Sensitivity Analysis in Multiple Imputation for Missing Data

Yang Yuan, SAS Institute Inc.

ABSTRACT

Multiple imputation, a popular strategy for dealing with missing values, usually assumes that the data are missing at random (MAR). That is, for a variable **Y**, the probability that an observation is missing depends only on the observed values of other variables, not on the unobserved values of **Y**. It is important to examine the sensitivity of inferences to departures from the MAR assumption, because this assumption cannot be verified using the data.

The pattern-mixture model approach to sensitivity analysis models the distribution of a response as the mixture of a distribution of the observed responses and a distribution of the missing responses. Missing values can then be imputed under a plausible scenario for which the missing data are missing not at random (MNAR). If this scenario leads to a conclusion different from that of inference under MAR, then the MAR assumption is questionable.

This paper reviews the concepts of multiple imputation and explains how you can apply the pattern-mixture model approach in the MI procedure by using the MNAR statement, which is new in SAS/STAT® 13.1. You can specify a subset of the observations to derive the imputation model, which is used for pattern imputation based on control groups in clinical trials. You can also adjust imputed values by using specified shift and scale parameters for a set of selected observations, which are used for sensitivity analysis with a tipping-point approach.

INTRODUCTION

Missing values are a problem in many statistical analyses. Most SAS statistical procedures exclude from analysis observations that have any missing variable values. These observations are called incomplete cases. Although using only complete cases is simpler, you lose the information that is in the incomplete cases. Excluding observations that have missing values also ignores the possibility of systematic differences between complete cases and incomplete cases, so the resulting inference might not apply to the entire population, especially when you have a small number of complete cases.

One strategy that you can use to handle missing values is multiple imputation, which replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute (Rubin 1976, 1987). You then analyze the multiply imputed data sets by using standard procedures for complete data and combining the results from these analyses.

Multiple imputation does not attempt to estimate each missing value through simulated values, but rather to represent a random sample of the missing values. This process results in valid statistical inferences that properly reflect the uncertainty that results from missing values, such as valid confidence intervals for parameters.

Multiple imputation inference involves three distinct phases:

1. The missing data are filled in m times to generate m complete data sets.

2. The m complete data sets are analyzed by using standard SAS procedures.

3. The results from the m complete data sets are combined for the inference.

The MI procedure is a multiple imputation procedure that creates multiply imputed data sets for incomplete p-dimensional multivariate data. It uses methods that incorporate appropriate variability across the m imputations. Which imputation method you choose depends on the patterns of missingness in the data and the type of the imputed variable.

A data set that contains the variables $\mathbf{Y}_1$, $\mathbf{Y}_2$, ..., $\mathbf{Y}_p$ (in that order) is said to have a *monotone missing pattern* when the event that a variable $\mathbf{Y}_j$ is missing for a particular individual implies that all subsequent variables $\mathbf{Y}_k$, $k > j$, are missing for that individual.

For data sets that have monotone missing patterns, the variables that contain missing values can be imputed sequentially using covariates constructed from their corresponding sets of preceding variables. You can use a regression method or a predictive mean matching method to impute missing values for a continuous variable; a logistic regression method to impute missing values for a classification variable that has a binary, nominal, or ordinal response; and a discriminant function method to impute missing values for a classification variable that has a binary or nominal response.

For data sets that have arbitrary missing patterns, you can use either of the following methods to impute missing values: a Markov chain Monte Carlo (MCMC) method (Schafer 1997) that assumes multivariate normality, or a fully conditional specification (FCS) method (Brand 1999; van Buuren 2007) that assumes the existence of a joint distribution for all variables. An FCS method uses a separate conditional distribution for each imputed variable. The applicable methods are similar to the methods available for data sets that have monotone missing patterns.

After you analyze the m complete data sets by using standard SAS procedures, you can use the MIANALYZE procedure to generate valid statistical inferences about these parameters by combining results from the m analyses.

Multiple imputation usually assumes that the data are missing at random (MAR). That is, for a variable $\mathbf{Y}$, the probability that an observation is missing depends only on the observed values of other variables, not on the unobserved values of $\mathbf{Y}$. The MAR assumption cannot be verified, because the missing values are not observed.

For a study that assumes MAR, the sensitivity of inferences to departures from the MAR assumption should be examined (National Research Council 2010, p. 111). If it is plausible that the missing data are not MAR, you can perform sensitivity analysis under the missing not at random (MNAR) assumption. That is, missing values are imputed under a plausible MNAR scenario, and the results are examined. If this scenario leads to a conclusion different from that of inference under MAR, then the MAR assumption is questionable.

The following section describes sensitivity analysis for the MAR assumption, followed by two examples: one to specify sets of observations for imputation models and the other to adjust imputed values for a subset of observations.

SENSITIVITY ANALYSIS FOR THE MAR ASSUMPTION

Multiple imputation usually assumes that the data are missing at random (MAR). For example, suppose the data contain a set of fully observed variables $\mathbf{X}$ and a variable $\mathbf{Y}$ that contains missing observations. Also suppose $\mathbf{R}$ is a response indicator whose element is 0 or 1, depending on whether $\mathbf{Y}$ is missing or observed. Then, the MAR assumption is that the probability that the $\mathbf{Y}$ value is missing for an observation can depend on the observed values of $\mathbf{X}$ for the observation, but not on the unobserved value of $\mathbf{Y}$. That is,

$$\mathrm{pr}(\mathbf{R} \mid \mathbf{X}, \mathbf{Y}) = \mathrm{pr}(\mathbf{R} \mid \mathbf{X})$$

It can be shown that

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{X}, \mathbf{R}) = \mathrm{pr}(\mathbf{Y} \mid \mathbf{X})$$

which implies

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{X}, \mathbf{R} = 0) = \mathrm{pr}(\mathbf{Y} \mid \mathbf{X}, \mathbf{R} = 1)$$

Thus the posterior distribution of observations that have observed $\mathbf{Y}$, $\mathrm{pr}(\mathbf{Y} \mid \mathbf{X}, \mathbf{R} = 1)$, can be used to create imputations for missing data under the MAR assumption.

A straightforward sensitivity analysis for the MAR assumption in multiple imputation is based on the pattern-mixture model approach (Little 1993; Molenberghs and Kenward 2007, pp. 30, 34–37), which models the distribution of a response as the mixture of a distribution of the observed responses and a distribution of the missing responses:

$$\mathrm{pr}(\mathbf{Y}, \mathbf{X}, \mathbf{R}) = \mathrm{pr}(\mathbf{Y}, \mathbf{X} \mid \mathbf{R})\,\mathrm{pr}(\mathbf{R}) = \mathrm{pr}(\mathbf{Y}, \mathbf{X} \mid \mathbf{R} = 0)\,\mathrm{pr}(\mathbf{R} = 0) + \mathrm{pr}(\mathbf{Y}, \mathbf{X} \mid \mathbf{R} = 1)\,\mathrm{pr}(\mathbf{R} = 1)$$

Under the MNAR assumption, the probability that the value of **Y** is missing for an observation can depend on the unobserved value of **Y**,

$$\mathrm{pr}(\mathbf{R} \mid \mathbf{X}, \mathbf{Y}) \neq \mathrm{pr}(\mathbf{R} \mid \mathbf{X})$$

which implies

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{X}, \mathbf{R} = 0) \neq \mathrm{pr}(\mathbf{Y} \mid \mathbf{X}, \mathbf{R} = 1)$$

For example, suppose the data in a clinical trial contain an indicator variable **Trt**, which has a value of 1 for patients in the treatment group and a value of 0 for patients in the control group; a variable **Y0** for the baseline efficacy score; and a variable **Y** for the efficacy score at a follow-up visit. Assume that **X** = (**Trt**, **Y0**) are fully observed and **Y** is not fully observed. Then, under the MAR assumption,

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 0,\ \mathbf{Y0} = y0,\ \mathbf{R} = 0) = \mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 0,\ \mathbf{Y0} = y0,\ \mathbf{R} = 1)$$

and

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 0) = \mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 1)$$

Under the MNAR assumption,

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 0,\ \mathbf{Y0} = y0,\ \mathbf{R} = 0) \neq \mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 0,\ \mathbf{Y0} = y0,\ \mathbf{R} = 1)$$

or

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 0) \neq \mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 1)$$

Thus, one MNAR scenario is that missing **Y** values in the treatment group can be imputed from a posterior distribution that is generated from observations in the control group,

$$\mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 0) = \mathrm{pr}(\mathbf{Y} \mid \mathbf{Trt} = 0,\ \mathbf{Y0} = y0,\ \mathbf{R} = 1)$$

Another MNAR scenario is that the imputed values can be adjusted to reflect a systematic difference between the distributions of missing and observed **Y** values,

$$\mathrm{pr}(\mathbf{Y} = y + \delta \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 0) = \mathrm{pr}(\mathbf{Y} = y \mid \mathbf{Trt} = 1,\ \mathbf{Y0} = y0,\ \mathbf{R} = 1)$$

where δ is the systematic difference.

In the MI procedure, the new MNAR statement imputes missing values by using the pattern-mixture model approach, assuming the missing data are MNAR. The two main options in the MNAR statement are MODEL and ADJUST. You use the MODEL option to specify a subset of observations from which imputation models are to be derived for specified variables. You use the ADJUST option to specify an imputed variable and adjustment parameters (such as shift and scale) for adjusting the imputed variable values of a specified subset of observations. The MNAR statement is applicable only if it is used along with a MONOTONE statement or an FCS statement.

SENSITIVITY ANALYSIS: SPECIFYING SETS OF OBSERVATIONS FOR IMPUTATION MODELS

Under the MAR assumption, all available observations are used to derive the imputation model. Under the MNAR assumption, you can specify the set of observations that are used to derive the imputation model. For each imputed variable, you can use the MODEL option in the MNAR statement in PROC MI to select the set of observations to construct the imputation model.

EXAMPLE: SENSITIVITY ANALYSIS WITH CONTROL-BASED PATTERN IMPUTATION

This example illustrates sensitivity analysis in multiple imputation under the MNAR assumption by creating control-based pattern imputation.

Suppose a pharmaceutical company is conducting a clinical trial to test the efficacy of a new drug. The trial studies two groups of equally allocated patients: a treatment group that receives the new drug and a control group that receives a placebo. The variable **Trt** is an indicator variable; it has a value of 1 for patients in the treatment group and a value of 0 for patients in the control group. The variable **Y0** is the baseline efficacy score, and the variable **Y1** is the efficacy score at a follow-up visit.

If the data set does not contain any missing values, then a regression model such as

$$Y1 = Trt \ Y0$$

can be used to test the treatment effect.

Suppose the variables **Trt** and **Y0** are fully observed and the variable **Y1** contains missing values in both the treatment and control groups, as shown in Table 1.

Table 1 Variables

Variables		
Trt	Y0	Y1
0	X	X
1	X	X
0	X	.
1	X	.

Suppose the data set Mono1 contains the data from the trial that have missing values in **Y1**. Figure 1 lists the first 10 observations.

Figure 1 Clinical Trial Data

First 10 Observations in the Trial Data

Obs	Trt	y0	y1
1	0	10.5212	11.3604
2	0	8.5871	8.5178
3	0	9.3274	.
4	0	9.7519	.
5	0	9.3495	9.4369
6	1	11.5192	13.2344
7	1	10.7841	.
8	1	9.7717	10.9407
9	1	10.1455	10.8279
10	1	8.2463	9.6844

The following statements use the MI procedure to impute missing values under the MAR assumption:

```
proc mi data=Mono1 seed=14823 nimpute=10 out=Out1;
   class Trt;
   monotone reg(y1);
   var Trt y0 y1;
run;
```

With a MONOTONE statement, the input data set Mono1 must have a monotone missing pattern with variables ordered in the VAR list. The REG(Y1) option specifies a regression method to impute missing values in **Y1**, and all preceding variables (**Trt** and **Y0**) are used as the covariates in the imputation model by default.

The following statements list the first 10 observations of the imputed data set Out1 in Figure 2:

```
proc print data=Out1(obs=10);
   title 'First 10 Observations of the Imputed Data Set';
run;
```

Figure 2 Imputed Data Set under MAR

First 10 Observations of the Imputed Data Set

Obs	_Imputation_	Trt	y0	y1
1	1	0	10.5212	11.3604
2	1	0	8.5871	8.5178
3	1	0	9.3274	9.1261
4	1	0	9.7519	9.6610
5	1	0	9.3495	9.4369
6	1	1	11.5192	13.2344
7	1	1	10.7841	12.4813
8	1	1	9.7717	10.9407
9	1	1	10.1455	10.8279
10	1	1	8.2463	9.6844

The variable **_Imputation_** indexes the 10 imputed data sets in Figure 2.

The following statements generate regression coefficients for each of the 10 imputed data sets:

```
proc reg data=Out1;
   model y1= Trt y0;
   by  _Imputation_;
   ods output parameterestimates=Parm1;
run;
```

The following statements use the MIANALYZE procedure to combine the 10 sets of regression coefficients:

```
proc mianalyze parms=Parm1;
   modeleffects Trt;
run;
```

The PARMS= option specifies the input data set that contains parameter estimates computed from imputed data sets. The MODELEFFECTS statement lists the effects to be analyzed.

The "Parameter Estimates" table in Figure 3 displays a combined estimate and standard error for the regression coefficient for **Trt**. The table shows a t statistic of 3.37, with an associated p-value of 0.0011 for the test that the regression coefficient is equal to 0. Thus, the treatment effect is significant at a two-sided Type I error level of 0.05.

Figure 3 Parameter Estimates

The MIANALYZE Procedure

									t for H0:	
Parameter	**Estimate**	**Std Error**	**95% Confidence**	**Limits**	**DF**	**Minimum**	**Maximum**	**Theta0**	**Parameter=Theta0**	**Pr > |t|**
Trt	0.893577	0.265276	0.366563	1.420591	90.029	0.624115	1.121445	0	3.37	0.0011

This conclusion is based on the MAR assumption. However, if missing **Y1** values for patients in the treatment group imply that these individuals no longer receive the treatment, then it is reasonable to assume that the conditional distribution of **Y1**, given **Y0** for patients who have missing **Y1** values in the treatment group, is similar to the corresponding distribution of patients in the control group.
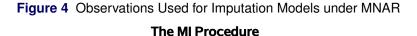
Ratitch and O'Kelly (2011) describe an implementation of the pattern-mixture model approach that uses a control-based pattern imputation. That is, an imputation model for the missing observations in the treatment group is constructed not from the observed data in the treatment group but rather from the observed data in the control group. This model is also the imputation model that is used to impute missing observations in the control group.

The following statements implement the control-based pattern imputation:

```
proc mi data=Mono1 seed=14823 nimpute=10 out=Out1X;
   class Trt;
   monotone reg(y1);
   mnar model( y1 / modelobs=(Trt='0'));
   var y0 y1;
run;
```

The MNAR statement imputes missing values for scenarios under the MNAR assumption. The MODEL option specifies that only observations in which TRT=0 are used to derive the imputation model for the variable **Y1**. Thus, **Y0** and **Y1** (but not **Trt**) are specified in the VAR list.

By default, for each imputed variable, all available observations are used in the imputation model. When you specify the MODEL option in the MNAR statement, the "Observations Used for Imputation Models Under MNAR Assumption" table, shown in Figure 4, lists the subset of observations that are used to construct the imputation model for **Y1**.

Figure 4 Observations Used for Imputation Models under MNAR

The MI Procedure

Observations Used for Imputation Models Under MNAR Assumption	
Imputed Variable	**Observations**
y1	Trt = 0

The following statements list the first 10 observations of the imputed data set Out1X in Figure 5:

```
proc print data=Out1X(obs=10);
   title 'First 10 Observations of the Imputed Data Set';
run;
```

Figure 5 Imputed Data Set under MNAR

First 10 Observations of the Imputed Data Set

Obs	_Imputation_	Trt	y0	y1
1	1	0	10.5212	11.3604
2	1	0	8.5871	8.5178
3	1	0	9.3274	9.5786
4	1	0	9.7519	9.6060
5	1	0	9.3495	9.4369
6	1	1	11.5192	13.2344
7	1	1	10.7841	10.7873
8	1	1	9.7717	10.9407
9	1	1	10.1455	10.8279
10	1	1	8.2463	9.6844

The following statements generate regression coefficients for each of the 10 imputed data sets:

```
proc reg data=Out1X;
   model y1= Trt y0;
   by _Imputation_;
   ods output parameterestimates=Parm1X;
run;
```

6

The following statements use the MIANALYZE procedure to combine the 10 sets of regression coefficients:

```
proc mianalyze parms=Parm1X;
    modeleffects Trt;
run;
```

The "Parameter Estimates" table in Figure 6 shows a t statistic of 2.24, with a p-value of 0.0292 for the test that the parameter is equal to 0. Thus, for a two-sided Type I error level of 0.05, the significance of the treatment effect is not reversed by control-based pattern imputation.

Figure 6 Parameter Estimates

The MIANALYZE Procedure

			Parameter Estimates						t for H0:	
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	Parameter=Theta0	Pr > |t|
Trt	0.664712	0.297378	0.069701	1.259724	59.197	0.329363	0.892285	0	2.24	0.0292

SENSITIVITY ANALYSIS: ADJUSTING IMPUTED VALUES FOR A SUBSET OF OBSERVATIONS

Under the MAR assumption, imputed values are not adjusted. Under the MNAR assumption, imputed values can be adjusted under different scenarios. When you impute continuous variables by using the regression and predictive mean matching methods, you can adjust the imputed values directly (Carpenter and Kenward 2013, pp. 237–239; van Buuren 2012, pp. 88–89). When you impute classification variables by using the logistic regression method, you can adjust the imputed classification levels by modifying the log odds ratios for the classification levels (Carpenter and Kenward 2013, pp. 240–241; van Buuren 2012, pp. 88–89). By modifying the log odds ratios, you modify the predicted probabilities for the classification levels.

For each imputed variable, you can use the ADJUST option in the MNAR statement to do the following:

- specify a subset of observations for which imputed values are adjusted. Otherwise, all imputed values are adjusted.

- adjust imputed continuous variable values by using the SHIFT=, SCALE=, and SIGMA= options. These options add a constant, multiply by a constant factor, and add a simulated value to the imputed values, respectively.

- adjust imputed classification variable levels by using the SHIFT= and SIGMA= options to adjust predicted probabilities for the classification levels. These options add a constant and add a simulated constant value, respectively, to the log odds ratios for the classification levels.

EXAMPLE: SENSITIVITY ANALYSIS WITH A TIPPING-POINT APPROACH

This example illustrates sensitivity analysis in multiple imputation under the MNAR assumption by searching for a tipping point that reverses the study conclusion.

Suppose a pharmaceutical company is conducting a clinical trial to test the efficacy of a new drug. The trial studies two groups of equally allocated patients: a treatment group that receives the new drug and a control group that receives a placebo. The variable **Trt** is an indicator variable; it has a value of 1 for patients in the treatment group and a value of 0 for patients in the control group. The variable **Y0** is the baseline efficacy score, and the variable **Y1** is the efficacy score at a follow-up visit.

If the data set does not contain any missing values, then a regression model such as

$$Y1 = Trt \ Y0$$

can be used to test the treatment effect.

Suppose the variables **Trt** and **Y0** are fully observed and the variable **Y1** contains missing values in both the treatment and control groups. Now suppose the data set Mono2 contains the data from a trial that have missing values in **Y1**. Figure 7 lists the first 10 observations.

Figure 7 Clinical Trial Data

First 10 Observations in the Trial Data

Obs	Trt	y0	y1
1	0	11.4826	.
2	0	10.0090	10.8667
3	0	11.3643	10.6660
4	0	11.3098	10.8297
5	0	11.3094	.
6	1	10.3815	10.5587
7	1	11.2001	13.7616
8	1	9.7002	10.3460
9	1	10.0801	.
10	1	11.2667	11.0634

The following statements use the MI procedure to impute missing values under the MAR assumption:

```
proc mi data=Mono2 seed=14823 nimpute=10 out=Out2;
   class Trt;
   monotone reg(y1);
   var Trt y0 y1;
run;
```

The following statements generate regression coefficients for each of the 10 imputed data sets:

```
proc reg data=Out2;
   model y1= Trt y0;
   by  _Imputation_;
   ods output parameterestimates=Parm2;
run;
```

The following statements use PROC MIANALYZE to combine the 10 sets of regression coefficients:

```
proc mianalyze parms=Parm2;
   modeleffects Trt;
run;
```

The "Parameter Estimates" table in Figure 8 displays a combined estimate and standard error for the regression coefficient for **Trt**. The table displays a 95% confidence interval (0.2865, 1.2261), which does not contain 0. The table also shows a t statistic of 3.19, with an associated p-value of 0.0019 for the test that the regression coefficient is equal to 0.

Figure 8 Parameter Estimates

The MIANALYZE Procedure

									t for H0:	
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	Parameter=Theta0	Pr > |t|
Trt	0.756280	0.236952	0.286493	1.226068	105.84	0.556144	0.964349	0	3.19	0.0019

Parameter Estimates

This result is based on the MAR assumption. However, if it is plausible that, for the treatment group, the distribution of missing **Y1** responses has a lower expected value than that of the distribution of observed **Y1** responses, then the result under the MAR assumption should be examined.

The %midata macro, shown in the Appendix, generates 10 imputed data sets for each of the specified shift parameters. The imputed values for observations in the treatment group (TRT=1) are adjusted using the shift parameter.

Assume that the tipping point that reverses the study conclusion is between –2 and 0. The following statement generates 10 imputed data sets for each of the shift parameters –2.0, –1.8, ... , 0:

```
%midata( data=Mono2, smin=-2, smax=0, sinc=0.2, out=OutX);
```

The following statements perform regression tests on each of the 10 imputed data sets:

```
proc reg data=OutX;
   model y1= Trt y0;
   by  Shift _Imputation_;
   ods output parameterestimates=ParmX;
run;
```

The following statements use PROC MIANALYZE to combine the 10 sets of regression coefficients for each shift parameter:

```
proc mianalyze parms=ParmX;
   modeleffects Trt;
   by  Shift;
   ods output parameterestimates=MiParmX;
run;
```

The following statements display the p-values that are associated with the shift parameters (as shown in Figure 9):

```
proc print label data=MiParmX;
   var Shift Probt;
   title 'P-values for Shift Parameters';
   label Probt='Pr > |t|';
   format Probt 8.4;
run;
```

Figure 9 Finding Tipping Point for Shift Parameter between –2 and 0

P-values for Shift Parameters

Obs	Shift	Pr > |t|
1	-2.0	0.1861
2	-1.8	0.1328
3	-1.6	0.0916
4	-1.4	0.0611
5	-1.2	0.0395
6	-1.0	0.0248
7	-0.8	0.0152
8	-0.6	0.0091
9	-0.4	0.0054
10	-0.2	0.0032
11	0.0	0.0019

For a two-sided Type I error level of 0.05, the tipping point for the shift parameter is between –1.4 and –1.2. The following statement generates 10 imputed data sets for each of the shift parameters –1.40, –1.39, ... , –1.20:

```
%midata( data=Mono2, smin=-1.4, smax=-1.2, sinc=0.01, out=OutX);
```

The following statements perform regression tests on each of the 10 imputed data sets:

```
proc reg data=OutX;
   model y1= Trt y0;
   by  Shift _Imputation_;
   ods output parameterestimates=ParmX;
run;
```

The following statements use PROC MIANALYZE to combine the 10 sets of regression coefficients for each shift parameter:

```
proc mianalyze parms=ParmX;
   modeleffects Trt;
   by  Shift;
   ods output parameterestimates=MiParmX;
run;
```

The following statements display the p-values that are associated with the shift parameters (as shown in Figure 10):

```
proc print label data=MiParmX;
   var Shift Probt;
   title 'P-values for Shift Parameters';
   label Probt='Pr > |t|';
   format Probt 8.4;
run;
```

Figure 10 Finding Tipping Point for Shift Parameter between −1.40 and −1.20

P-values for Shift Parameters

Obs	Shift	Pr > |t|
1	-1.40	0.0611
2	-1.39	0.0598
3	-1.38	0.0586
4	-1.37	0.0573
5	-1.36	0.0561
6	-1.35	0.0549
7	-1.34	0.0538
8	-1.33	0.0526
9	-1.32	0.0515
10	-1.31	0.0504
11	-1.30	0.0493
12	-1.29	0.0482
13	-1.28	0.0472
14	-1.27	0.0462
15	-1.26	0.0452
16	-1.25	0.0442
17	-1.24	0.0432
18	-1.23	0.0423
19	-1.22	0.0413
20	-1.21	0.0404
21	-1.20	0.0395

The study conclusion under MAR is reversed when the shift parameter is −1.31. Thus, if the shift parameter −1.31 is plausible, the conclusion under MAR is questionable.

SUMMARY

Multiple imputation usually assumes that the data are missing at random (MAR). This assumption cannot be verified, because the missing values are not observed. Although the MAR assumption becomes more plausible as more variables are included in the imputation model (Schafer 1997, pp. 27–28), it is important to examine the sensitivity of inferences to departures from the MAR assumption (National Research Council 2010, p. 111).

In the MI procedure, you can use the MNAR statement, which is new in SAS/STAT 13.1, to impute missing values under various MNAR scenarios. You can then generate inferences and examine the results. If the results under MNAR differ from the results under MAR, then you should question the conclusion under the MAR assumption.

APPENDIX

The following %midata macro generates 10 imputed data sets for each of the specified shift parameters. The imputed values for observations in the treatment group (TRT=1) are adjusted using the shift parameter.

```
/*----------------------------------------------------------------*/
/*--- Generate imputed data set for specified shift parameters ---*/
/*--- data= input data set                                    ---*/
/*--- smin= min shift parameter                               ---*/
/*--- smax= max shift parameter                               ---*/
/*--- sinc= increment of the shift parameter                  ---*/
/*--- out=  output imputed data set                           ---*/
/*----------------------------------------------------------------*/
%macro midata( data=, smin=, smax=, sinc=, out=);

data &out;
   set _null_;
run;

/*------------ # of shift values ------------*/
%let ncase= %sysevalf( (&smax-&smin)/&sinc, ceil );

/*------- Imputed data for each shift -------*/
%do jc=0 %to &ncase;
   %let sj= %sysevalf( &smin + &jc * &sinc);

   proc mi data=&data seed=14823 nimpute=10 out=outmi;
      class Trt;
      monotone reg(y1);
      mnar adjust( y1 / shift=&sj  adjustobs=(Trt='1') );
      var Trt y0 y1;
   run;

   data outmi;
      set outmi;
      Shift= &sj;
   run;

   data &out;
      set &out outmi;
   run;

%end;
%mend midata;
```

REFERENCES

Brand, J. P. L. (1999), *Development, Implementation, and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets*, Ph.D. thesis, Erasmus University.

Carpenter, J. R. and Kenward, M. G. (2013), *Multiple Imputation and Its Application*, New York: John Wiley & Sons.

Little, R. J. A. (1993), "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association*, 88, 125–134.

Molenberghs, G. and Kenward, M. G. (2007), *Missing Data in Clinical Studies*, New York: John Wiley & Sons.

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

Ratitch, B. and O'Kelly, M. (2011), "Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures," in *Proceedings of PharmaSUG 2011 (Pharmaceutical Industry SAS Users Group)*, SP04, Nashville.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman & Hall.

van Buuren, S. (2007), "Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification," *Statistical Methods in Medical Research*, 16, 219–242.

van Buuren, S. (2012), *Flexible Imputation of Missing Data*, Boca Raton, FL: Chapman & Hall/CRC.

ACKNOWLEDGMENT

CONTACT INFORMATION

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

Yang Yuan
SAS Institute Inc.
111 Rockville Pike, Suite 1000
Rockville, MD 20850
301-838-7030
310-838-7410 (Fax)
Yang.Yuan@sas.com