PharmaSUG 2019 - Paper ST-103

SAS® V9.4 MNAR statement for multiple imputations for missing not at random in longitudinal clinical trials

Lingling Li

ABSTRACT

Missing data is a common problem in longitudinal clinical trials. The primary analysis used in clinical trials usually rely on the untestable assumption of missing at random (MAR) and the sensitivity analysis under plausible assumption of missing not at random (MNAR) is needed to examine the robustness of the statistical inference obtained from the primary analysis. Multiple imputations with pattern-mixture models (PMMs) is widely used for implementing sensitivity analysis under the assumption of MNAR. Two MNAR assumptions, control-based pattern imputation and delta-adjusted pattern imputation are regarded as clinically plausible, transparent, and easy to implement. SAS® Version 9.4 PROC MI provides a MNAR statement, with two options MODEL and ADJUST, that allows implementation of the two assumptions conveniently. MNAR statement is used along with MONOTONE statement to handle monotone missing data and FCS statement to handle arbitrary missing data. This paper discusses the implementation of sensitivity analysis under MNAR assumptions within PMMs framework in longitudinal clinical trial data using MNAR statement. A simulated longitudinal clinical trial data with sample SAS code is described for illustration.

INTRODUCTION

In longitudinal clinical trials, missing data is a common problem that can potentially result in loss of statistical power and biased estimation. The missing data is present in two types of missing patterns: monotone missing and non-monotone missing. For monotone missing pattern, if an outcome is not observed in a particular visit for a subject, the outcome will not be observed in all subsequent visits for that subject. Monotone missing is caused by patients dropping out of the study before completion and is the major source of missingness in longitudinal clinical trial data. Patients drop-out is commonly seen in clinical trials for a variety of the reasons, for example, adverse events, lack of efficacy, death, relocation, or sponsor decision. Non-monotone missing pattern is also called intermittent missing, in which an unobserved outcome can be followed by observed outcomes. Non-monotone missing is caused by reasons such as patients skipping visits but returning for the following visits, or occasional measurement failure. In psychiatric clinical trials where the endpoint is total score summed over several individual item scores in a questionnaire, patients might forget or choose not to answer some of the individual items thus leading to total score missing. Another source of non-monotone missing is due to data manipulation for the preparation of analysis. If we analyze the data using analysis visits by windowing the study day (usually defined as the distance between the assessment date and the first dose date) instead of using actual nominal visits, a value at a visit might be windowed out, thus causing the value at that visit missing. Compared to monotone missing, the amount of non-monotone missing is small. Data sets that have both monotone and non-monotone missing patterns are said to have an arbitrary missing data pattern.

Missing data mechanism, which expresses the process causing the missing data, is classified to three categories (Little and Rubin, 2002): missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR happens when the probability of outcome being missing is simply due to chance and is not related to observed or unobserved data. The missing data with MCAR can be considered as a random sample of the complete data and the statistical methods used on complete data can be used on data with MCAR without compromising statistical estimation and inference. However, MCAR is often an unrealistically strong assumption in practice. In the case of MAR, the probability of outcome being missing depends on the observed data rather than the missing outcome itself. MAR is often considered as ignorable missing on the condition that the parameters describing the missing data mechanism and the parameters to be estimated are distinct - a condition that is unlikely to be violated in practice. The implication of MAR being ignorable is that the missing data mechanism can be ignored and does not have to be modelled in maximal likelihood-based analysis. MAR is untestable and might not be plausible. MNAR happens when the probability of outcome being missing depends on at

least one unobserved outcome. A number of scenarios in clinical trials can lead to MNAR. For example, if a patient discontinued study because of an unrecorded worsening outcome, the resulting missing data is related to at least one unobserved outcome therefore is MNAR. Many standard statistical methods for analyzing the longitudinal clinical trial data require MAR. If a primary analysis, for example mixed models repeated measures (MMRM) using PROC MIXED, relies on the assumption of MAR, sensitivity analysis that consider various MNAR scenarios is needed to test the robustness of the statistical inference against departure from the MAR assumption. If the inference is maintained under reasonable MNAR assumptions then the results are considered to be robust; otherwise, the results are questionable.

Two classic modeling families have been proposed to implement MNAR assumptions: pattern-mixture models (PMMs) and selection models (Little and Rubin, 2002; National Research Council, 2010). In randomized clinical trials it is important that the MNAR assumptions are transparent to researchers and clinicians. PMMs express the joint probability distribution of outcomes and missingness as the product of probability distribution of missing data patterns and the conditional distribution of responses given each missing data pattern. Within each missing data pattern, the unknown distribution of missing values could be explicitly specified by imposing assumptions on the missing data, therefore PMMs gain the advantages that the assumptions are more transparent and accessible to clinicians and more suitable for imputations. Two common strategies to formulate MNAR assumptions which are regarded as clinical plausible are control-based pattern imputation (Little and Yau, 1996; O'Kelly and Ratitch, 2014; Ratitch et al., 2013) and delta-adjusted pattern imputation (O'Kelly and Ratitch, 2014; Ratitch et al., 2013). Both approaches are more conservative with the intention to reduce the difference between experimental and control groups. SAS® Version 9.4 PROC MI added a MNAR statement that has two options: MODEL and ADJUST, which can be used to directly implement control-based pattern imputation and delta-adjusted pattern imputation respectively. MNAR statement works with MONOTONE statement to handle monotone missing patterns and FCS statement to handle arbitrary missing patterns. This paper presents the details for implementation of sensitivity analysis under MNAR assumptions within the PMMs framework in longitudinal clinical trial data using MNAR statement. A simulated longitudinal clinical trial data with sample SAS code is described for illustration.

PATTERN-MIXTURE MODELS

PMMs express the joint probability distribution of outcome and missing patterns, given the explanatory variables, as the product of probability distribution of missing data patterns and the conditional distribution of outcome given each missing data pattern as shown in the below formula (National Research Council, 2010; Ratitch et al., 2013), where Y denotes the outcome vector with Y_{obs} representing observed outcomes and Y_{mis} representing missing outcomes; R is the missing data indicator which takes the value 0 if Y is observed and 1 if Y is missing; X denotes the treatment groups and baseline covariates such as age and sex which are collected before randomization and are fully observed and conditioned on in the model fitting. The formula is as follows:

$$[Y_{obs}, Y_{mis}, R|X] = [Y_{obs}, Y_{mis}|R, X] \times [R|X]$$

The above formula can be further factored as below:

$$[Y_{obs}, Y_{mis}, R|X] = [Y_{mis}|Y_{obs}, R, X] \times [Y_{obs}|R, X] \times [R|X]$$

$[Y_{obs}|R, X]$ represents probability distribution of the observed data and $[Y_{mis}|Y_{obs}, R, X]$ represents the unknown distribution of missing values given the observed data; both are conditioned on the covariate X and missing data patterns represented by R. PMMs are said to be under-identified due to the unknown distribution of missing values. One way to handle under-identified of PMMs are imposing certain assumption to the missing data by specifying a relationship to the observed data, such as the difference in mean. Multiple imputations can be carried out to fill in the missing values with multiple sets of plausible values based on the observed data. The imputed values can be further adjusted for MNAR assumptions.

MULTIPLE IMPUTATIONS UNDER MAR FOR MONOTONE MISSING DATA

Monotone missing patterns caused by patients' drop-outs prematurely are the majority of the missingness in the clinical trial data. Under MAR, the probability distribution of the missing values does not depend on the missing patterns. Suppose in a clinical trial, T visits are scheduled. Let 0 denote the baseline visit and

L denotes the last visit. The missing pattern is defined by the last visit L which takes the value 0 to T. At a given visit j,

$$[Y_{j+1}|Y_0, \dots Y_j, X, L = j] = [Y_{j+1}|Y_0, \dots Y_j, X, L > j]$$

Thus, the missing Y_{j+1} in pattern j can be imputed using the observed Y_{j+1} in patterns >= $j + 1$. Take an example of linear regression method to impute continuous outcomes in PROC MI, first a statistical model is fitted with observed Y_{j+1} as response variable, and the covariates and the preceding outcomes as explanatory variables, yielding the posterior distribution of the parameters from which the new parameters are randomly drawn. The new parameters and the corresponding predictors are used to generate the predicted outcome values, which is then used to produce the imputed values by adding a simulated error term. Next, observed Y_{j+2} from patterns >= $j + 2$ is used to impute the missing Y_{j+2}. Continuing fitting a series of parametric models, missing Y_{j+1} and beyond can be imputed sequentially in the chronologic order. NRC (National Research Council, 2010) recommend collecting auxiliary information which might not be the interest of inference to evaluate treatment effect in the final analysis model but are believed to be associated with the reason of drop-outs or the values of missing outcome when substantial missing values are anticipated. An advantage of MI is that the imputation model allows including these auxiliary variables as explanatory variables to improve the accuracy of the imputation, in addition to the explanatory variables used in the final analysis model.

Sequential imputation is readily implemented with MONOTONE statement and can be used for a flexible type of responses, for example, the regression and predictive mean matching methods for continuous responses, the logistic regression method for binary or ordinal responses and the discriminant function method for binary or nominal responses. Also, since univariate model is fitted sequentially with one outcome variable taken at a time, distinct methods can be used for different variables within one imputation model.

MULTIPLE IMPUTATIONS FOR MNAR ASSUMPTIONS

For MNAR, the probability of outcome being missing after conditioning on the observed data depends on the unobserved outcomes that should have been obtained. Because missing values are not observed and cannot be verified, assumptions are needed to impute the missing values and there are potentially countless MNAR assumptions. Two MNAR assumptions within the PMMs framework, control-based pattern imputation and delta-adjustment, gained wide acceptance and usage due to being clinically plausible, transparent and easy to implement.

The control-based pattern imputation was initially proposed by Little R et al. (Little and Yau, 1996) based on the idea of “as treated” model, which imputes missing values based on the actual treatment and dosage patients receive after drop-out and one of the assumptions explored are patients revert to control group after drop-out. Thus, it is assumed that after drop-out, the unobserved values in experimental group follow the path of observed values in control group. Based on this assumption, only the observed values in control group are used to derive the posterior distribution of the parameters from which the missing values in both control and experimental group are imputed. This approach is conservative as it tends to reduce the difference between experimental group and control group, but not extremely conservative since it still allows the carry over treatment effect by using the prior observed values in experimental group as predictors.

The option MODEL within MNAR statement allows specification of the subset of observations that are used to model the distribution of missing values. Control-based pattern imputation is implemented by specifying control group as the subset of observations.

The delta-adjusted pattern imputation is based on an assumption known as “nonfuture dependence” (Kenward et al., 2003, National Research Council, 2010). Let L denote the last visit that has outcome observed, the assumption states the missingness at $L + 1$ depend on the observed values up to $L + 1$ and the possible missing values at $L + 1$, but not the future missing values after $L + 1$. Thus, the distribution of missing Y_{L+1} can be specified based on certain relationship to that of observed Y_{L+1} after controlling for the observed history prior to $L + 1$ and covariates. One way to do this is assuming the mean of the missing outcome in experimental group deviate a certain amount to their observed counterparts in the direction of worsening the outcome but the mean in control group is the same between missing values and observed

values (Ratitch et al., 2013). Thus, the missing values within each treatment group can be imputed using the observed values in their respective treatment groups, and after imputation, the imputed values in experimental group are added or subtracted a certain amount in the direction of worsening the outcomes but the imputed values in the control group remain unchanged. Three variants to this approach have been proposed. The first variant is that the delta-adjusted pattern imputation is only applied in the first visit that has outcome missing, and the adjusted values are used as predictor to impute the subsequent missing values which will not be adjusted. In this approach, the worsening effect is propagated to the imputed outcomes at later visits by correlation. The second variant is that the adjustment is applied at each visit and the adjusted values are used as predictors to impute the subsequent missing values. The third variant is that the adjustment is applied to each visit after all missing values are imputed under MAR assumption. The delta-adjusted pattern imputation can be used as a stand-alone analysis in which a single offset is specified or tipping-point analysis in which multiple offsets within a plausible range are applied. When a range of offsets are applied, the offset that turns the conclusion from favoring experimental group to favoring control group is called “tipping point”. If the underlying assumption of the tipping point is clinically plausible, then the conclusion from MAR is questionable.

MNAR statement, when used together with the option ADJUST, allows the imputed variables to be adjusted for a subset of observations, and if experimental group is chosen as the subset of observations, the imputed variables in experimental group will be deviated at a pre-specified amount, which allows the implementation of delta-adjusted pattern imputation.

MNAR statement works with MONOTONE statement to handle monotone missing patterns and FCS statement to handle arbitrary missing patterns. Missing patterns in longitudinal clinical data are arbitrary with small amount of intermittent missingness and majority of monotone missingness. The most common strategy is using Markov Chain Monte Carlo (MCMC) to fill in the small amount of intermittent missing values under the assumption of MAR, resulting in an intermediate data with monotone missing patterns. After the data is turned into monotone missing patterns, the MONOTONE statement can be applied and MNAR statement can be used together to implement the MNAR assumptions, with option MODEL for control-based pattern imputation and option ADJUST for delta-adjusted pattern imputation.

There are several limitations to this approach. Conditioning on the missing patterns, the intermittent missingness is usually assumed to be MAR. The rationale is since the subjects stay on the treatment, the distribution of missing values should follow the observed values. However, treating the intermittent missing as MAR lack theoretical support and the appropriateness of the assumption needs to be further examined. Even so, some departure from MAR should not jeopardize the validity of the results because of the small amount of intermittent missing. In addition, due to the multivariate normal assumption, MCMC is more suitable for imputing continuous outcomes. However, MCMC is robust to small departure from the multivariate normal assumption when the missing information is not large. Categorical responses to be imputed and treated as categorical variables should be coded with numerical values; if a categorical variable need to be included as predictor in the imputation model, a common practice is to code the categorical variable to several dummy variables and use the dummy variables as predictors (O’Kelly and Ratitch, 2014).

Another strategy is using MNAR statement together with fully conditional specification (FCS) which can handle both monotone and non-monotone missing patterns. Similar with monotone methods for monotone missing patterns, FCS specifies a series of univariate models that impute variables sequentially, therefore allowing different types of responses and distinct model for each imputed variable within an imputation model. FCS involves two phases referred to as the filled-in phase and the imputation phase. In the filled-in phase, the missing values are initially filled in by randomly drawing from the conditional distribution of the observed outcome given the preceding variables. In the imputation phase, the missing values are drawn from the posterior distribution of the observed outcome given the remaining variables in the VAR statement. The imputation phase is an iterative process, and the imputed values are used after a number of iterations to achieve stationary distribution. The variables are imputed sequentially in the order specified in the VAR statement. The observations used to derive posterior distribution contain the observed response variable, and the covariates can be observed or filled in (filled-in phase) or imputed (imputation phase).

The theoretical basis of FCS needs more validation although results are generally regarded as robust and reliable. In addition, when used with MNAR statements, the intermittent missing values will be imputed under the assumption of MNAR together with the monotone missing values.

THE NUMBER OF IMPUTATIONS AND OTHER OPTIONS

To account for the uncertainty of the missing values, MI is repeated multiple times which generate multiple versions of the complete data, with the observed values remain unchanged in each version and the imputed values different due to random drawn. The number of imputations has been recommended varying from 5 to 100 based on relative efficiency, power, standard error and p value for estimated coefficient (see review in SAS/STAT® 14.1 User's Guide, 2015). One way of informally verifying the sufficiency of the number of imputations is checking the stability of estimation results when changing the number of imputation or the random seeds (O'Kelly and Ratitch, 2014). It was suggested that increasing number of imputations to thousands stabilize the estimation results regarding to changing random seeds and the results are closer to the maximal likelihood estimation with an equivalent model, while not bringing technical difficulty with modern statistical applications (O'Kelly and Ratitch, 2014).

The number of imputations can be specified by the option NIMPUTE. Also, the options MIN, MAX, and ROUND are provided to make the imputed values more consistent with the observed values. MIN and MAX constrain the range of the imputed values, and when an imputed value is out of the specified range, the procedure will redraw another value. Constraining the imputed values to a specified range avoid generating impossible and unmeaningful values, but also potentially bringing bias. ROUND is generally not recommended when the outcomes take a large range of discrete values and are treated as continuous variables in the analysis; however, in the situation where MCMC are used for partial imputation to achieve monotone missing patterns for categorical variables, the imputed values should be rounded (O'Kelly and Ratitch, 2014).

COMBINE ESTIMATION

After multiple completed data sets are generated, the method that is the same as primary analysis, such as MMRM using PROC MIXED, is used to analyze each data set separately. The analysis results, including the estimate and the associated standard error from each imputed data set, are combined into a single estimation with standard error using Rubin's rule (Little and Rubin, 2002) which is directly implemented by PROC MIANALYZE in SAS. The single point estimate is the average of parameter estimates obtained from each imputed data set. The overall variance incorporates within-imputation variance and between-imputation variance. The within-imputation variance is the average of the variances across the data sets; the between-imputation variance is the sample variance of the parameter estimates across the data sets. The formula to derive the overall variance is " $Var_{\text{within}} + (1 + \frac{1}{M}) \text{Var}_{\text{Between}}$ " where M is the number of imputed data sets. The point estimate and overall variance can be used to construct t-statistic for hypothesis test about the estimated parameter.

ILLUSTRATION

For illustration, a data set resembling a typical randomized schizophrenia clinical trial was simulated. The primary efficacy endpoint is the difference in the Positive and Negative Syndrome Scale (PANSS) total score between experimental and control group in terms of mean change from baseline to Week 6. The data set is named "NON_MONO" and contains 380 observations and the following variables: USUBJID, TRT01PN, REGION, Y0, Y1, Y2, Y3, Y4, Y5 and Y6. Variable USUBJID is unique subject ID. Variable TRT01PN takes the value 1 and 0 with 1 indicating experimental group and 0 indicating control group. Variable REGION takes the value 1, 2 and 3 which represent three different regions. Y0 is the baseline PANSS total score, and Y1 to Y6 correspond to the PANSS total score collected from Week 1 to Week 6. The primary analysis is MMRM with change from baseline to Week 6 in PANSS total score as response variable, treatment group, region, baseline values of PANSS total score, visit, baseline values-by-visit interaction as fixed effects. The primary analysis is carried out using PROC MIXED assuming MAR. Sensitivity analysis under the assumptions of MNAR is performed to test the robustness of statistical inference against departure from MAR.

The following statements are used to examine the missing patterns of the data:

```
proc mi data=non_mono nimpute=0;
    var trt01pn region y0-y6;
    ods output missPattern=pattern;
run;
```

By specifying NIMPUTE=0, PROC MI does not impute missing values and only display the missing patterns of the data. VAR statement specifies the variables in the order the missing patterns are created for. The missing patterns can be outputted as a SAS data set using ODS OUTPUT statement.

The procedure created Output 1 as below. The covariates TRT01PN and REGION are fully observed. The baseline outcome Y0 is fully observed and post baseline outcomes Y1 to Y6 are partially observed. A mixture of non-monotone and monotone missing patterns exist in the data with majority of monotone missing patterns.

Output 1 shows the missing data patterns of the data “NON_MONO”.

Group	trt01pn	region	y0	y1	y2	y3	y4	y5	y6	Freq	Percent
1	X	X	X	X	X	X	X	X	X	250	65.79
2	X	X	X	X	X	X	X	X	.	28	7.37
3	X	X	X	X	X	X	X	.	X	7	1.84
4	X	X	X	X	X	X	X	.	.	19	5
5	X	X	X	X	X	X	.	X	X	3	0.79
6	X	X	X	X	X	X	.	X	.	2	0.53
7	X	X	X	X	X	X	.	.	.	22	5.79
8	X	X	X	X	X	.	X	X	X	6	1.58
9	X	X	X	X	X	.	X	.	.	1	0.26
10	X	X	X	X	X	.	.	.	.	24	6.32
11	X	X	X	X	.	X	X	X	X	3	0.79
12	X	X	X	X	.	.	X	X	X	1	0.26
13	X	X	X	X	.	.	.	.	.	1	0.26
14	X	X	X	.	.	.	.	.	.	13	3.42

Output 1. Output from PROC MI NIMPUTE=0 statement

Below statements invoke MCMC procedure and specify IMPUTE=MONOTONE to turn the arbitrary missing patterns to monotone missing patterns under MAR assumption:

```
proc mi data = non_mono out = monotone nimpute = 1000 seed = 123
    min = . . . 30 30 30 30 30 30 30
    max = . . . 210 210 210 210 210 210 210;
    var trt01pn regn1 regn2 y0 y1 y2 y3 y4 y5 y6;
    mcmc chain = multiple impute = monotone;
run;
```

PROC MI reads the data set “NON_MONO” and produces the data set “MONOTONE”. VAR statement specifies the covariates and outcomes in the order for which the monotone missing patterns are created. REGION is a categorical variable that has three levels, so two dummy variables were created with REGN1 takes the value 1 when REGION=1 and 0 otherwise, and REGN2 takes the value 1 when REGION=2 and 0 otherwise. REGION=3 serves as reference group. The option NIMPUTE specifies the number of imputations as 1000 which creates 1000 copies of monotone missing data and an index variable _IMPUTATION_ with the value 1 to 1000 is automatically created to distinguish each copy of the data set. The numbers specified in the options MIN and MAX correspond to the variables in the VAR statement. Since no imputation is needed for TRT01PN, REGN1 and REGN2, MIN and MAX are specified as a period for these variables which indicates no restriction is applied. The missing patterns of data set “MONOTONE” were examined and shown in Output 2.

Output 2 shows the missing patterns of the data set “MONOTONE”.

Group	trt01pn	region	y0	y1	y2	y3	y4	y5	y6	Freq	Percent
1	X	X	X	X	X	X	X	X	X	270000	71.05
2	X	X	X	X	X	X	X	X	.	30000	7.89
3	X	X	X	X	X	X	X	.	.	20000	5.26
4	X	X	X	X	X	X	.	.	.	22000	5.79
5	X	X	X	X	X	.	.	.	.	24000	6.32
6	X	X	X	X	.	.	.	.	.	1000	0.26
7	X	X	X	.	.	.	.	.	.	13000	3.42

Output 2. monotone missing patterns of the data set ‘MONOTONE’

The below two code fragments implement control-based pattern imputation and are equivalent:

```
proc mi data=monotone out=imputed1 seed=124 nimpute=1
   min = . . 30 30 30 30 30 30 30
   max = . . 210 210 210 210 210 210 210;
   by _imputation_;
   class trt01pn region;
   var region y0 y1 y2 y3 y4 y5 y6;
   monotone reg(/details);
   mnar model(y1 y2 y3 y4 y5 y6/ modelobs= (trt01pn='0'));
run;

proc mi data=monotone out=imputed2 seed=124 nimpute=1
   min = . . 30 30 30 30 30 30 30
   max = . . 210 210 210 210 210 210 210;
   by _imputation_;
   class trt01pn region;
   var region y0 y1 y2 y3 y4 y5 y6;
   monotone reg(/details);
   mnar model(y1/ modelobs= (trt01pn='0'))
        model(y2/ modelobs= (trt01pn='0'))
        model(y3/ modelobs= (trt01pn='0'))
        model(y4/ modelobs= (trt01pn='0'))
        model(y5/ modelobs= (trt01pn='0'))
        model(y6/ modelobs= (trt01pn='0'));
run;
```

MONOTONE and MNAR MODEL statements are used to impute missing outcomes in the order listed in the VAR statement. Treatment variable TRT01PN is in CLASS statement but not included as a model effect. MODELOBS= (TRT01PN=’0’) means only the control group is used to derive the imputation model. Variable Y1, Y2, Y3, Y4, Y5 and Y6 are specified after MODEL statement, which means all 6 outcomes from Y1 to Y6 will be imputed using the observed values from control group. If omitted from MODEL statement and specified in VAR statement, the outcome value will be imputed using observed values from both treatment groups. Y1 to Y6 can be listed together after a single MODEL option as in the first code fragment, or be listed separately using multiple MODEL options as in the second. MONOTONE REG specify Y1-Y6 be imputed using regression method. One note is when specifying options MIN and MAX, although TRT01PN does not appear in VAR statement, the first number is for TRT01PN. Since 8 variables are listed after VAR statements, there are 9 min and max specified. The data set “MONOTONE” already contains 1000 copies of the data, therefore, only one imputation (indicate by “NIMPUTE=1”) is performed for each data set (indicate by “BY _IMPUTATION_ “).

The following statements illustrate the imputation with delta-adjusted pattern imputation:

```
proc mi data=monotone out=imputed seed=125 nimpute=1
```

```
      min =  . . 30 30 30 30 30 30 30
      max =  . . 210 210 210 210 210 210 210;
      by _imputation_;
      class trt01pn region;
      var trt01pn region y0-y6;
      monotone reg(/details);
      mnar    adjust(y1 / shift=1 adjustobs=(trt01pn='1'))
              adjust(y2 / shift=1 adjustobs=(trt01pn='1'))
             adjust(y3 / shift=1 adjustobs=(trt01pn='1'))
             adjust(y4 / shift=1 adjustobs=(trt01pn='1'))
             adjust(y5 / shift=1 adjustobs=(trt01pn='1'))
             adjust(y6 / shift=1 adjustobs=(trt01pn='1'));

   run;
```

The variables to be imputed are listed in VAR statement. The variables to be adjusted are specified in MNAR ADJUST statement. By default, all the imputed observations for the specified variables in ADJUST option will be adjusted, by using the suboption ADJUSTOBS to specify a sub-level of classification variables, the adjustment will be applied to the subset of the imputed observations determined by the sub-level of the classification variable. The classification variable should also be specified in the CLASS statement. Several suboptions are available to specify the amount of the adjustment, for example, for imputed continuous variables there are suboption SHIFT to add a constant, SCALE to multiply by a constant factor, and SIGMA to add a simulated value. The above code increases the imputed outcomes Y1 to Y6 in the experimental group by the value of 1. The adjustment is applied to each visit until the end of the study and the adjusted values are used as predictors to impute the subsequent responses, which is the second variant proposed by Ratitch et al. (Ratitch et al., 2013). Note treatment variable TRT01PN is specified in the CLASS statement, and also included as a model effect.

The first variant proposed by Ratitch et al. applies the adjustment only to the first unobserved outcome from the experimental group for each subject and the adjusted imputed values are used as predictor to impute the subsequent missing values. The below statements illustrate the first variant:

```
%macro adjfvis(in=,lastvis=,seed=,adjvar=,out=);
    proc mi data=&in.(where=(lastvis>=&lastvis.)) out=temp seed=&seed.
nimpute=1
       min =  . . 30 30 30 30 30 30 30
       max =  . . 210 210 210 210 210 210 210;
          by _imputation_;
          class trt01pn region;
          var trt01pn region y0-y6;
          monotone reg(/details);
          mnar adjust( &adjvar.  / shift=1 adjustobs=(trt01pn='1'));
    run;

    data &out.;
        set &in. (where=(lastvis<&lastvis.)) temp;
    run;

    proc sort data=&out.;by _imputation_;run;
%mend;

%adjfvis(in=monotone,lastvis=5,seed=1261,adjvar=y6,out=y6);
%adjfvis(in=y6,lastvis=4,seed=1262,adjvar=y5,out=y5);
%adjfvis(in=y5,lastvis=3,seed=1263,adjvar=y4,out=y4);
%adjfvis(in=y4,lastvis=2,seed=1264,adjvar=y3,out=y3);
%adjfvis(in=y3,lastvis=1,seed=1265,adjvar=y2,out=y2);
%adjfvis(in=y2,lastvis=0,seed=1266,adjvar=y1,out=y1);
```

The variable LASTVIS is the last visit that a subject has the outcome value observed. The input data sets used by PROC MI contains two distinct records: records with all outcomes filled and records with a particular LASTVIS, and only the missing outcome next to LASTVIS is adjusted. For example, to adjust variable Y6 the records with LASTVIS=6 which is complete case and LASTVIS=5 are used, and the imputed data set is assembled with the rest unused records to form data set "Y6". To adjust variable Y5, the records with LASTVIS >= 4 are selected from data set "Y6".

The third variant proposed by Ratitch et al. applies the adjustment to the imputed values from the experimental group at all visits after the all the missing values are imputed under the assumption of MAR. The below commands illustrate the third variant:

```
proc mi data=monotone out=temp seed=127 nimpute=1
   min = . . 30 30 30 30 30 30 30
   max = . . 210 210 210 210 210 210 210;
   by _imputation_;
   class trt01pn region;
   var trt01pn region y0-y6;
   monotone reg(/details);
run;

data imputed;
    set temp;
    array m(6) y1-y6;
    if trt01pn=1 then do i=1 to 6;
      if i> lastvis then m(i)=m(i)+1;
    end;
run;
```

In addition to MONOTONE statements, MNAR can also be used together with FCS statement. When used with ADJUST option, the adjustment is applied at P-phase and the adjusted values are used as predictor to impute other variables in P-phase. The following statements implement the control-based pattern imputation and delta-adjust pattern imputation using MNAR together with FCS statement:

```
proc mi data=non_mono out=imputed seed=128 nimpute=1000
   min = . . 30 30 30 30 30 30 30
   max = . . 210 210 210 210 210 210 210;
   class trt01pn region;
   var region y0 y1 y2 y3 y4 y5 y6;
   fcs reg(/details);
   mnar model(y1 y2 y3 y4 y5 y6/ modelobs= (trt01pn='0'));
run;

proc mi data= non_mono out=imputed seed=129 nimpute=1000
   min =  . . 30 30 30 30 30 30 30
   max =  . . 210 210 210 210 210 210 210;
   class trt01pn region;
   var trt01pn region y0-y6;
   fcs reg(/details);
   mnar  adjust(y1 / shift=1 adjustobs=(trt01pn='1'))
         adjust(y2 / shift=1 adjustobs=(trt01pn='1'))
        adjust(y3 / shift=1 adjustobs=(trt01pn='1'))
        adjust(y4 / shift=1 adjustobs=(trt01pn='1'))
        adjust(y5 / shift=1 adjustobs=(trt01pn='1'))
        adjust(y6 / shift=1 adjustobs=(trt01pn='1'));
run;
```

The option and suboption for MNAR are specified in the same way as in the MONOTONE statement. Since FCS can handle arbitrary missing pattern, there is no need to turn the data set to monotone missing patterns first and the imputation can be done in one step.

The imputation generated 1000 complete data sets. The method that is the same as primary analysis is used to analyze each imputed data set separately. After imputation, the data sets are in the wide format in which different variables (Y0-Y6) represent the outcome measured at different occasions. Before analyzing using PROC MIXED, the data sets need to be converted into long format in which one variable represents all outcome with different values of AVISITN differentiating different occasions.

The following statements analyze each complete data set using PROC MIXED:

```
  proc mixed data=imputed;
      by _imputation_;
      class region trt01pn(ref='0') usubjid avisitn;
      model chg = region trt01pn avisitn trt01pn*avisitn base base*avisitn
             /ddfm=kr2;
      repeated avisitn / subject=usubjid type=un;
      lsmeans trt01pn*avisitn /pdiff cl;
      ods output lsmeans=lsmeans diffs=diffs;
run;
```

BASE is Y0 and CHG is change from baseline (Y1 - Y0 to Y6 - Y0). The analysis results are stored as SAS data set using ODS OUTPUT statement. The least square means (LSMs) for each treatment group for each visit are outputted to data set “LSMEANS”. The differences of LSMs between experimental group and control group for each visit are outputted to data set “DIFFS”.

In the final step, the analysis results obtained from PROC MIXED procedure are combined into a single estimation with standard error using PROC MIANALYZE. The code is shown in the below:

```
  proc sort data=lsmeans;by trt01pn avisitn _imputation;run;

  proc mianalyze parms=lsmeans;
       modeleffects trt01pn*avisitn;
       ods output ParameterEstimates=lsm;
       by trt01pn avisitn;
   run;

  proc sort data=diffs;by avisitn _imputation;run;

  proc mianalyze parms=diffs;
       modeleffects trt01pn*avisitn;
       ods output ParameterEstimates=dif;
       by avisitn;
  run;
```

The data sets “LSMEANS” and “DIFFS” are passed to PROC MIANALYZE using PARMS option for combination. The MIANALYZE procedure reads the effect names contained in the variable EFFECT, the estimates contained in the variable ESTIMATE, the standard errors contained in the variable STDERR and differentiates each imputation by the variable _IMPUTATION_. The statement MODELEFFECTS declares the effect to be analyzed is treatment by visit interaction term TRT01PN*AVISITN. The BY statement is used to ensure the LSMs are combined per treatment group per visit and differences of LSMs between treatment groups are combined per visit. The results are outputted to data sets “LSM” and “DIF” using ODS statements.

CONCLUSION

Missing data is unavoidable in longitudinal clinical trial data and the primary analysis often operates under the assumption of MAR. The sensitivity analysis under various MNAR scenarios is required by regulatory

agency to examine the sensitivity of the statistical inferences against departure from the MAR assumption. SAS® Version 9.4 PROC MI provides a MNAR statement, with two options MODEL and ADJUST, that conveniently implements the sensitivity analysis using multiple imputations with pattern-mixture models. MNAR statement can be used along with MONOTONE statement to handle monotone missing data patterns and FCS statement to handle arbitrary missing data patterns. In this paper, the implementation of sensitivity analysis using MNAR statement is discussed and the SAS code is presented.

REFERENCE

Kenward M, Molenberghs G, Thijs H. 2003. Pattern-mixture models with proper time dependence. *Biometrika*, 90(1): 53-71.

Little R, Yau L.1996. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*, 52:1324-1333.

Little, RJA. and Rubin, DB. 2002. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

National Research Council, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Panel on Handling Missing Data in Clinical Trials. 2010. *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academic Press.

O'Kelly, M and Ratitch, B. 2014. *Clinical Trial with Missing Data A Guide for Practitioners*. 1st ed. West Sussex, UK: John Wiley & Sons.

Ratitch, B and O'Kelly, M and Tosielloc, R. 2013. "Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models". *Pharmaceutical Statistics*, 12:337-347.

SAS/STAT® 14.1 User's Guide. Accessed March 29, 2019. http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_mi_details55.htm

CONTACT INFORMATION

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

Lingling Li

Linglingli9907@gmail.com