PharmaSUG 2021 - Paper DS-102

ADaM-Like Dataset: how to do big things in a short time

Oksana Mykhailova, Andrii Klekov Quartesian Europe LLC

ABSTRACT

There are long-term studies where either interim, futility analysis, conference membership with an intermediate result, or both of them are applicable. Let's name these requests as an Ad Hoc request. In the case of such requests for the study, there is no concrete scope of work during the trial and we should do repeat work from time to time. By definition of CDISC, ADaM dataset is a type of analysis dataset that follows to fundamental principles described in ADaM model document. As you know ADaM dataset goes hand in hand with SDTM datasets but frequently SDTMs weren't finished and Ad Hoc requests coming. In this case, we propose to generate ADaM-like datasets. This paper describes the concept of ADaM-like datasets, different types of them and approaches to create them for Ad Hoc requests.

INTRODUCTION

The main purpose of using ADaM is to save traceability and prepare data for analysis. To avoid programming of same algorithms throw the outputs and reduce the time for the creation of codes. But sometimes in Long-term studies, occurs some problems related to the unstable structure of data, so the creation and update of ADaM for each request may require a lot of time. So to avoid extra time spending, we can avoid the implementation of some CDISC requirements, implement them slowly during study lifetime. It will help us create stable SDTMs only once without big changes and also will request a chance of typos/mistakes. In Table 1 we can see types of ADaM likes datasets. Figure 1 showed the process of transformation, as Caterpillar transforms into the butterfly, in the same way, ADaM likes to transform into the ADaM.

Short name of ADaM-like	**Main difference beetween ADaM-Like and ADaM**	**SDTMs/ADaM availability**	**Study design stability**	**Scope of Work**
Listing based	We connect different raw data and transpose it. so that it vaguely resembles ADaM.	Only core SDTMs produced. Outputs/ADaM-likes based on raw data.	Many of raw data is empty	Safety analyses/ 1-4 outputs per ADaM-like.
Transition ADaM-like	Almost all variables are SDTMs based, some are still coming from rawdata.	ADSL is partially available, some of variables can be based on rawdata.	Some changes can be applied. No new CRF forms expected.	Safety analyses/ Simple Efficacy 5-10 outputs per ADaM-like.
Special request ADaM-like	Various	Various	Various	Various
Fully ADaM-like	This is almost ADaM, but some variables which aren't used for outputs are still pending.	All Adam based only on SDTMs.	No changes expected or minor changes	Efficacy part scope/ 10-15 outputs per ADaM-like.

Table 1. Types of ADaM-likes

Since ADaM-like is just a recommendation, how to simplify the working process on long-term studies, you can avoid using all of the types of the ADaM-like, and make the transition period shorter, or longer, you can directly go from Listing-Based ADaM-like to ADaM, in case of need.

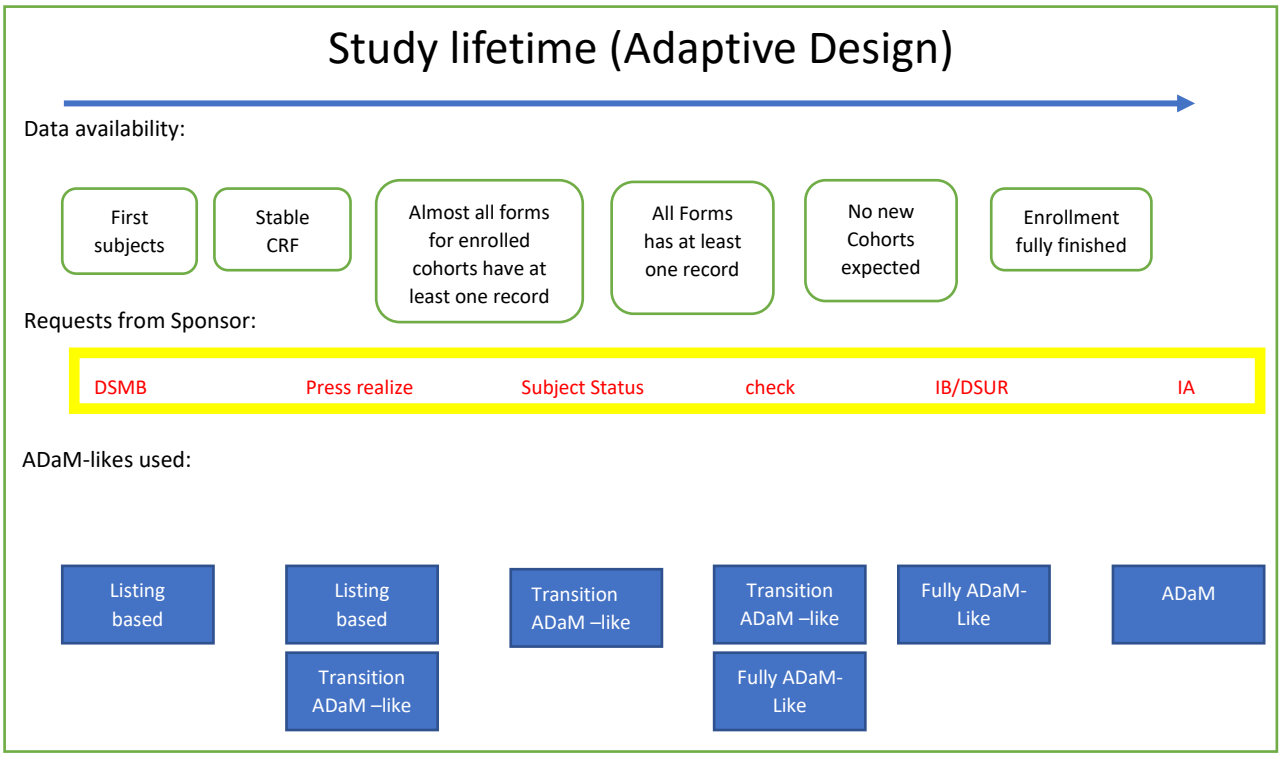

Figure 1. ADaM-like types change during study lifetime.

ADAM-LIKE - TO DO OR NOT TO DO

We need very carefully choose, which ADaM rules can be avoided. Here is listed some of the rules:

- If we do not need to generate xpt – length of character variables can be more than 200.
- If we use ADSL only to derive populations, there is no need to add other required variables, which will not be used in Analysis.
- Traceability is important, but if there are parameter which comes for separate for, which should be mapped to SDTM, which is still pending, there is no need to create partially filled SDTM, pull this parameter direct from raw data. In this case, we can save data, as it used in Listings.

When you avoiding, some of these rules, never forget, to add additional checks into codes. So in case, you used a length bigger than 200, so you will not truncate data when it comes to the creation of xpt. Also, remember that it better to implement those rules, so it will be easier to transform one type to another and make, this transformation process faster when it will come to specification/codes updates, but if you want to misbehave and break the rules, it's okay this is why you chose to create ADaM-like instead of AdaM.

The main Purpose of ADaM-like is to simplify the process, so when you have a stable data structure, no change expected, or the study will be finished shortly, there is no reason, to use ADaM-like. But when it comes to cases when ADaM-like can be used, the next step is to understand, which type of ADaM-like should be used.

INVESTIGATION OF DIFFERENT TYPE OF ADAM-LIKE DATASETS

In this section we will explore the process of creation of different types ADaM-like datasets.

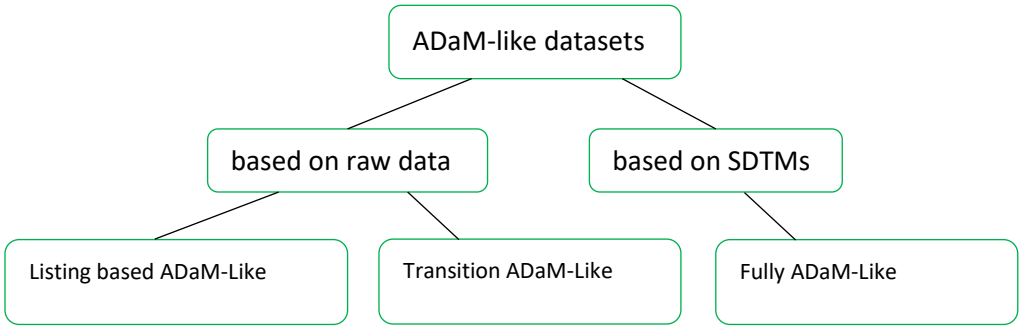

Figure 2. Classification of ADaM-like datasets in a relationship to approaches of creation.

Suppose there is a request for interim analysis on an oncology study and the scope of outputs includes efficacy analysis. The table and listing below should be generated:

Table 14.2.1
Summary of Best Overall Unconfirmed Response
Efficacy-evaluable Population

Parameter	Cohort 1 (N=XX) (n %)	Cohort 2 (N=XX) (n %)	Cohort 3 (N=XX) (n %)	Total (N=XX) (n %)
Best Overall Unconfirmed Response				
Complete Response	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Partial Response	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Stable Disease	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Progressive Disease	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Not Evaluable	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Not Available	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Objective Response Rate (ORR)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
Disease Control Rate (DCR)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
Disease Control Rate 16 Weeks (-7 Days)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
Disease Control Rate 24 Weeks (-7 Days)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)

N = Number of subjects within population and treatment cohort. Percentages are based on N.
Best Overall Unconfirmed Response (RECIST v1.1) is defined as the best response recorded from the start of the study treatment until disease progression, defined in order of descending significance: CR (Complete Response), PR (Partial Response), SD (Stable Disease), PD (Progressive Disease), NE (Not Evaluable).
Objective Response Rate (ORR) is defined as the percentage of subjects with a Best Overall Unconfirmed Response of CR or PR.
Disease Control Rate (DCR) is defined as the percentage of subjects with a Best Overall Unconfirmed Response of CR, PR, or SD.
Disease Control Rate 16 Weeks (-7 Days) is defined as the percentage of subjects with a Best Overall Unconfirmed Response of CR, PR, or continuous SD calculated as (last SD date before the first PD - treatment start date + 1)/7 >= 15 Weeks.
Disease Control Rate 24 Weeks (-7 Days) is defined as the percentage of subjects with a Best Overall Unconfirmed Response of CR, PR, or continuous SD calculated as (last SD date before the first PD - treatment start date + 1)/7 >= 23 Weeks.
Reference: Listing 16.2.2

Figure 3. Mock-Shell for Table 14.2.1.

Listing 16.2.2
Listing of Best Overall Unconfirmed Response and Binary Endpoints
Efficacy-evaluable Population

Treatment Cohort: <Cohort>

	Best Overall Unconfirmed Response [a]		Subject has Best Overall Unconfirmed Response of			
Subject	Response	Visit	CR or PR	CR, PR or SD	CR, PR, or Continuous SD > 16 Weeks (-7 Days)	CR, PR, or Continuous SD > 24 Weeks (-7 Days)
xxxxxx-xxx	xxx	DA #XX	Yes/No	Yes/No	Yes/No	Yes/No
…etc.						

DA = Disease Assessment.
[a] Best Overall Unconfirmed Response is defined as the best response recorded from the start of the study treatment until disease progression, defined in order of descending significance: CR (Complete Response), PR (Partial Response), SD (Stable Disease), PD (Progressive Disease), NE (Not Evaluable).

Figure 4. Mock-Shell for Listing 16.2.2.

ADAM-LIKE DATASETS BASED ON RAW DATA

There are two types of ADaM-like datasets based on raw data. The first one is the transition ADaM-like dataset and the second one is listing based ADaM-like dataset.

Transition ADaM-like dataset

The first approach to the creation of ADaM-like more convenient with the creation of traditional ADaMs but there is one major difference – the source of data is raw data instead of SDTMs. Also, this approach might be applicable in case when core SDTMs are available.

During this approach recommended to create ADSL dataset that includes analysis population, treatment start/end date, treatment cohort, and other subject-level necessary information. Assume, ADSL for the example was created.

Our aim is to generate an analysis dataset for oncology-related efficacy endpoints - Efficacy Analysis Dataset Oncology (ADEFON).

Here are the required steps for creation such ADaM-like dataset:

1. To find the necessary information for the analysis in raw data. For convenience, use Case Report Form (CRF) for the study.

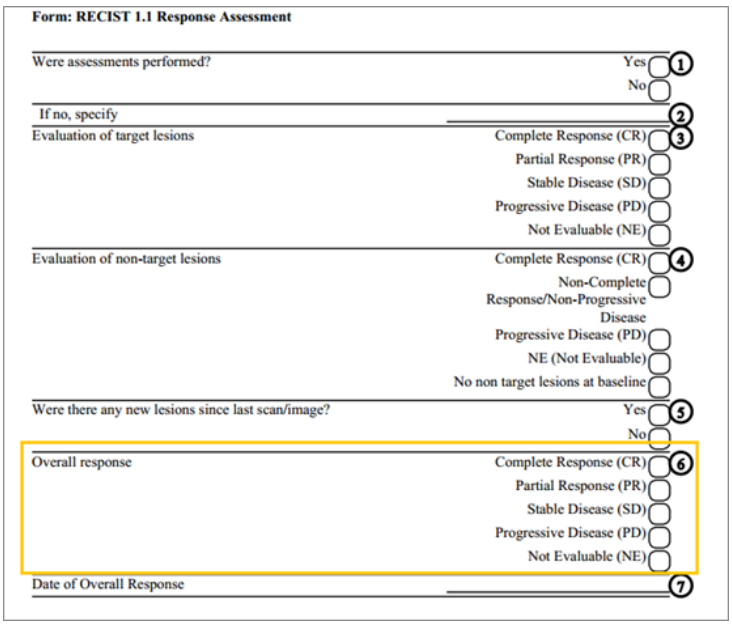

Figure 5. Case Report Form: RECIST 1.1 Response Assessment.

The table contains the frequency of Best Overall Unconfirmed Response, the proportions of subjects with objective response and disease control and their exact binomial 2-sides 95% confidence intervals. Jumping ahead, for the derivation of these parameters Overall Responses per RECIST v1.1 would be used.

Subject ID	Disease Assesment	Were assessments performed?	Specify	Evaluation of target lesions	Evaluation of non-target lesions	Overall response	Date if Overall Response	New lesions since last scan/image?
77777-777	DA01	Yes		Stable Disease (SD)	Non-Complete Response/Non-Progressive Disease	Stable Disease (SD)	08JAN2020:00:00:00.000	No
77777-777	DA02	Yes		Stable Disease (SD)	Complete Response (CR)	Stable Disease (SD)	10FEB2020:00:00:00.000	No
77777-777	DA03	Yes		Stable Disease (SD)	Non-Complete Response/Non-Progressive Disease	Stable Disease (SD)	09MAR2020:00:00:00.000	No
77777-777	DA04	Yes		Progressive Disease (PD)	Non-Complete Response/Non-Progressive Disease	Progressive Disease (PD)	29APR2020:00:00:00.000	No
77777-776	DA01	Yes		Partial Response (PR)	Non-Complete Response/Non-Progressive Disease	Partial Response (PR)	28JAN2020:00:00:00.000	No
77777-776	DA02	Yes		Partial Response (PR)	Non-Complete Response/Non-Progressive Disease	Partial Response (PR)	24MAR2020:00:00:00.000	No
77777-776	DA03	Yes		Complete Response (CR)	Complete Response (CR)	Complete Response (CR)	13MAY2020:00:00:00.000	No
77777-776	DA04	Yes		Complete Response (CR)	Complete Response (CR)	Complete Response (CR)	30JUN2020:00:00:00.000	No
77777-776	DA05	Yes		Complete Response (CR)	Complete Response (CR)	Complete Response (CR)	25AUG2020:00:00:00.000	No
77777-776	DA06	Yes		Complete Response (CR)	Complete Response (CR)	Progressive Disease (PD)	20OCT2020:00:00:00.000	Yes
77777-776	DA07	Yes		Complete Response (CR)	Complete Response (CR)	Progressive Disease (PD)	30NOV2020:00:00:00.000	No

Figure 6. An appropriate raw data.

2. Create a specification for ADEFON based on appropriate raw data and ADSL.

ADEFON is Basic Data Structure (BDS) dataset, which means it contains one or more records per subject, per analysis parameter, per analysis timepoint. Hence, for outputs above we should derive the next parameters for subjects which included in the efficacy-evaluable population: Best Overall Unconfirmed Response, Objective Response (Unconfirmed), Disease Control (Unconfirmed), Disease Control (Unconfirmed) 16 Weeks (-7 Days), Disease Control (Unconfirmed) 24 Weeks (-7 Days). The footnote in table contains derivation rules for parameters.

Both Disease Control (Unconfirmed) 16 Weeks (-7 Days) and Disease Control (Unconfirmed) 24 Weeks (-7 Days) parameters were integrated with Disease Control (Unconfirmed) using criterion flags (see specification below).

DATASET	PARAMCD	PARAM	PARAMTYP	PARCAT1	VARIABLE	DERIVATION
ADEFON	UBOR	Best Overall Response (Unconfirmed)	DERIVED	RECIST 1.1	AVALC	Select sub-set of responses and keep all the Assessment up to first rawdata.RAOVER = "Progressive Disease (PD)", then choose maximum significance within USUBJID. If maximum significance response repeated, then choose the first one (earliest). The order of significance: 1. CR (Complete Response), 2. PR (Partial Response), 3. SD (Stable Disease), 4. PD (Progressive Disease), 5. NE (Not Evaluable) If Subject hasn't any Disease Assessment but the study drug was administered at least one time (EFFFL = "Y") then set to "NA".
ADEFON	UOBJRESP	Objective Response (Unconfirmed)	DERIVED	RECIST 1.1	AVALC	If PARAMCD = "UBOR" and AVALC in ("CR", "PR") then set to "Y".
ADEFON	UDCRESP	Disease Control (Unconfirmed)	DERIVED	RECIST 1.1	AVALC	Select sub-set of responses rawdata.RAOVER and keep all the Assessment up to first rawdata.RAOVER = "Progressive Disease (PD)", then choose maximum significance within USUBJID. If maximum significance response is repeated, then choose the last one (latest) and set to "Y". The order of significance: 1. CR (Complete Response), 2. PR (Partial Response), 3. SD (Stable Disease). If Subject hasn't any "CR", "PR", "SD" but the study drug was administered at least one time (EFFFL = "Y") then set to "N".

Figure 7. Value Level Metadata for ADEFON.

Dataset	**ADEFON**
Description	Efficacy Analysis Dataset Oncology
Structure	One record per efficacy parameter per subject
Purpose	Analysis
Keys	STUDYID, USUBJID, PARCAT1, PARAMCD, ADT
Raw Datasets	raw.RA

Variable	Label	Type	Length/ Significant Digits	Format	Programming Algorithm
STUDYID	Study Identifier	Char	~ As per Data		ADSL.STUDYID
USUBJID	Unique Subject Identifier	Char	~ As per Data		ADSL.USUBJID
SUBJID	Subject Identifier for the Study	Char	~ As per Data		ADSL.SUBJID
SITEID	Study Site Identifier	Char	~ As per Data		ADSL.SITEID
Analysis variables					
ADT	Analysis Date	Num	8	DATE9.	Convert rawdata.RAORDAT date part in date9. format
ADY	Analysis Relative Day	Num	8		ADT - ADSL.TRTSDT + (ADT >= ADSL.TRTSDT)
AVISIT	Analysis Visit	Char	~ As per Data		For PARAMCD = "UBOR" set to raw.RA.RADISASS
ARDURW	Accumulative Response Duration (Week)	Num	8		If PARAMCD = "UDCRESP" and AVALC = "Y" then calculate as: (ADT - ADSL.TRTSDT + 1) / 7 and round up to one decimal point.
PARAM	Parameter	Char	~ As per Data		See VLM sheet
PARAMCD	Parameter Code	Char	~ As per Data		See VLM sheet
PARAMTYP	Parameter Type	Char	~ As per Data	PARAMTYP.	See VLM sheet
PARCAT1	Parameter Category 1	Char	~ As per Data		See VLM sheet
AVALC	Analysis Value (C)	Char	~ As per Data		See VLM sheet
CRIT1	Analysis Criterion 1	Char	~ As per Data		If PARAMCD = "UDCRESP" then set to "CR, PR, or SD > 16 Weeks (-7 days)".
CRIT1FL	Criterion 1 Evaluation Result Flag	Char	1	YN.	If PARAMCD = "UDCRESP" then do as follows: - If subset's calculated response is "SD" and ARDURW >= 15 then set to "Y"; - else if subset's calculated response isn't "SD" and AVALC = "Y" then set to "Y"; - Else set to "N".
CRIT2	Analysis Criterion 2	Char	~ As per Data		If PARAMCD = "UDCRESP" then set to "CR, PR, or SD > 24 Weeks (-7 days)".
CRIT2FL	Criterion 2 Evaluation Result Flag	Char	1	YN.	If PARAMCD = "UDCRESP" then do as follows: - If subset's calculated response is "SD" and ARDURW >= 23 then set to "Y"; - else if subset's calculated response isn't "SD" and AVALC = "Y" then set to "Y"; - Else set to "N".
All the core variables from ADSL dataset					

Figure 8. Specification for ADEFON.

SUBJID	AVISIT	ARDURW	PARAM	AVALC
77777-777	DA01		Best Overall Response (Unconfirmed)	SD
77777-777		15.9	Disease Control (Unconfirmed)	Y
77777-777			Objective Response (Unconfirmed)	N
77777-776	DA03		Best Overall Response (Unconfirmed)	CR
77777-776		38	Disease Control (Unconfirmed)	Y
77777-776			Objective Response (Unconfirmed)	Y

Figure 9. The main part of ADEFON dataset that will be used for efficacy analysis.

Listing based ADaM-like dataset

The second type of ADaM-like based on raw data is a listing based ADaM-like. The listing based ADaM-like dataset contains information from raw data that was used for generating output. In general, there is separate ADaM-like dataset for each output. In case, when a request contains both table and referenced listing the good idea will be generate one ADaM-like dataset for them. If there is a just a table in scope of outputs, listing based ADaM-like dataset contains information that was used for calculating statistics.

In our example, there are both tables and referenced listing. Let's see on listing based ADaM-like for these outputs. The dataset contains information that displays in the listing and will be using for calculating statistics in the table.

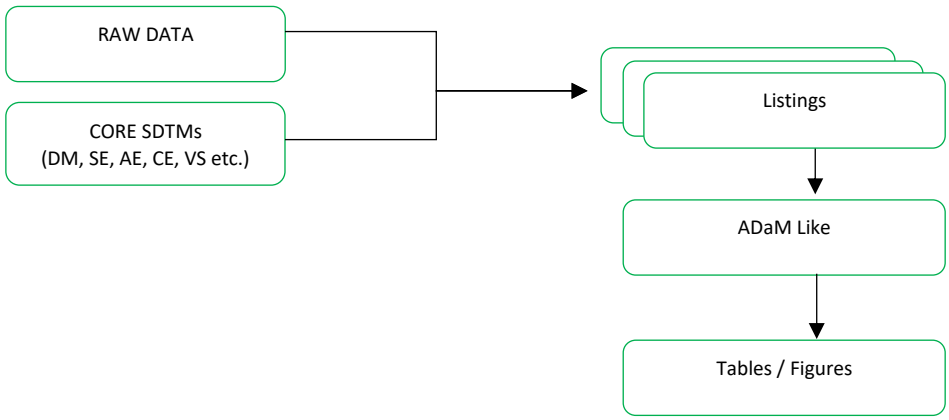

Figure 10. Listing based ADaM-like dataset for Listing 16.2.2 and Table 14.2.1.

	trt01a	subjid	efffl	UBOR	UOBJRESP	UDCRESP	UDCRESP16	UDCRESP24
1	Cohort 1	77777-721	Y	NA	N	N	N	N
2	Cohort 1	77777-773	Y	SD	N	Y	Y	Y
3	Cohort 1	77777-777	Y	NA	N	N	N	N
4	Cohort 1	77777-778	Y	SD	N	Y	Y	Y
5	Cohort 2	77777-713	Y	SD	N	Y	Y	Y
6	Cohort 2	77777-714	Y	SD	N	Y	Y	Y
7	Cohort 2	77777-717	Y	SD	N	Y	Y	Y
8	Cohort 2	77777-728	Y	SD	N	Y	Y	Y
9	Cohort 2	77777-729	Y	SD	N	Y	Y	Y
10	Cohort 2	77777-737	Y	SD	N	Y	Y	Y
11	Cohort 3	77777-710	Y	SD	N	Y	Y	Y
12	Cohort 3	77777-711	Y	SD	N	Y	Y	Y
13	Cohort 3	77777-712	Y	PD	N	N	N	N
14	Cohort 3	77777-716	Y	SD	N	Y	Y	Y
15	Cohort 3	77777-718	Y	PR	Y	Y	Y	Y
16	Cohort 3	77777-719	Y	SD	N	Y	Y	Y

Figure 11. Listing based ADaM-like dataset for Listing 16.2.2 and Table 14.2.1.

CONCLUSION

There are different approaches to the creation of ADaM-like datasets. You should choose one of them in relation to the type of request.

In case, a request contains a scope of Adverse Event outputs the better way would be to create ADAE transition ADaM-like dataset with analysis flags, categories and use it for all outputs instead of creating a separate listing based ADaM-like for each output. On the other hand, if a request includes many distinct types of information, for example, Prior Medications, Concomitant Medications, Medical History, Protocol Deviation would be more rational to use listing based ADaM-like datasets to avoid wasting time for writing specification at least.

REFERENCES

The Analysis Data Model (ADaM) Prepared by the CDISC Analysis Data Model (ADaM) Team, Version 2.1, available at http://www.cdisc.org

CONTACT INFORMATION

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

Name: Oksana Mykhailova
E-mail: oksana.mykhailov@gmail.com
Name: Andrii Klekov
E-mail: andryklekov@gmail.com