Journal of Modern Applied Statistical Methods

Volume 9 | Issue 1 Article 8

5-1-2010

The Performance of Multiple Imputation for Likert-type Items with Missing Data

Walter Leite
University of Florida, Walter.Leite@coe.ufl.edu

S. Natasha Beretvas
The University of Texas at Austin, tasha.beretvas@mail.utexas.edu

Journal of Modern Applied Statistical Methods
May 2010, Vol. 9, No. 1, 64-74

1538 – 9472/10/$95.00

The Performance of Multiple Imputation for Likert-type Items with Missing Data

Walter Leite
University of Florida

S. Natasha Beretvas
The University of Texas at Austin

The performance of multiple imputation (MI) for missing data in Likert-type items assuming multivariate normality was assessed using simulation methods. MI was robust to violations of continuity and normality. With 30% of missing data, MAR conditions resulted in negatively biased correlations. With 50% missingness, all results were negatively biased.

Key words: Multiple imputation, missing data, missing at random, Likert-type items

Introduction

Missing values introduce several problems to statistical analyses and researchers have tried many methods to ameliorate these problems. A few popular methods have become popular and have been implemented in statistical software, which has boosted their usage. These methods include listwise deletion, pairwise deletion, mean substitution, regression imputation, maximum-likelihood methods and multiple imputation. Among these procedures, multiple imputation (MI), together with maximum likelihood estimation, is becoming one of the preferred techniques for dealing with missing data; due to its increasing popularity, this study focuses on the performance of MI.

MI was first proposed by Rubin (1987) as a way to handle missing data in public survey datasets. Research about MI in the statistical literature is abundant, however, only a handful of studies apply MI to missing data in datasets consisting of Likert-type items. This may be partially explained by the fact that MI depends on the extensive use of computer software and specialized software has only recently become easily accessible.

The MI method that best fits a set of data depends on the distribution assumed for the variables in the dataset. MI is most often performed under the assumption that the variables are multivariate normally distributed; cases exist, however, where this assumption may not be appropriate. In particular, surveys or scales used in organizational research frequently contain dichotomous or Likert-type items whose responses are not normally distributed. Very little research has been done concerning missing data in Likert-type scales and there are no studies evaluating the use of MI under a multivariate normal model with ordinal variables. Although Schafer (1997) argued that MI under the multivariate normal model is robust to departures from normality, extensive investigation of this issue does not currently exist in the literature. Thus, the objective of this study is to examine the performance of MI with datasets composed of Likert-type items.

Types of Missing Data

The existence of missing data in a dataset can result in loss of statistical power and biased parameter estimates. Causes of missing values in data are varied, for example: the refusal of some subjects to answers certain questions, data-entry errors and attrition (Little & Rubin, 1989). Missing data can be classified

Walter Leite is an Assistant Professor at the Research and Evaluation Methodology program in the College of Education. His research addresses structural equation modeling, longitudinal studies, causal inference and missing data. Email: Walter.Leite@coe.ufl.edu. S. Natasha Beretvas is an Associate Professor and Chair of the Quantitative Methods program in the Educational Psychology Department at UT, Austin. Her research focuses on extensions to the multilevel model for handling complex data structures and on meta-analytic techniques. Email: tasha.beretvas@mail.utexas.edu.

according to its pattern within the dataset. Little and Rubin (1989) adopted four classifications for patterns of missing data: general pattern of missingness, univariate missing data, unit non-response and monotone missing data. A general pattern of missingness is characterized when values are missing in many variables without any special arrangement. If the data are missing in just one variable of the dataset, the missing data are univariate. Unit non-response is a pattern where a block of variables has missing values for the same set of cases, but data for those cases for all other variables is complete. Monotone missing data describe a pattern where complete cases in a variable that has X missing values will also be complete in a variable that has $(X - 1)$ values.

Whether a procedure to deal with missing data will result in unbiased estimates of parameters depends on the relationships between the missing values, the incomplete variable and the other variables in the dataset. These relationships allow classification of missing data into three types (Rubin, 1976; Little & Rubin, 1987; Sinharay, Stern & Russell, 2001) commonly referred to as missing data mechanisms: data missing completely at random (MCAR), data missing at random (MAR) and data missing not at random (MNAR) or non-ignorable missingness.

Data are MCAR for a variable X when the missing values in this variable are independent of both the variable X and the other variables in the dataset. In this case, the observed variables can be considered a random sub-sample of the hypothetical complete data. Missing values for a variable are considered MAR when they depend on the other variables in the dataset, but not on the variable itself. MNAR or non-ignorable missingness occurs when the probability of the missing values for a variable X is dependent on the underlying value of X (Little & Rubin, 1987; Sinharay, Stern & Russell, 2001).

Multiple Imputation

The most common procedure to deal with missing data is deleting cases with incomplete data, called listwise deletion. However, listwise deletion results in unbiased parameter estimates only when (1) the data can be assumed MCAR, and (2) the fraction of missing data is very small (e.g., 5%) (Graham & Hofer, 2000). Other methods, such as person and item mean imputation, hot-deck imputation (Huisman, 2000), regression imputation and the expectation maximization algorithm (EM) can be used with MAR data, but they reduce the variability of the dataset and produce artificially small standard errors (McDonald, Thurston & Nelson, 2000) Among the many procedures that have been developed to cope with missing data, full-information maximum likelihood estimation and multiple imputation (MI) are the most sophisticated methods, and are also the ones likely to yield the least biased parameter estimates (Sinharay, Stern & Russell, 2001; Graham & Hofer, 2000).

The results of a missing-data procedure are affected by the type of missingness (MCAR, MAR or MNAR) and also by characteristics of the sample and variables being analyzed. These characteristics include sample size, scale of measurement of the variables, range of data points and distribution of the observed variables. In the case where the dataset contains scores of a psychometric scale, the reliability and validity of scores on the instrument are also important (Raaijmakers, 1999).

The MI method consists of creating a vector of possible values for every missing value in the database. It represents a step forward from regression-based single imputation and the EM algorithm because the multiply imputed values reflect the uncertainty of the imputed values. MI restores two sources of variability: the variability of each variable and the variability of the sample covariance matrix. The variability of each variable is restored because the imputed values do not fall exactly on the regression line. This is accomplished by adding error variability to the imputed missing values. These errors are sampled from the distribution of known errors. The variability of the sample covariance matrix is restored by sampling many covariance matrices from a simulated population. Due to the restoration of these sources of variability, the resulting imputed values will include a component of within-imputation and a between-imputation variance.

Schafer (1997) developed methods to execute MI by cycling through two steps. In the

first step, missing values are imputed, and in the second step unknown parameters are estimated. After the second step, the estimated parameters are used to impute missing values and the cycle is repeated until reaching a criterion of convergence. The process begins with an initial estimate of the parameters given by the expectation maximization (EM) algorithm. Schafer (1997) calls the first step of MI the I (imputation) step. It consists of replacing missing data points by randomly drawn values from the conditional distribution of the missing data given the observed data and the parameter estimates provided by the EM algorithm. The second step is termed P (posterior) and consists of estimating parameters. The estimated parameters are then used in another I step, and this process is repeated until the distribution of covariance matrices stops changing substantially. The EM algorithm is used to calculate the initial parameter estimates for the first imputation step. After this initial estimate, missing values are imputed and parameters are estimated using the data augmentation method.

Data augmentation is an iterative procedure that imputes missing data under assumed values of the parameters and then draws new parameters from a posterior distribution based on the complete data (Schafer & Olsen, 1998). This process of imputing values and estimating parameters creates a Markov chain. When the Markov chain stabilizes, the data augmentation process has reached convergence. This state is characterized by a stable distribution of parameters. After convergence, multiple imputations are generated based on independent draws from this distribution. Any number of imputed data-sets can be obtained by repeating the data augmentation algorithm; consequently, each set of imputed values will be different from the others.

MI has been shown to depend on three assumptions to generate unbiased parameter estimates. The first assumption specifies what types of missing data can be addressed using MI. The other two assumptions are necessary due to the Bayesian nature of MI. The first assumption of MI is that the data are MCAR or MAR. This assumption is important because using MI with MNAR data may result in biased parameter estimates (Little & Rubin, 1989). The second assumption is the prior distribution; because MI is a Bayesian method, a prior distribution is used to represent the state of knowledge about the data before it is available. Usually a non-informative prior (Sinharay, Stern & Russell, 2001) is chosen corresponding to ignorance about the distribution of the data. Such a prior is ambiguous as to the location of the likelihood's maximum, allowing a wide range of values. In some cases, it is adequate to specify an informative prior distribution. This distribution is chosen from a family of distributions and it is combined with the likelihood to generate a posterior distribution from the same family. These assumptions are essential because multiple imputations involve random draws from the posterior probability distribution of the unknown parameters given the observed values.

Finally, MI requires an assumption about the complete-data model. Each multiple imputation method uses a specific probability model to generate the imputed values. The distribution of the observed values should match this imputation model. MI software usually uses the multivariate normal model to impute numeric data and the loglinear model for categorical data. The multivariate normal is the most common model for multivariate statistical analysis. Schafer (1997) argues that the normal model is robust to departures of normality when the proportion of missing data is not large. The reason for this robustness is that the model only affects the missing values, leaving the observed values unchanged. In addition, Schafer & Olsen (1998) indicate that it is often acceptable to impute values of categorical variables under the normality assumption and round off the continuous imputed values to the nearest category.

MI allows the researcher to improve the quality of the imputed values by using information from variables that predict the missing values or correlate with the variables containing missing values. These variables may be of no interest for the data analysis itself, therefore, they can be included in the dataset during the multiple imputation procedure and then excluded in the data analysis. The variables that may help with the imputation process can be detected through an examination of correlations

and contingency tables between these variables and the variables that have missing values. However, the inclusion of an exaggerated number of variables may result in multicollinearity problems and variance inflation (Wayman & Swaim, 2002).

Five to ten imputations are typically recommended because this number has been found to provide adequate estimates (Rubin, 1987; Collins, Schafer & Kam, 2001). After multiple imputed datasets are obtained the analysis of interest to the researcher should be conducted with each imputed data set. For example, a researcher might be interested in conducting a multiple regression analysis. Assume the researcher obtained ten multiply imputed datasets containing imputations replacing each missing value. The researcher would run the regression analysis using each data set, and the resulting parameter estimates (the regression coefficient estimates, for example) can then be combined across the m = 10 imputed datasets to obtain the single best estimate of the relevant parameter (Rubin, 1987). Specifically, the mean of the parameter estimates across the m imputed datasets, $\overline{q}$, can be calculated as:

$$\overline{\mathrm{q}} = \frac{1}{\mathrm{m}} \sum_{\mathrm{i=1}}^{\mathrm{m}} \hat{\mathrm{q}}_{\mathrm{i}} \qquad (1)$$

where $\hat{q}_i$ is the parameter estimate from the i^{th} imputed dataset and m represents the number of imputed datasets being combined.

To calculate the variance of each parameter estimate, two sources of variability should be combined (Schafer & Olsen, 1998): the variability within and between imputed datasets. The within-imputation variance, $\overline{u}$, is the mean of the variance estimates from each imputed dataset:

$$\overline{\mathrm{u}} = \frac{1}{\mathrm{m}} \sum_{\mathrm{i=1}}^{\mathrm{m}} \hat{\mathrm{u}}_{\mathrm{i}} \qquad (2)$$

where $\hat{u}_i$ is the variance estimate for the relevant parameter estimated for imputed dataset i. In the example described in which multiple regression analyses were conducted with each of the ten imputed data sets, the square of the standard error estimated for one of the predictor's unstandardized regression coefficients for imputed dataset i would provide that imputation's $\hat{u}_i$.

The between-imputations variance, B, is the variance of the parameter estimates across the set of imputations:

$$B = \frac{1}{(m-1)} \sum_{i=1}^{m} (\hat{q}_i - \overline{q})^2 \qquad (3)$$

where $\hat{q}_i$ represents the parameter estimate for imputation i.

The total variance, T, associated with the multiply imputed parameter estimate, $\overline{q}$, is the sum of the within- and the between-imputations variances. This sum is corrected to account for the simulation error in $\overline{q}$ (Schafer & Olsen, 1998) using the formula:

$$T = \overline{u} + \left(1 + \frac{1}{m}\right) B \qquad (4)$$

This total variance provides the advantage of MI over other methods for dealing with missing data. The within-imputations variance component represents sampling variability while the between-imputations variance represents missing data uncertainty. These two components prevent the missing values from creating an artificial precision in the parameter estimates, resulting in negatively biased standard errors and associated test statistic p-values that are too low (Schafer, 1997).

Recently, many computer programs have become available to perform MI (e.g., NORM, S-Plus, R, SAS). NORM 2.02 is a stand-alone multiple imputation program developed by Schafer (1999) that executes MI under the multivariate normal model. The freely-available R software (R development core team, 2008) contains the norm library, which is an implementation of MI similar to the NORM software. Different implementations of MI in the R software can be found in the CAT, Mix,

Amelia and Mice packages. S-Plus (Insightful Corp., 2001) has a library that performs MI under the Gaussian, loglinear and conditional Gaussian models. The statistical package SAS Version 8.2 incorporated functions for MI but it has the disadvantage of allowing little control over the imputation model (Horton & Lipsitz, 2001).

Multiple Imputation of Likert-Type Items

Little research has been conducted concerning missing data in Likert-type scales. For example, Downey and King (1998) investigated missing data in Likert-type variables but only evaluated mean substitution methods (person mean and item mean). Roth, Switzer and Switzer (1999) investigated missing data in multiple item scales, but only examined listwise deletion, regression imputation, hot-deck imputation, person mean substitution and item mean substitution.

MI has been most frequently conducted under the assumption that the variables are multivariate-normally distributed. However, surveys and scales commonly contain non-normally distributed Likert-type items, whose distributions may only approximate normality. Although Schafer (1997) developed a MI method for categorical data based on the loglinear model, he argued that multivariate normal MI could be used for categorical variables. However, evaluation of this claim has yet to be conducted. If MI, under the assumption of normality, works sufficiently well with typically non-normal Likert-type (ordinal) variables/items, the analysis of this type of data would be simplified.

Methodology

The performance of MI wasassessed using simulation methods assuming multivariate normality in the commonly occurring scenario in which some of the responses to Likert-type items are missing. The impact of the following factors on the performance of MI were assessed: the underlying distribution of the item responses (normal versus non-normal), the magnitude of the variables' inter-correlations ($\rho = 0.2$, $\rho = 0.8$), the bluntness of the categorization of the data into discrete item scores (3, 5 and 7), the missing data mechanism (MCAR and MAR) and the degree of missingness (10%, 30% and 50%). Recovery of the true correlations will be used in the evaluation of MI's performance.

Responses to a set of 10 items were generated to fit either multivariate normal or non-normal distributions with a known correlational structure. To simplify the generating correlation matrix, each variable was modeled to have the same correlation with each of the others (0.8 or 0.2). Next, each interval-scaled item score was discretized to match the Likert-scale format of relevance to the condition (3, 5 or 7), and the condition's pattern (MCAR or MAR) and degree of missingness were built into the generated data. Three degrees of missingness were investigated (10%, 30%, 50%) and MI was used to impute missing data. For each iteration (and condition), the imputed datasets were summarized using Equations 1 - 4 to assess recovery of the generating correlation values. Due to their importance in methods such as multiple regression and factor analysis, correlations were the parameters of interest in this study.

Simulation of Item Data

The software, *S-Plus* (Insightful, 2001) was used to conduct the simulation. To represent items on a 10-item scale or survey, 10 continuous random variables were generated with normal and non-normal distributions. Each variable was sampled from a multivariate normal distribution with a mean of zero and standard deviation of one. The multivariate normal random values were created using the function RMVNORM of S-Plus, which generates pseudo-random numbers given a correlation matrix, vector of means and standard deviations and a random seed.

The skewness and kurtosis was introduced into the data using the method originated by Valle and Maurelli (1983), which produces multivariate non-normal distributions with a given value of skewness and kurtosis by combining Kaiser and Dickman's method (1962) with one proposed by Fleishman (1978) to simulate univariate non-normal distributions with specified degrees of skewness and kurtosis. Fleishman's method uses the transformation

$$Y = a + bX + cX^2 + dX^3$$

where *a*, *b*, *c*, and *d* are constants, to convert variable X into variable Y with the desired degree of skewness and kurtosis. Fleishman (1978) provides equations and tables detailing values for these constants along with their associated skewness and kurtosis levels. When applied together, Kaiser and Dickman's and Fleishman's methods interact such that the correlations between the simulated non-normal variables differ from those specified in the population correlation matrix. Vale & Maurelli (1983) solved this problem by adjusting the values of the population correlations using the formula:

$$\begin{aligned} r_{y1y2} = \rho_{x1x2}(b_1b_2 + 3b_1d_2 + 3d_1b_2 + 9d_1d_2) \\ +\rho^2_{x1x2}(2c_1c_2) + \rho^3_{x1x2}(6d_1d_2) \end{aligned} \quad (5)$$

where ρ_{x1x2} is the population correlation between variables X_1 and X_2, r_{y1y2} is the adjusted correlation between the non-normal variables Y_1 and Y_2, and b_1, b_2, c_1, c_2, d_1, d_2, are Fleishman's coefficients for Y_1 and Y_2. After adjusting the population correlations, non-normal random variables are obtained by first executing Kaiser and Dickman's method and then using Fleishman's method. The resulting variables will have the desired degrees of skewness, kurtosis and inter-correlations. However, with Likert-type variables this method has the limitation that the transformation of continuous variables into categorical variables results in a slight change of the degrees of skewness and kurtosis originally simulated.

For each of the conditions, 1,000 samples of 400 cases were generated and the variables were converted into Likert-type scores. Datasets with three types of Likert-type items were created (with scales ranging from 1 to 3, 1 to 5, and 1 to 7) by dividing the total range of the scores into *k* segments of equal size, where *k* is the desired number of categories. This resulted in discrete distributions that better approximated the shape of their continuous, generating distributions. The correlation matrices for each replication sample and condition were also calculated to allow an assessment of the change resulting from the categorization process and to serve as a baseline for later evaluations.

Simulation of Missing Data

Two types of missing data were introduced: MCAR and MAR. Three overall proportions of missing values were simulated (10%, 30% and 50%). MCAR missing data was obtained through random deletion of values from the datasets. To simulate the MAR condition, one variable in the dataset, *Z*, was used to predict the missing values in the other nine variables. The predictor *Z* was the only variable in the dataset with no missing values. Data points were deleted according to the MAR-linear condition described by Collins, Schafer and Kam (2001). In the MAR-linear condition (perhaps better described as monotonically increasing rather than linear), the proportion of missing values is approximately linearly related to the value of *Z*. To simulate this condition, the cases were grouped according to the value of *Z*, and subgroups of cases with larger values of *Z* were assigned a higher probability of being missing.

Analyses

Values for the missing data were imputed assuming the multivariate normal model using the functions of the missing library (Schimert, et al., 2000) implemented in S-PLUS version 6.0 (Insightful, 2001). Ten imputations were created for each dataset and the correlation between each pair of variables was calculated for each imputed data set. When correlation estimates are the unit of analysis, Fisher's (1928) normalizing and variance-stabilizing *r*-to-Z_r transformation is frequently used to correct the non-normality of the sampling distribution of *r*. This transformation was used; specifically, each correlation was transformed to a Z_r using the formula:

$$Z_r = (1/2)\ln\left[\frac{1+r}{1-r}\right]. \quad (6)$$

These 10 transformed correlations (one per imputation) for each pair of variables were combined using Rubin's (1987) rules as outlined in Equation 1 to provide an overall transformed correlation estimate, $\overline{q}$ combined across imputations for each sample and condition. (This was repeated for each of the transformed correlations between the variables). In addition, the between-imputations variance, *B*, of the transformed correlation estimates (see Equation 3) was also calculated for each multiply imputed estimate, $\overline{q}$.

The criterion used to judge the performance of MI involved an assessment of the recovery of the correlations (conducted using the transformed correlations). While the original generating value for the correlations was either 0.2 or 0.8, this value applied only to the continuous distributions. It should be noted that, for the non-normal distributions, although data were transformed to have a slight degree of kurtosis and skewness, the transformations were chosen to maintain the generating correlation values. However, the categorization of the continuously scaled scores into ordinal-scaled data resulted in correlations between pairs of variables that differed from the original generating values. The values of the correlations were compared after categorization - but before missingness had been introduced - with the correlations estimated after MI had been used to compensate for the missingness. The correlations after categorization were transformed using Fisher's r-to-Z_r transformation to provide the average of the sampling distribution of Z_rs for categorized variables. For each dataset simulated, the Z_r values calculated after MI were compared with the values describing the categorized distributions without missingness.

The comparisons were performed using relative bias averaged across replications. The relative bias (Hoogland & Boomsma, 1998) compares the average value of the parameter estimated Z_r with the population value, ζ_ρ, using the formula:

$$B(\hat{Z}_r) = \frac{\bar{\hat{Z}}_r - \zeta_\rho}{\zeta_\rho}. \qquad (7)$$

The relative bias of the parameter estimate was considered acceptable if its magnitude was less than 0.05 (Hoogland & Boomsma, 1998).

Because one of the benefits of using MI is that it provides better standard error estimates, this study also summarized the efficiency of the parameter estimates. Note that the variance associated with the multiply imputed parameter estimate, $\overline{q}$, is a function of the average within-imputation variance, $\overline{u}$, and the between-imputation variance, *B*. (see Equations 2, 3 and 4). When the parameter estimate of interest is the Z_r-transformed correlation, its within-imputation variance is solely a function of sample size ($\hat{u} = \frac{1}{n-3}$). Because sample size was not varied in any of the conditions of this study, the average within-imputation variance, $\overline{u}$, was consistently equal to $\frac{1}{n-3}$, regardless of condition and replication. However, the between-imputations variance associated with Z_r did vary across conditions and provided the source of resulting differences in the total variance associated with $\overline{q}$. For this reason, the efficiency of the Z_r-transformed correlations was summarized by calculating the average between-imputation variances by condition.

Results

The relative biases of correlation estimates with normally and non-normally distributed data are presented in Table 1. This table shows that that MI of Likert-type data assuming continuous multivariate-normal data can yield acceptable parameter estimates with different types of missing data (MCAR and MAR) if the percentage of missing data is approximately 10%. However, with 30% of missing data, only the MCAR conditions resulted in acceptable relative bias. With 50% of missing data, acceptable relative biases were not obtained in any of the conditions. MI, assuming continuous data, showed robustness to categorization. Only slight differences in relative biases were identified between the three types of Likert scales. MI was also found to be robust to violations of normality. The relative biases of the skewed and normal conditions were similar.

The magnitude of correlations between variables (i.e., 0.8 or 0.2) also did not affect the performance of MI. The biases of parameter estimates obtained with MI were found to be consistently negative across all conditions. This leads to the conclusion that the presence of imputed data in datasets results in systematic reduction of the values of correlation coefficient estimates.

With MI, the variance associated with the multiply imputed parameter estimate is a function of the variability between estimates from each multiply imputed dataset as well as the variance of each estimate. (see Equations 2 - 4). This accounts for the extra amount of error introduced by the imputation process. Table 2 shows the average between-imputations variance summarized across generating conditions. The proportion of missing data had the strongest effect on the between-imputation variance. More specifically, as the overall proportion of missingness increased so did the between-imputation variance. A smaller effect was also identified: With the exception of conditions with 10% of missing data, the between-imputation variances were larger with correlation equal to 0.8 than 0.2. Furthermore, the conditions with 50% of missing data and correlation of 0.8 produced somewhat higher between-imputation variances, which increased as the number of points in the Likert scale increased. It is possible that this is the result of a three-way interaction between percentage of missing data, correlation between variables and number of points of the Likert scale. Additional studies expanding the levels of these three conditions would be needed to confirm the interaction.

Conclusion

Study results show that multiple imputation is robust to violations of both continuity and normality. This supports the assertion by Schafer (1997) that multiple imputation assuming the normal model works well even with ordered categorical data. However, it seems that resulting statistical tests will be less powerful because the sampling variance of the correlation estimates tends to increase and the values of the

Table 1: Relative Bias of the Z_r Estimates

Percentage of Missing Data Likert Scale	Type	Correlation = 0.8*			Correlation = 0.2		
		k = 3	*k* = 5	*k* = 7	*k* = 3	*k* = 5	*k* = 7
Normally-Distributed Data							
10%	MCAR	-0.002	-0.006	-0.004	-0.002	-0.001	0.003
	MAR	-0.003	-0.007	-0.005	-0.004	-0.002	0.002
30%	MCAR	-0.032	-0.037	-0.039	-0.044	-0.031	-0.041
	MAR	-0.041	**-0.053**	**-0.052**	**-0.054**	**-0.046**	**-0.042**
50%	MCAR	**-0.118**	**-0.129**	**-0.134**	**-0.156**	**-0.157**	**-0.137**
	MAR	**-0.163**	**-0.183**	**-0.176**	**-0.202**	**-0.207**	**-0.182**
Non-Normally Distributed Data							
10%	MCAR	-0.003	-0.007	-0.004	-0.001	0.007	-0.002
	MAR	-0.016	-0.012	-0.009	-0.004	-0.008	-0.009
30%	MCAR	-0.035	-0.038	-0.040	-0.040	-0.026	-0.036
	MAR	**-0.057**	**-0.064**	**-0.063**	**-0.058**	**-0.053**	**-0.039**
50%	MCAR	**-0.118**	**-0.130**	**-0.139**	**-0.160**	**-0.154**	**-0.164**
	MAR	**-0.200**	**-0.212**	**-0.179**	**-0.212**	**-0.224**	**-0.170**

*Bold numbers indicate unacceptable bias

correlations themselves tend to be negatively biased as the proportion of missing data increases. It should be noted that this decrease in power is a somewhat desirable feature of multiple imputation given that it adds a suitable degree of uncertainty to the resulting imputed datasets. Consequently, significance tests performed after MI will tend to be conservative compared with tests using complete data. Table 2 presents the average between-imputations variances for each condition in which missing data had been introduced. When no missingness exists, the between-imputations variance is zero and the resulting total variance for an estimate based on a dataset without missingness will be smaller with a concomitant increase in power.

For multiply imputed datasets, although the significance tests have less power, they will also meet the desired nominal α-levels; this is not the case when other missing data procedures such as mean and/or regression imputation are used. While the power of associated statistical tests under mean or regression imputation is maintained at the level of a complete data set, inflated Type I error rates can occur. Alternative missing data procedures such as listwise and/or pairwise deletion, similar to MI, are also known to result in decreased power. In addition, these deletion procedures have also been known to result in biased estimates given large degrees of missingness and non-MCAR patterns of missingness (Roth, Stwitzer & Switzer, 1999).

Based on results of the many different conditions simulated herein, it is possible to conclude that MI can be safely used to estimate parameters if the overall proportion of missing data is small (i.e., approximately 10%). If the data is missing completely at random, it was observed that as much as 30% of missing data does not result in inadequate parameter estimates. However, the major difficulty for applied researchers dealing with missing data is that it is not possible to know with certainty whether the missing values in a dataset are missing completely at random.

Table 2: Average Between-Imputation Variance of the Z_r Estimates For Normally Distributed Data

Percentage of Missing Data Likert Scale	Type	Correlation = 0.8			Correlation = 0.2		
		$k = 3$	$k = 5$	$k = 7$	$k = 3$	$k = 5$	$k = 7$
Normally Distributed Data							
10%	MCAR	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
	MAR	0.0005	0.0005	0.0006	0.0006	0.0005	0.0005
30%	MCAR	0.0033	0.0050	0.0063	0.0021	0.0021	0.0021
	MAR	0.0045	0.0081	0.0096	0.0022	0.0023	0.0023
50%	MCAR	0.0150	0.0235	0.0290	0.0046	0.0048	0.0049
	MAR	0.0188	0.0311	0.0369	0.0049	0.0053	0.0052
Non-Normally Distributed Data							
10%	MCAR	0.0004	0.0004	0.0005	0.0005	0.0005	0.0005
	MAR	0.0004	0.0004	0.0004	0.0005	0.0005	0.0004
30%	MCAR	0.0032	0.0051	0.0070	0.0021	0.0021	0.0021
	MAR	0.0038	0.0050	0.0063	0.0021	0.0019	0.0019
50%	MCAR	0.0148	0.0237	0.0318	0.0046	0.0048	0.0050
	MAR	0.0140	0.0237	0.0283	0.0045	0.0050	0.0047

For MAR conditions, this study did not omit the variable that caused the missing data (i.e., variable Z) from the datasets, which improves the performance of MI (Collins, Schafer & Kam, 2001). However, in real datasets, it is common that the researcher does not know or does not include the variables causing the missing data in the dataset. It can be expected that biases in the parameter estimates due to missing data would be larger if the variable causing missingness was omitted. A limitation of this study is that all datasets had a sample size of 400; different results might be obtained if smaller or larger sample sizes were used.

The datasets used in this study contained 10 inter-correlated variables. This type of dataset approximates a measurement situation where there is a scale or survey containing similar items. MI can benefit from the presence of inter-correlated variables, because the inter-correlations provide some of the missing information. The results of this study may have been different if uncorrelated variables were used; however, datasets containing uncorrelated variables are unlikely in measurement settings. Conversely, this study used some conditions where variable inter-correlations were probably weaker (i.e., 0.2) or stronger (i.e., 0.8) than those that would be found for responses to real scales or surveys. Items correlated at 0.2 would be realized in surveys, but would be somewhat lower that what would be expected for a psychometric scale measuring a single construct. These correlations were used in order to simulate distinct conditions.

Many unknowns exist regarding the ability of MI to generate acceptable estimates with large amounts of missing data. The question: What is the maximum amount of missing data that can be adequately imputed? has no easy solution, due to the interaction between the proportion of missing data and the pattern of correlations between variables in the dataset. Future research should address the effects of predictors included in the dataset to increase the accuracy of MI estimates in situations where the proportion of missing data is large. Another point deserving further investigation is the quality of correlation estimates when MI is used with a large imputation model containing several covariates. Sinharay, Stern and Russell (2001) found that MI of datasets with 20 covariates under the MAR assumption resulted in negatively biased correlation estimates. Additional research could address the effect of the covariates in MI of both continuous and categorical data.

Currently, MI together with full information maximum likelihood estimation are the frontrunners among missing data methods in terms of providing the most adequate estimates in the presence of MCAR and MAR missing data. Despite the fact that MI is available in many statistical programs, it has not become common practice in applied research. This may be due to the complex specification of the MI model that some software require (e.g., S-PLUS and R) or to the time consuming task of combining multiple imputed datasets. To promote an increase in use of MI among applied researchers, more automatic handling of imputed datasets by software is needed.

Although it was found that the appropriateness of MI to deal with missing data depends on whether data is MCAR or MAR as well as the proportion of missing data, Schafer and Olsen (1998) pointed out that it is misleading to classify the missing data in a dataset according to just one type of relationship between missing values and variables, because missing values can occur for many reasons within the same dataset. Furthermore, situations exist where neither the MCAR nor the MAR assumptions are plausible. Unfortunately, current missing data methods cannot handle MNAR data. Care should be taken to ensure that the procedure used to deal with missing data is appropriate for the missing data mechanism for a particular dataset.

References

Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, *6*(*4*), 330-351.

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *Journal of General Psychology*, *125*, 175-191.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, *8*(*1*), 128-141.

Fisher, R. A. (1928). *Statistical methods for research workers* (*2nd Ed.*). London: Oliver & Boyd.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*(*4*), 521-352.

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In: T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples*, p. 201-218. Hillsdale, NJ: Erlbaum.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods & Research*, *26*(*3*), 329-367.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician*, *55*, 244-254.

Huisman, M. (2000). Imputation of Missing Item Responses: Some Simple Techniques. *Quality & Quantity*, *34*, 331-351

Insightful. (2001). *S-PLUS* (Version 6.0) [Computer software]. Seattle, WA: Insightful.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika*, *27*(*2*), 179-182.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley & Sons.

Little, R., & Rubin, D. (1989). The analysis of social science data with missing values. *Sociological methods and research*, *18*(*2-3*), 292-326.

McDonald, R. A., Thurston, P. W., & Nelson, M. R. (2000). A Monte Carlo study of missing item methods. *Organizational research methods*, *3*(*1*), 70-91.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raaijmakers, Q. A. (1999). Effectiveness of different missing data treatments in surveys with likert-type data: Introducing the relative mean substitution approach. *Educational and Psychological Measurement*, *59*(*5*), 725-748.

Roth, P. L., Stwitzer, F. S., & Switzer, D. M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods*, *2*, 211-232.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(*3*), 581-92.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schafer, J. L. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model* (Version 2.02). [Computer software]. Retrieved from http://www.stat.psu.edu/~jls/misoftwa.html.

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.

Schimert, J., Schafer, J. L., Hesterberg, T., Fraley, C. & Clarkson, D.B. (2000). *Analyzing Data with Missing Values in S-PLUS*. Seattle, WA: Insightful Corporation,

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*(*4*), 317-329.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*(*3*), 465-471.

Wayman, J. C., & Swaim, R. C. (2002, April). *Practical considerations in constructing a multiple imputation model – A data example.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.