A Complex ADaM dataset? Three different ways to create one.

Kevin Lee, Cytel, Inc., Chesterbrook, PA

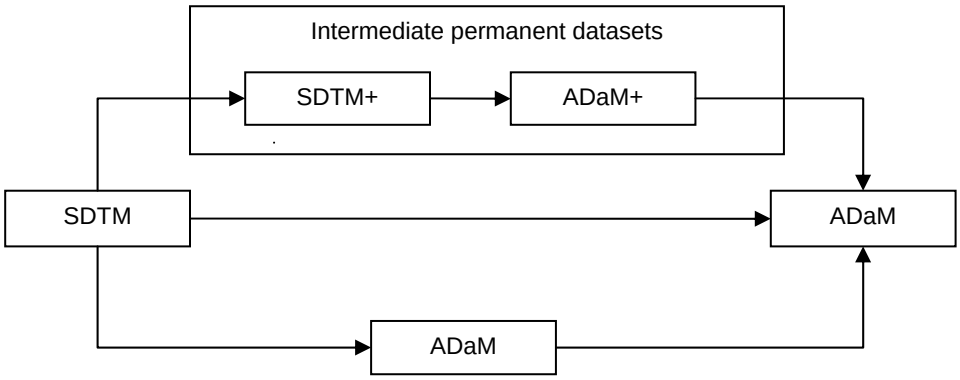
ABSTRACT

The paper is intended for Clinical Trial SAS® programmers who create and validate a complex ADaM dataset. Some ADaM datasets require the use of complex algorithms. These algorithms could require several steps of data manipulation and more than one SDTM datasets. It can be very challenging to create a complex ADaM dataset in accordance with ADaM data structures and standards. Furthermore, it can be equally as challenging to validate those ADaM datasets. The paper will introduce three different ways to create a complex ADaM dataset. The first way is to create ADaM from SDTM directly without any intermediate permanent datasets. The second way is to create ADaM through the intermediate permanent datasets like SDTM+ or ADaM+ from SDTM. The third way is to create the final ADaM through the intermediate ADaM from SDTM. The paper will discuss the benefits and limitations of each method and also show some examples.

THREE METHODS

There are three methods to create a complex ADaM dataset.

- SDTM to ADaM
- SDTM through intermediate permanent datasets to ADaM
- SDTM through ADaM to final ADaM



SDTM to ADaM

In most cases, a SAS programmer creates ADaM datasets from SDTM datasets without creating any intermediate permanent datasets. In this process, a SAS programmer creates a complex ADaM dataset directly from a SDTM dataset.

Its benefits are

- Simple process

Its limitations are

- A lack of data point traceability (Traceability will be provided with Define.xml)
- Difficult to troubleshoot issues if development SAS programmer and validation SAS programmer do not agree on issues in the final ADaM dataset.

SDTM through intermediate permanent datasets to ADaM

To illustrate each step, a SAS programmer can create intermediate permanent datasets from SDTM during ADaM dataset creation process. These intermediate permanent datasets such as SDTM+ dataset or ADaM+ dataset do not need to follow any standard such as CDISC, but they are saved so that one can reference them during the review process. This process can be very useful to demonstrate how the complex algorithms are applied and how data evolve through each algorithm.

Its benefits are

- Easy to follow each step and to validate
- Flexibility of the data structure of intermediate datasets (A programmer does not need to follow CDISC standards in the intermediate permanent datasets)

Its limitations are

- A lack of data point traceability, especially for the reviewers.
- Creation of the intermediate permanent datasets take additional time.

SDTM through ADaM to ADaM

In an effort to achieve improved data point traceability, a SAS programmer can create intermediate ADaM datasets from SDTM datasets and the final ADaM dataset from the intermediate ADaM datasets. Both ADaM datasets follow ADaM data structures and standards and both will be part of the final database submitted.

Its benefits are

- Easy to follow each step
- Great data point traceability

Its limitations are

- Need to create and validate all ADaM datasets including the intermediate ADaM datasets
- Not much flexibility of ADaM datasets as the intermediate datasets

EXAMPLE

In the following example, the SAS programmers need to create a complex ADaM dataset from a SDTM dataset. They need to compare the average of daily drinking rate between placebo and study drug. The average daily drinking rate is derived from the number of drinks per day as captured from a patient diary over a specified time period. After a patient is hospitalized, the baseline rate will be calculated for the 21 days (3 weeks) from the hospitalization date. For the treatment period, each patient has to provide data for a certain number of days – 42 days (6 weeks) from the first dose date. If the subjects discontinue prior to 42 days or miss any assessments during the treatment period, the rest of days will be imputed with the baseline rate.

SDTM to ADaM

Table 1.1 SDTM dataset of SU

STUDYID	DOMAIN	USUBJID	SUSEQ	SUTRT	SUSTAT	SUDOSE	SUDOSU	SUSTDTC	SUSTDY	VISIT
001	SU	001-01-001	1	ALCOHOL		0	DRINKS	2011-02-08	-21	Screening
001	SU	001-01-001	2	ALCOHOL	NOT DONE		DRINKS	2011-02-09	-20	Screening
001	SU	001-01-001	3	ALCOHOL		5	DRINKS	2011-02-10	-19	Screening
…….										
001	SU	001-01-001	18	ALCOHOL		3	DRINKS	2011-02-25	-4	Screening
001	SU	001-01-001	19	ALCOHOL	NOT DONE		DRINKS	2011-02-26	-3	Screening
001	SU	001-01-001	20	ALCOHOL		0	DRINKS	2011-02-27	-2	Screening
001	SU	001-01-001	21	ALCOHOL		0	DRINKS	2011-02-28	-1	Screening
001	SU	001-01-001	22	ALCOHOL		0	DRINKS	2011-03-01	1	Visit 1 (Day 1 to 14)
001	SU	001-01-001	23	ALCOHOL	NOT DONE		DRINKS	2011-03-02	2	Visit 1 (Day 1 to 14)
001	SU	001-01-001	24	ALCOHOL		0	DRINKS	2011-03-03	3	Visit 1 (Day 1 to 14)
001	SU	001-01-001	25	ALCOHOL		2	DRINKS	2011-03-04	4	Visit 1 (Day 1 to 14)
001	SU	001-01-001	26	ALCOHOL	NOT DONE		DRINKS	2011-03-05	5	Visit 1 (Day 1 to 14)
001	SU	001-01-001	27	ALCOHOL		5	DRINKS	2011-03-06	6	Visit 1 (Day 1 to 14)
001	SU	001-01-001	28	ALCOHOL		0	DRINKS	2011-03-07	7	Visit 1 (Day 1 to 14)
……..										
001	SU	001-01-001	57	ALCOHOL		2	DRINKS	2011-04-05	36	Visit 3 (Day 29 to 42)
001	SU	001-01-001	58	ALCOHOL	NOT DONE		DRINKS	2011-04-06	37	Visit 3 (Day 29 to 42)
001	SU	001-01-001	59	ALCOHOL		4	DRINKS	2011-04-07	38	Visit 3 (Day 29 to 42)
001	SU	001-01-001	60	ALCOHOL		0	DRINKS	2011-04-08	39	Visit 3 (Day 29 to 42)
001	SU	001-01-001	61	ALCOHOL		2	DRINKS	2011-04-09	40	Visit 3 (Day 29 to 42)
001	SU	001-01-001	62	ALCOHOL		1	DRINKS	2011-04-10	41	Visit 3 (Day 29 to 42)
001	SU	001-01-001	63	ALCOHOL		4	DRINKS	2011-04-11	42	Visit 3 (Day 29 to 42)
……										

The table 1.1 shows the original SDTM SU dataset. All other examples will be derived from this SDTM dataset.

Table 1.2. Analysis Dataset Metadata for ADDR

Dataset Name	Dataset Description	Dataset Location	Dataset Structure	Key Variables of Dataset	Class of Dataset	Documentation
ADDR	Drinking Rate Analysis Data	addr.xpt	one record per subject per parameter per analysis visit	USUBJID, PARAMCD, AVISITN	BDS	c-addr.txt

Table 1.3. Analysis Variable Metadata including Analysis Parameter Value-Level Metadata for ADDR

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADDR	*ALL*	USUBJID	Unique Subject Identifier	text	$20		ADSL.USUBJID
ADDR	*ALL*	SITEID	Site ID	text	$20		ADSL.SITEID
ADDR	*ALL*	SEX	Sex	text	$20	M, F	ADSL.SEX
ADDR	*ALL*	FASFL	Full Analysis Set Population Flag	text	$1	Y, N	ADSL.FASFL
ADDR	*ALL*	TRTPN	Planned Treatment (N)	integer	1.0	1 = Placebo, 2 = Study Drug	ADSL.TRTPN
ADDR	*ALL*	TRTP	Planned Treatment	text	$20	Placebo, Study Drug	ADSL.TRTP
ADDR	PARAMCD	PARAMCD	Parameter Code	text	$8	ADDRATE	
ADDR	*ALL*	PARAM	Parameter	text	$50	Average Daily Drinking Rate	
ADDR	*ALL*	PARAMTYP	Parameter Type	text	$20	DERIVED	
ADDR	*ALL*	AVISITN	Analysis Visit (N)	integer	3.0	1 = Baseline, 2 = Treatment Period	
ADDR	*ALL*	AVISIT	Analysis Visit	text	$20	Baseline, Treatment Period	‘Baseline’ when SU.VISIT = ‘Screening’ ‘Treatment Period’ when SU.VISIT in (‘VISIT 1’, ‘VISIT 2’, ‘VISIT 3’)
ADDR	*ALL*	AVAL	Analysis Value	float	8.2		Average Daily Drinking within analysis visit. If a patient discontinues early or have missing records, impute with baseline rate. See pg 20 in SAP for further explanation.
ADDR	*ALL*	ABLFL	Baseline Record flag	text	$1	Y	‘Y’ at AVISIT = “Baseline”

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADDR	*ALL*	BASE	Baseline Value	float	8.2		AVAL of AVISIT = “Baseline”
ADDR	*ALL*	CHG	Change from Baseline	float	8.2		AVAL - BASE

The table 1.2 and 1.3 are the metadata for ADDR.

Table 1.4. Final ADaM dataset of ADDR

USUBJID	SITEID	SEX	FASFL	TRTPN	TRTP	PARAMCD	PARAM	AVISITN	AVIST	ABLFL	AVAL	BASE	CHG
001-01-001	01	M	Y	2	Study Drug	ADDRATE	Average Daily Drinking Rate	1	Baseline	Y	4.40		
001-01-001	01	M	Y	2	Study Drug	ADDRATE	Average Daily Drinking Rate	2	Treatment Period		2.72	4.40	-1.68
001-01-002	01	F	Y	1	Placebo	ADDRATE	Average Daily Drinking Rate	1	Baseline	Y	4.26		
001-01-002	01	F	Y	1	Placebo	ADDRATE	Average Daily Drinking Rate	2	Treatment Period		3.10	4.26	-1.16

Key points to note in the example are:
Row 2: There are 3 missing assessments during the treatment period for the subject of 001, so the baseline rate imputation method was applied as follow

$$\frac{2.60 * 39 + 4.40 * (42-39)}{42} = 2.72$$

In table 1.4, please note that we do not include all the variables in the metadata of table 1.3 because of space constraints.

The final ADaM dataset of ADDR in table 1.4 is created from the SDTM dataset of SU in the table 1.1 according to the metadata in the table 1.2 and 1.3. As explained above, the baseline rate imputation was applied to subject 001 for the treatment period, but looking at the dataset, we do not know whether the baseline rate imputation method was used or how it was applied. In the other methods, we are able to see whether or not the baseline rate imputation method was used and how it was applied.

SDTM through intermediate permanent datasets to ADaM

We created SDTM+ dataset and ADaM+ dataset as intermediate permanent datasets. The table 2.1 shows SDTM+ dataset and table 2.2 shows ADaM+ dataset.

Table 2.1. SDTM+ dataset of _SU

STUDYID	DOMAIN	USUBJID	SUSEQ	SUTRT	SUSTAT	SUDOSE	SUDOSU	SUSTDTC	SUSTDY	VISIT	_HOSEQ	_SDSEQ
001	SU	001-01-001	1	ALCOHOL		0	DRINKS	2011-02-08	-21	Screening	1	
001	SU	001-01-001	2	ALCOHOL	NOT DONE		DRINKS	2011-02-09	-20	Screening		
001	SU	001-01-001	3	ALCOHOL		5	DRINKS	2011-02-10	-19	Screening	2	
…….												
001	SU	001-01-001	18	ALCOHOL		3	DRINKS	2011-02-25	-4	Screening	17	
001	SU	001-01-001	19	ALCOHOL	NOT DONE		DRINKS	2011-02-26	-3	Screening		
001	SU	001-01-001	20	ALCOHOL		0	DRINKS	2011-02-27	-2	Screening	18	
001	SU	001-01-001	21	ALCOHOL		0	DRINKS	2011-02-28	-1	Screening	19	
001	SU	001-01-001	22	ALCOHOL		0	DRINKS	2011-03-01	1	Visit 1 (Day 1 to 14)		1
001	SU	001-01-001	23	ALCOHOL	NOT DONE		DRINKS	2011-03-02	2	Visit 1 (Day 1 to 14)		
001	SU	001-01-001	24	ALCOHOL		0	DRINKS	2011-03-03	3	Visit 1 (Day 1 to 14)		2
001	SU	001-01-001	25	ALCOHOL		2	DRINKS	2011-03-04	4	Visit 1 (Day 1 to 14)		3
001	SU	001-01-001	26	ALCOHOL	NOT DONE		DRINKS	2011-03-05	5	Visit 1 (Day 1 to 14)		
001	SU	001-01-001	27	ALCOHOL		5	DRINKS	2011-03-06	6	Visit 1 (Day 1 to 14)		4
001	SU	001-01-001	28	ALCOHOL		0	DRINKS	2011-03-07	7	Visit 1 (Day 1 to 14)		5
……..												
001	SU	001-01-001	57	ALCOHOL		2	DRINKS	2011-04-05	36	Visit 3 (Day 29 to 42)		34
001	SU	001-01-001	58	ALCOHOL	NOT DONE		DRINKS	2011-04-06	37	Visit 3 (Day 29 to 42)		
001	SU	001-01-001	59	ALCOHOL		4	DRINKS	2011-04-07	38	Visit 3 (Day 29 to 42)		35
001	SU	001-01-001	60	ALCOHOL		0	DRINKS	2011-04-08	39	Visit 3 (Day 29 to 42)		36
001	SU	001-01-001	61	ALCOHOL		2	DRINKS	2011-04-09	40	Visit 3 (Day 29 to 42)		37
001	SU	001-01-001	62	ALCOHOL		1	DRINKS	2011-04-10	41	Visit 3 (Day 29 to 42)		38
001	SU	001-01-001	63	ALCOHOL		4	DRINKS	2011-04-11	42	Visit 3 (Day 29 to 42)		39

Key points to note in the example are:
Rows 2, 6, 10, 13 and 18: when SUSTAT = 'NOT DONE', _HOSEQ and _SDSEQ are not increased by 1.

The SDTM plus variables, _HOSEQ and _SDSEQ, are added to the raw SDTM dataset of SU. _HOSEQ is the sequence number of non-missing drinking assessment from the hospitalization date (2011-02-08) and _SDSEQ is the sequence number of non-missing drinking assessment from the first dose date (2011-03-01). In table 2.1, _HOSEQ and _SDSEQ are SDTM plus variables which help us see how some of the algorithms are applied. Using the SDTM plus variables of _HOSEQ and _SDSEQ, we understand that the "NOT DONE" records will not be used for further algorithm calculations.

Table 2.2. ADaM + dataset of _ADDR

USUBJID	SITEID	SEX	FASFL	TRTPN	TRTP	PARAMCD	PARAM	AVISITN	AVIST	ABLFL	AVAL	BASE	CHG	_TOTAL	_DAYS	_AVAL
001-01-001	01	M	Y	2	Study Drug	ADDRATE	Average Daily Drinking Rate	1	Baseline	Y	4.40			83.6	19	4.40
001-01-001	01	M	Y	2	Study Drug	ADDRATE	Average Daily Drinking Rate	2	Treatment Period		2.72	4.40	-1.68	101.2	39	2.60
001-01-002	01	F	Y	1	Placebo	ADDRATE	Average Daily Drinking Rate	1	Baseline	Y	4.26			89.4	21	4.26
001-01-002	01	F	Y	1	Placebo	ADDRATE	Average Daily Drinking Rate	2	Treatment Period		3.10	4.26	-1.16	130.2	42	3.10

Key points to note in the example are:
Row 1 to 4: _TOTAL, _DAYS and _AVAL are the ADaM plus variables. _TOTAL is the sum of SUDOSE within analysis visits and _DAYS is the number of days where drinking was assessed. _AVAL is _TOTAL/_DAYS.
Row 1 and 3: At baseline visit, AVAL = _AVAL
Row 2 and 4: At treatment period visit, we applied the algorithm of AVAL = (_AVAL * _DAYS + BASE * (42 - _DAYS)) / _DAYS
Row 1: _DAYS indicates that there are 19 drinking days assessed at baseline visit. At the baseline visit, we do not apply the imputation method even though some of the days did not have an assessment of drinking.
Row 2: _DAYS= 39 indicates that there are 3 missing assessments. So, the baseline rate imputation was applied like below

$$\frac{2.60 * 39 + 4.40 * (42-39)}{42} = 2.72$$

Row4: For subject 002, since _DAYS is equal to 42, _AVAL is equal to AVAL if the AVAL algorithm was applied

$$\frac{3.10 * 42 + 4.46 * (42-42)}{42} = 3.10$$

If we remove plus variables (_TOTAL, _DAYS and _AVAL), the final ADaM dataset of ADDR is finished like that in the table 1.4. For this process, we do not provide the metadata and the final ADaM dataset examples since they are same as those in the table 1.2, 1.3 and 1.4.

We also establish some business rules for SDTM plus and ADaM plus datasets.

- Plus datasets – we create them in the same SAS program as the final ADaM dataset development program. We do not have separate dataset programs for the intermediate permanent datasets.
- Plus variables – we keep them as the temporary variables by adding the prefix '_', so we can easily remove them if needed.
- No Standard for plus variables – we assign the labels, but do not follow any CDISC standards.
- Same number of the records – we keep the same number of records between SDTM datasets and SDTM plus datasets and also ADaM datasets and ADaM plus datasets.

SDTM through ADaM to ADaM

The table 3.1 shows the intermediate ADaM dataset of ADSU and table 3.3 shows the final ADaM dataset of ADDR.

Table 3.1. Intermediate ADaM dataset of ADSU (Daily Drinking Rate Analysis Dataset)

USUBJID	PARAMCD	PARAM	AVAL	ADT	AVISIT	VISIT	DTYPE	ASEQ	SUSEQ
001-01-001	DDRATE	Daily Drinking Rate	0	2011-02-08	Baseline	Screening		1	1
001-01-001	DDRATE	Daily Drinking Rate	5	2011-02-10	Baseline	Screening		2	3
…..									
001-01-001	DDRATE	Daily Drinking Rate	3	2011-02-25	Baseline	Screening		17	18
001-01-001	DDRATE	Daily Drinking Rate	0	2011-02-27	Baseline	Screening		18	20
001-01-001	DDRATE	Daily Drinking Rate	0	2011-02-28	Baseline	Screening		19	21
001-01-001	DDRATE	Daily Drinking Rate	4.4		Baseline		AVERAGE	20	
001-01-001	DDRATE	Daily Drinking Rate	0	2011-03-01	Treatment Period	Visit 1 (Day 1 to 14)		21	22
001-01-001	DDRATE	Daily Drinking Rate	0	2011-03-03	Treatment Period	Visit 1 (Day 1 to 14)		22	24
001-01-001	DDRATE	Daily Drinking Rate	2	2011-03-04	Treatment Period	Visit 1 (Day 1 to 14)		23	25
001-01-001	DDRATE	Daily Drinking Rate	5	2011-03-06	Treatment Period	Visit 1 (Day 1 to 14)		24	27
001-01-001	DDRATE	Daily Drinking Rate	0	2011-03-07	Treatment Period	Visit 1 (Day 1 to 14)		25	28
…..									
001-01-001	DDRATE	Daily Drinking Rate	2	2011-04-05	Treatment Period	Visit 3 (Day 29 to 42)		54	57
001-01-001	DDRATE	Daily Drinking Rate	4	2011-04-07	Treatment Period	Visit 3 (Day 29 to 42)		55	59
001-01-001	DDRATE	Daily Drinking Rate	0	2011-04-08	Treatment Period	Visit 3 (Day 29 to 42)		56	60

USUBJID	PARAMCD	PARAM	AVAL	ADT	AVISIT	VISIT	DTYPE	ASEQ	SUSEQ
001-01-001	DDRATE	Daily Drinking Rate	2	2011-04-09	Treatment Period	Visit 3 (Day 29 to 42)		57	61
001-01-001	DDRATE	Daily Drinking Rate	1	2011-04-10	Treatment Period	Visit 3 (Day 29 to 42)		58	62
001-01-001	DDRATE	Daily Drinking Rate	4	2011-04-11	Treatment Period	Visit 3 (Day 29 to 42)		59	63
001-01-001	DDRATE	Daily Drinking Rate	2.6		Treatment Period		AVERAGE	60	

The row at SUSEQ of 2, 26, 30, 33 and 65 from SU were not included in ADSU since the assessments were not done, so for baseline visit, we only include 19 records.
We used DYPTE='AVERAGE' to achieve the average of assessed doses at baseline visit at ASEQ = 20.
ASEQ is "Analysis Sequence Number" that is introduced in the new ADaM documents of Time to Event Analysis. It will be used for the data point traceability of ADaM datasets to ADaM datasets.

Table 3.2. Analysis Variable Metadata including Analysis Parameter Value-Level Metadata for ADDR

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADDR	*ALL*	USUBJID	Unique Subject Identifier	text	$20		ADSL.USUBJID
ADDR	*ALL*	SITEID	Site ID	text	$20		ADSL.SITEID
ADDR	*ALL*	SEX	Sex	text	$20	M, F	ADSL.SEX
ADDR	*ALL*	FASFL	Full Analysis Set Population Flag	text	$1	Y, N	ADSL.FASFL
ADDR	*ALL*	TRTPN	Planned Treatment (N)	integer	1.0	1 = Placebo, 2 = Study Drug	ADSL.TRTPN
ADDR	*ALL*	TRTP	Planned Treatment	text	$20	Placebo, Study Drug	ADSL.TRTP
ADDR	PARAMCD	PARAMCD	Parameter Code	text	$8	ADDRATE	
ADDR	*ALL*	PARAM	Parameter	text	$50	Average Daily Drinking Rate	
ADDR	*ALL*	AVISITN	Analysis Visit (N)	integer	3.0	1 = Baseline, 2 = Treatment Period	
ADDR	*ALL*	AVISIT	Analysis Visit	text	$20	Baseline, Treatment Period	

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADDR	*ALL*	AVAL	Analysis Value	float	8.2		ADSU.AVAL at ADSU.DTYPE="AVERAGE" for ADSU.AVISIT = 'Baseline' ADSU.AVAL at ADSU.DTYPE="AVERAGE" for ADSU.AVISIT = "Treatment Period" if there are 42 assessment. If there are less than 42 assessments, impute with baseline rate. See pg 20 in SAP for further explanation.
ADDR	*ALL*	ABLFL	Baseline Record Flag	text	$1	Y	'Y' at AVISIT = "Baseline"
ADDR	*ALL*	BASE	Baseline Value	float	8.2		AVAL of AVISIT = "Baseline"
ADDR	*ALL*	CHG	Change from Baseline	float	8.2		AVAL - BASE
ADDR	*ALL*	SRCDOM	Source Data	text	$8	ADSU	"ADSU" if there is no imputation
ADDR	*ALL*	SRCSEQ	Source Sequence Number	integer	8.0		ASEQ if there is no imputation

Table 3.3. Final ADaM dataset of ADDR

USUBJID	SITEID	SEX	FASFL	TRTPN	TRTP	PARAMCD	PARAM	AVISITN	AVIST	ABLFL	AVAL	BASE	CHG	SRCDOM	SRCSEQ
001-01-001	01	M	Y	2	Study Drug	ADDRATE	Average Daily Drinking Rate	1	Baseline	Y	4.40			ADSU	20
001-01-001	01	M	Y	2	Study Drug	ADDRATE	Average Daily Drinking Rate	2	Treatment Period		2.72	4.40	-1.68		
001-01-002	01	F	Y	1	Placebo	ADDRATE	Average Daily Drinking Rate	1	Baseline	Y	4.26			ADSU	22
001-01-002	01	F	Y	1	Placebo	ADDRATE	Average Daily Drinking Rate	2	Treatment Period		3.10	4.26	-1.16	ADSU	65

Key points to note in the example are:
Row 1, 3 and 4: Since there is no baseline rate imputation, AVAL comes straight from ADSU.
Row 2: Since there are 3 missing assessments, we applied the baseline rate imputation method as

$$\frac{2.60 * 39 + 4.40 * (42-39)}{42} = 2.72$$

The final ADaM dataset of ADDR in the table 3.3 is created from the intermediate ADaM dataset of ADSU in the table 3.1 according to the metadata in the table 3.2.

CONSIDERATION

The methods of SDTM to ADaM and SDTM through intermediate datasets to ADaM provide only SDTM dataset and final ADaM dataset in the submission. Meanwhile, the process of SDTM through ADaM to final ADaM provides SDTM dataset, intermediate ADaM dataset and final ADaM dataset. From the reviewer's perspective, first two methods provide the same outputs since we do not submit the SDTM+ and ADaM+ datasets, but the last one provides the intermediate ADaM datasets for data point traceability along with SDTM and final ADaM dataset.

CONCLUSION

Some analyses in the clinical trial require complex algorithms. The complex algorithms could require several data manipulations steps and more than one SDTM dataset for creation of a final ADaM dataset. A SAS programmer could create a complex ADaM dataset in three different methods - SDTM to ADaM, SDTM through the intermediate permanent datasets such as SDTM+ and ADaM+ to ADaM and SDTM through the intermediate ADaM to the final ADaM. Each method has its strengths and limitations as shown above. Which method ends up being used will depend on the type and complexity of the analysis. If a SAS programmer understands how each process works, he or she will be in a much better position to decide which data set creation method will be best fulfill the requirements of the specified analyses.

REFERENCES

The ADaM Implementation Guide, Version V 1.0 (ADaMIG v1.0)
The Analysis Data Model, Version 2.1 (ADaM 2.1)
The ADaM Basic Data Structure for Time-to-Event Analyses

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
Cytel, Inc.
Chesterbrook, PA
(610) 994 - 9840
Email:Kevin.lee@cytel.com

TRADEMARKS

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 registered trademarks or trademarks of their respective companies.