

Missing Data and Complex Sample Surveys Using SAS®: The Impact of Listwise Deletion vs. Multiple Imputation on Point and Interval Estimates when Data are MCAR and MAR

Anh P. Kellermann, DeAnn Trevathan, and Jeffrey D. Kromrey
University of South Florida

ABSTRACT

Social scientists from many fields use secondary data analysis of complex sample surveys to answer research questions and test hypotheses. Despite great care taken to obtain the data needed, missing data are frequently found in such samples. Even though missing data is a ubiquitous problem, the methodological literature has provided little guidance to inform the appropriate treatment for such missingness. This Monte Carlo study used SAS to investigate the impact of missing data treatment (multiple imputation versus listwise deletion) when data are MCAR and MAR. By using 10% to 70% of missing data (along with complete sample conditions as a reference point for interpretation of results), the research focused on the parameter estimates in multiple regression analysis in complex sample data. Results are presented in terms of statistical bias in the parameter estimates and both confidence interval coverage and width.

INTRODUCTION

COMPLEX SAMPLE SURVEY DATA

Secondary data analysis of nationally representative surveys is commonly conducted by researchers and can be extremely useful when investigating a variety of social and behavioral outcomes. By providing access to a vast array of variables on large numbers of people and their environments such as schools or neighborhoods, the nature of large-scale social science data is enticing to many researchers. The degree to which unbiased estimates and accurate inferences can be made from complex samples depends, however, on the care researchers take when analyzing the data.

Data from complex sample surveys differ from those obtained via simple random sampling in several ways that impact how statistical analyses should be conducted. For example, the probabilities of selection of the observations are often not equal leading to the need to incorporate sample weights. Further, multi-stage sampling yields clustered observations in which the variance among units within each cluster is less than the variance among units in general (i.e., intraclass correlation), which complicates the estimation of sampling error. In addition, stratification in the sampling design (e.g., geographical stratification) insures appropriate sample representation on the stratification variable(s), but yields negatively biased estimates of the population variance when not considered in the analysis. Finally, adjustments can be applied to the sample for unit nonresponse and other post-stratification to allow unbiased estimates of population characteristics (Brick & Kalton, 1996).

Sample Weights

Observations from complex sample surveys are typically weighted such that an observation's weight is based on the reciprocal of the observation's probability of being selected. That is, observations more likely to be selected (e.g., from oversampling) receive smaller weights than observations less likely to be selected. In data available from large-scale surveys, weights may be provided such that the sum of the weights equals the sample size (relative weights)

$$\sum_{k=1}^s \sum_{j=1}^{p_k} \sum_{i=1}^{n_{jk}} w_{ijk} = n$$

where w_{ijk} is the weight assigned to the i^{th} individual in the j^{th} primary sampling unit (PSU) of the k^{th} stratum in a study with s strata where the k^{th} stratum has p_k PSUs and the j^{th} PSU within the k^{th} stratum has a sample of n_{jk} individuals. Alternatively, weights may be provided such that the sum of the weights equals the population size (raw weights)

$$\sum_{k=1}^s \sum_{j=1}^{p_k} \sum_{i=1}^{n_{jk}} w_{ijk} = N$$

Sample weights are then applied in the computation of statistics from the sample observations. For example, the sample mean is computed as

$$\bar{X} = \frac{\sum_{k=1}^s \sum_{j=1}^{p_k} \sum_{i=1}^{n_{jk}} w_{ijk} X_{ijk}}{\sum_{k=1}^s \sum_{j=1}^{p_k} \sum_{i=1}^{n_{jk}} w_{ijk}}$$

where X_{ijk} is the value for the i^{th} individual in the j^{th} PSU of the k^{th} stratum, and other symbols are defined as they were previously. When researchers omit sample weights from the analysis of complex survey data, parameter estimates are typically biased and incorrect inferences can be drawn. Further, when sample weights are not used, findings are generally not representative of the larger population of interest.

Estimation of Variances

The estimation of sampling error is a critical component of survey analysis. Sampling error provides an index of the precision of point estimates (e.g., sample means or regression coefficients) and is used in the calculation of confidence intervals and hypothesis tests. For complex sample surveys (involving stratification, multi-stage sampling, and unequal probabilities of selection) the calculation of sampling error differs from the calculation used in simple random sampling. As an example of multi-stage sampling consider a context where schools are sampled within geographically defined strata, and then students are sampled within schools. In this sort of situation, the sampling variance of a statistic can be obtained by focusing on the between-cluster variance estimate of the statistic (Williams, 2000). Specific variance estimates can be obtained through Taylor series linearization, or through replication methods, such as balanced repeated replications, jackknife, and bootstrap methods (Skinner, Holt, & Smith, 1989).

Taylor series linearization is used in many statistical applications to obtain approximate values of functions that are difficult to calculate precisely. Because most statistical estimates from complex sample surveys are not simple linear functions of the observations, a Taylor series expansion may be used to obtain an approximation of the estimate based on the linear (first-order) part of the Taylor series. The variance of this approximation may then be used to estimate the variance of the original statistic of interest. The Taylor series approach tends to be computationally fast (in comparison with replication methods) but carries the limitation that a separate formula must be developed for each estimate of interest.

As an example, consider estimating the variance of the sample mean. Graubard and Korn (1996) show that Taylor linearization leads to

$$\hat{\sigma}_{\bar{X}}^2 = \frac{\sum_{k=1}^s \frac{p_k}{p_k - 1} \sum_{j=1}^{p_k} \left[w_{jk} (\bar{X}_{jk} - \bar{X}) - \frac{1}{p_k} \sum_{t=1}^{p_k} w_{tk} (\bar{X}_{tk} - \bar{X}) \right]^2}{\left(\sum_{k=1}^s \sum_{j=1}^{p_k} w_{jk} \right)^2}$$

where W_{jk} is the sum of the weights in the j^{th} PSU in the k^{th} stratum, $\bar{X}_{jk}$ is the mean for the j^{th} PSU in the k^{th} stratum, and the other symbols are as previously defined.

MISSING DATA

Missing data are a ubiquitous problem in social science research. As most common statistical procedures assume complete data, failing to address the presence of missing values can lead to a multitude of issues, ranging from decreased power to heightened bias and inaccurate Type I error control. Among the many methods developed to deal with missing data problems, multiple imputation (MI) has become one of the most accepted. Provided the assumptions are met, MI has shown excellent qualities (see, for example, reviews by Graham, Cumsille, & Elek-Fisk, 2003; Graham & Hoffer, 2000; Shafer & Olsen, 1998).

Missing data problems become even more complicated when we consider variation in the missing data mechanism. While most statistical analyses assume data are missing completely at random (MCAR) or missing at random (MAR), this is often not an assumption that can be tested using sample data. In order to determine how robust our methodological choices are under different mechanisms for missingness, the missing data techniques were investigated using data that were MCAR and MAR.

Rubin's Missing Data Taxonomy

When applying a missing data treatment, e.g., MI, the process or mechanism that causes missingness might need to be incorporated into the statistical model because incorporating the missing data mechanism leads to efficient data modeling for fitting the data adequately. In other words, the missing data treatment should reflect the underlying uncertainty of the process or mechanism that best explains why data are missing (Rubin, 2005). Rubin's (1976) theoretical framework is the most widely accepted for missing data research.

Let $Y = (Y_1, Y_2, \dots, Y_p)$ be a random vector variable for p -dimensional multivariate data ($y \times p$ matrix), and θ and δ are the parameter of the data and the parameter of the conditional probability of the observed pattern of missing data, respectively. In the presence of missing data, ϕ , inferences about θ are conditional on δ , the mechanism that causes missingness. That is, given the data at hand, the purpose of the missing data process is to allow making valid inferences about θ but these inferences depend upon or are conditional on δ , the mechanism that generated missing data.

Missing Completely at Random (MCAR). When the probability of a missing value on a variable is independent of the observed or unobserved value for any variable, that is, the probability that y_i is missing is independent of the probability of missingness for any y_j , data are missing completely at random or MCAR.

Missing at Random (MAR). When the probability that y_i is missing for variable Y is dependent on the data of any other variables but not on the variable Y of interest, data are missing at random or MAR. In addition, MAR requires data on a variable to be missing randomly within subgroups (Roth, 1994).

Missing not at Random (MNAR). When the probability of a missing value depends on unobserved data or data that could have been observed, data are missing not at random or MNAR. That is, the missing value of y_i depends on the value of y_i . Under this scenario, why data are missing is not ignorable, thus requiring that the missing data mechanism is modeled to make valid inferences about the model parameter.

Treatment of Missing Data

A variety of methods have been proposed for the treatment of missing data (Roth, 1994; Schafer & Graham, 2002). This research compares two of the most popular methods: Multiple imputation and listwise deletion. Multiple imputation (MI) creates m imputed data sets for an incomplete p -dimensional multivariate data matrix. That is, missing values are replaced with a set of plausible values that represent a random sample and thus characterize the uncertainty about the missing values. Each imputed data set is analyzed independently and the results are combined to provide parameter estimates (point and interval estimates) that incorporate the uncertainty resulting from the missing data.

In contrast, listwise deletion (LW) of cases with missing values simply deletes any observations with one or more missing values and provides a single analysis of the remaining data. This method of missing data treatment is the common default in most software packages.

PURPOSE

The purpose of this study was to investigate the impact of missing data treatment (listwise deletion and MI), when the data are missing completely at random (MCAR) and missing at random (MAR), in the context of multiple regression analysis of complex sample data (Little & Rubin, 1987; Rubin, 1987; Collins, Schafer, & Kam, 2001; Schafer & Graham, 2002). Data that are MCAR represent missingness that is related to neither the variable presenting missing data nor other variables in the analysis. In contrast, MAR represents missing data that are unrelated to value of the variable presenting missingness, but are related to other measured variables.

METHOD

For this Monte Carlo study, complex samples were generated from multivariate populations and each sample was analyzed using both listwise deletion and MI approaches. The sample simulation included both stratification and cluster sampling. Specifically, observations from ten strata were simulated in which each stratum differed in population means on all variables, with the maximum difference in stratum means being twice as large as the between PSU standard deviation. From each stratum, the sampling of PSUs and subsequently, observations within PSUs was simulated, controlling the relative variance between and within PSUs to produce target values of the intraclass correlation.

The Monte Carlo study included four factors in the design. The number of PSUs sampled from each of the ten strata was linked with the number of observations sampled from each PSU to provide low density (100 PSUs per stratum with 10 – 30 observations per PSU) and high density (20 PSUs per stratum with 50 – 150 observations per PSU) samples. To obtain realistic samples in the Monte Carlo study, the number of observations per PSU was a random factor in the simulations. This combination of the number of PSUs with the average sample size per PSU provided consistent overall sample sizes across these two factors (i.e., a mean of 2000 observations per stratum for an average total sample size of 20,000 for each complex sample). In addition to the number of PSUs and the sample size per PSU, the intraclass correlation was manipulated to investigate the effects of different degrees of observation

clustering. Three levels of intraclass correlation were simulated (.00, .05, and .10) by controlling the ratio of the between PSU variance to the within PSU variance. Finally, four levels of missing data were simulated: 10%, 30%, 50%, and 70%. Within each of these levels, 50% of the missing data were selected at the observation level and 50% at the PSU level. Through this process, not only were entire PSUs completely removed from the simulated samples when listwise deletion was used, but the structure of the remaining PSUs was also altered. For example, some of the remaining PSUs lost some, but not all, observations, thus resulting in a reduced clustering effect, while some PSUs retained their original structure and number of observations. Two missing data mechanisms were simulated to produce data that were either MCAR or MAR (Little & Rubin, 1987; Schafer & Graham, 2002).

Within each PSU, multivariate normal data were generated using a correlation matrix derived from an actual matrix obtained from the NELS-88 survey (National Center for Educational Statistics [NCES], 2007). Specifically, the intercorrelations between eight predictor variables were taken directly from the NELS-88 results. Zero-order correlations with a hypothetical criterion variable were calculated so that the predictors would provide a range of effect sizes in the eight predictor regression equation. Two predictors were generated to provide small (X7 and X8), medium (X1 and X2), and large (X3 and X4) effect sizes, respectively, as well as two predictors (X5 and X6) which were approximately null (i.e., regression coefficients were generated to be practically zero) in the multiple regression equation. The correlation matrix used in the simulations is provided in Table 1. Observations within each sample were weighted so that the sample weight was proportional to the inverse probability of selection (taking into account the probability of PSU selection from the stratum and observation selection from the PSU) and the sample weights were incorporated in subsequent analyses.

	Y	X1	X2	X3	X4	X5	X6	X7	X8
Y	1.00000								
X1	0.29354	1.00000							
X2	0.28902	0.03716	1.00000						
X3	0.33003	-0.02342	-0.08097	1.00000					
X4	0.42926	0.02039	0.05139	-0.15033	1.00000				
X5	0.17179	0.04689	0.07601	-0.14001	0.40799	1.00000			
X6	0.05367	0.07268	0.11877	-0.21079	0.16350	0.25853	1.00000		
X7	0.10842	0.09224	-0.06382	0.11601	-0.05750	-0.10975	-0.20160	1.00000	
X8	0.15151	0.05810	0.21698	-0.13668	0.10849	0.17502	0.34115	-0.21985	1.00000

Table 1. Correlation Matrix Used as Template for Data Simulation

For both the MCAR and MAR design factors, the data generation was conducted using SAS/IML version 9.4, and each sample for both missing data factors was analyzed separately, using both the listwise deletion approach and MI in SAS (SAS Institute, 2004). The available packaged procedure for complex sample survey analysis (i.e., PROC SURVEYREG in SAS) used the Taylor Series approximation to estimate the sampling variances (Kiecolt & Nathan, 1985). Conditions for the study were run under Windows and Unix platforms. Normally distributed random variables were generated using the RANNOR random number generator in SAS. A different seed value for the random number generator was used in each execution of the program. The program code was verified by hand-checking results from benchmark datasets.

For each condition investigated in this study, 5,000 samples were generated. The use of 5,000 estimates provides adequate precision for the investigation of the sampling behavior of point and interval estimates of the regression coefficients. For example, 1,000 samples provide a maximum 95% confidence interval width around an observed proportion that is .03 (Robey & Barcikowski, 1992). The outcomes of interest in this simulation study included both point estimates (the bias and sampling error of the regression coefficients) and interval estimates (confidence interval coverage and width for the coefficients).

Multiple imputation using SAS package involves three steps: (1) imputing missing data using PROC MI, (2) analyzing the imputed dataset using standard statistical procedures, and (3) pooling the results using PROC MIANALYZE. In this study, the PROC MI creates multiply imputed data sets for the simulated incomplete multivariate data. The following statements invoke the MI procedure and impute missing values for the “miss” data set. The SAS variable ‘decode’ is a categorical variable that represents the stratum number and PSU as a single composite variable. The FCS statement requests multivariate imputation by fully conditional specification methods. The five imputed datasets are concatenated and stored in the dataset named “mi_out” with the variable –Imputation– indicating the imputation numbers.

```
proc mi data = miss out = mi_out;
  class descodes;
  FCS;
  var y x1 - x8 descodes new_wt;
run;
```

In the 2nd step, the SURVEYREG procedure was used to do regression analysis on the imputed sample data. The PROC SURVEYREG can handle complex survey sample designs with stratification, clustering, and unequal weighting. It computes regression coefficients and provides significance tests for any specified estimable parameters. The WEIGHT statement is used to correctly weigh the population estimates; if an observation has a missing value or a nonpositive value for the WEIGHT variable, then PROC SURVEYREG excludes that observation from the analysis. An observation is also excluded if it has a missing value for any STRATA variable, CLUSTER variable, dependent variable, or any variable used in the independent effects. PROC SURVEYREG uses the Output Delivery System (ODS) to create output data sets. In the below code snippet, the BY _IMPUTATION_ statement was used to request separate analyses of each MI _IMPUTATION_ repetition.

```
proc surveyreg data = mi_out;
  cluster school_id;
  strata stratum;
  weight new_wt;
  model y = x1 x2 x3 x4 x5 x6 x7 x8 / clparm deff;
  by _imputation_;
  ods output parameterestimates = complete;
run;
```

The MIANALYZE procedure combines the results of analyses of imputations and generates statistical inferences. The MIANALYZE procedure reads parameter estimates and associated standard errors or covariance matrix that are computed by the PROC SURVEYREG for each imputed data set. The MIANALYZE procedure then derives valid univariate inference for these parameters. In the following code snippet, The EDF option was used to ensure a correct determination of the complete data degrees of freedom used in the construction of MI confidence intervals for the population statistics of interests ($df = \#PSUs - \#Strata$).

```
proc mianalyze data = complete EDF = &error_df;
  modeleffects estimate;
  stderr stderr;
  ods output parameterestimates=mianalyze_out;
run;
```

RESULTS

MCAR

Across the MCAR conditions, the listwise deletion point estimates showed virtually no statistical bias, and the estimated bias for listwise deletion and for the complete data samples were nearly identical; however, the MI point estimates revealed both positive and negative bias, with the negative bias showing greater variability (Figure 1). In addition, the RMSE estimates associated with the listwise deletion method were small and close to those associated with the complete data samples while the RMSE associated with MI were much larger than RMSE associated with listwise deletion and varied substantially across the simulation conditions (Figure 2).

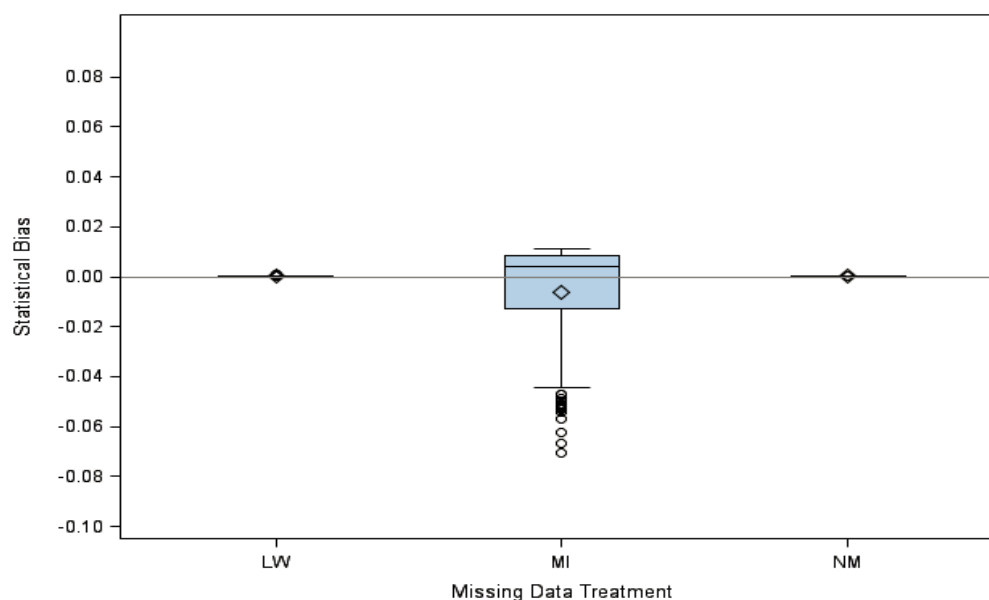


Figure 1. Distributions of Bias by Missing Data Treatment Method for MCAR

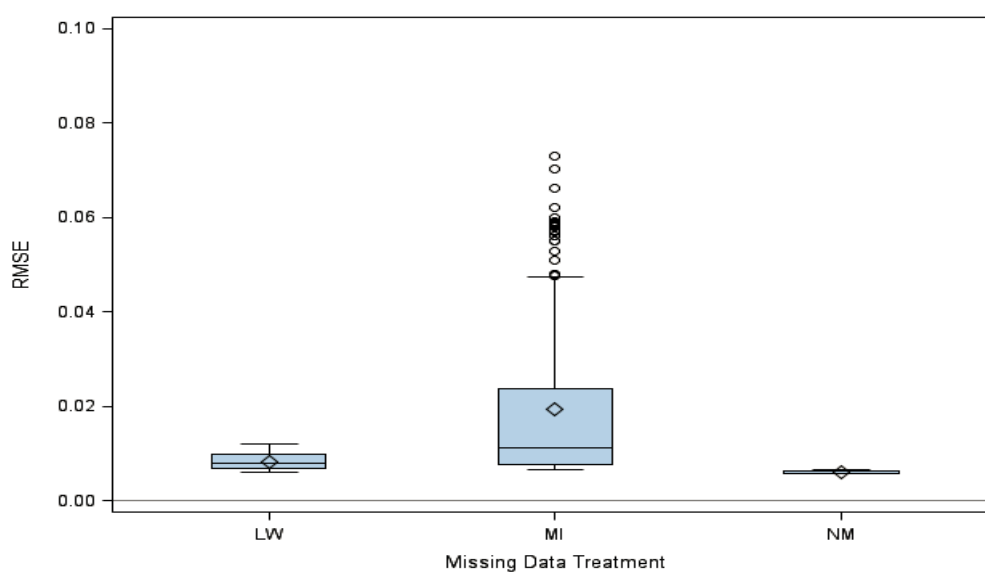


Figure 2. Distributions of RMSE by Missing Data Treatment Method for MCAR

An eta squared analysis revealed that differences among the parameter estimates for the eight regressors accounted for most of the variability in bias and in RMSE for MI (η^2 for bias = 95%; η^2 for RMSE = 93%), and the rest of the variability in bias and in RMSE were mostly accounted for by the first-order interaction between parameter and the intraclass correlation between the between cluster variance and total variance. Parameters associated with X7 and X8, the two variables that had missing data, produced a large negative bias while parameters associated with variables that had no missing data produced a small positive bias (Figure 3). In addition, for variables without missing data, intraclass correlation appeared to have no impact on the point estimate bias and RMSE, but for the two variables with missing data, the higher level of variability between the clusters, the smaller the bias (Figure 4) and the smaller RMSE (Figure 5).

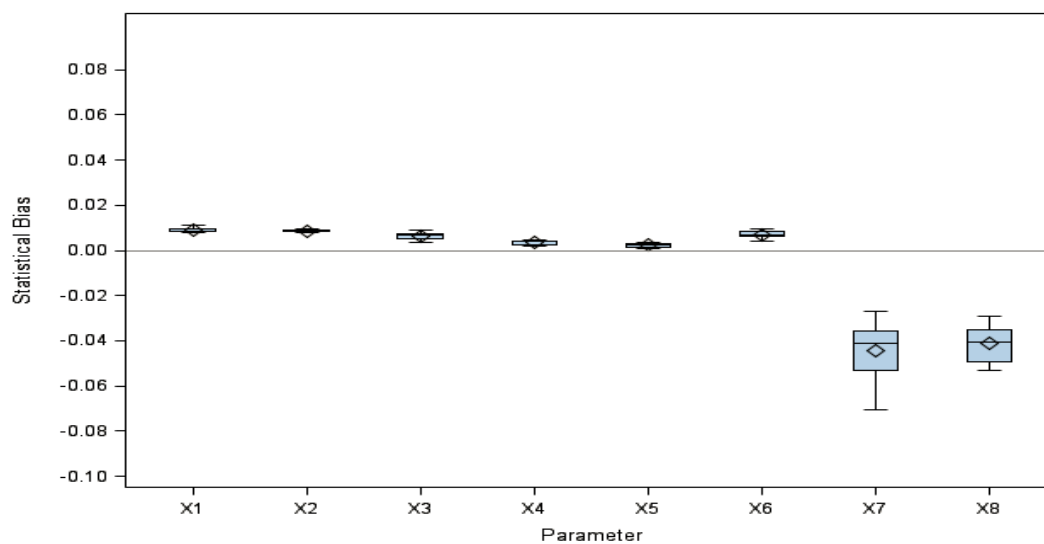


Figure 3. Distributions of Bias Estimates for Multiple Imputation by Parameter for MCAR

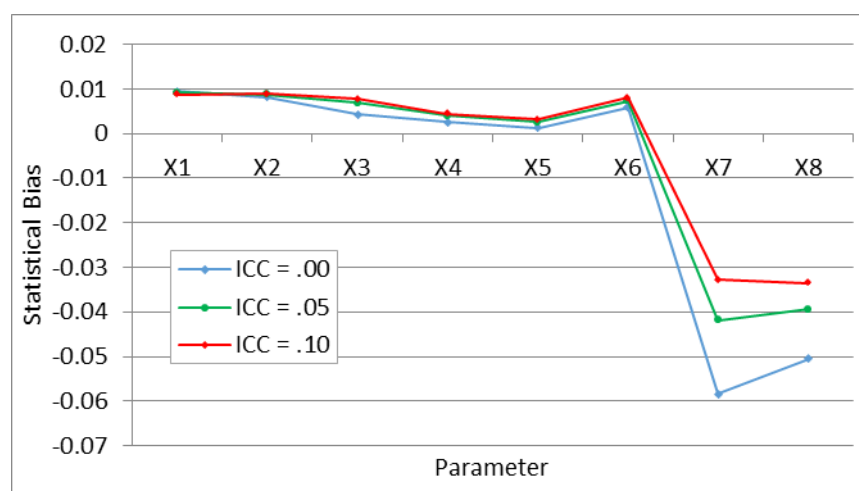


Figure 4: Bias Estimates for Multiple Imputation by Parameter and ICC in MCAR

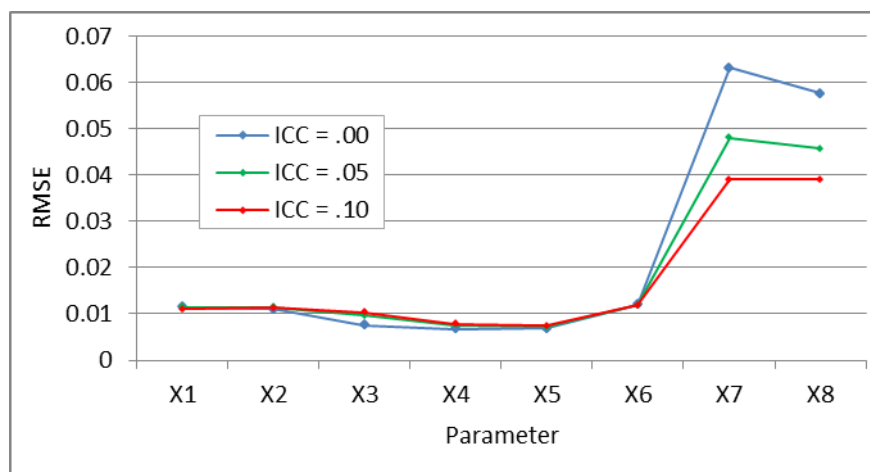


Figure 5: RMSE Estimates for Multiple Imputation by Parameter and ICC in MCAR

Confidence interval coverage and confidence interval width estimates produced by listwise deletion also showed a superior performance compared to the MI method. While confidence interval coverage estimated by listwise deletion is around 95% (similar to that estimated from the complete samples), MI confidence intervals undercovered the regression parameters by a considerable amount, and the coverage varied substantially across simulation conditions (Figure 6). In addition, the confidence interval width for MI was much larger than the confidence interval width for listwise deletion (Figure 7).

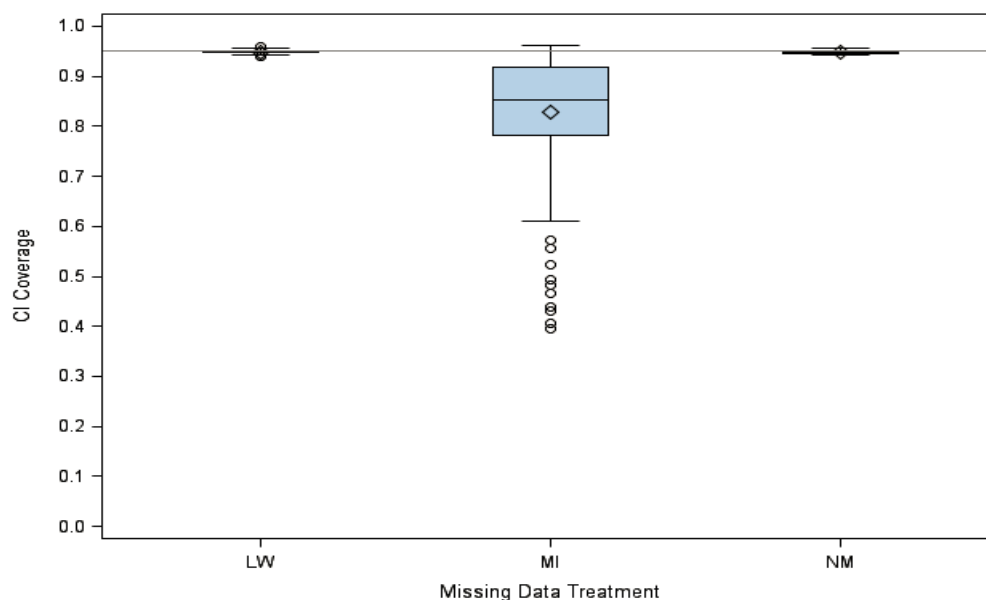


Figure 6. Distributions of Confidence Interval Coverage Estimates by Missing Data Treatment Method for MCAR

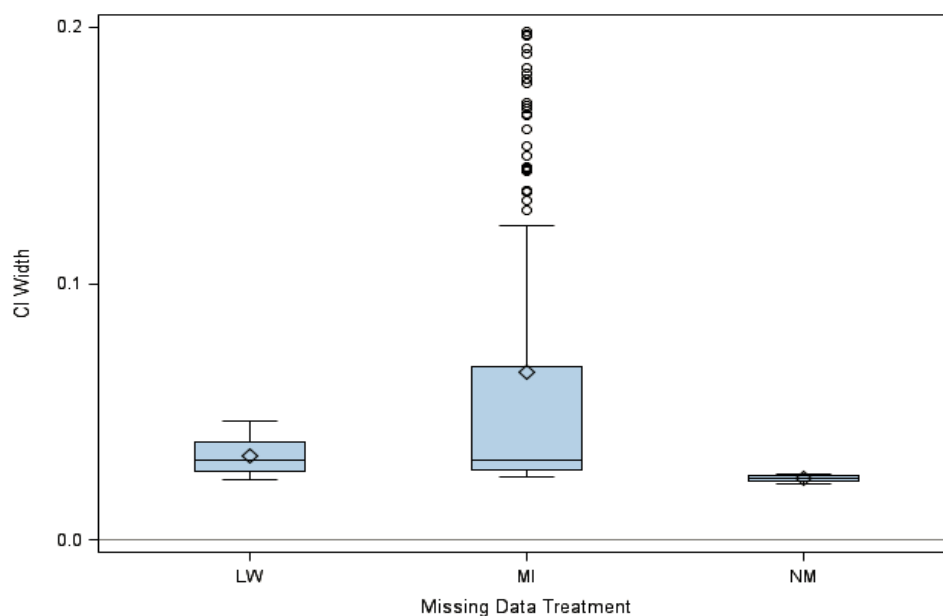


Figure 7: Distributions of Confidence Interval Width Estimates by Missing Data Treatment Method for MCAR

Similar to bias and RMSE, the variability in confidence interval coverage for MI was influenced mostly by the differences among the eight regressors ($\eta^2 = 68\%$). It was also modestly influenced by density level ($\eta^2 = 10\%$) and

the first-order interaction of density level and parameter ($\eta^2 = 13\%$). The estimated probability coverage for each parameter showed undercoverage of all parameter estimates except that for X5, particularly lowest for the parameter associated with X7 (Figure 8).

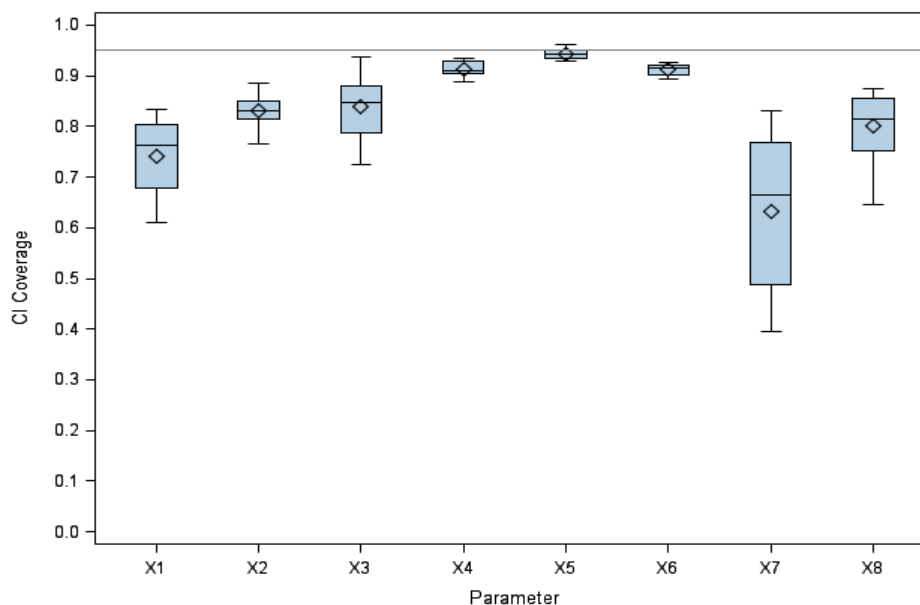


Figure 8. Distributions of Confidence Interval Coverage Estimates for Multiple Imputation by Parameter for MCAR

Regarding to the impact of density level on the CI coverage, high density levels were associated with lower probability coverage for most parameters; and this impact is emphasized in variables with missing data (Figure 9).

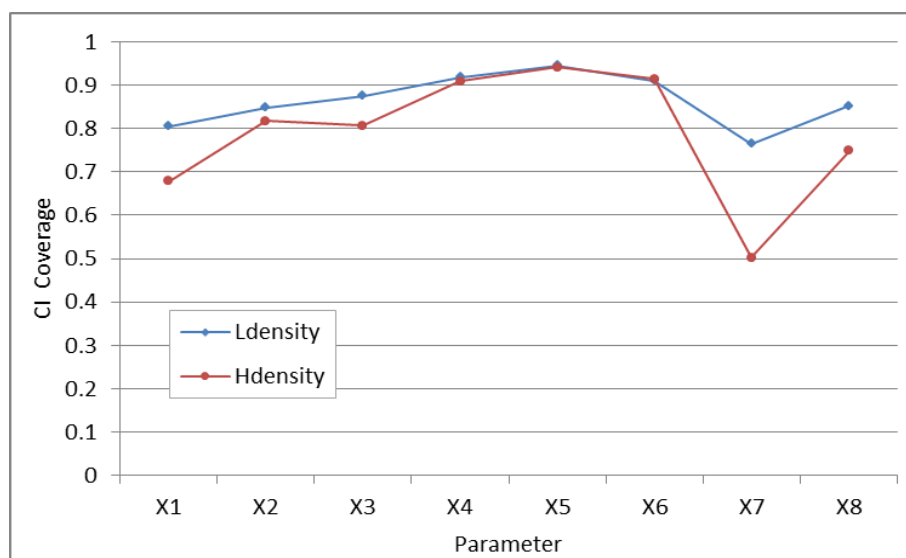


Figure 9. Confidence Interval Coverage Estimates for Multiple Imputation by Density Level and Parameters for MCAR

MAR

Across the MAR conditions, the measures of statistical bias, RMSE, CI coverage, and CI width also showed clear evidence of listwise deletion superior performance compared to MI. Statistical bias obtained from the LW sample data was close to that obtained from the complete sample data (although it also displayed some level of variability for negative bias which was not seen in MCAR); the bias obtained from the MI data was seen in both directions, with the negative bias showing greater variability as seen in MCAR (Figure 10). RMSE obtained from the LW sample data

was also much smaller than that obtained from the MI sample data, but it also showed variety of outliers which were not seen in MCAR (Figure 11).

The confidence interval coverage estimate obtained from the LW sample data for MAR was around 95% while the confidence interval coverage obtained from the MI data undercovered the point estimates as found in MCAR; however, the LW coverage also showed broad variety of undercover outliers which was not seen MCAR (Figure 12). In addition, the LW confidence interval width was much narrower for MI confidence interval width (Figure 13).

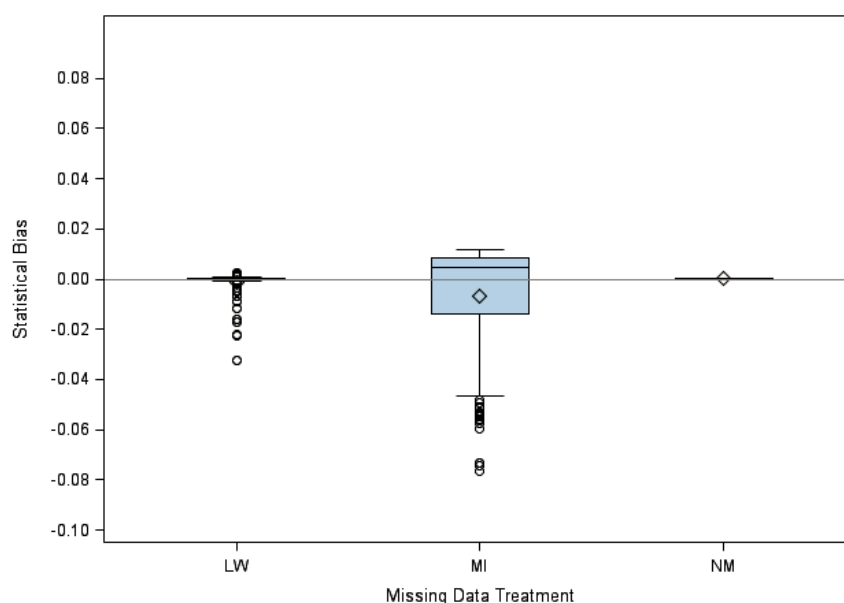


Figure 10. Distributions of Bias by Missing Data Treatment Method for MAR

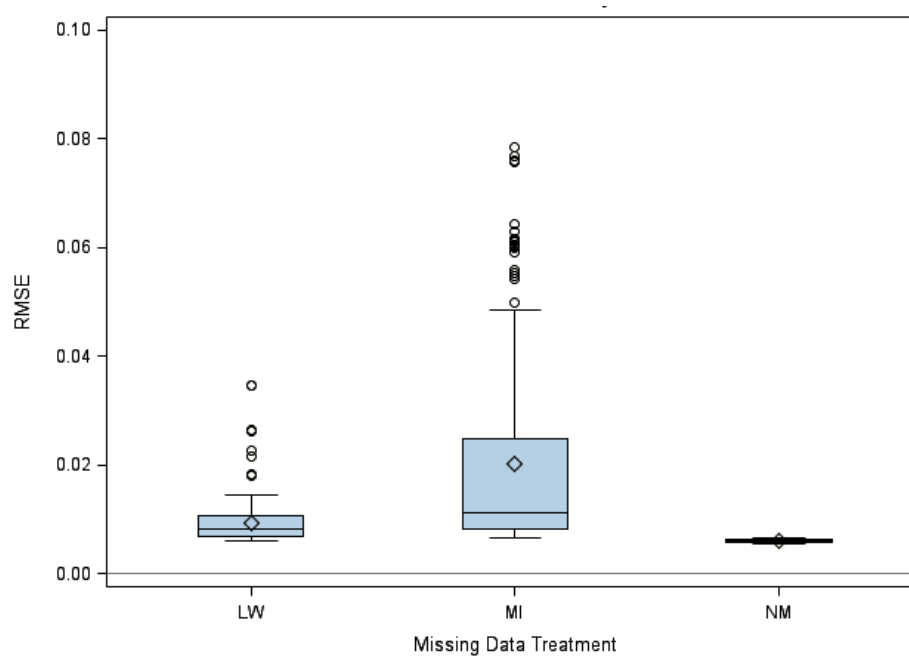


Figure 11: Distributions of RMSE by Missing Data Treatment Method for MAR

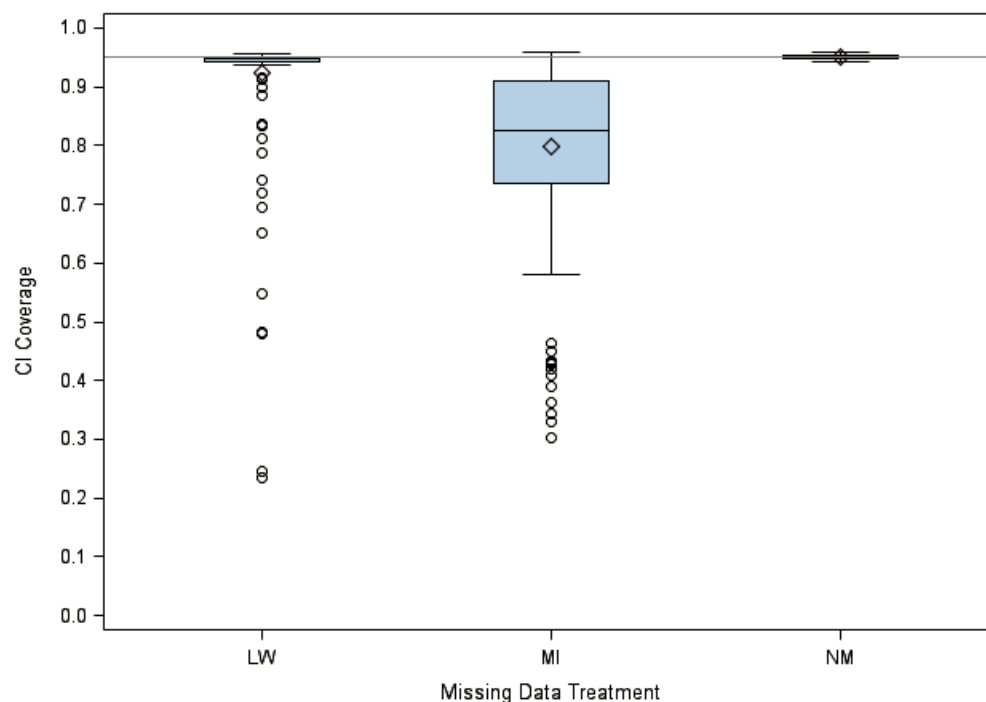


Figure 12. Distributions of Confidence Interval Coverage Estimates by Missing Data Treatment Method for MAR

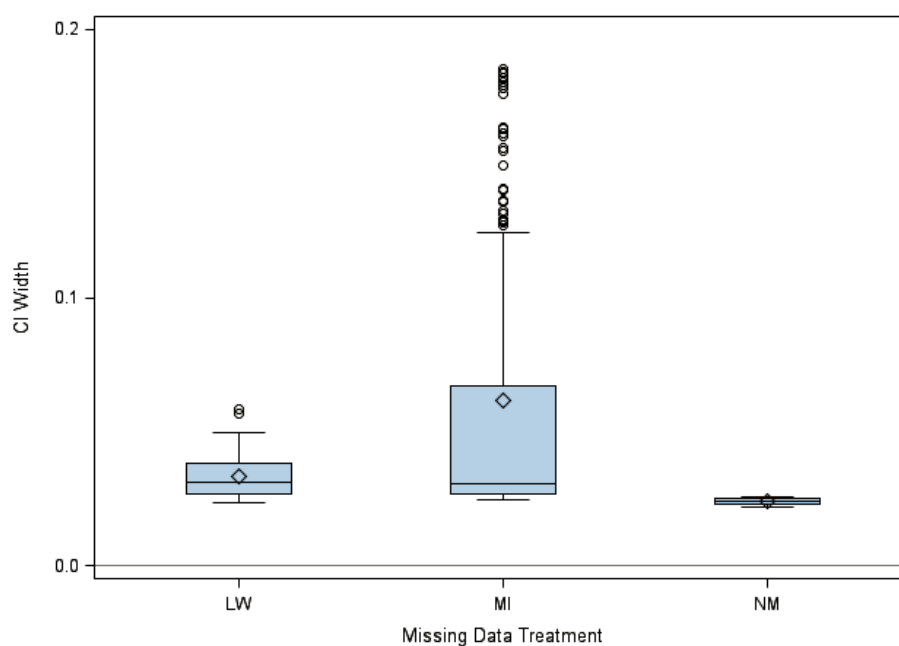


Figure 13. Distributions of Confidence Interval Width Estimates by Missing Data Treatment Method for MAR

CONCLUSION

For both MCAR and MAR conditions, the performance indicators (statistical bias, RMSE, CI coverage, and CI width) showed clear evidence of listwise deletion's superiority compared to multiple imputation. The performance indicators for listwise deletion were very close to those for the complete data samples. The findings contradict the literature that suggests that listwise deletion methods for complex sample data are inappropriate. Specifically, listwise deletion was found to be less biased than multiple imputation for both MCAR and MAR; so if you are a "listwise deleter," no need to feel guilty especially with the popularity of big data, statistical power is no longer a major issue.

REFERENCES

- Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5, 215 – 238.
- Berglund, P., & Heeringa, S. G. (2014). *Multiple imputation of missing data using SAS*. SAS Institute.
- Collins, L. M., Schafer, J. L., & Kam, C. M (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6, 330-351.
- Graham, J. W., Cumsille, P. E. & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Research Methods in Psychology* (pp. 87-114). New York: Wiley.
- Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schanabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201-218). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Graubard, B. I., & Korn, E. L. (1996). Survey inference for subpopulations. *American Journal of Epidemiology*, 144, 102-106.
- Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.
- National Center for Education Statistics. n.d. "Surveys and Programs". Retrieved January 15, 2007, from <http://nces.ed.gov/surveys/nels88/>
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47, 537-560.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
- Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469), 322-331.
- SAS Institute. (2004). *SAS/STAT® 9.4 User's Guide*. Cary, NC: SAS Institute Inc.
- Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33(4), 545-571.
- Schafer, J. L. & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of Complex Surveys: Wiley Series in Probability and Mathematical Statistics*. New York: John Wiley & Sons.
- Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56, 645-646.

CONTACT INFORMATION

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

Anh P. Kellermann
University of South Florida
4202 East Fowler Ave., EDU 105
Tampa, FL 33620
Fax: (813) 974-4495
Email: napham@mail.usf.edu

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.