PharmaSUG 2019 – Paper ST-160

Experiences in Building CDISC Compliant ADaM Dataset to Support Multiple Imputation Analysis for Clinical Trials

Xiangchen (Bob) Cui, Alkermes, Inc, Waltham, MA

ABSTRACT

Multiple imputation (MI) is becoming an increasingly popular method to address the missing data problem in regulatory clinical trials, especially when the outcome variables come from repeated assessments. SAS® procedures, PROC MI and PROC MIANALYZE, apply the multiple imputation techniques to generate multiple imputations for incomplete multivariate data and to analyze results from multiply imputed data sets, respectively.

How to use PROC MI to build CDISC compliant ADaM dataset to support MI analysis is a new ADaM programming technique. This paper illustrates how to apply ADaM BDS data structure to build such one through an example. We will not present how to use PROC MI procedure, for it has been very well explained in SAS® user manual and other papers. However we do provide some tutorial of related statistical concepts to help Statistical Programmers to better understand this procedure and apply it in ADaM programming. We mainly focus on ADaM programming logic flow, key variable derivations for the imputed data including ADaM specification writing, and programming independent validation process. Some tips and pitfalls provided in this paper could be time-saving ones, and assist you in your programming to achieve technical accuracy and operational efficiency.

The sharing of hands-on experiences in this paper is intended to assist readers to prepare CDISC compliant ADaM dataset to facilitate MI analysis in regulatory clinical trials, and further to support FDA submission.

INTRODUCTION

The paper [1] reviews the basic concepts and applications of multiple imputation techniques for analyzing missing data, and introduces the SAS procedures, PROC MI and PROC MIANALYZE, which generate multiple imputations for incomplete multivariate data and analyze results from multiply imputed data sets, respectively. It provides very detailed examples to illustrate these procedures. The paper [2] describes the 3-step process in order to perform MI analyses, and presents ADaM programming technique by the ADaM BDS structure.

Building a CDISC compliant ADaM dataset to support MI analysis is very critical for FDA submission, in addition to the analysis of clinical trials. This new ADaM programming technique is illustrated through a hypothetical example in this paper. The ADaM programming logic flow, key variable derivations for the imputed data including ADaM specification writing, and programming independent validation process are presented for the tutorial. The tips presented in this paper could benefit the readers if they work on the ADaM programming for MI to support analysis of regulatory clinical trials and/or FDA submission.

A HYPOTHETICAL EXAMPLE OF ADAM BDS DATA STRUCTURE FROM BODY WEIGHT FOR PRIMARY EFFICACY ANALYSIS

An ADaM BDS dataset, named as **ADWT**, stored body weights across the clinical study visits: V2 Week 0, V3 Week1, …, V17 Week 24/ET, V18 week 28_FU. Please refer to Display 1 for an example. **ADWT** had been developed from STDM.VS dataset per SAP and its shell to support TFLs, except for multiple imputation analysis. The treatment period of this study was from V2 Week 0 to V17 Week 24. The follow-up visit for body weights was conducted after the last dose of study drug, either completion of treatment or earlier discontinuation of treatment. The body weights at an early termination (ET) visit, V17 Week 24_ET, were mapped to the scheduled visit for analysis per SAP. The variables **AVISIT** and **ANL01FL,** which were used to select records for statistical analysis, had been derived in **ADWT** programming.

Display 1 below shows an example from a subject (USUBJID='xxx-001') who completed the treatment. The subject had an unscheduled visit after V9 Week 8, which was excluded from multiple imputation (MI)

per SAP, and **ANL01FL** was set to **''**. Display 2 below shows an example from a subject who early discontinued the treatment after V13 Week 16. Due to the windowing, the ET record could not be mapped into a "scheduled" visit, and **ANL01FL** was set to **''** in ADWT programing. Hence the record at this visit was considered "missing" for efficacy analysis, even though it did have value collected in the study!

The first subject had complete data from baseline to V17 Week 24 with ten (10) scheduled visits during the treatment period. The second subject had missing data at V15 Week 20 and V17 Week 24.

USUBJID	VISIT	VISITNUM	APHASE	AVISITN	AVISIT	ADT	AVAL	ANL01FL
xxx-001	V2 Week 0	2	PRE-TREATMENT	0	Baseline	2016-07-08	80	Y
xxx-001	V3 Week 1	3	TREATMENT	8	V3-Day 8	2016-07-14	81.2	Y
xxx-001	V4 Week 2	4	TREATMENT	15	V4-Day 15	2016-07-21	80.2	Y
xxx-001	V6 Week 4	6	TREATMENT	29	V6-Day 29	2016-08-05	82.1	Y
xxx-001	V8 Week 6	8	TREATMENT	43	V8-Day 43	2016-08-19	80.8	Y
xxx-001	V9 Week 8	9	TREATMENT	57	V9-Day 57	2016-09-02	81.8	Y
xxx-001	Unscheduled 9.1	9.1	TREATMENT	.		2016-09-16	83.1	
xxx-001	V11 Week 12	11	TREATMENT	85	V11-Day 85	2016-09-30	81.9	Y
xxx-001	V13 Week 16	13	TREATMENT	113	V13-Day 113	2016-10-28	81.4	Y
xxx-001	V15 Week 20	15	TREATMENT	141	V15-Day 141	2016-11-28	81	Y
xxx-001	V17 Week 24_ET	17	TREATMENT	169	V17-Day 169	2016-12-22	80.6	Y
xxx-001	V18 Week 28_FU	18	FOLLOW-UP	197	V18-Day 197-Safety Follow-up	2017-01-20	78.9	

Display 1. An Example of Body Weight Data from ADaM.ADWT, Who Completed the Treatment and Had Complete Data

USUBJID	VISIT	VISITNUM	APHASE	AVISITN	AVISIT	ADT	AVAL	ANL01FL
xxx-003	V2 Week 0	2	PRE-TREATMENT	0	Baseline	2016-09-27	89.4	Y
xxx-003	V3 Week 1	3	TREATMENT	8	V3-Day 8	2016-10-04	88.6	Y
xxx-003	V4 Week 2	4	TREATMENT	15	V4-Day 15	2016-10-11	89	Y
xxx-003	V6 Week 4	6	TREATMENT	29	V6-Day 29	2016-10-25	89.4	Y
xxx-003	V8 Week 6	8	TREATMENT	43	V8-Day 43	2016-11-07	90.1	Y
xxx-003	V9 Week 8	9	TREATMENT	57	V9-Day 57	2016-11-22	91.4	Y
xxx-003	V11 Week 12	11	TREATMENT	85	V11-Day 85	2016-12-19	91.8	Y
xxx-003	V13 Week 16	13	TREATMENT	113	V13-Day 113	2017-01-17	93.1	Y
xxx-003	V15 Week 20	15		.				
xxx-003	V17 Week 24_ET	17	TREATMENT	.		2017-01-23	93.2	

Display 2. An Example of Body Weight Data from ADaM.ADWT, Who Early Discontinued the Treatment and Had Missing Data with a Monotone Missing Patten

INTRODUCTION OF MISSING DATA PATTEN

To better understand PROC MI procedure and apply it in the ADaM programming, understanding ***missing data pattern*** is one of the most important aspects. The ***missing data pattern*** can be classified as arbitrary or monotone.

If the miss of data occurs in a random fashion in between the visits then the data set is said to have an ***arbitrary missing pattern***. A data set is said to have a ***monotone missing pattern*** when the data is missing at a certain visit for a subject, as well as all subsequent ones.

Display 3 below shows an example from a subject who early discontinued the treatment and had missing data at on-treatment visits V9 Week 8, V15 Week 20, and V17 Week 24. Hence the subject had an arbitrary missing pattern. Display 2 above shows an example from a subject who early discontinued the treatment and had complete data from V2 Week 0 to V13 Week 16 and had missing data at on-treatment visits V15 Week 20 and V17 Week 24. Hence the subject had a monotone missing pattern.

Please note that both subjects had "missing data" at V17 Week 24 per analysis perspective. For subject xxx-003, the data was collect at V17 Week 24, but it was excluded from efficacy analysis per SAP due to the fact that it could not be mapped to a scheduled visit. For subject: xxx-002, the record was not collected at V17 Week 24.

USUBJID	VISIT	VISITNUM	APHASE	AVISITN	AVISIT	ADT	AVAL	ANL01FL
xxx-002	V2 Week 0	2	PRE-TREATMENT	0	Baseline	2016-12-15	82.7	Y
xxx-002	V3 Week 1	3	TREATMENT	8	V3-Day 8	2016-12-22	82.9	Y
xxx-002	V4 Week 2	4	TREATMENT	15	V4-Day 15	2016-12-29	83.1	Y
xxx-002	V6 Week 4	6	TREATMENT	29	V6-Day 29	2017-01-09	83	Y
xxx-002	V8 Week 6	8	TREATMENT	43	V8-Day 43	2017-01-25	84.1	Y
xxx-002	V9 Week 8	9	.	.	.	.	.	
xxx-002	V11 Week 12	11	TREATMENT	85	V11-Day 85	2017-03-07	83.5	Y
xxx-002	V13 Week 16	13	TREATMENT	113	V13-Day 113	2017-04-04	84	Y
xxx-002	V15 Week 20	15	.	.	.	.	.	
xxx-002	V17 Week 24_ET	17	.	.	.	.	.	

Display 3. An Example of Body Weight Data from ADaM.ADWT with an Arbitrary Missing Patten

INTRODUCTION OF PROGRAMMING LOGIC FLOW FOR MI DATA DERIVATION

Figure 1 below shows the programming logic flow for generating CDISC compliant ADaM dataset with multiple imputation.

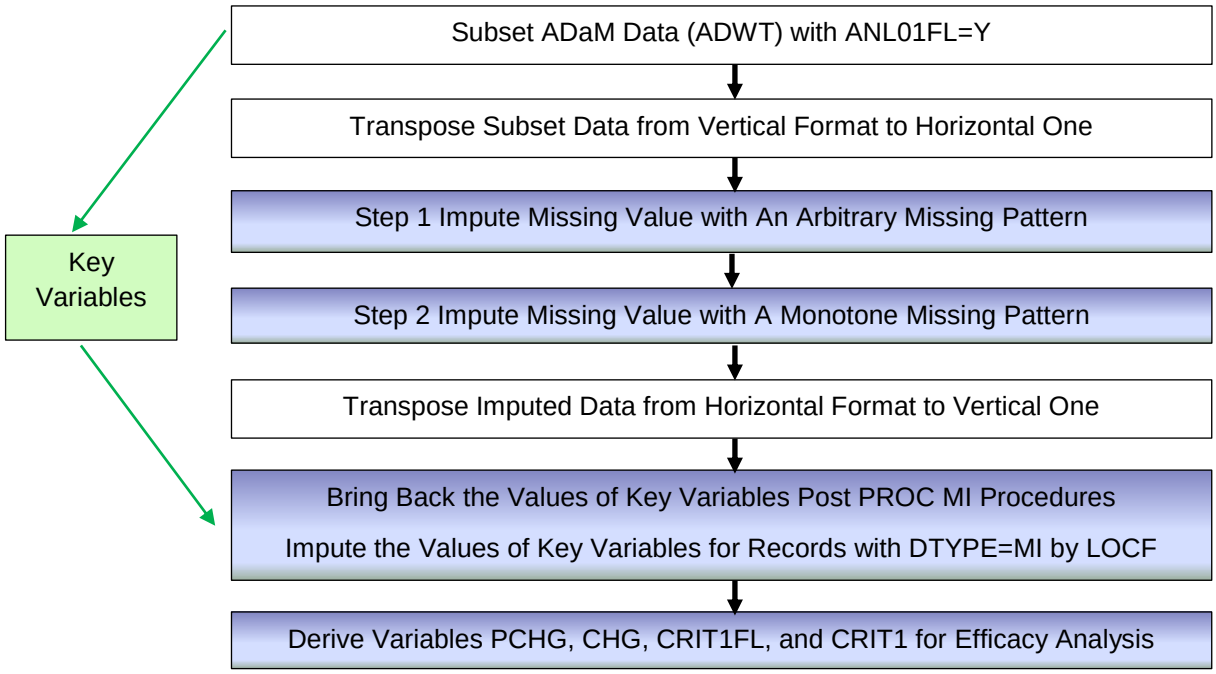

Figure 1. Flow Chart of Generating CDISC Compliant ADaM Dataset with Multiple Imputation

PREPARE A SUBSET OF ADWT WITH ANL01FL='Y' INCLUDING VARIABLES FOR STATISTICAL MODELS FOR EFFICACY ANALYSIS PER SAP

Since only the records of these body weights from scheduled on-treatment assessments with ANL01FL='Y' are used in efficacy analysis, a subset of **ADWT** with the condition, ANL01FL='Y', is needed to further derive MI dataset. Display 4 shows an example of these three subjects' data with the condition: ANL01FL='Y'. This is the first step for SAS programming to prepare the input dataset for multiple imputation (MI) derivation.

In the hypothetical example, the statistical models for efficacy analysis per SAP include the treatment group (ADWT.TRTP), race group (ADWT.RACEGR1), and age group (ADWT.AGEGR1) as factors, and the baseline weight as the covariate. TRTP, RACEGR1, and AGEGR1 should be merged into the subset of ADWT for the following two-step for PROC MI procedures. Display 5 shows an example of these factors from these three subjects. Note that the numerical versions of these variables (TRTPN, RACEGR1, and AGEGR1) are also included for the easiness of programming further down the road.

USUBJID	VISIT	VISITNUM	APHASE	AVISITN	AVISIT	ADT	AVAL	ANL01FL
xxx-001	V2 Week 0	2	PRE-TREATMENT	0	Baseline	2016-07-08	80	Y
xxx-001	V3 Week 1	3	TREATMENT	8	V3-Day 8	2016-07-14	81.2	Y
xxx-001	V4 Week 2	4	TREATMENT	15	V4-Day 15	2016-07-21	80.2	Y
xxx-001	V6 Week 4	6	TREATMENT	29	V6-Day 29	2016-08-05	82.1	Y
xxx-001	V8 Week 6	8	TREATMENT	43	V8-Day 43	2016-08-19	80.8	Y
xxx-001	V9 Week 8	9	TREATMENT	57	V9-Day 57	2016-09-02	81.8	Y
xxx-001	V11 Week 12	11	TREATMENT	85	V11-Day 85	2016-09-30	81.9	Y
xxx-001	V13 Week 16	13	TREATMENT	113	V13-Day 113	2016-10-28	81.4	Y
xxx-001	V15 Week 20	15	TREATMENT	141	V15-Day 141	2016-11-28	81	Y
xxx-001	V17 Week 24_ET	17	TREATMENT	169	V17-Day 169	2016-12-22	80.6	Y
xxx-002	V2 Week 0	2	PRE-TREATMENT	0	Baseline	2016-12-15	82.7	Y
xxx-002	V3 Week 1	3	TREATMENT	8	V3-Day 8	2016-12-22	82.9	Y
xxx-002	V4 Week 2	4	TREATMENT	15	V4-Day 15	2016-12-29	83.1	Y
xxx-002	V6 Week 4	6	TREATMENT	29	V6-Day 29	2017-01-09	83	Y
xxx-002	V8 Week 6	8	TREATMENT	43	V8-Day 43	2017-01-25	84.1	Y
xxx-002	V11 Week 12	11	TREATMENT	85	V11-Day 85	2017-03-07	83.5	Y
xxx-002	V13 Week 16	13	TREATMENT	113	V13-Day 113	2017-04-04	84	Y
xxx-003	V2 Week 0	2	PRE-TREATMENT	0	Baseline	2016-09-27	89.4	Y
xxx-003	V3 Week 1	3	TREATMENT	8	V3-Day 8	2016-10-04	88.6	Y
xxx-003	V4 Week 2	4	TREATMENT	15	V4-Day 15	2016-10-11	89	Y
xxx-003	V6 Week 4	6	TREATMENT	29	V6-Day 29	2016-10-25	89.4	Y
xxx-003	V8 Week 6	8	TREATMENT	43	V8-Day 43	2016-11-07	90.1	Y
xxx-003	V9 Week 8	9	TREATMENT	57	V9-Day 57	2016-11-22	91.4	Y
xxx-003	V11 Week 12	11	TREATMENT	85	V11-Day 85	2016-12-19	91.8	Y
xxx-003	V13 Week 16	13	TREATMENT	113	V13-Day 113	2017-01-17	93.1	Y

Display 4. An Example of Body Weight Data from ADaM.ADWT with ANL01FL='Y'

USUBJID	TRTP	TRTPN	RACEGR1	RACEGR1N	AGEGR1	AGEGR1N
xxx-001	ABC	2	White	1	>=40 Years	2
xxx-002	Placebo	1	Non-White	2	<40 Years	1

USUBJID	TRTP	TRTPN	RACEGR1	RACEGR1N	AGEGR1	AGEGR1N
xxx-003	Placebo	1	White	1	>=40 Years	2

Display 5. An Example of the Treatment group (ADWT.TRTP), Race group (ADWT.RACEGR1), and Age group (ADWT.AGEGR1) as Factors for Statistical Models per SAP

TRANSPOSE A VERTICAL ADAM BDS INTO A HORIZONTAL FORMAT

PROC MI requires a horizontal data format, i.e., one record per subject and all values from different visits are presented in columns. The vertical ADaM data above should be transposed to a horizontal format.

TIPS & TRICKS:

1. To ease the programming in converting a vertical data format to a horizontal one and the usage of PROC MI procedures, a new variable, named as **N**, was created by the one-to-one mapping, shown below in Display 6.
2. The keys to sort the dataset for PROC TRANSPOSE must be the "the treatment group (ADWT.TRTP), race group (ADWT.RACEGR1), and age group (ADWT.AGEGR1)" to keep them in the new dataset, in addition to AVISIT and AVISITN.

	Analysis Visit (N)	N
1	0	1
2	8	2
3	15	3
4	29	4
5	43	5
6	57	6
7	85	7
8	113	8
9	141	9
10	169	10

Display 6. One-to-one Mapping to Create a New variable N for converting a vertical format to a horizontal format

```
proc sort data=adwt out=adwt01;
     by avisitn usubjid paramcd;
     where fasfl='Y' and not missing(trtp) and aval>.Z and
     0<=avisitn<=169 and anl01fl = 'Y' and dtype ne 'LOCF';
run;
proc sort data=adwt01 out=allavst(keep=avisitn) nodupkey;by avisitn;run;
data allavst;
     set allavst;
     by avisitn;
     n=_n_;
run;
data adwt02;
     merge adwt01 allavst;
     by avisitn;
run;
```

USUBJID	PARAMCD	TRTP	TRTPN	RACEGR1	RACEGR1N	AGEGR1	AGEGR1N	N	AVISITN	AVISIT	AVAL
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	1	0	Baseline	80
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	2	8	V3-Day 8	81.2
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	3	15	V4-Day 15	80.2

USUBJID	PARAMCD	TRTP	TRTPN	RACEGR1	RACEGR1N	AGEGR1	AGEGR1N	N	AVISITN	AVISIT	AVAL
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	4	29	V6-Day 29	82.1
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	5	43	V8-Day 43	80.8
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	6	57	V9-Day 57	81.8
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	7	85	V11-Day 85	81.9
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	8	113	V13-Day 113	81.4
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	9	141	V15-Day 141	81
xxx-001	DMWEIGHT	ABC	2	White	1	>=40 Years	2	10	169	V17-Day 169	80.6

Display 7. An Example of Body Weight Data in Vertical Format after Merging with New variable N

```
proc sort data=adam0;by usubjid trtp trtpn racegr1 racegr1n agegr1 agegr1n
avisitn avisit;run;
proc transpose data=adam0 out=adam1(drop=_name_ _label_) prefix=v;
     by usubjid trtp trtpn racegr1 racegr1n agegr1 agegr1n;
     id n;
     var aval;
run;
```

Display 8 shows the horizontal data from these three subjects with the horizontal format.

USUBJID	TRTP	TRTPN	RACEGR1	RACEGR1N	AGEGR1	AGEGR1N
xxx-001	ABC	2	White	1	>=40 Years	2
xxx-002	Placebo	1	Non-White	2	<40 Years	1
xxx-003	Placebo	1	White	1	>=40 Years	2

USUBJID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
xxx-001	80	81.2	80.2	82.1	80.8	81.8	81.9	81.4	81	80.6
xxx-002	82.7	82.9	83.1	83	84.1	.	83.5	84	.	.
xxx-003	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	.	.

Display 8. An Example of Body Weight Data from Transposing the Data from Vertical Format to Horizontal One in Display 4

KEEP THE VALUES FOR KEY VARIABLES FOR POST PROC MI PROCEDURES

The original values in dataset ADWT01 should be kept for the key variables: PARAMCD PARAM TRTA TRTAN AVISITN AVISIT BASE ADT ATM ADTM ASEQ VISIT VISITNUM ADY APHASE APHASEN AWTDIFF for the post PROC MI procedures. These variables will be dropped from PROC MI procedures, and are needed in the final ADaM dataset from MI for **CDISC ADaM compliance**.

Below SAS code shows the dataset, named as WT_AVISIT, which would be used to bring back the values for above variables, except AVISITN and AVISIT.

```
proc sort data=adwt02 out=wt_avisit(keep=usubjid trta trtan avisitn avisit
base adt atm adtm aseq visit visitnum ady aphase aphasen awtdiff);
     by usubjid avisitn avisit;
     where paramcd in ('DMWEIGHT');
run;
```

IMPUTATION STEP 1: IMPUTE THE MISSING VALUE WITH AN ARBITRARY MISSING PATTERN

The first step is to impute the missing value with an arbitrary missing pattern.

Below SAS code shows the first step of two-step PROC MI in the hypothetical example.

```
*** imputation step 1---- MI: MCMC for an arbitrary missing pattern ***;
proc mi data=adam1 out=mi_mono nimpute=500 seed=1104078;
     var trtpn racegr1n agegr1n v1-v10;
     mcmc chain=single impute=monotone PRIOR=JEFFREYS NBITER=200 NITER=100
     INITIAL=EM;
run;
```

TIPS & TRICKS:

1. PROC MI Options must be provided by study Biostatistician, for the different choices of options would generate different values of the imputed variables. The options should be documented in both SAP and ADaM specification, which is used for the independent ADaM validation from SAS programming and/or Study Biostatistician, and/or to support FDA submission.
2. The **VAR** statement above listing the variables to be analyzed, should match the statistical models for efficacy analysis per SAP, which included TRTPN, RACEGR1N, AGEGR1N, and all outcome variables coming from repeated assessments (V1-V10) in this hypothetical example.
3. To ease this ADaM programming, variables TRTP, RACEGR1, and AGEGR1 were dropped for the time being, and they will be added back at the end of ADaM programming.
4. Renaming the variable names from the transposed dataset to V1, V2, …., V10 can ease the programing in this step!
5. The value for **NIMPUTE** (the number of imputations) should be clearly specified in both SAP and ADaM Specification.
6. The value for **SEED** (the seed to begin random number generator) should be clearly specified in ADaM specification to support ADaM programming validation, and/or FDA submission.

The subject xxx-002 had missing data from variable V6 in Display 8 among the three subjects. Display 9 shows the imputed values for variable V6 for the subject (xxx-002) from the first 10 imputations. Note the first column with column name _**IMPUTATION**_ was automatically generated from PROC MI procedure.

For the subject (xxx-001), who had complete data, and the subject (xxx-003), who had the monotone missing pattern, their values were kept the same as before after Imputation Step 1.

IMPUTATION	USUBJID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	xxx-002	82.7	82.9	83.1	83	84.1	81.323044901	83.5	84	.	.
2	xxx-002	82.7	82.9	83.1	83	84.1	85.350664843	83.5	84	.	.
3	xxx-002	82.7	82.9	83.1	83	84.1	86.184785572	83.5	84	.	.
4	xxx-002	82.7	82.9	83.1	83	84.1	81.812767079	83.5	84	.	.
5	xxx-002	82.7	82.9	83.1	83	84.1	83.219090898	83.5	84	.	.
6	xxx-002	82.7	82.9	83.1	83	84.1	83.548999378	83.5	84	.	.
7	xxx-002	82.7	82.9	83.1	83	84.1	81.810641299	83.5	84	.	.
8	xxx-002	82.7	82.9	83.1	83	84.1	84.37896536	83.5	84	.	.
9	xxx-002	82.7	82.9	83.1	83	84.1	82.519305502	83.5	84	.	.
10	xxx-002	82.7	82.9	83.1	83	84.1	83.367566225	83.5	84	.	.

Display 9. An Example of Body Weight Data with Imputed Value for an Arbitrary Missing Pattern for 10 Imputations

IMPUTATION STEP 2: IMPUTE THE MISSING VALUE WITH A MONOTONE MISSING PATTERN

After Imputation Step 1, the data should ***ONLY*** have a ***monotone missing pattern*** if it has any missing data at this moment.

The second step is to impute the missing value with a monotone missing pattern from Step 1 output dataset, named as mi_mono.

Below SAS codes shows the second step of two-steps of PROC MI in this hypothetical example.

```
*** imputation step 2---- MI: regression method for a monotonic missing
patten ***;
proc mi data=mi_mono out=mi_reg nimpute=500 seed=1104078;
     by _imputation_;
     var trtpn racegr1n agegr1n v1-v10;
     class trtpn racegr1n agegr1n;
     monotone regression(v10/details);
run;
```

TIPS & TRICKS:

1. Same as Item 1 in Step 1 above!
2. Same as Item 2 in Step 1 above!
3. Same as Item 3 in Step 1 above!
4. Same as Item 4 in Step 1 above!
5. The ordering of factors in Step 1, for example, treatment group (ADWT.TRTPN), race group (ADWT.RACEGR1N), and age group (ADWT.AGEGR1N), has an effect on the generation of the imputed values for the missing values in Step 2, i.e., different orderings of these factors will generate different imputed values for the monotone missing pattern from PRCO MI procedure above.
6. The ordering of subjects in the dataset also has the effect on the generation of the imputed values for the missing values in Step 2.
7. There is no “data manipulation” from the output of Step 1 to Step 2 per Tip 5 and 6!
8. The VAR statements from these two steps should be the same in both PD and QC of ADaM programming, which should be documented in ADaM Specification.

Tips 5 and 6 are due to the SAS PROC MI algorithm for Step 2. The lesson was learned from our independent programming validation of this ADaM dataset in order to have the 100% match between the production and validation.

The subjects (xxx-002 and xxx-003) had missing values at V9 and V10 in Display 9 above. The second PROC MI procedure call generated the imputation for them. Display 10 below shows the imputed values for V9 and V10 from the first 10 imputations.

USUBJID	_IMPUTATION_	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
xxx-002	1	82.7	82.9	83.1	83	84.1	81.323044901	83.5	84	82.122979957	84.047660998
xxx-002	2	82.7	82.9	83.1	83	84.1	85.350664843	83.5	84	86.681516159	88.451147043
xxx-002	3	82.7	82.9	83.1	83	84.1	86.184785572	83.5	84	83.946084815	84.076909047
xxx-002	4	82.7	82.9	83.1	83	84.1	81.812767079	83.5	84	84.161109924	87.025908824
xxx-002	5	82.7	82.9	83.1	83	84.1	83.219090898	83.5	84	85.193002945	84.068936456
xxx-002	6	82.7	82.9	83.1	83	84.1	83.548999378	83.5	84	81.804072031	83.227144955
xxx-002	7	82.7	82.9	83.1	83	84.1	81.810641299	83.5	84	87.176039288	85.908500528
xxx-002	8	82.7	82.9	83.1	83	84.1	84.37896536	83.5	84	87.18278964	84.057085173
xxx-002	9	82.7	82.9	83.1	83	84.1	82.519305502	83.5	84	85.195781186	83.136138655

USUBJID	_IMPUTATION_	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
xxx-002	10	82.7	82.9	83.1	83	84.1	83.367566225	83.5	84	83.537948631	84.59719923
xxx-003	1	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	89.563776764	88.669968516
xxx-003	2	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	95.50885283	94.93717963
xxx-003	3	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	91.122020921	89.077957944
xxx-003	4	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	93.547025544	96.913999227
xxx-003	5	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	92.798910796	92.535564757
xxx-003	6	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	91.637832607	88.77465652
xxx-003	7	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	90.133277538	90.080456622
xxx-003	8	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	91.631178694	89.844892346
xxx-003	9	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	96.098291908	97.112774989
xxx-003	10	89.4	88.6	89	89.4	90.1	91.4	91.8	93.1	97.797713682	96.212670122

Display 10. An Example of Body Weight Data with Imputed Values for a Monotone Missing Pattern

TRANSPOSE HORIZONTAL FORMAT TO VERTICAL FORMAT FOR ADAM COMPLIANCE

After two steps from PROC MI, the vertical format should be converted to a horizontal one to be compliant with ADaM BDS to further support both efficacy analysis per SAP and FDA submission.

Below SAS code shows this data step. AVISIT, AVISTN, ABLFL, and AWTARGET were generated in ADaM2 dataset.

```
data adam2;
     length avisit $80.;
     set mi_reg;
     aval=v1;avisit='Baseline';avisitn=0;ablfl='Y';awtarget=0;output;
     ablfl='';
     aval=v2;avisit='V3-Day 8';avisitn=8;awtarget=8;output;
     aval=v3;avisit='V4-Day 15';avisitn=15;awtarget=15;output;
     aval=v4;avisit='V6-Day 29';avisitn=29;awtarget=29;output;
     aval=v5;avisit='V8-Day 43';avisitn=43;awtarget=43;output;
     aval=v6;avisit='V9-Day 57';avisitn=57;awtarget=57;output;
     aval=v7;avisit='V11-Day 85';avisitn=85;awtarget=85;output;
     aval=v8;avisit='V13-Day 113';avisitn=113;awtarget=113;output;
     aval=v9;avisit='V15-Day 141';avisitn=141;awtarget=141;output;
     aval=v10;avisit='V17-Day 169';avisitn=169;awtarget=169;output;
run;
proc sort data=adam2;by usubjid avisitn avisit;run;
```

Display 11 below shows an example of a “standard” ADaM BDS dataset with the vertical format after multiple imputation.

USUBJID	_IMPUTATION_	TRTP	AVISITN	AVISIT	AVAL	ABLFL
xxx-002	1	Placebo	0	Baseline	82.7	Y
xxx-002	1	Placebo	8	V3-Day 8	82.9	
xxx-002	1	Placebo	15	V4-Day 15	83.1	
xxx-002	1	Placebo	29	V6-Day 29	83	
xxx-002	1	Placebo	43	V8-Day 43	84.1	
xxx-002	1	Placebo	57	V9-Day 57	81.323044901	
xxx-002	1	Placebo	85	V11-Day 85	83.5	
xxx-002	1	Placebo	113	V13-Day 113	84	
xxx-002	1	Placebo	141	V15-Day 141	82.122979957	

USUBJID	_IMPUTATION_	TRTP	AVISITN	AVISIT	AVAL	ABLFL
xxx-002	1	Placebo	169	V17-Day 169	84.047660998	
xxx-002	2	Placebo	0	Baseline	82.7	Y
xxx-002	2	Placebo	8	V3-Day 8	82.9	
xxx-002	2	Placebo	15	V4-Day 15	83.1	
xxx-002	2	Placebo	29	V6-Day 29	83	
xxx-002	2	Placebo	43	V8-Day 43	84.1	
xxx-002	2	Placebo	57	V9-Day 57	85.350664843	
xxx-002	2	Placebo	85	V11-Day 85	83.5	
xxx-002	2	Placebo	113	V13-Day 113	84	
xxx-002	2	Placebo	141	V15-Day 141	86.681516159	
xxx-002	2	Placebo	169	V17-Day 169	88.451147043	

Display 11. An Example of ADaM BDS Dataset for Body Weight with Vertical Format after Multiple Imputation

BRING BACK THE VALUES OF KEY VARIABLES POST PROC MI PROCEDURES

The variables PARAMCD PARAM TRTA TRTAN BASE ADT ATM ADTM ASEQ VISIT VISITNUM ADY APHASE APHASEN AWTDIFF were dropped from the dataset during PROC MI procedures. They were needed to be brought back to ADaM dataset for **CDISC ADaM compliance** after PROC MI procedures. To build ADaM traceability of these records with imputed values, the value of ADaM standard variable DTYPE is set to 'MI'.

```
data adam3;
     length dtype $10. avalc $40.;
     merge adam2(in=a) wt_avisit(in=b);
     by usubjid avisitn avisit;
     if a;
     if not b then do;imput=1; dtype='MI';end;
     avalc=strip(put(aval,best.));
     drop v1-v10;
run;
```

Display 12 below shows the values for **DTYPE** and **AVAL**, along with the missing values highlighted in yellow for the variables listed above.

USUBJID	_IMPUTATION_	ASEQ	TRTP	TRTA	ADT	VISIT
xxx-002	1	1002	Placebo	Placebo	2016-12-15	V2 Week 0
xxx-002	1	1003	Placebo	Placebo	2016-12-22	V3 Week 1
xxx-002	1	1004	Placebo	Placebo	2016-12-29	V4 Week 2
xxx-002	1	1005	Placebo	Placebo	2017-01-09	V6 Week 4
xxx-002	1	1006	Placebo	Placebo	2017-01-25	V8 Week 6
xxx-002	1	.	Placebo		.	
xxx-002	1	1008	Placebo	Placebo	2017-03-07	V11 Week 12
xxx-002	1	1009	Placebo	Placebo	2017-04-04	V13 Week 16
xxx-002	1	.	Placebo		.	
xxx-002	1	.	Placebo		.	
xxx-002	2	1002	Placebo	Placebo	2016-12-15	V2 Week 0
xxx-002	2	1003	Placebo	Placebo	2016-12-22	V3 Week 1
xxx-002	2	1004	Placebo	Placebo	2016-12-29	V4 Week 2
xxx-002	2	1005	Placebo	Placebo	2017-01-09	V6 Week 4
xxx-002	2	1006	Placebo	Placebo	2017-01-25	V8 Week 6

USUBJID	_IMPUTATION_	ASEQ	TRTP	TRTA	ADT	VISIT
xxx-002	2	.	Placebo		.	
xxx-002	2	1008	Placebo	Placebo	2017-03-07	V11 Week 12
xxx-002	2	1009	Placebo	Placebo	2017-04-04	V13 Week 16
xxx-002	2	.	Placebo		.	
xxx-002	2	.	Placebo		.	

USUBJID	_IMPUTATION_	APHASE	AVISITN	AVISIT	DTYPE	AVAL	BASE
xxx-002	1	PRE-TREATMENT	0	Baseline		82.7	82.7
xxx-002	1	TREATMENT	8	V3-Day 8		82.9	82.7
xxx-002	1	TREATMENT	15	V4-Day 15		83.1	82.7
xxx-002	1	TREATMENT	29	V6-Day 29		83	82.7
xxx-002	1	TREATMENT	43	V8-Day 43		84.1	82.7
xxx-002	1		57	V9-Day 57	MI	81.323044901	.
xxx-002	1	TREATMENT	85	V11-Day 85		83.5	82.7
xxx-002	1	TREATMENT	113	V13-Day 113		84	82.7
xxx-002	1		141	V15-Day 141	MI	82.122979957	.
xxx-002	1		169	V17-Day 169	MI	84.047660998	.
xxx-002	2	PRE-TREATMENT	0	Baseline		82.7	82.7
xxx-002	2	TREATMENT	8	V3-Day 8		82.9	82.7
xxx-002	2	TREATMENT	15	V4-Day 15		83.1	82.7
xxx-002	2	TREATMENT	29	V6-Day 29		83	82.7
xxx-002	2	TREATMENT	43	V8-Day 43		84.1	82.7
xxx-002	2		57	V9-Day 57	MI	85.350664843	.
xxx-002	2	TREATMENT	85	V11-Day 85		83.5	82.7
xxx-002	2	TREATMENT	113	V13-Day 113		84	82.7
xxx-002	2		141	V15-Day 141	MI	86.681516159	.
xxx-002	2		169	V17-Day 169	MI	88.451147043	.

Display 12. An Example of ADaM BDS Compliant Dataset for Body Weight after Bringing Back Key Variables Post PROC MI Procedures

IMPUTE THE VALUES WITH DTYPE='MI' BY LOCF FOR ADAM COMPLIANCE

For the records with DTYPE='MI', i.e., with imputed values of AVAL , the values for **ADT ATM ADTM ADY BASE ASEQ APHASEN APHASE TRTAN TRTA** were never collected during the study, even though the values of AVAL from these visits had non-missing values, which were imputed by PROC MI procedures. However their values were still missing.

To build a CDISC compliant ADaM dataset and to further support efficacy analysis, and/or FDA submission, the values of these variables listed above should also be "imputed". Otherwise, Pinnacle 21 would report error messages if the current ADaM dataset was uploaded to it for compliance checking. To avoid these error messages and further ease ADRG (Analysis Data Reviewer's Guide) writing, the best approach or solution is to "impute" these "missing" values.

The approach we used is LOCF (last observation carried forward) by following traditional LOCF method for imputing missing values of post baseline efficacy endpoint(s).

Below SAS code shows this data step for LOCF for the missing values of ADT ATM ADTM ADY BASE ASEQ APHASEN APHASE TRTAN TRTA.

```
proc sort data=adam3;by _imputation_ paramcd agegr1 agegr1n racegr1
racegr1n trtpn trtp usubjid avisitn avisit dtype;run;
data adam4;
      length _trta aphase $40.;
      retain cnt _adt _atm _adtm _ady _base _aseq _aphasen _trtan _trta;
      set adam3;
      by _imputation_ paramcd agegr1 agegr1n racegr1 racegr1n trtpn trtp
         usubjid avisitn avisit dtype;
      if first.usubjid then cnt=0;
      if adt>.Z then do;skip=1;cnt=0;end;
      else skip=0;
      if adt>.Z then do;
         _adt=adt;_atm=atm;_adtm=adtm;_ady=ady;_base=base;_aseq=aseq;
         _aphasen=aphasen;_trtan=trtan;_trta=trta;
      end;
      if adt=. and not first.usubjid and imput=1 then do;
         cnt=cnt+0.1;
         adt=_adt;atm=_atm;adtm=_adtm;ady=_ady;base=_base;aseq=_aseq;
         aphasen=_aphasen;trtan=_trtan;trta=_trta;aseq=_aseq+cnt;
      end;
      if adt=. then adt=_adt;
      if atm=. then atm=_atm;
      if adtm=. then adtm=_adtm;
      if aseq=. then aseq=_aseq;
      if aphasen=. then aphasen=_aphasen;

      if trta='' then trta=_trta;
      if trtan=. then trtan=_trtan;
      if aphasen=1 then aphase='TREATMENT';
      else  if aphasen=2 then aphase='FOLLOW-UP';
      if aphase='' and avisitn>0 then do;
            aphase='TREATMENT';aphasen=1;
      end;
      format _adt  yymmdd10. _adtm datetime20.;
      drop _adt _atm _adtm _ady _base _aseq _aphasen _trtan _trta;
run;
```

Display 13 below shows the imputed values of ADT ATM ADTM ADY BASE ASEQ APHASEN APHASE TRTAN TRTA, which are highlighted in yellow.

USUBJID	_IMPUTATION_	ASEQ	TRTP	TRTA	ADT	VISIT
xxx-002	1	1002	Placebo	Placebo	2016-12-15	V2 Week 0
xxx-002	1	1003	Placebo	Placebo	2016-12-22	V3 Week 1
xxx-002	1	1004	Placebo	Placebo	2016-12-29	V4 Week 2
xxx-002	1	1005	Placebo	Placebo	2017-01-09	V6 Week 4
xxx-002	1	1006	Placebo	Placebo	2017-01-25	V8 Week 6
xxx-002	1	1006.1	Placebo	Placebo	2017-01-25	
xxx-002	1	1008	Placebo	Placebo	2017-03-07	V11 Week 12
xxx-002	1	1009	Placebo	Placebo	2017-04-04	V13 Week 16
xxx-002	1	1009.1	Placebo	Placebo	2017-04-04	
xxx-002	1	1009.2	Placebo	Placebo	2017-04-04	
xxx-002	2	1002	Placebo	Placebo	2016-12-15	V2 Week 0
xxx-002	2	1003	Placebo	Placebo	2016-12-22	V3 Week 1
xxx-002	2	1004	Placebo	Placebo	2016-12-29	V4 Week 2

USUBJID	_IMPUTATION_	ASEQ	TRTP	TRTA	ADT	VISIT
xxx-002	2	1005	Placebo	Placebo	2017-01-09	V6 Week 4
xxx-002	2	1006	Placebo	Placebo	2017-01-25	V8 Week 6
xxx-002	2	1006.1	Placebo	Placebo	2017-01-25	
xxx-002	2	1008	Placebo	Placebo	2017-03-07	V11 Week 12
xxx-002	2	1009	Placebo	Placebo	2017-04-04	V13 Week 16
xxx-002	2	1009.1	Placebo	Placebo	2017-04-04	
xxx-002	2	1009.2	Placebo	Placebo	2017-04-04	

USUBJID	_IMPUTATION_	APHASE	AVISITN	AVISIT	DTYPE	AVAL	BASE
xxx-002	1	PRE-TREATMENT	0	Baseline		82.7	82.7
xxx-002	1	TREATMENT	8	V3-Day 8		82.9	82.7
xxx-002	1	TREATMENT	15	V4-Day 15		83.1	82.7
xxx-002	1	TREATMENT	29	V6-Day 29		83	82.7
xxx-002	1	TREATMENT	43	V8-Day 43		84.1	82.7
xxx-002	1	TREATMENT	57	V9-Day 57	MI	81.323044901	82.7
xxx-002	1	TREATMENT	85	V11-Day 85		83.5	82.7
xxx-002	1	TREATMENT	113	V13-Day 113		84	82.7
xxx-002	1	TREATMENT	141	V15-Day 141	MI	82.122979957	82.7
xxx-002	1	TREATMENT	169	V17-Day 169	MI	84.047660998	82.7
xxx-002	2	PRE-TREATMENT	0	Baseline		82.7	82.7
xxx-002	2	TREATMENT	8	V3-Day 8		82.9	82.7
xxx-002	2	TREATMENT	15	V4-Day 15		83.1	82.7
xxx-002	2	TREATMENT	29	V6-Day 29		83	82.7
xxx-002	2	TREATMENT	43	V8-Day 43		84.1	82.7
xxx-002	2	TREATMENT	57	V9-Day 57	MI	85.350664843	82.7
xxx-002	2	TREATMENT	85	V11-Day 85		83.5	82.7
xxx-002	2	TREATMENT	113	V13-Day 113		84	82.7
xxx-002	2	TREATMENT	141	V15-Day 141	MI	86.681516159	82.7
xxx-002	2	TREATMENT	169	V17-Day 169	MI	88.451147043	82.7

Display 13. An Example of Body Weight Data with Imputed Values for ADT ASEQ TRTA, ect., for Records with DTYPE='MI'.

The ADaM Variable ASEQ was derived for the records with DTYPE='MI' for ADaM traceability. Its specification is shown below.

Variable Name	Variable Label	Type	Length/ Display Format	Controlled Terms or Format	Source/Derivation/Comment	Core
ASEQ	Analysis Sequence Number	float		8.4	Derived: For the records from ADWT, ASEQ=ADWT.ASEQ; For records with DTYPE='MI', ASEQ=ASEQ+(# of imputation Iteration)*0.1, where ASEQ is from the last non-missing record of ADWT.ASEQ per subject.	Perm

DERIVE OTHER VARIABLES TO FURTHER SUPPORT ANALYSIS OF EFFICACY

The variables CHG PCHG CRIT1FL CRIT1, etc., should be derived to further support efficacy analysis.

```
data adam5;
     length crit1fl $1. crit1 $50.;
     set adam4;
     imputnm=_imputation_;
     if aphasen>0 and aval>.Z and base>.Z then do;
        chg=aval-base;
        if base>0 then pchg=100*(chg/base);
     end;
     if pchg>=5 then do; crit1fl='Y';crit1='>=5% Weight Gain'; end;
run;
```

Display 14 below shows the derived values of these variables: CHG PCHG CRIT1FL CRIT1, ect., for the records with DTYPE='MI'.

USUBJID	_IMPUTATION_	AVISIT	DTYPE	AVAL	BASE	CHG	PCHG	CRIT1FL	CRIT1
xxx-002	1	Baseline		82.7	82.7	.	.		
xxx-002	1	V3-Day 8		82.9	82.7	0.2	0.2418379686		
xxx-002	1	V4-Day 15		83.1	82.7	0.4	0.4836759371		
xxx-002	1	V6-Day 29		83	82.7	0.3	0.3627569528		
xxx-002	1	V8-Day 43		84.1	82.7	1.4	1.6928657799		
xxx-002	1	V9-Day 57	MI	81.323044901	82.7	-1.376955099	-1.66500012		
xxx-002	1	V11-Day 85		83.5	82.7	0.8	0.9673518742		
xxx-002	1	V13-Day 113		84	82.7	1.3	1.5719467956		
xxx-002	1	V15-Day 141	MI	82.122979957	82.7	-0.577020043	-0.697726775		
xxx-002	1	V17-Day 169	MI	84.047660998	82.7	1.3476609976	1.6295779899		
xxx-002	2	Baseline		82.7	82.7	.	.		
xxx-002	2	V3-Day 8		82.9	82.7	0.2	0.2418379686		
xxx-002	2	V4-Day 15		83.1	82.7	0.4	0.4836759371		
xxx-002	2	V6-Day 29		83	82.7	0.3	0.3627569528		
xxx-002	2	V8-Day 43		84.1	82.7	1.4	1.6928657799		
xxx-002	2	V9-Day 57	MI	85.350664843	82.7	2.6506648433	3.2051570052		
xxx-002	2	V11-Day 85		83.5	82.7	0.8	0.9673518742		
xxx-002	2	V13-Day 113		84	82.7	1.3	1.5719467956		
xxx-002	2	V15-Day 141	MI	86.681516159	82.7	3.9815161588	4.8144088982		
xxx-002	2	V17-Day 169	MI	88.451147043	82.7	5.7511470427	6.9542285886	Y	>=5% Weigh t Gain

Display 14. An Example of Body Weight Data with Derived Variables: CHG PCHG CRIT1FL CRIT1, ect., for Records with DTYPE='MI'.

Note: _IMPUTATION_ in the second column above will be renamed as IMPUTNM in final ADMIWT.

AN EXAMPLE OF USE OF PROC MIANALYZE FOR EFFICACY ANALYSIS

For the hypothetical example, the primary analyses will be carried out using a logistic regression model based on MI for missing data. The logistic regression model will include the treatment group (ADWT.TRTPN), race (White and Non-white (ADWT.RACEGR1N)) and baseline age (<40, ≥40 years (ADWT.AGEGR1N)) as factors, and the baseline weight as the covariate.

The logistic regression model approach is demonstrated by the following SAS codes.

```
**** logistic regression ****;
ods output diffs=logistic;
proc sort data=admiwt; by avisitn imputnm; run;
proc genmod data=admiwt descending;
     by avisitn imputnm;
     class trtp (ref='Control') agegr1 racegr1/param=glm;
     model myaval=trtp agegr1 trtp*agegr1 racegr1 base/link=logit dist=bin;
     lsmeans trtp*agegr1/diff cl exp;
     where avisitn>15;
run;
proc sort data=logistic out=logistic2;by agegr1 avisitn;
     where (trtp ne _trtp) and (agegr1=_agegr1);
run;
*** combine results ***;
proc mianalyze data=logistic2;
     by agegr1 avisitn;
     modeleffects estimate;
     stderr stderr;
     ods output ParameterEstimates=p_diff;
run;
ods pdf close;
ods listing;
```

AVISITN	IMPUTNM	StmtNo	Effect	TRTP	_TRTP	Estimate	StdErr
8	1	1	TRTP	ABC	Placebo	0.5625	0.8913
8	2	1	TRTP	ABC	Placebo	0.5625	0.8913
8	3	1	TRTP	ABC	Placebo	0.5625	0.8913
8	4	1	TRTP	ABC	Placebo	0.5625	0.8913
8	5	1	TRTP	ABC	Placebo	0.5625	0.8913
8	6	1	TRTP	ABC	Placebo	0.5625	0.8913
8	7	1	TRTP	ABC	Placebo	0.5625	0.8913
8	8	1	TRTP	ABC	Placebo	0.5625	0.8913
8	9	1	TRTP	ABC	Placebo	0.5625	0.8913
8	10	1	TRTP	ABC	Placebo	0.5625	0.8913
15	1	1	TRTP	ABC	Placebo	-0.4127	0.3876
15	2	1	TRTP	ABC	Placebo	-0.2529	0.3727
15	3	1	TRTP	ABC	Placebo	-0.3188	0.3801
15	4	1	TRTP	ABC	Placebo	-0.3289	0.3791
15	5	1	TRTP	ABC	Placebo	-0.3438	0.3920
15	6	1	TRTP	ABC	Placebo	-0.2371	0.3736
15	7	1	TRTP	ABC	Placebo	-0.2351	0.3739
15	8	1	TRTP	ABC	Placebo	-0.5194	0.3814
15	9	1	TRTP	ABC	Placebo	-0.3232	0.3792
15	10	1	TRTP	ABC	Placebo	-0.4038	0.3882

AVISITN	zValue	Probz	Alpha	Lower	Upper	ExpEstimate	LowerExp	UpperExp
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
8	0.63	0.5280	0.05	-1.1844	2.3095	1.7551	0.3059	10.0692
15	-1.06	0.2871	0.05	-1.1724	0.3471	0.6619	0.3096	1.4149
15	-0.68	0.4974	0.05	-0.9835	0.4776	0.7765	0.3740	1.6123
15	-0.84	0.4015	0.05	-1.0638	0.4261	0.7270	0.3452	1.5313
15	-0.87	0.3856	0.05	-1.0719	0.4141	0.7197	0.3424	1.5130

AVISITN	zValue	Probz	Alpha	Lower	Upper	ExpEstimate	LowerExp	UpperExp
15	-0.88	0.3805	0.05	-1.1121	0.4245	0.7091	0.3289	1.5289
15	-0.63	0.5256	0.05	-0.9693	0.4951	0.7889	0.3793	1.6406
15	-0.63	0.5295	0.05	-0.9680	0.4978	0.7905	0.3798	1.6450
15	-1.36	0.1733	0.05	-1.2670	0.2282	0.5949	0.2817	1.2564
15	-0.85	0.3941	0.05	-1.0665	0.4201	0.7238	0.3442	1.5221
15	-1.04	0.2983	0.05	-1.1647	0.3571	0.6678	0.3120	1.4292

Output 1. Output from Logistic Regression Model by Imputation and Visits

Finally, **PROC MIANALYZE** combines the results, showed in Output 1 above. Output 2 shows the combined results from Output 1 by visits.

AVISITN	NImpute	Parm	Estimate	StdErr	LCLMean	UCLMean
8	10	estimate	0.562536	0.891313		
15	10	estimate	-0.337573	0.392154	-1.1065	0.4314
29	10	estimate	-0.161591	0.266711	-0.6848	0.3616
43	10	estimate	-0.160994	0.236744	-0.6265	0.3045
57	10	estimate	-0.011139	0.225571	-0.4551	0.4328
85	10	estimate	0.002623	0.206747	-0.4032	0.4084
113	10	estimate	0.047133	0.201520	-0.3490	0.4433
141	10	estimate	0.126438	0.207483	-0.2824	0.5353
169	10	estimate	0.125184	0.205877	-0.2806	0.5309

AVISITN	NImpute	DF	Min	Max	Theta0	tValue	Probt
8	10		0.562536	0.562536	0		
15	10	2786.6	-0.519371	-0.235114	0	-0.86	0.3894
29	10	1484.8	-0.256786	-0.055123	0	-0.61	0.5447
43	10	385.9	-0.297272	-0.025874	0	-0.68	0.4969
57	10	289.66	-0.154258	0.105357	0	-0.05	0.9607
85	10	837.33	-0.072831	0.139358	0	0.01	0.9899
113	10	411.37	-0.057593	0.183819	0	0.23	0.8152
141	10	227.89	0.025043	0.275687	0	0.61	0.5429
169	10	219.64	-0.021330	0.232391	0	0.61	0.5438

Output 2. Combined Results from Output 1 by Visits

ADAM SPECIFICATION WRITING

The ADaM specification writing is very critical for the study Biostatistician to review and approve it, in addition to the programming validation. It is also critical for FDA reviewers to review ADaM and TFLs if the FDA submission occurs down the road.

This newly derived ADaM dataset from ADWT is called **ADMIWT**, and is labeled as "***Body Wt. Anal. Dataset with MI***". The specification of some key variables from it, IMPUTNM ASEQ AVAL ADT DTYPE, along with dataset information is provided as an example in Appendix I for your reference when you are working on the specification for your study.

CONCLUSION

This paper introduces a tutorial for building a CDISC compliant ADaM dataset to support multiple imputation analysis for clinical trials by an example. Some SAS sample codes is provided for your reference. We hope that these tips and SAS codes, as well as the specification of some key variables can make your life a little easier when you are working on ADaM programming for MI analysis for a clinical study report and/or FDA submission.

REFERENCES

[1] Yang C. Yuan, SAS Institute Inc., Rockville, MD, "Multiple Imputation for Missing Data: Concepts and New Development (Version 9.0)". Available at https://support.sas.com/rnd/app/stat/papers/multipleimputation.pdf.

[2] Chris Smith and Scott Kosten, “Multiple Imputation: A Statistical Programming Story”, Proceedings of the Pharmaceutical SAS® Users Group Conference, PharmaSUG 2017

[3]. CDISC Analysis Data Model Team. “Analysis Data Model (ADaM) Implementation Guide Version 1.1”. February 2016.
https://www.cdisc.org/system/files/members/standard/foundational/adam/ADaMIG_v1.1.pdf

[4] Little RJ, D’Agostino R, Cohen M, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med, 367(14): 1355-1360.

[5] O’Neill RT and Temple R (2012). The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther, 91: 550-554.

ACKNOWLEDGMENTS

Appreciation goes to Min Chen for her valuable review and comments.

CONTACT INFORMATION

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

Name: Xiangchen (Bob) Cui, Ph.D.
Enterprise: Alkermes, Inc.
Address: 852 Winter Street
City, State ZIP: Waltham, MA 02451
Work Phone: 781-609-6038
Fax: 781-609-5855
E-mail:xiangchen.cui@alkermes.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.

APPENDIX I

ADMIWT: Body Weight Analysis Dataset with MI

Dataset Name	ADMIWT
Dataset Description	Body Wt. Anal. Dataset with MI
Class of Dataset	BASIC DATA STRUCTURE (BDS)
Dataset Structure	One record per subject per parameter per analysis time point per imputation number
Key Variables of Dataset	STUDYID, IMPUTNM, USUBJID, PARAM, ADT
Input Datasets	ADSL, ADWT
Documentation	Includes records from FASFL subjects with the following conditions: ADWT.DTYPE=‘’ and ADWT.PARAMCD=’DMWEIGHT’ and ADWT.ANL01FL=’Y and ADSL.FASFL=’Y’
Notes	This document is for Study xxx

Variable Name	Variable Label	Type	Length/ Display Format	Controlled Terms or Format	Source/Derivation/Comment	Core
IMPUTNM	Imputation Number	integer	8		Derived: It was derived from PROC MI procedure from the setting: nimpute=100 and seed=1104078. The maximum for this variable is 100.	Req
ASEQ	Analysis Sequence Number	float		8.4	Derived: For the records from ADWT, ASEQ=ADWT.ASEQ; For records with DTYPE=’MI’, ASEQ=ASEQ+(# of imputation Iteration)*0.1, where ASEQ is from the last non-missing record of ADWT.ASEQ per subject.	Perm
AVAL	Analysis Value	float		8.1	Derived: Equal to ADWT.AVAL for the records from ADWT with non-	Req

					missing values; For records with DTYPE='MI', it was derived by multiple imputation through SAS PROC MI procedure by two steps. First, sort by **USUBJID TRTP TRTPN RACEGR1 RAEGR1N AGEGR1 AGEGR1N,** then transpose the data into "horizontal" format with column variable names: BASE v3 v4 v6 v8 v9 v11 v13 v15 v17, where V[i] stands for transposed values of AVAL for visit i=3, 4, 5, 8, 9,11,13,15, and 17, and rename them as V1 to V10, respectively. The dataset is named as DATAIN. Step 1: handle a dataset with an arbitrary missing pattern (having missing data in between visits): **PROC MI DATA=DATAIN OUT=MI_MONO NIMPUTE=500 SEED=1104078;** **VAR TRTPN RACEGR1N AGEGR1N V1-V10;** **MCMC CHAIN=SINGLE IMPUTE=MONOTONE PRIOR=JEFFREYS NBITER=200 NITER=100 INITIAL=EM;** **RUN;** Step 2: handle a dataset with a monotone missing pattern (the data is missing at the current visit, and the data at all following visits are missing.): **PROC MI DATA=MI_MONO OUT=MI_REG NIMPUTE=1 SEED=1104078;** **BY _IMPUTATION_;** **VAR TRTPN RACEGR1N AGEGR1N V1-V10;** **CLASS TRTPN RACEGR1N AGEGR1N;** **MONOTONE REGRESSION(V10/DETAILS);** **RUN;** Note: The dataset **MI_REG** is transposed back to vertical format to fit ADaM BDS.	
ADT	Analysis Date	integer	8	YYMMDD10.	Derived: ADWT.ADT for the records from ADWT; For records with DTYPE='MI', ADT was carried forward from the	Perm

					last non-missing post-baseline record of ADWT.ADT per subject.	
DTYPE	Derivation Type	text	10	(1) MI	Derived: DTYPE='MI' for the records which were missing and were imputed by PROC MI procedure.	Perm