pp.p)
No	nn (pp.p)	nn (pp.p)
History of Comorbidity #2		
Yes	nn (pp.p)	nn (pp.p)
No	nn (pp.p)	nn (pp.p)
Organ/System Involved		
<Locations>	YY (pp.p)	YY (pp.p)
Resolved		
Yes	nn/YY (pp.p)	nn/YY (pp.p)
No	nn/YY (pp.p)	nn/YY (pp.p)
Severity of Flare		
Mild	nn/YY (pp.p)	nn/YY (pp.p)
Moderate	nn/YY (pp.p)	nn/YY (pp.p)
Severe	nn/YY (pp.p)	nn/YY (pp.p)

OCCDS AND BDS OVERVIEW

Occurrence Data Structure (OCCDS) is used for occurrence analyses. These are analyses for "the counting of subjects with a given record or term, and often includes a structured hierarchy of dictionary coding categories. Examples of data that fit into this structure include those used for typical analysis of adverse events, concomitant medications, and medical history" (CDISC Analysis Data Model Team, 2021). In other words, an OCCDS dataset is not a parameter (PARAM) with a response (AVAL/AVALC) collection type, but rather a medical/medication term collection type. Variables such as dose, location, route, frequency, severity, start and end dates, etc. are usually found with occurrence data. OCCDS datasets typically represent data that is only collected if an event of interest, such as an adverse event, happens. These datasets are wide in nature.

Basic Data Structure (BDS) is used for response level analyses which include, but are not limited to, summary of responses to a question, change from baseline, percent change from baseline, or time-to-event analyses. BDS is used for parameter level information where a response/result is provided. "A BDS contains one or more records per subject, per analysis parameter, per analysis timepoint… This structure contains a central set of variables that describe the analysis parameter (e.g., PARAM and related variables) and contain the value being analyzed (e.g., AVAL and AVALC and related variables)" (CDISC Analysis Data Model Team, 2009). In other words, BDS is used to support analyses in a vertical/tall manner such as data collected from a questionnaire across multiple visits. Examples of data that are typically analyzed using BDS include questionnaires, laboratory results, and vital signs.

FUNDAMENTAL PRINCIPLES AND TRACEABILITY

Section *3.1 Fundamental Principles* of ADaM v2.1 outline the following as fundamental principles of ADaM (CDISC Analysis Data Model Team, 2009):

- Analysis datasets and their associated metadata must:
    - Facilitate clear and unambiguous communication
    - Provide traceability between the analysis data and its source data (ultimately SDTM)
    - Be readily useable by commonly available software tools
- Analysis datasets must:
    - Be accompanied by metadata
    - Be analysis-ready

These principles must be applied when looking at designing datasets to achieve the statistical analysis of interest. One of the largest components of the fundamental principles is traceability. All ADaM datasets require traceability. Traceability provides "an understanding of where an analysis value (whether an analysis result or an analysis variable) came from, i.e., the data's lineage or relationship between an analysis value and its predecessor(s)" (CDISC Analysis Data Model Team, 2009). Traceability allows the reviewer to identify where the source of the analysis result comes from, such as SDTM or other ADaMs, along with any methods used to create the analysis result (derivations, imputations).

There are two types of traceability: metadata and data point.

- Metadata traceability "is established by describing (via metadata) the algorithm used or steps taken to derive or populate an analysis value from its immediate predecessor" (CDISC Analysis Data Model Team, 2009). In other words, the traceability is provided in the define and/or the reviewer's guide.
- Data point "…traceability enables the user to go directly to the specific predecessor record(s)… This traceability is established by providing clear links in the data (e.g., via use of --SEQ variable) to the specific data values used as input for an analysis value" (CDISC Analysis Data Model Team, 2009). In other words, the traceability is provided directly in the dataset.

Providing both types of traceability are most ideal to enable "… quick and efficient reviews of analysis data and results and thus is a cornerstone of a quality submission" (CDISC Analysis Data Model Team, 2022).

If data point traceability is feasible but the analysis data sources from more than one input (e.g., multiple SDTM domains, SDTM domain and an ADaM dataset) then special sourcing variables are available. SRCDOM, SRCVAR, and SRCSEQ can be added to any style of ADaM dataset to provide data point traceability back to the source domain, variable, and sequence. A subset of *Table 3.3.9.1 Datapoint Traceability Variables* from ADaM IG v1.3 is below to provide more detail on these variables (CDISC Analysis Data Model Team, 2009).

Variable Name	**Variable Label**	**CDISC Notes**
SRCDOM	Source Data	The SDTM domain name or ADaM dataset name that relates to the analysis value (i.e., AVAL or AVALC in a BDS dataset). If the source data is a supplemental qualifier in SDTM, this variable will contain the value of RDOMAIN in SUPP-- or SUPPQUAL.
SRCVAR	Source Variable	The name of the column (in the domain or dataset identified by SRCDOM) that relates to the analysis value (i.e., AVAL or AVALC in a BDS dataset). In the event that SRCDOM is a SUPPQUAL, then SRCVAR will be populated with the value of the related QNAM.
SRCSEQ	Source Sequence Number	The sequence number --SEQ or ASEQ of the row (in the domain or dataset identified by SRCDOM) that relates to the analysis value (i.e., AVAL or AVALC in a BDS dataset). In the event that SRCDOM is a SUPPQUAL, then this variable will contain the sequence number of the relevant related domain record.

For the purposes of this exercise, all 3 sourcing variables would be utilized in the BDS dataset but not in the OCCDS dataset.

ADAM OCCDS VERSION

The data collected on the Disease *x* History CRF could be brought into an OCCDS ADaM dataset. Our three example subjects are below showing how their data would be used in an occurrence data structure. The data from MH are contained in the MH variables directly copied from SDTM.MH (green). The FA data is transformed into ATERMy and ACATy variables. Each ACAT-ATERM combination contains the values of FATEST (sometimes concatenated with FAOBJ) and FAORRES, respectively, per FALNKGRP (purple). Essentially, the MH data would remain appearing as it was in SDTM while the FA data is transformed into various columns. ANL01FL is used to flag the records that meet the analysis requirements, specifically those around the selection of records for the organ/system section (where subjects can report multiple records per organ/system). Subject 1001's record where MHSEQ = 5 does not have ANL01FL populated. This subject has multiple records for the same organ/system so the one with the maximum severity is selected per the analysis requirements.

Data point traceability variables such as MHSEQ, MHLNKGRP, MHTERM, MHOCCUR, MHSEV, MHSTDTC, and MHENRTPT help provide connection back to the SDTM source data for the MH records. However, for the FA records there are no data point traceability variables provided. Traceability for these records would be achieved via metadata traceability provided in the define and/or reviewer's guide. The FA data is linked to the MH data via the --LNKGRP variables. The multiple source data point traceability variables cannot be utilized because when FA data is merged back to the MH records, every record with data in both would cause the ADaM SRC variables to be null. SRC variables relate a given record back to just one domain, whereas for this data two domains concurrently are contained in one row of data.

DISEASE X HISTORY ANALYSIS DATASET (ADDISX- OCCDS)

USUBJID	MHSEQ	MHLNKGRP	MHTERM	MHOCCUR	MHSEV	MHSTDTC
1001	1	DIS01	DISEASE X	Y		
1001	2		DISEASE X FLARES	Y		
1001	3	COM01	COMORBIDITY #1	Y		
1001	4		COMORBIDITY #2	Y		
1001	5		DISEASE X FLARES: RENAL		MILD	2024-01-01
1001	6		DISEASE X FLARES: RENAL		SEVERE	2020-10-10
1002	1	DIS01	DISEASE X	Y		
1002	2		DISEASE X FLARES	N		
1002	3		COMORBIDITY #1	N		
1002	4		COMORBIDITY #2	N		
1003	1		DISEASE X	N		

(continued)

USUBJID	MHSEQ	MHENRTPT	ACAT1	ATERM1	ACAT2	ATERM2	ANL01FL
1001	1		Status	CHRONIC ACTIVE			Y
1001	2						Y
1001	3				Class of comorbidity disease #1	CLASS II	Y
1001	4						Y
1001	5	ONGOING					
1001	6	BEFORE					Y
1002	1		Status	RELAPSING-REMITTING			Y
1002	2						Y
1002	3						Y
1002	4						Y
1003	1						Y

There are some strengths and weaknesses to using OCCDS.

Strength 1: Typical Occurrence Data Presentation

The strength of using OCCDS for this data resides in the resulting analysis dataset representing that it is collected mainly as occurrence data in SDTM. The main data is of an occurrence nature: occurrence of disease x, occurrence of comorbidities #1 and #2. Agency reviewers are accustomed to seeing medical history data using OCCDS so they would be familiar with this representation. This method would also be

easier to explain to statisticians or other stakeholders, as there are limited modifications to the occurrence data from SDTM (FA data is transposed not MH).

Strength 2: Ease of Programming

OCCDS implementation is easier to program for this analysis. MH and FA data would need to be merged by USUBJID and --LNKGRP variables in order to get all connected FA data on the proper corresponding rows in MH. The FA data is then stored in new columns which are pairs of ACAT-ATERM variables. New columns are faster to add programmatically than new rows.

Weakness 1: Only Metadata Traceability for FA records

For the FA records, only metadata traceability would be achieved. Whenever possible, datapoint traceability along with metadata traceability is preferential. The Analysis Data Reviewer's Guide (ADRG) along with the define should provide details on the source of each ACAT-ATERM combination.

For the MH records both metadata and datapoint traceability is achieved. Metadata traceability is obtained through define, where sourcing per MH variable would be described, and the ADRG which should explain the MH and FA portions of the dataset. Datapoint traceability is obtained by MHSEQ.

Weakness 2: Inferring Responses

In the *Organ/System Involved* portion of the analysis table, the *Resolved* rows with responses of *Yes* or *No* would not be able to be directly pulled from this dataset. MHENRTPT will have values of "BEFORE" if the event was resolved or "ONGOING" if the event was not resolved. In the table program, a format could be used to display the values of "BEFORE" and "ONGOING" as *Yes* and *No*, respectively. In this case, the values of *Yes* and *No* displayed in the table are not directly in the dataset itself, but rather inferred based on MHENRTPT. The ADRG could be utilized to provide additional details on how the MHENRTPT values relate to the *Resolved* display values.

Weakness 3: Improper use of ACAT-ATERM

ACAT and ATERM, in this example, are being used in a manner analogous to the PARAM-AVAL relationship in a BDS dataset. ACATy in the *Other categorization Variables* from OCCDS IG v1.1 is supposed to be utilized for categorizing records like special interests (ex. prohibited medications) that are not captured within the CAT or SCAT variables from SDTM (CDISC Analysis Data Model Team, 2021). The ACAT contains the "category" of interest, but that category corresponds to additional result information about MHTERM instead of categorizing/grouping the MHTERMs. The use of ACATy in this manner is not an appropriate usage of the variable per the variable's CDISC notes.

ATERM is supposed to represent the analysis reported term from the SDTM data. ATERM is generally used when OCCDS sources from several SDTM/ADaM datasets, where the reported term for analysis differs depending on the source. In the application outlined in *Table 14.2.3 Sample ADaM Value-level Metadata* from OCCDS IG v1.1, ATERM would be the focus of the occurrence reporting (CDISC Analysis Data Model Team, 2021). That would not be the case in this implementation, as MHTERM is still the focus of the occurrence reporting, not ATERM1, ATERM2, or ATERM3. In this illustration, ATERM contains the result of the category. It is also unusual to have several ATERMs in one occurrence dataset. It would be more typical to have one record per ATERM not several ATERMs for one record. The use of ATERM in this way is not an appropriate use of this variable based on the variable's CDISC notes.

Additionally, for submission, Pinnacle 21 (P21) is required to validate datasets and defines. These validation rules identify CDISC implementation issues and conduct cross domain/dataset content validity assessments. As of the writing of this paper, P21 does not have any validation rules around ACAT or ATERM. In this case, relying solely on P21 to catch an implementation issue, such as the use of ACAT-ATERM in this manner, would be problematic.

ADAM BDS VERSION

The data collected on the Disease x History CRF could be brought into a BDS ADaM dataset. The method of creating Occurrence parameters in a BDS dataset, outlined below, is adapted from the 2020 PharmaSUG paper *Is Your Data Set Analysis Ready?* written by Brucken and Patel. Each of the parameters in this BDS dataset would represent each of the summary rows in the Analysis Requirements section. AVALC is populated with the response text and AVAL is the numeric code of AVALC (only populated to ensure row ordering in the table creation).

Analysis Requirements	Parameter	AVALC Population	AVAL Population
Diagnosed with Disease x	Occurrence of Disease x	Y N	1 = Y 2 = N
History of Flares	Occurrence of History of Disease x Flares	Y N	1 = Y 2 = N
Current Course of Disease	Current Course of Disease x	Chronic Active Relapsing-Remitting	1 = Chronic Active 2 = Relapsing-Remitting
History of Comorbidity #1	Occurrence of Comorbidity Disease #1	Y N	1 = Y 2 = N
Comorbidity #1 Class	Comorbidity Disease #1 Class	Class I Class II Class III	1 = Class I 2 = Class II 3 = Class III
History of Comorbidity #2	Occurrence of Comorbidity Disease #2	Y N	1 = Y 2 = N
<Organ/System involved selection> Occurrence	Occurrence of <Organ/System involved selection> Flare	Y	1 = Y
<Organ/System involved selection> Resolution	<Organ/System involved selection> Flare Resolution	Y N	1 = Y 2 = N
<Organ/System involved selection> Severity of Flare	<Organ/System involved selection> Flare Severity	Mild Moderate Severe	1 = Mild 2 = Moderate 3 = Severe

In order to obtain data point traceability when multiple SDTM domains are used as sources, the following sourcing variables would be utilized: SRCDOM, SRCVAR, and SRCSEQ. Each variable is defined as:

- SRCDOM would contain values of MH or FA
- SRCVAR per parameter would be populated with either MHOCCUR, MHENRTPT, or MHSEV for the MH records or FAORRES for the FA records
- SRCSEQ would contain the values of MHSEQ or FASEQ.

Our three example subjects are below showing how their data would be used in a BDS ADaM dataset. Each MH record in the *Organ/System* section would be transformed into 3 different parameters (rows): Occurrence for the organ/system, resolution for the organ/system, and severity of the organ/system. Otherwise, the other record count is equal to the count of records in SDTM that are being used.

A key item to point out is the variable contained in SRCVAR per record does not mean that AVAL or AVALC contain the value contained in this variable outright, but rather that variable relates to the values contained in AVAL or AVALC. For the FA records, AVALC is populated with the value of FAORRES,

meaning that the value contained within FAORRES is directly utilized as AVALC. However, for the Organ/System records, AVALC is based on the value of the variable in SRCVAR but is not the value itself necessarily. Where SRCVAR = “MHENRTPT”: AVALC relates to the values contained in MHENRTPT where

- AVALC = “Y” relates to MHENRTPT = “BEFORE”
- AVALC = “N” relates to MHENRTPT = “ONGOING”

The table below is color coded to aid in the explanation per subject.

- Subject 1003 only has one MH record, for the occurrence of disease x, and 0 FA records so there is only one record in the BDS dataset (blue).
- Subject 1002 has 4 MH records (yellow) and 1 FA record (orange) which is why there are 5 records in the BDS dataset.
- Subject 1001 has the largest transformation of SDTM to ADaM, as this subject had 6 records in MH (purple) and 2 in FA (green) which results in 12 records in the BDS dataset. For each organ/system response, 3 different BDS records were created. Otherwise, each MH and FA record has one corresponding BDS record. Just like the OCCDS method, subject 1001’s set of records where SRCSEQ = 5 does not have ANL01FL populated. This subject has multiple records for the same Organ/System so the set of records pertaining to the maximum severity of the organ/system are selected per the analysis requirements.

DISEASE X HISTORY ANALYSIS DATASET (ADDISX- BDS)

USUBJID	SRCDOM	SRCVAR	SRCSEQ	ASTDT	PARAM	PARAMCD	AVAL	AVALC	ANL01FL
1001	MH	MHOCCUR	1		Occurrence of Disease x	OCCURDSX	1	Y	Y
1001	MH	MHOCCUR	2		Occurrence History of Disease x Flares	OCCURFLR	1	Y	Y
1001	FA	FAORRES	1		Current Course of Disease x	STATUS	1	CHRONIC ACTIVE	Y
1001	MH	MHOCCUR	3		Occurrence of Comorbidity Disease #1	OCCURCM1	1	Y	Y
1001	FA	FAORRES	2		Comorbidity Disease #1 Class	CLASS	2	CLASS II	Y
1001	MH	MHOCCUR	4		Occurrence of Comorbidity Disease #2	OCCURCM2	1	Y	Y
1001	MH	MHOCCUR	5	2024-01-01	Occurrence of Renal Flare	OCCURREN	1	Y	
1001	MH	MHENRTPT	5	2024-01-01	Renal Flare Resolution	RENRES	2	N	
1001	MH	MHSEV	5	2024-01-01	Renal Flare Severity	RENSEV	1	MILD	
1001	MH	MHOCCUR	6	2020-10-10	Occurrence of Renal Flare	OCCURREN	1	Y	Y
1001	MH	MHENRTPT	6	2020-10-10	Renal Flare Resolution	RENRES	1	Y	Y
1001	MH	MHSEV	6	2020-10-10	Renal Flare Severity	RENSEV	3	SEVERE	Y
1002	MH	MHOCCUR	1		Occurrence of Disease x	OCCURDSX	1	Y	Y
1002	MH	MHOCCUR	2		Occurrence History of Disease x Flares	OCCURFLR	2	N	Y
1002	FA	FAORRES	1		Current Course of Disease x	STATUS	1	RELAPSING-REMITTING	Y
1002	MH	MHOCCUR	3		Occurrence of Comorbidity Disease #1	OCCURCM1	2	N	Y
1002	MH	MHOCCUR	4		Occurrence of Comorbidity Disease #2	OCCURCM2	2	N	Y
1003	MH	MHOCCUR	1		Occurrence of Disease x	OCCURDSX	2	N	Y

There are some strengths and weaknesses to using BDS.

Strength 1: Data point and Metadata Traceability for both MH and FA

The data needed for the table to be created comes from both MH and FA domains. Data point traceability is contained within the sourcing variables: SRCDOM, SRCVAR, and SRCSEQ. This allows the reviewer to directly trace the analysis variables back to SDTM. Metadata traceability is also achieved through the define value level metadata (VLM) and the ADRG should provide details on the organ/system section (where 3 BDS records are created per MH record).

Strength 2: Directly Meets Analysis Requirements

The values for the table are contained within AVALC per parameter without having to do any transformations. AVAL is also populated serving as the numeric code for AVALC only to be utilized for ease of the table creation rows per parameter (to ensure proper ordering of rows).

Weakness 1: Increased Programming Time

This method will require more programming time to be allotted to the programming team. This is not an incredibly challenging BDS dataset but does require more effort than the OCCDS method. For the Organ/System section, each MH record creates a maximum of 3 BDS records: occurrence of the organ/system flare, organ/system flare resolution, and the organ/system flare severity.

Weakness 2: Increased Time for Define Creation and Parameter Repository

This method requires the define to have numerous VLM records. AVALC and AVAL would both require VLM to fully describe the methodology utilized to obtain the values contained in these variables along with the appropriate assigning of codelists. This method also requires the team to ensure the parameter repository of possible PARAM-PARAMCD combinations is kept up to date. There would be a maximum of 33 PARAM-PARAMCD combinations: Organ/system MH section: 9 locations x 3 parameters, all other MH data: 4 occurrence parameters, FA data: 2 parameters. This repository would not be required in the OCCDS method.

Weakness 3: Requires Explanation

Unlike the OCCDS method, this method will require more of an explanation to other functional groups as well as an explanation in the ADRG for the agency reviewers to fully understand the method. Both MH and FA domains are transformed: MH data is parameterized and, for the *Organ/System* section, multiple ADaM records are created from one SDTM record; FA data in SDTM is already in a pseudo parameterized arrangement (FATESTCD-FAORRES) so limited modifications are made, in terms of structure, to this domain. The concept of taking one record from SDTM and creating several records in ADaM is not as straightforward as the OCCDS method.

CONCLUSION

The protocol determines what data is required to be collected. The CRF design dictates how the protocol is executed and the appearance/structure of the raw data. SDTM then formats the data into specific requirements. ADaM sources from SDTM and is based on analysis needs. Depending on the analysis requirements, multiple ADaM dataset structures may be appropriate. Medical history data normally is presented in an OCCDS ADaM dataset but, depending on analysis needs, a BDS dataset could also achieve the analysis objectives. Each method results in meeting the analysis requirements, but both have strengths along with some weaknesses.

The OCCDS method is the quickest to program and most straightforward implementation of the analysis requirements. Each FA record is merged back with its corresponding MH record. The FA data is put into columns to be used for the table. This method does lack data point traceability but has metadata traceability. This method also would require a format in the table program in order to map the results in MHENRTPT to the table requirements (aka inferring responses). Additionally, the usage of the paired ACAT-ATERM variables may not be proper CDISC implementation necessarily and could be missed if P21 is used as the sole validation tool.

The BDS method provides both data point and metadata traceability to SDTM along with the table being able to be directly sourced from the values in AVALC per parameter. This method does require more programming time along with requiring more work in terms of explanations to other groups in order for a full understanding of the method to be obtained, define creation for VLM, and keeping the PARAM-PARAMCD repository up to date.

The OCCDS version is a wide dataset structure versus the BDS which is a long dataset structure. Programmatically it is easier to add new columns to an existing dataset (OCCDS ACAT-ATERM) than it is to create new rows based off of a singular SDTM record. The datapoint traceability allotted in the BDS method provides any reviewer with a direct link between SDTM and ADaM in the ADaM dataset itself, without referring to any additional submission documentation. The table below shows the summary comparison of the two methods.

ADaM Structure	**Strengths**	**Weaknesses**
OCCDS	1. Typical occurrence data presentation a. Reflects the main data collection type which is occurrence b. Agency reviewers are used to receiving medical history data in OCCDS 2. Easier to program and understand a. Adding columns for the FA data is faster than adding additional rows.	1. Only metadata traceability a. Lack of data point traceability to FA domain. 2. Inferring Responses a. Yes or No response for if the organ/system resolved is needing to be created in the table program based on values in MHENRTPT. 3. Improper use of ACAT-ATERM a. P21 does not have validation rules around ACAT or ATERM
BDS	1. Data point and metadata traceability a. Traceability to both MH and FA domains contained within the SRC variables. 2. Directly meets analysis requirements a. Table values source directly from AVALC (AVAL available for proper row order).	1. Increased programming time a. It is more laborious to create additional rows in a dataset than additional columns. 2. Increased time for define creation and parameter repository a. VLM required for AVAL and AVALC b. 33 possible PARAM-PARAMCD combinations 3. Requires more explanation to fully understand

Per the FDA Technical Conformance Guide section 4.1.2.3 *Key Efficacy and Safety Data* only requires ADaM to support efficacy (primary and secondary) and safety analyses. For these required ADaMs, the analysis requirements outlined in the SAP and outputs can allow for multiple different ADaM dataset structures to be utilized. In such cases, the strengths, and weaknesses of the differing ADaM dataset structures should be scrutinized. Allow yourself to think through the challenging questions in diverse ways in order to ensure that you are choosing the best approach for the given situation. Ensure the method selected works for your team and adheres to the ADaM fundamental principles.

REFERENCES

Brucken, N., & Patel, K. (2020). Is Your Data Set Analysis Ready? PharmaSUG. Retrieved 2024, from https://www.lexjansen.com/pharmasug/2020/DS/PharmaSUG-2020-DS-133.pdf

CDISC Analysis Data Model Team. (2009, Dec 17). Analysis Data Model (ADaM). *(2.1)*. Retrieved 2024, from https://www.cdisc.org/standards/foundational/adam

CDISC Analysis Data Model Team. (2021, November 29). *ADaM Structure for Occurrence Data (OCCDS) Implementation Guide.* (1.1, Ed.) Retrieved 2024, from https://www.cdisc.org/standards/foundational/adam

CDISC Analysis Data Model Team. (2022, May 12). Analysis Data Model (ADaM) Examples of Traceability. *1.0*. Retrieved 2024, from https://www.cdisc.org/standards/foundational/adam/adam-examples-traceability-v1-0

U.S. Department of Health and Human Services Food and Drug Administration. (2024, March). Study Data Technical Conformance Guide. FDA. Retrieved September 06, 2024, from https://www.fda.gov/media/153632/download

ACKNOWLEDGMENTS

A special thanks to Jennifer McGrogan, Joe Lorenz, Mario Widel, and Priyanka Pollarine for reviewing this paper. Also, thank you to Nancy Brucken and Kapila Patel as your paper, in the recommended reading section, inspired me to think about this real work application in a different manner which led to this paper.

RECOMMENDED READING

Barrick, M., & Troxell, J. (2014). A Guide to the ADaM Basic Data Structure for Dataset Designers. PharmaSUG. Retrieved from https://www.lexjansen.com/pharmasug/2014/DS/PharmaSUG-2014-DS11.pdf

Freimark, N., & et all. (2012). Common Misunderstandings about ADaM Implementation. PharmaSUG. Retrieved from https://www.lexjansen.com/pharmasug/2012/DS/PharmaSUG-2012-DS16.pdf

Minjoe, S., & Widel, M. (2013). Using the ADaM ADAE Structure for non-AE Data. PharmaSUG. Retrieved from https://support.sas.com/resources/papers/proceedings13/177-2013.pdf

Minjoe, S. (2015). It Depends On Your Analysis Need. PharmaSUG.Retrieved from https://www.lexjansen.com/pharmasug/2015/DS/PharmaSUG-2015-DS11.pdf

Minjoe, S., & et all. (2018). Traceability: Some Thoughts and Examples for ADaM Needs. PharmaSUG. Retrieved from https://www.lexjansen.com/pharmasug/2018/DS/PharmaSUG-2018-DS01.pdf

Shostak, J. (2017). ADaM Grouping: Groups, Categories, and Criteria. Which Way Should I Go? PharmaSUG. Retrieved from https://www.pharmasug.org/proceedings/2017/DS/PharmaSUG-2017-DS17.pdf

Troxell, J. (2017). Clarifications About ADaM Implementation Provided in ADaMIG Version 1.1. PharmaSUG.Retrieved from https://www.pharmasug.org/proceedings/2017/DS/PharmaSUG-2017-DS18.pdf

U.S. Department of Health and Human Services Food and Drug Administration. (2024, March). Study Data Technical Conformance Guide. FDA. Retrieved September 06, 2024, from https://www.fda.gov/media/153632/download

Watson, R., & Miller, K. (2018). Avoiding Sinkholes: Common Mistakes During ADaM Data Set Implementation. MWSUG. Retrieved from https://www.mwsug.org/proceedings/2018/HS/MWSUG-2018-HS-160.pdf

CONTACT INFORMATION

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

Veronica Gonzalez
Biogen, Inc
vee.gonzalez@biogen.com
Or LinkedIn: www.linkedin.com/in/veronica-vee-gonzalez