

## Applying ADaM Principles in Developing a Response Analysis Dataset

Mei Dey, Merck & Co., Inc  
Lisa Pyle, Merck & Co., Inc

### ABSTRACT

The Clinical Data Interchange Standards Consortium (CDISC) Analysis Data Model (ADaM) is an emerging new industry standard for submitting analysis datasets to regulatory agencies, such as the U.S. Food and Drug Administration (FDA). While it provides many solutions for construction of analysis datasets, it is subject to interpretation by users during implementation.

This paper explores an early application of ADaM principles in developing a response analysis dataset to meet the analysis requirements for reporting. It describes the process of constructing an analysis dataset: obtaining trace records, developing endpoint records and concatenating data into one final ADaM dataset. A specific example is provided, based on the draft ADaM Implementation Guide (IG) using oncology data but the solutions presented could be expanded to other therapeutic areas as well.

### RESPONSE DATA

The response data are one of the key efficacy measurements for oncology trials. There are generally two types of efficacy analyses for oncology trials that require response endpoint data: response analysis and time-to-event analysis.

The response endpoint is typically an overall patient assessment performed by the investigator and/or independent radiologist. It is typically a categorical variable indicating the assessment of disease status usually based on certain well established clinical methods, for example, the Response Evaluation in Solid Tumors (RECIST) method.

According to the NCI dictionary of cancer terms, the 4 categories of response for solid tumors are:

Complete Response (CR)	The disappearance of all signs of cancer in response to treatment. This does not always mean the cancer has been cured. Also called complete remission.
Partial Response (PD)	A decrease in the size of a tumor or in the extent of cancer in the body, in response to treatment. Also called partial remission.
Stable disease (SD)	Cancer that is neither decreasing nor increasing in extent or severity.
Disease Progression (PD)	Cancer that continues to grow or spread.

In many late stage oncology trials, the independent radiologist assessment of disease status is obtained by expert medical imaging vendors and in addition to the investigator's assessment is the source for the primary efficacy endpoint analyses.

### PROPOSED SDTM RESPONSE DATA – RS

Although there is no accepted CDISC SDTM domain for response data, a sample domain is provided below based on a findings domain model to facilitate the discussion of developing an ADaM analysis dataset (Dey and Pyle, 2009).

The response domain has one record per response observation per time point per visit per subject.

	RSSEQ	SUBJID	RSGRPID	RSDTC
Row 1	2345678913	0001	RAD	28-Jan-2008
Row 2	2345678914	0001	RAD	5-Mar-2008
Row 3	2345678915	0001	RAD	11-Mar-2008
Row 4	2345678916	0001	RAD	22-Apr-2008
Row 5	2345678917	0001	RAD	2-May-2008

	RSTESTCD	RSTEST	RSOBJ
Row 1 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 2 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 3 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 4 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 5 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN

	RSORRES	RSMETHOD	RSNAM	RSEVAL
Row 1 (cont)	PARTIAL RESPONSE	RECIST	XHumanImaging Co.	Reader 1
Row 2 (cont)	STABLE DISEASE	RECIST	XHumanImaging Co.	Reader 1
Row 3 (cont)	COMPLETE RESPONSE	RECIST	XHumanImaging Co.	Reader 1
Row 4 (cont)	NOT EVALUABLE	RECIST	XHumanImaging Co.	Reader 1
Row 5 (cont)	PROGRESSIVE DISEASE	RECIST	XHumanImaging Co.	Reader 1

RSDTC is the exact date the image was taken. Date of response is crucial to be associated with each response for time-to-response analysis. This should not be the date the image is read by the radiologist as it can take days even weeks in some cases to transfer the images to the reader for assessment. All subsequent programming for analyses should be based on this date

RSTESTCD is limited to 8 characters in length to be SDTM compliant. At this time, there is no CDISC dictionary of test codes for response, so they are sponsor defined. In this trial, OVERALL identifies the Overall Response at each time point RSTEST is the full description of the test and is a one-to-one relationship with the RSTESTCD.

RSORRES is the result of the response of assessment of disease. Example shows the NCI dictionary of cancer response terms. RSORRES of 'NOT EVALUABLE (NE)' are patients whose disease status can't be evaluated (e.g. missing scans, poor quality images).

## ADAM IG (DRAFT V1.0) BACKGROUND

Submissions by a sponsor to regulatory agencies, such as the FDA include the derived data used in the safety and efficacy analyses. In addition to having an understanding of the CDISC Study Data Tabulation Model (SDTM), it is also important to understand the ADaM IG. Since the SDTM datasets are not intended to support statistical analyses, the ADaM model further describes the organization, format and structure for analysis datasets that are used for statistical modeling, efficacy and safety reporting.

The ADaM IG outlines four principles to consider when developing an efficacy analysis dataset: traceability to the raw data to facilitate clear and unambiguous communication, be useable by currently available software tools, be linked to metadata and be analysis ready.

The two principles of the ADaM model that were most important for determining what response data to map were the components of analysis ready and traceability. The final analysis dataset for statistical modeling needs to be

analyzed with little or no additional pre-programming to prepare data. The response analysis dataset needs to be ready or in a format to submit to a SAS® procedure to produce results from statistical methods. For the response criteria, the analysis dataset should support response rate and time-to-event analyses.

Since most analysis datasets are derived from SDTM datasets, it is expected that there be some level of traceability between the SDTM dataset(s) and analysis dataset(s). In general, the CDISC ADaM standards recommend including as much supporting data in the traceability records and variables as possible, except in instances where it is not practical to do so.

According to the draft ADaM IG version 1.0 documentation, there exist many solutions for concepts for developing analysis datasets. Our paper explores a solution for developing a time-to-event analysis structure. We developed an ADaM response analysis dataset following the steps below

## DEVELOPING THE ADAM RESPONSE ANALYSIS DATASET

### 1. OBTAINING TRACE RECORDS

The first step in developing the dataset is to obtain the trace data or records. We include input SDTM data that are not analyzed but support a derivation in the analysis dataset as recommended by the implementation guide. From the RS domain, we include all trace RS data for determining the best response. This is done for investigator assessments and then again for radiologist assessments. A quick utility macro can be developed to obtain response data at each time point. A sample is listed below.

```
***** bring in investigator data *****;
%extract_rs( input_dataset      =lptss.rs
             ,output_dataaset   =rs_invest
             ,param_value       =OVERALL ASSESSMENT-INVESTIGATOR
             ,paramcd_value      =OVIN
             ,paramn_value       =1
             ,rsgrpidd           =OPA
             ,rstestcd           =OVERALL
             ,rscat              =
             );

***** bring in radiology data *****;
%extract_rs( input_dataset      =lptss.rs
             ,output_dataaset   =rs_rad
             ,param_value       =OVERALL ASSESSMENT-RADIOLOGY
             ,paramcd_value      =OVRA
             ,paramn_value       =2
             ,rsgrpidd           = RAD
             ,rstestcd           = OVERALL
             ,rscat              = RECIST
             );
```

In the %extract\_rs utility, PARAM, PARAMCD and PARAMN are assigned per the ADAM specifications to differentiate the response records from each data source (investigator response vs. radiologic response). Values for all 3 of these variables are user-defined as there are no current CDISC specifications for these values. The RSGRPID, RSTESTCD and RSCAT (if needed) are the minimum required columns needed to filter the SDTM RS findings domain for the necessary analysis data. SRCSEQ and SRCDOM are assigned when data from the RS findings domain is filtered and are retained on the trace rows.

**A partial set of sample trace records after executing %extract\_rs.**

SUBJID	PARAM	PARAMCD	PARAMN	AVISIT	AVISITN	AVALC	ADY
02362	OVERALL ASSESSMENT- INVESTIGATOR	OVIN	1	Treatment	1	STABLE DISEASE	41
02362	OVERALL ASSESSMENT- INVESTIGATOR	OVIN	1	Treatment	1	PARTIAL RESPONSE	85
02362	OVERALL ASSESSMENT- INVESTIGATOR	OVIN	1	Treatment	1	PARTIAL RESPONSE	125
02362	OVERALL ASSESSMENT- INVESTIGATOR	OVIN	1	Treatment	1	COMPLETE RESPONSE	167
02362	OVERALL ASSESSMENT- INVESTIGATOR	OVIN	1	Treatment	1	COMPLETE RESPONSE	209
02362	OVERALL ASSESSMENT- RADIOLOGY	OVRA	2	Treatment	1	STABLE DISEASE	41
02362	OVERALL ASSESSMENT- RADIOLOGY	OVRA	2	Treatment	1	STABLE DISEASE	85
02362	OVERALL ASSESSMENT- RADIOLOGY	OVRA	2	Treatment	1	STABLE DISEASE	125
02362	OVERALL ASSESSMENT- RADIOLOGY	OVRA	2	Treatment	1	STABLE DISEASE	167

Note: The permissible supportive variables SRCSEQ and SRCDOM are not depicted due to space limitation.

## 2. DERIVE ENDPOINT DATA

The next step after pulling in trace data records is to derive the endpoint data. The dataset structure needs to support efficacy analyses of summary of response and time to objective response.

**Objective Response** is defined as patients in the analysis population who have confirmed complete response (CR) or confirmed partial response (PR) during the course of the study as determined by the independent radiologist(s) or investigator.

To support the analyses, the derived endpoint records should contain information about whether a subject had an objective response based on trace records and time-to-event information.

First, it is helpful to initialize the analysis endpoint parameter variables that are required and some supportive permissible variables to support traceability to the SDTM data per ADAM IG standards. We choose to add some additional timing variables as well to facilitate analysis review.

```
%macro add_vars( paramn_add      =
                    );

%*****
Add ADaM variables for the analysis rows and blank out the trace
rows. This macro is placed within a data step.
%*****;

length  PARAMCD $20 PARAM $100;
        PARAMN  =&paramn_add;
        PARAMCD ='TTEOVIN';
        PARAM   ='TIME TO OBJECTIVE RESPONSE - INVESTIGATOR (DAYS)';

%*** start of blank out domain variables ***;
AVISITN =.;
AVISIT  ='Endpoint';
SRCSEQ  =.;
SRCDOM  =' ';
EPOCH   =' ';
visit   =' ';
visitnum=.;
dtype   ='Endpoint';
%mend add_vars;
```

The utility macro %add\_vars assigns missing to the permissible variables SRCSEQ and SRCDOM on the endpoint record since the data contained on this row is derived data. In addition, the PARAM, PARAMCD and PARAMN are user-defined. EPOCH, ADY and timing variables are assigned to missing as well on the endpoint row.

We recommend that the response outcome is stored in avalc and the time to event (in days) is stored in AVAL. Additionally, we recommend adding analysis descriptor variables like event description (EVENTDSC), event (EVENTFN), censoring reason (CENSREAS), etc. that are not described in the draft ADaM IG but are needed for analysis and review.

### Derived endpoint data using draft ADaM IG, v1.0

SUBJID	PARAM	PARAMCD	PARAMN	AVISIT	AVAL	EVENTDSC	EVENTFN	CENSREAS	DTYPE
02362	TIME TO OBJECTIVE RESPONSE - INVESTIGATOR (DAYS)	TTEOVIN	11	Endpoint	85	Partial Response	1		Endpoint
02362	TIME TO OBJECTIVE RESPONSE - RADIOLOGY (DAYS)	TTEOVRA	12	Endpoint	167	No Objective Response	0	Last Efficacy Measurement	Endpoint

Note: The recently approved ADaM IG, v1.0 identifies variables CNSR, EVNTDESC and STARTDT as time-to-event variables. EVNTDESC consolidates variables EVENTDSC and CENSREAS into one per the approved ADaM IG.

In cases where the endpoint is a composite derivation, the EVENTDSC can identify the component that was achieved first for the endpoint derivation. In our example, the composite endpoint is time to objective response and as a result, the EVENTDSC variable will indicate a Partial Response (PR) or a Complete Response (CR) event depending on which occurred first. For an event depicted by EVENTFN=1, the AVAL value will be the relative day at the start of the first PR or CR event. For no event, the AVAL will be the relative day of the last assessment.

### 3. PREPARE FINAL DATASET

The final step is to concatenate the trace data to the endpoint data. To facilitate response summary or time-to-response analyses, other variables may be added from the ADSL dataset.

#### Key Variable Definition for Analysis Overall Patient Assessment dataset ADOPA.SAS7BDAT

VARIABLE	LABEL	TYPE	CONTROLLED TERMS OR FORMAT OR COMMENTS
USUBJID	Unique Subject Identifier	Char	
SUBJID	Study Id	Char	
TRTP	Planned Treatment Group	Char	
TRTPN	Planned Treatment Group	Num	
ITTFL	Intention-to-Treat	Char	Y, N, null
FASFL	Full Analysis Set population flag	Char	Y, N, null
PARAM	Analysis Parameter name	Char	OVERALL-INVESTIGATOR, OVERALL-RADIOLOGY, TIME TO OBJECTIVE RESPONSE- INVESTIGATOR (DAYS), TIME TO OBJECTIVE RESPONSE – RADIOLOGY (DAYS)

PARAMCD	Analysis parameter code	Char	OVIN= OVERALL-INVESTIGATOR OVRA =OVERALL-RADIOLOGY TTEOVIN= TIME TO OBJECTIVE RESPONSE – INVESTIGATOR (DAYS) TTEOVRA= TIME TO OBJECTIVE RESPONSE – RADIOLOGY(DAYS)
PARAMN	Analysis Parameter number	Num	OVIN = 1 (trace data row) OVRA = 2 (trace data row) TTEOVIN = 11 (endpoint derivation) TTEOVRA = 12 (endpoint derivation)
AVISIT	Analysis visit	Char	For PARAMN in (11,12) code 'Objective Response', 'Unconfirmed Response' Otherwise missing
AVISITN	Analysis visit number	Num	For derived rows code to blank.
ADY	Analysis study day	Num	For PARAMN in (1,2) use RS.RSDY Otherwise missing
AVALC	Analysis value	Char	For PARAMN =1, use RS.RSSTRESC where RS:RSTESTCD = 'OVERALL' and RS:RSTEST = ' Overall Response' and RS:RSCAT= ' ' (investigator)  For PARAMN =2, use RS.RSSTRESC where RS:RSTESTCD = 'OVERALL' and RS:RSTEST = ' Overall Response' and RSCAT= 'RECIST' (central vendor).
AVAL	Numeric Analysis Value	Num	For derived rows code to blank For trace rows code to blank. For derived time to event rows, represents the relative start day of event
EVENTDSC	Event Description	Char	Response description for the derived time to event endpoint rows.  Objective Response No Objective Response
EVENTFN	Event Flag	Num	For PARAMN in (11,12) 1- Event occurred or Objective Response 0- Event did not occurred or No Objective Response Otherwise missing.
CENSREAS	Censor Reason	Char	When EVENTFN =0, fill in the censor reason such as 'Last Efficacy Date', 'Treatment Failure' etc.
DTYPE	Derivation Type	Char	"Endpoint" for PARAMN in (11,12) Otherwise missing
SRCSEQ	Sequence Number	Num	For trace data rows, use RS.RSSEQ. For derived endpoint rows, code to missing
SRCDOM	Source Domain	Char	For trace data rows, indicates the 2-letter source domain code. Otherwise missing.
EPOCH	Epoch time period	Char	For trace data rows, use RS.EPOCH. For objective response, code to missing
TAETORD	Epoch ordering variable	Num	For trace data rows, use RS.TAETORD. For objective response, code to missing
VISIT	Visit Name	Char	For trace data rows, use RS.VISIT. For objective response, code to missing.
VISITNUM	Visit Name	Num	For trace data rows, use RS.VISITNUM. For objective response, code to missing.

## ANALYSIS DATASET SOLUTION BENEFITS

This dataset approach consolidates the trace data and derived data which enables the efficient review of potential data issues as recommended by the ADaM IG. The consolidation also allows for a more thorough statistical review ensuring accuracy of the endpoint derivation. Previously, trace data records would have been abandoned in the analysis dataset but now they are retained. The inclusion of the variables PARAM and PARAMCD helps to facilitate endpoint filtering or filtering by trace and endpoint data combinations. The added descriptor variables for the time-to-event analysis provide detailed insight into the derivation of the endpoint. This ADaM structure also enables integration of investigator and radiologic responses into one dataset. Lastly, it includes the analysis values for results generation in AVAL and AVALC. A sample report call is included below.

```
%pr0ef0objective0response(  
    debug=N,  
    population_from=,  
    population_where=and adsl.fasfn=1,  
    observation_from=,  
    observation_where=,  
    rename_output=,  
    subtitle=(Full Analysis Set Population),  
    therapy_data_var=adsl.trtp2,  
    paramcd_type=TTEOVIN  
);
```

## CONCLUSION

In closing, an analysis dataset built using ADaM IG principles is powerful enough and flexible enough such that one can filter on various combinations of PARAM and PARAMN to view, review and analyze results. The ADaM IG provides adequate examples and solutions for adding and identifying derived records that can be applied to time-to-event analysis structures. With the additional programming required to create an ADaM IG compliant dataset, we feel that developing a series of utility macros will enable users to readily extract and develop consistent datasets that meet analysis and reporting requirements.

## REFERENCES

Analysis Data Model, Prepared by the CDISC ADaM Team, Version 2.1

ADaM Implementation Guide, Prepared by the CDISC ADaM Team, Draft Version 1.0, May 2008

ADaM Implementation Guide, Prepared by the CDISC ADaM Team, Final Version 1.0, December 2009

Dey, Mei and Pyle, Lisa. "CDISC SDTM Implementation of Response Data for Solid Tumor Trials in Oncology " Proceedings of the 22nd Annual NESUG Conference. September 2009.  
<http://www.nesug.org/Proceedings/nesug09/ph/ph04.pdf>

New Guidelines to Evaluate the Response to Treatment in Solid Tumors, Journal of the National Cancer Institute, Vol. 92, No. 3, February 2, 2000

New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1), European Journal of Cancer 45 (2009)

Study Data Tabulation Model, Prepared by the CDISC Submission Data Standards Team, Version 1.1

SDTM Implementation Guide, Prepared by the CDISC Submission Data Standards Team, Version 3.1.1

SDTM Implementation Guide, Prepared by the CDISC Submission Data Standards Team, Version 3.1.2

## ACKNOWLEDGEMENTS

The authors would like to thank Mary Varughese for her careful review and helpful comments. The author also thanks the management team for their review of this paper and assistance with the company publication process.

## CONTACT INFORMATION

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

Mei Dey  
Merck & Co., Inc.  
P.O. Box 1000  
351 North Sumneytown Pike  
North Wales, PA 19454  
Phone: (267) 305-6814  
Fax: (267) 305-6474  
Email: mei\_dey@merck.com

Lisa Pyle  
Merck & Co., Inc.  
P.O. Box 1000  
351 North Sumneytown Pike  
North Wales, PA 19454  
Phone: (267) 305-6887  
Fax: (267) 305-6474  
Email: lisa\_pyle@merck.com

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

Other brand and product names are trademarks of their respective companies.