ABSTRACT

Sydney R. Siver, METHODS FOR HANDLING MISSING DATA FOR MULTIPLE-ITEM QUESTIONNAIRES (Under the direction of Dr. Alexander Schoemann) Department of Psychology, August 2017

Missing data is a common problem, especially in the social and behavioral sciences. Modern missing data methods are underutilized in the industrial/organizational psychology and human resource management literature. Recommendations for handling missing data and default options in software packages often use outdated, suboptimal methods for missing data. Resulting analyses tend to be biased, underpowered, or both. Best practice recommendationss for the handling of missing data includes the use of multiple imputation (MI) methods. However, this method is often ignored in favor of more convenient methods. For industrial/organizational psychologists, missing data is particularly problematic on multiple-item questionnaires, such as the Survey of Perceived Organizational Support (SPOS). Person mean imputation is one of the most common methods used to handle missing data on multiple-item questionnaires. However, it makes strong assumptions about the missing data mechanism and the underlying factor structure of a measure and should be avoided, particularly if there is a high rate of non-response. MI does not make the same assumptions as person mean imputation and may be a superior method when items are missing from a multiple-item questionnaire. Results indicate that PMI and MI provide similar results, however PMI may outperform MI when the number of variables is large.

Keywords: missing data, multiple imputation, person mean imputation, Monte Carlo

METHODS FOR HANDLING MISSING DATA FOR MULTIPLE-ITEM QUESTIONNAIRES

A Thesis

Presented to

The Faculty of the Department of Psychology

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in Psychology

by

Sydney R. Siver

August, 2017

METHODS FOR HANDING MISSING DATA FOR MULTIPLE-ITEM QUESTIONNAIRES

By

Sydney R. Siver

APPROVED BY:

DIRECTOR OF THESIS ______________________________
Alexander M. Schoemann, PhD

COMMITTEE MEMBER ______________________________
Karl L. Wuensch, PhD

COMMITTEE MEMBER ______________________________
Mark C. Bowler, PhD

CHAIR, DEPARTMENT OF PSYCHOLOGY ______________________________
Susan L. McCammon, PhD

DEAN OF THE GRADUATE SCHOOL ______________________________
Paul Gemperline, PhD

Acknowledgements

This work could not have been completed without the direction of Dr. Alex Schoemann as well as Dr. Karl Wuensch and Dr. Mark Bowler who all contributed invaluably to the learning experience of completing this research. I specifically want to thank Alex Schoemann for his patience, constructive feedback, and positivity regarding this work. Additionally, I want to thank the members of my cohort for their constant support and encouragement during the past two years.

Table of Contents

List of Tables

List of Figures

List of Equations

CHAPTER I : INTRODUCTION

Missing data creates difficulty in scientific research in both academia and applied settings and is frequently found in the social and behavioral sciences (Enders, 2010; Rubin, 1976; Schlomer, Bauman, & Card, 2010; Schafer & Olsen, 1998; West, 2001). In organizational settings, missing data commonly occurs when participants fail to respond to individual items on a survey or when data are simply not available (e.g. someone is absent during the time of his or her performance evaluation). Missing data can also occur if records are lost, discarded, or erased (Fichman, 2003). Missing data is hard to prevent in many situations, including research with multiple-item questionnaires, and it has been described as "one of the most important statistical and design problems in research" (Baraldi & Enders, 2010, p. 5).

Multiple-item questionnaires are commonly used in organizatinal settings to measure complex constructs. For example, job satisfaction can be measured with the Job Satisfaction Survey (JSS) which includes 36 items to measure overall satisfaction and includes nine facet scores: pay, promotion, supervision, fringe benefits, contingent rewards, operating procedures, coworkers, nature of work, and communication (Spector, 1985). These types of questionnaires are popular in work settings because they can be distributed to a large population easily and their completion does not take away from major duties during the work day. However, missing data tends to be a problem with these scales due to a high level of nonresponse. Therefore, problems arise when researchers analyze such data with suboptimal methods such as listwise deletion or mean substitution, which lead to biased parameter estimates (when data are missing at random) or underpowered analyses (when data are missing completely at random) (Enders, 2010; Newman, 2009; Schafer & Olsen, 1998;).

A typical method when data are missing on these scales is person mean imputation (PMI). PMI occurs when the researcher averages the scores for all items that correspond to a particular dimension for a single participant and substitutes that average for the missing data (Enders, 2010). This method is sometimes referred to as averaging across available items, and by doing so researchers are using a technique that is equivalent to PMI. PMI is probably the most common approach for dealing with item-level missing data on questionnaires, even though little is known about the biases that can result from this method.

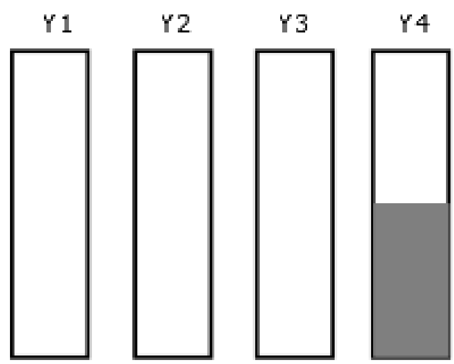
Best practice recommendations for missing data analyses are to use multiple imputation (MI) to handle missing data, however MI is often ignored out of convenience (Enders, 2010; Newman, 2014). MI is superior to other traditional techniques in that MI provides one general tool to address the problem of missing data, is unbiased under both MAR and MCAR conditions, and works with standard statistical software. Additionally, MI is advantageous for analyzing data from multi-item questionnaires because it provides a mechanism for dealing with item-level missingness.

Missing Data Theory

Researchers often confuse missing data patterns with missing data mechanisms. A missing data pattern refers to the configuration of observed and missing values in a data set. Missing data mechanisms established by Rubin (1976) describe relationships between measured variables and the probability of missing data. Missing data mechanisms postulate as to the causes of missing data whereas missing data patterns only point to the location of the missing data. However, missing data patterns serve as the basis for missing data mechanisms, so understanding them is crucial to missing data analyses. Multiple imputation methods are well suited for most missing data patterns and mechanisms, so distinguishing between them is not absolutely necessary if using this method (Enders, 2010).

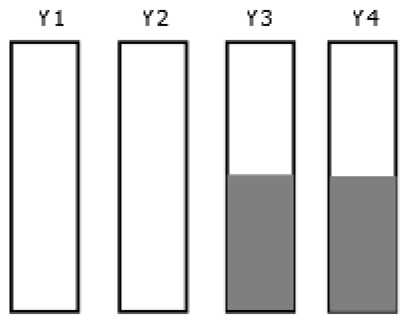
Missing data patterns. Enders (2010) describes four prototypical missing data patterns: univariate, unit nonresponse, monotone, and general. The univariate pattern has missing values that are isolated to a single variable as shown in Figure 1 such that Y_1 through Y_3 are manipulated variables, Y_4 is the outcome variable, complete data is represented by white and missing data is represented by gray. It also includes situations in which Y represents a group of items that is either entirely observed or entirely missing for each unit (Schafer & Graham, 2002). This pattern is more common in experimental studies.

Figure 1. *Univariate Pattern*



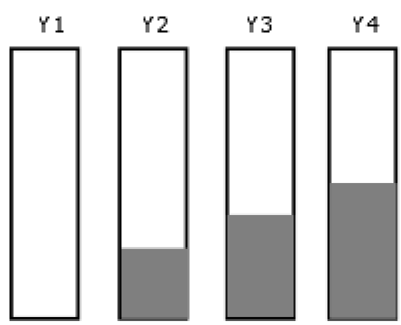
The unit nonresponse pattern is common in survey research (Enders, 2010; Fichman, 2003; Little & Rubin, 2002; Schafer & Graham, 2002). As shown in Figure 2, this occurs when data are completely missing for Y_1 and Y_2 such that Y_1 and Y_2 are characteristics that are available for every member of the sample. When this happens, missing data will appear in a haphazard pattern (Little & Rubin, 2002). If the sampled person is not at home or refuses to answer the surveys in their entirety, data will result in a unit nonresponse pattern. For example, this pattern can arise when an organization uses two inexpensive assessments (Y_1 and Y_2) and two expensive assessments (Y_3 and Y_4) in its selection process; candidates will have dropped out of the selection process before getting to the expensive assessments. Unit nonresponse has been traditionally handled by reweighting, however multiple imputation methods will produce more accurate analyses (Shafer & Graham, 2002).

Figure 2. *Unit Nonresponse Pattern*



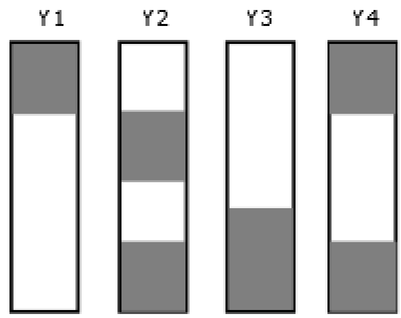
Monotone patterns are most easily understood through attrition, or participants leaving a longitudinal study, which undoubtedly results in a loss of a substantial amount of information (Little & Rubin, 2002). For example, in a clinical trial of a new medication participants may choose to stop treatment or may no longer be eligible to participate due to health reasons in various stages of the study. As shown in Figure 3, this pattern resembles a staircase. Mathematically, items or item groups may be ordered in such a way that if Y_j is missing for a unit, then $Y_{j+1} \dots Y_p$ are missing as well (Schafer & Graham, 2002).

Figure 3. *Monotone Pattern*



A general pattern is the most common combination of missing values (Enders, 2010; Shafer & Graham, 2002). In a general pattern any set of variables may be missing for any unit, however this does not mean that the values are not systematically missing; the missing data might be able to be further separated into specific patterns with differing reasons for missingness.

Figure 4. *General Pattern*



Levels of missing data. In addition to missing data patterns, researchers should also be familiar with missing data levels to appropriately choose the best missing data method for multiple-item questionnaires. There are three main levels of missingness: construct-level, item-level, and person-level (Newman, 2014; Parent, 2012). Construct-level, or scale-level, occurs when all the items for a participant for a particular measure are missing therefore omitting an entire construct. This level of missingness can be handled with advanced methods that consider the available data and correlations among observed variables for all cases such as MI or full information maximum likelihood (FIML).

Item-level missingness is the most common level of missing data concerning multiple-item questionnaires, and that which this study aims to address (Parent, 2012). Item-level missingness occurs when the participant leaves a few items blank without completely missing any scales. Due to the haphazard nature of this level of missingness, choosing the appropriate statistical method can be challenging.

Person-level is the hardest level to analyze because it involves failure by an individual to respond to any part of the survey. As a result of the complete lack of data for a participant at this level, missingness at this level is best addressed through survey design (Parent, 2012). For example, a survey administered only in English excludes those who do not understand English from participating, or a survey administered by mail excludes those who do not have a mailing address.

Distribution of missingness. To understand missing data mechanisms, one must first understand the missing data distributions. For any data set, one can define indicator variables, R, which identify what is known and what is missing. Rubin (1976) defines R as binary (i.e. $R = 1$ if a score is known and $R = 0$ if a score is missing). In the case of multivariate data, R becomes a matrix and has the same number of rows and columns as the data matrix when every variable has missing values. R is treated as a set of random variables with a joint probability distribution, and it is this distribution that differentiates between the missing data mechanisms.

Missing data mechanisms. As previously mentioned, there are three main mechanisms for classifying missing data: missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). From a practical standpoint, Rubin's (1976) mechanisms are essentially assumptions that govern the performance of different analytic techniques. Most traditional methods have strict assumptions about MCAR mechanisms and subsequently suffer from biased results as missing data are rarely MCAR in practice (Enders, 2010; Graham, 2009; Little & Rubin, 2002). However, more advanced methods such as multiple imputation require less restrictive assumptions and therefore produce more accurate results. Understanding the mechanisms is the first step for psychological researchers in choosing analyses since the mechanisms determine the nature and magnitude of missing data bias and imprecision.

Missing at random. MAR data occurs when missingness may be related to another variable in the data. That is, a systematic relationship exists between one or more measured variables and the probability of missing data (Rubin, 1976; Schafer & Graham, 2002). For example, respondents may be hesitant to report their salaries if they are in high-level positions. Assuming job position was measured in the same survey, this situation would produce MAR data. In modern missing data theory, the MAR distribution can be expressed as,

$$p(R \mid Y_{obs}, Y_{mis}, \phi) = p(R \mid Y_{obs}, \phi) \quad (1)$$

where p is the probability distribution, R is the missingness mechanism, Y_{obs} is the observed data, Y_{mis} is the missing part of the data, and ϕ is a parameter that describes the relationship between R and the data. In other words, the probability of missingness for MAR data depends on observed data, but not on missing data (Allison, 2001). MAR is often described as ignorable missingness because there is no need to estimate the distribution parameters, ϕ, when performing analyses such as multiple imputation. This is because multiple imputation does not require parameter information if the data are MAR or MCAR. However, it is not possible to test the mechanism to verify the scores are MAR (Enders, 2010; Littvay, 2009).

Missing completely at random. Despite its name, MAR does not mean that the missing values are a simple random sample of all data values. This occurs in a special case of MAR called missing completely at random (MCAR) (Fichman, 2003; Little & Rubin, 2002; Schafer, 1997). In other words, the reason for missingness in MCAR is not systematic; it is truly random and haphazard. For example, MCAR data can occur if data are lost, equipment malfunctioned, or data were incorrectly entered. MCAR is more restrictive than MAR because MAR allows the probability that a value is missing to depend on that value itself through observed quantities whereas MCAR assumes that missingness is completely unrelated to the data:

$$p(R \mid Y_{obs}, Y_{mis}, \phi) = p(R \mid \phi) \quad (2)$$

This idea can be tested by separating the missing and complete cases and then examining group mean differences. If the missing data patterns are randomly equivalent, then the means between the groups should be the same. Fichman (2003) describes the distribution for this test as:

$$Y_{obs} \mid R_{k1k \neq j} = 1 \text{ to } Y_{obs} \mid R_{k1k \neq j} = 0 \tag{3}$$

Because the nature of MCAR is truly random, Wayman (2003) asserts "the only real penalty in failing to account for [MCAR] missing data is loss of power" (p. 3). Nevertheless, methods such as multiple imputation can account for a loss of power because they use all of the available data in analyses by filling in missing values prior to analysis.

Missing not at random. Missing not at random (MNAR) data occur when the probability of missing data is related to the values of the data itself after controlling for other variables (Enders, 2010; Newman, 2009; Schafer & Graham, 2002). That is, MNAR data are missing due to some reason not captured in the data i.e. it depends on unobserved data. The MNAR distribution can be expressed as:

$$p(R \mid Y_{obs}, Y_{mis}, \phi) \tag{4}$$

Unlike MCAR, there is no way to verify the MNAR mechanism without knowing the values of the missing variables. MNAR data is often difficult to analyze because it produces substantial bias with most techniques. Additionally, methods that are used to analyze MNAR data are rarely used because they require strict assumptions and therefore are not practical (Graham, 2009; Newman, 2009). However, one can hope to negate the use of MNAR methods by properly designing the study to measure variables related to missingness so that the missing data become MAR instead of MNAR.

Traditional Methods to Analyze Missing Data

Traditional methods are still commonly used due to their convenience, researchers' familiarity with the techniques, and because they are the default in standard statistical software. However, results are often biased and underpowered powered due to incorrect assumptions of the missing data mechanism (Enders, 2010; Fichman, 2003; Newman, 2009). The methods outlined in this section can be categorized into either deletion methods or single imputation methods. Deletion methods are by far the most popular approaches to missing data in the social and behavioral sciences due to their ease of use. The drawbacks to these approaches include biased parameters and loss of power due to an assumed MCAR mechanism. Single imputation methods fill in missing values prior to analyses (Enders, 2010; Newman, 2014). They are more useful than deletion methods because they produce a complete data set without reducing sample size. Nevertheless, they are still problematic in that they produced biased estimates with each of the three mechanisms, attenuate standard errors, and underestimate sampling error. For these reasons, researchers should move to multiple imputation techniques even though they are more computationally complex.

Listwise deletion. Listwise deletion, also known as complete-case analysis, discards cases for which there is missing data on one or more variables in an effort provide complete data for all of the variables surveyed (Enders, 2010; Graham, 2009; Newman, 2009). This method is probably the most widely used due to convenience; eliminating the missing data removes the need to use any special analyses or software, and it is the default in most standard statistical software. However, it does have drawbacks. Listwise deletion decreases sample size, thus reducing power. It also requires data to be MCAR and produces biased, inaccurate parameters if the data are MAR or MNAR (Enders, 2010; Newman, 2009).

Pairwise deletion. Pairwise deletion, also known as available-case analysis, requires that individuals with enough information for any calculation are used. It is usually employed in conjunction with a correlation matrix in which each correlation is estimated based on the cases that have data for all variables (Enders, 2010; Graham, 2009; Newman, 2009; Schafer & Graham, 2002). The main problem with this method is that it uses different subsets of cases for correlations and variance estimates. It is noteworthy that pairwise deletion is more powerful than listwise deletion, especially when the correlations between the variables in the data set are low, however it does have problems similar to those of listwise deletion. Pairwise deletion requires data to be MCAR and produces complications when computing standard errors due to differing sample sizes within the variables. Similar to listwise deletion, pairwise deletion produces biased, inaccurate parameters if the data are MAR or MNAR.

Arithmetic mean imputation. Arithmetic mean imputation, or mean substitution, occurs when missing values for a variable are filled in with the mean of all available cases (Enders, 2010; Sinharay, Stern, & Russell, 2001). As with listwise deletion arithmetic mean imputation produces a complete data set relatively easily, however this method severely biases the parameter estimates even when the data are MCAR. The variability of the data is greatly reduced therefore underestimating variance and covariance patterns, correlations, and regression coefficients. Additionally, this approach leads to an underestimate of error because the sample size is increased without adding any new information (Howell, 2015a).

Similar response pattern imputation. This method replaces missing values with the score from another participant who has a similar response pattern on the same variables (Enders, 2010). If no such case exists, imputation does not take place. Consequently, this method does not necessarily produce a complete data set. This approach can produce relatively accurate parameter estimates when the data are MCAR, but not MAR. Additionally, similar response pattern imputation has no theoretical justification therefore researchers should refrain from using this method.

Hot-deck imputation. Hot-deck imputation assumes the distribution is the most appropriate source of missing data (Brown, 1994). Missing values are replaced with the scores from another respondent who scored similarly on a set of matching variables, usually demographics. While hot-deck imputation preserves the univariate distributions of data, this method will increase the size of the variance estimates as well as bias correlations and regression coefficients and underestimate standard errors. As Howell (2015a) points out, this hot-deck imputation was developed in the 1940's by statisticians at the Census Bureau for use with public data sets when the percentage of missing data was rather small. Therefore, hot-deck imputation is not commonly used by social and behavioral researchers anymore.

Regression imputation. Regression imputation replaces missing values with predicted scores from a regression equation obtained from the observed cases (Enders, 2010; Howell, 2015a; Newman, 2014). Regression imputation is similar to maximum likelihood (ML) and multiple imputation practices in that it borrows information from the sample. However, regression imputation is not preferable to these techniques because single imputation methods, such as regression imputation, are biased under MCAR and lead to an underestimation of the variance and an overestimation of the correlation due to multicollinearity (Newman, 2014). Regression imputation runs into the same problem as arithmetic mean imputation in that no new information is added to the study therefore reducing variability and increasing error. While regression imputation is superior to arithmetic mean imputation, modern methods to analyze missing data such as ML and MI have similar advantages to regression imputation without any of the biases.

Stochastic regression imputation. Stochastic regression imputation is an attempt to improve upon regression imputation in that it adds a normally distributed residual term to the regression imputation method to account for the lack of variability in the data that occurs from multicollinearity (Baraldi & Enders, 2010; Enders, 2010; Newman, 2014; Roth, Switzer, & Switzer, 1999). While this method is unbiased under both MAR and MCAR conditions, Newman (2014) does not recommend stochastic regression imputation due to the inability to calculate accurate standard errors for hypothesis testing and the increased probability of a type I error. Furthermore, this method can be complicated with multivariate data as each regression equation needs its own residual distribution. Researchers have proposed corrections to the biases and inaccuracies of this method, however applying these corrections to stochastic regression imputation tends to be more difficult than the use of modern methods such as maximum likelihood and multiple imputation.

Person mean imputation. Person mean imputation (PMI), or averaging the available items, is like arithmetic mean imputation in that missing values are replaced with the mean of a set of scores. However, in PMI the imputed value is the average of the scores of all the items for the participant for which there are missing values, and this technique is equivalent to averaging the available items. Roth, Switzer, and Switzer (1999) point out that PMI is a special case of regression imputation in which equal variances are assumed for the independent variables. This method is commonly employed when computing scale scores for a specific construct. PMI is probably the most common approach for dealing with data that are missing on an item-level, however empirical studies have only investigated PMI in the context of internal consistency reliability analyses (Downey & King, 1998; Enders, 2003). Furthermore, an investigation into missing data techniques for multiple item questionnaires by Roth, Switzer, and Switzer (1999) found that PMI yielded the most unbiased regression coefficients when compared to other deletion and single imputation techniques.

Enders (2010) and Roth, Switzer, and Switzer (1999) caution about the use of PMI when the rate of item nonresponse is high because not much is known about the potential problems with this method. Conversely, Newman (2014) advocates for the use of this method even when the participant has only answered one item per construct because ML and MI techniques do not always work for item-level missingness even though PMI is not unbiased under MAR. Additionally, PMI leads to less reliable scale scores due to the use of fewer items which then increases the observed effect size whenever the item-level missingnesss is not MCAR (Newman, 2014).

Modern Methods to Analyze Missing Data

Modern methods such as maximum likelihood and multiple imputation are widely regarded as "state of the art" missing data techniques because they produce unbiased parameter estimates under MAR and MCAR data. Additionally, these methods tend to be more accurate than deletion and single imputation methods since none of the data are discarded, yielding higher sample sizes and more accurate parameters (Enders, 2010; Schafer & Olsen, 1998). Nevertheless, Baraldi and Enders (2010) note that these methods will still yield biased parameter estimates under the MNAR condition, but they will be minor compared to those obtained with lesser methods such as deletion and single imputation methods.

Full information maximum likelihood. Full information maximum likelihood estimation (FIML) uses all the available data to estimate parameter values that have the highest probability of producing the sample data. Multiple iterations of the data are computed until they converge upon a set that most closely resembles the sample data (i.e., the distance from the mean to the data is minimized as much as possible, or the highest log-likelihood value is produced) (Baraldi & Enders, 2010; Howell, 2015b; Roth, Switzer, & Switzer, 1999). The log-likelihood for sample scores is shown in equation 5.

$$logL = \sum_{i=1} N\, log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5\left(\frac{y_i-\mu}{\sigma}\right)}\right] \quad (5)$$

Score estimates substituted into this equation that are close to the mean produce a small z-score and a large log-likelihood which are indicative of a better fit (Enders, 2010). This calculation is repeated for the entire sample and the individual log likelihood values are summed to create the sample log likelihood (Baraldi & Enders, 2010). In multivariate data, matrices replace the scalar values in equation 5 so the equation becomes

$$(Y_i\text{-}\mu)^T\Sigma^{-1}(Y_i\text{-}\mu) \quad (6)$$

where μ is the mean population vector, Σ is the population covariance matrix, Y_i is a vector that contains the sores for a single individual, *T* represents the matrix transpose, and -1 signifies the inverse. This equation yields a set of scores for an individual and the population means, whereas equation 5 quantifies the joint probability of drawing the sample of scores from a normal distribution.

FIML is considered superior to other missing data techniques because it requires a less restrictive MAR assumption and therefore produces unbiased parameter estimates under both MAR and MCAR (Enders & Bandalos, 2001). Graham, Hofer, and MacKinnon (1996) found that ML estimates were unbiased under both MAR and MCAR and were more accurate than those from deletion methods.

One of the problems with maximum likelihood is that it is specific to the model being applied (Sinharay, Stern, & Russell, 2001). If causes or correlates of missingness are excluded from the model, parameter estimates may be biased under FIML (Collins, Schafer, & Kam, 2001). Additionally, FIML produces biased estimates under the MNAR mechanism (Schafer & Graham, 2002). Furthermore, FIML requires relatively advanced statistical software to which researchers may not have access.

Multiple imputation. Multiple imputation (MI) is an alternative to FIML in that it attempts to fill in the missing values prior to analysis rather than estimating the parameters directly from the available data. Additionally, both techniques have the same assumptions, MAR and multivariate normality, and yield the same results under infinite imputations (Enders, 2010; Newman, 2014). However, MI is superior to FIML in situations with item-level missingness and many auxiliary variables related to missingness because FIML is model specific whereas MI can include any number of auxiliary variables (Enders, 2010; Sinharay, Stern, & Russell, 2001).

MI is commonly described in three stages: the imputation phase, the analysis phase, and the pooling phase. In the imputation phase, multiple copies of the data are created with different estimates of missing values (Enders, 2010; Horton & Lipsitz, 2001). The data sets are analyzed in the analysis phase once per each imputed data set using the same procedures one would have used had the data been complete, and then combined into a single set of results in the pooling stage. The parameter estimates are averaged across the number of imputed data sets, and the standard errors are combined using equation 7 where $\frac{1}{M}\sum_{m=1}^{M} S.E._m^2$ is the average squared standard error across imputations and $\left(1+\frac{1}{M}\right)$ is a correction factor that converges to 1 as the number of imputations increases.

$$SE = \sqrt{\frac{1}{M}\sum_{m=1}^{M} S.E._m^2 + \left(1+\frac{1}{M}\right)\left(\frac{1}{M-1}\right)\sum_{m=1}^{M}(b_m - b)^2} \qquad (7)$$

The pooled MI parameter estimates are unbiased under both MAR and MCAR and the pooled MI standard errors are accurate due to the second term in the equation, $\left(\frac{1}{M-1}\right)\sum_{m=1}^{M}(b_m - b)^2$, which is the variance of the parameter estimates between imputations.

When analyzing data with item-level missingness, Newman (2014) generally recommends the use of MI when conducting construct-level analyses. However, Newman (2014) recommends the use of PMI for a construct-level analysis from a multi-item scale. Additionally, Newman (2009) points out that PMI works well if the scale items are parallel (i.e. the factor loadings are equal). Furthermore, Schafer and Graham (2002) investigated the use of PMI when dealing with missing data for scale scores under MCAR for both 30% missing and 5% missing and found that PMI may be a reasonable alternative to MI, and that bias in the scales tends to decrease as the scales become more correlated with each other.

Hypothesis 1: When missing data are MAR, MI should outperform PMI and this difference will be stronger with more missing data.

Hypothesis 2: When missing data are MCAR and factor loadings are equal, MI and PMI will have the same results regardless of percent missing or scale size.

Hypothesis 3: When missing data are MCAR and factor loadings are not equal, MI will outperform PMI, and this will be more apparent with more missing data, smaller sample sizes, and smaller numbers of items.

The Current Study

Newman (2014) recommends the use of PMI instead of MI techniques when conducting a construct-level analysis using a multi-item scale because using "MI techniques on item-level data is often difficult to do" (p. 392). Therefore, this study seeks to refute this claim while adding to the current body of missing data literature.

Using previous missing data research, this study examines the performance of PMI and MI for varying conditions of sample size, factor loadings, percent missing data, type of missing data, and scale length. Additionally, this study aims to provide recommendations for the use of MI methods or PMI when investigating the relationship among scale scores for multi-item questionnaires that measure a single construct.

CHAPTER II: METHODS

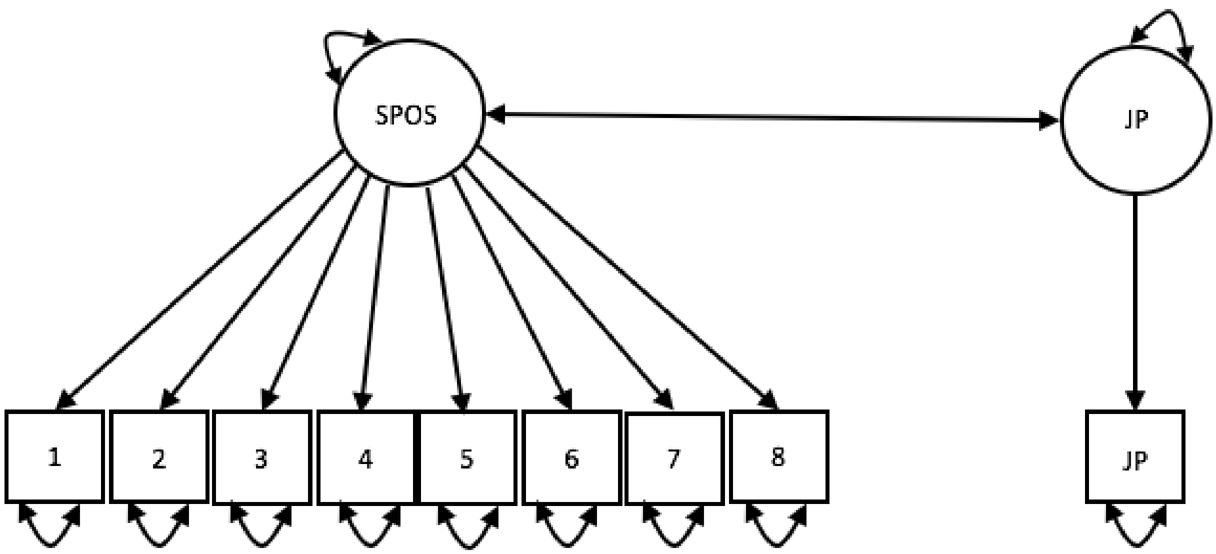
Measures

Two versions of the Survey for Perceived Occupational Support (SPOS) were used as an example of a multi-item questionnaire in this study. The SPOS long version consists of 36 items that measure overall perceived occupational support (Eisenberger et al., 1986). It asks respondents to indicate their level of agreement or disagreement with the various statements, and it can be used to indicate how an employee perceives the extent to which an organization values their contribution and cares about their well-being. The short version contains eight of the highest loading items that measure overall perceived occupational support (Eisenberger et al., 1986). Sample items include: “The organization values my contribution to its well-being” and “The organization would ignore any complaint from me.” Items are rated on a 0-6 Likert scale ranging from “strongly disagree” to “strongly agree.”

Data Generating Model

The data generating model consists of data of varying sample sizes from populations for both the 36-item and the 8-item SPOS. The population model for the 8-item SPOS is shown in Figure 5. The baseline factor loadings are 0.7. The correlation coefficient for SPOS and job performance (JP) is set to 0.3 and is assessed as the parameter of interest. All latent variances were set to 1. The residual variances of the SPOS items were set to 0.51, and the means of each item were set to 0.

Figure 5. *Population Model for 8-Item SPOS*



Conditions

The simulation included five different variables: sample size, percent missing data, missing data mechanism, inequality of factor loadings, and scale length. Sample sizes consist of either 50, or 200 respondents. Data consists of either 2%, 5%, or 10% missing values under either the MAR or MCAR mechanism. The MNAR mechanism was not used due to its biased nature in both PMI and MI. Both the 36-item and the 8-item SPOS were used, and the items loaded on a single factor, perceived occupational support, with half of the factor loadings differing from the baseline factor loading with differences of 0. 0.2, 0.4, i.e. half of all factor loadings are always 0.7, the other half of the factor loadings are either 0.7, 0.5, or 0.3. A 3 X 2 X 2 X 2 X 2 design was employed with 500 replications in each condition.

Data Analysis

The current study uses Monte Carlo simulations to determine recommendations for the use of MI methods or PMI. Data were generated with the *simsem* package for R.

(Pornprasertmanit, Miller, & Schoemann, 2016; R Core Team, 2016). Missing data were analyzed using both PMI and MI methods. MI was performed with the *mice* package (van Buuren & Groothuis-Oudshoorn, 2011). Scale scores were computed for the SPOS measures, and the correlation coefficient between scale scores and job performance was assessed. Factorial ANOVAs were computed to investigate the effects of each of the conditions described in the previous section. Outcomes include bias in parameters, mean squared error, and power. Parameter bias was computed as the population value subtracted from the observed parameter estimate divided by the population value for both the PMI and MI conditions; the resulting value was then multiplied by 100 to create a percent bias scale. Mean squared error was computed as the raw bias of the estimate squared plus the standard deviation of the estimates within condition. Power was computed as the proportion of outcomes for PMI and MI with statistical significance, $p < 0.05$. Statistical significance tests are not reported; instead, the partial η^2 for each effect is reported and interpreted for those which exceed 0.01 which is a small effect per Cohen (1973).

CHAPTER III: RESULTS

Data were simulated according to the conditions listed above with a resulting sample size of 35,000 cases. However, MI only converged in 96.2% of cases overall. After eliminating those cases in which MI did not converge the resulting sample size was 33,346. Cases were considered to not converge if the estimated correlation was an extreme value, greater than 0.9 or less than -0.6, as well as cases in which MI did not return an estimate. The percentages of cases that converged in each condition are shown in Table 1; convergence was 100% for those conditions not listed in the table. MI was more likely to converge in conditions that had a sample size of 200, 10% missing, 8 items, and MAR data. Difference in factor loadings did not appear to affect the percentage of convergence. Convergence was very poor in situations with many items relative to the sample size; in this study, this occurred when sample size was 50, 10% of data were missing, and there were 36 items with MCAR data.

Table 1. *Percentage of Convergence for MI*

		Conditions			**Percent Converged**
Sample Size	*Percent Missing*	*Difference in Factor Loadings*	*Number of Items*	*Missing Mechanism*	
50	5%	0	36	MCAR	99.8%
50	5%	0.2	36	MCAR	99.2%
50	5%	0.4	36	MCAR	99.2%
50	10%	0	36	MCAR	24.8%
50	10%	0.2	36	MCAR	21.2%
50	10%	0.4	36	MCAR	19.2%
200	10%	0	36	MCAR	89%
200	10%	0.2	36	MCAR	88.8%
200	10%	0.4	36	MCAR	89.6%

Bias in Parameter Estimates

Parameter bias was computed separately for each set of data. A mixed-design ANOVA with method as a within-subjects factor and sample size, percent missing, difference in factor loadings, missing data mechanism, and scale length as between-subjects factors revealed multiple effects with partial η^2 greater than 0.01. Results for within-subjects effects and estimated marginal means are shown in Tables 2 and 3 respectively. Between-subjects effects can be found in Appendix A.

Table 2. *Within-Subjects Effects for Parameter Bias*

Effects	***F***	η_p^2
Method	$F(1, 33276) = 1104.842, p < .001$	0.032
Method * Sample Size	$F(1, 33276) = 521.658, p < .001$	0.015
Method * Percent Missing	$F(2, 33276) = 11.628, p < .001$	0.001
Method * Factor Loading	$F(1, 33276) = 5.284, p = 0.005$	0.000
Method * Number of Items	$F(1, 33276) = 27.239, p < .001$	0.001
Method * Missing Mechanism	$F(1, 32276) = 12.983, p < .001$	0.000
Method * Sample Size * Percent Missing	$F(2, 33276) = 0.304, p = 0.738$	0.000
Method * Sample Size * Factor Loading	$F(2, 33276) = 2.959, p = 0.052$	0.000
Method * Sample Size * Number of Items	$F(1, 33276) = 5.698, p = 0.017$	0.000
Method * Sample Size * Missing Mechanism	$F(1, 33276) = 1.798, p = 0.180$	0.00
Method * Percent Missing * Factor Loading	$F(4, 33276) = 4.062, p = 0.063$	0.000
Method * Percent Missing * Number of Items	$F(2, 33276) = 0.716, p = 0.489$	0.00
Method * Percent Missing * Missing Mechanism	$F(2, 33276) = 4.683, p = 0.009$	0.000
Method * Factor Loadings * Number of Items	$F(2, 33276) = 3.750, p = 0.024$	0.000
Method * Factor Loading * Missing Mechanism	$F(2, 33276) = 5.849, p = 0.003$	0.000
Method * Number of Items* Missing Mechanism	$F(1, 33276) = 8.873, p = 0.003$	0.000
Method * Sample Size * Percent Missing* Factor Loadings	$F(4, 33276) = 1.425, p = 0.223$	0.000

Effects	***F***	η_p^2
Method * Sample Size * Percent Missing * Number of Items	$F(2, 33276) = 0.724$, $p = 0.485$	0.000
Method * Sample Size * Percent Missing * Missing Mechanism	$F(2, 33276) = 0.885$, $p = 0.413$	0.000
Method * Sample Size * Factor Loadings * Number of Items	$F(2, 33276) = 2.089$, $p = 0.124$	0.000
Method * Sample Size * Factor Loadings * Missing Mechanism	$F(2, 33276) = 1.930$, $p = 0.145$	0.000
Method * Sample Size * Number of Items * Missing Mechanism	$F(1, 33276) = 0.748$, $p = 0.387$	0.000
Method * Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 3.195$, $p = 0.012$	0.000
Method * Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 2.575$, $p = 0.036$	0.000
Method * Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 3.295$, $p = 0.037$	0.000
Method * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 5.376$, $p = 0.005$	0.000
Method * Sample Size * Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 1.582$, $p = 0.176$	0.000
Method * Sample Size * Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 0.861$, $p = 0.486$	0.000
Method * Sample Size * Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 0.578$, $p = 0.561$	0.000
Method * Sample Size * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 1.709$, $p = 0.181$	0.000
Method * Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(4, 33276) = 3.267$, $p = 0.011$	0.000
Method * Sample Size * Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 1.976$, $p = 0.139$	0.000

A small effect was found for the main effect of method, $\eta_p^2 = .032$, with estimated marginal means of -5.607 for PMI and -7.964 for MI indicating PMI is slightly less biased than MI. The interaction between method and sample size yielded a small effect, $\eta_p^2 = .015$, with estimated marginal means indicating PMI and MI had similar bias when the sample size was 200, but PMI showed less bias than when sample size was 50, this effect may be related to the lack of convergence for MI when the sample size is 50. All other interactions fell short of a meaningful effect.

Table 3. *Relevant Estimated Marginal Means for Parameter Bias*

Conditions			**Mean**	**Std. Error**	**95% Confidence Interval** Lower Bound	**95% Confidence Interval** Upper Bound
Method	PMI		-5.607	0.223	-6.044	-5.170
	MI		-7.964	0.222	-8.400	-7.529
Method * Sample Size	PMI	50	-5.122	0.365	-5.837	-4.407
		200	-6.064	0.263	-6.581	-5.548
	MI	50	-9.223	0.363	-9.935	-8.511
		200	-6.775	0.263	-7.290	-6.261

Mean Squared Error

Mean squared error was computed separately for each set of data. A mixed-design ANOVA with method as a within-subjects factor and sample size, percent missing, difference in factor loadings, missing data mechanism, and scale length as between-subjects factors did not reveal any effects with partial η^2 greater than 0.01. Results for within-subjects effects are shown in Table 4. Between-subjects effects can be found in Appendix B. All interactions fell short of a meaningful effect.

Table 4. *Within-Subjects Effects for Mean Squared Error*

Effects	***F***	η_p^2
Method	$F(1, 33276) = 37.408, p < .001$	0.001
Method * Sample Size	$F(1, 33276) = 137.929, p < .001$	0.004
Method * Percent Missing	$F(2, 33276) = 15.774, p < .001$	0.001
Method * Factor Loading	$F(2, 33276) = 74.729, p < .001$	0.004
Method * Number of Items	$F(1, 33276) = 289.686, p < .001$	0.009
Method * Missing Mechanism	$F(1, 33276) = 11.286, p = 0.001$	0.000
Method * Sample Size * Percent Missing	$F(2, 33276) = 23.091, p < .001$	0.001
Method * Sample Size * Factor Loading	$F(2, 33276) = 48.391, p < .001$	0.003
Method * Sample Size * Number of Items	$F(1, 33276) = 102.917, p < .001$	0.003
Method * Sample Size * Missing Mechanism	$F(1, 33276) = 3.642, p = 0.056$	0.000
Method * Percent Missing * Factor Loading	$F(4, 33276) = 9.649, p < .001$	0.001
Method * Percent Missing * Number of Items	$F(2, 33276) = 43.250, p < .001$	0.003
Method * Percent Missing * Missing Mechanism	$F(2, 33276) = 11.858, p < .001$	0.001
Method * Factor Loadings * Number of Items	$F(2, 33276) = 7.536, p = 0.001$	0.000
Method * Factor Loading * Missing Mechanism	$F(2, 33276) = 0.420, p = 0.657$	0.000
Method * Number of Items* Missing Mechanism	$F(1, 33276) = 13.079, p < .001$	0.000
Method * Sample Size * Percent Missing* Factor Loadings	$F(4, 33276) = 11.960, p < .001$	0.001
Method * Sample Size * Percent Missing * Number of Items	$F(2, 33276) = 30.522, p < .001$	0.002
Method * Sample Size * Percent Missing * Missing Mechanism	$F(2, 33276) = 9.815, p < .001$	0.001
Method * Sample Size * Factor Loadings * Number of Items	$F(2, 33276) = 1.590, p = 0.204$	0.000
Method * Sample Size * Factor Loadings * Missing Mechanism	$F(2, 33276) = 0.360, p = 0.698$	0.000
Method * Sample Size * Number of Items * Missing Mechanism	$F(1, 33276) = 14.991, p < .001$	0.000
Method * Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 9.435, p < .001$	0.001

Effects	***F***	**η_p^2**
Method * Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 1.816$, $p = 0.123$	0.000
Method * Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 11.288$, $p < .001$	0.001
Method * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.789$, $p = 0.454$	0.000
Method * Sample Size * Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 10.805$, $p < .001$	0.001
Method * Sample Size * Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 1.771$, $p = 0.132$	0.000
Method * Sample Size * Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 9.489$, $p < .001$	0.001
Method * Sample Size * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.148$, $p = 0.863$	0.000
Method * Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(4, 33276) = 1.134$, $p = 0.338$	0.000
Method * Sample Size * Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 1.530$, $p = 0.216$	0.000

Power

Power was computed separately for each set of data; data was coded as 1 if the p-value was less than 0.05, and as 0 if the p-value was greater than 0.05. Binary logistic regression was implemented separately for each set of data to predict power when either PMI or MI was employed using sample size, percent missing, difference in factor loadings, number of items, and missing mechanism as predictors.

A test of the full for PMI model versus a model with intercept only was statistically significant, $\chi^2(5) = 10693.098$, $p < 0.001$. The model was able to correctly classify power for significant PMI estimates 93.3% of the time and non-significant PMI estimates 23.2% of the time for an overall percentage of 77.2%.

Table 5 shows logistic regression coefficients, Wald tests, and odds ratios for each of the predictors. Employing a 0.05 criterion of statistical significance, sample size, difference in factor loadings, and number of items had significant partial effects. As sample size increased, the power also increased. The greater the different in factor loadings the more power decreased and the more items on the scale the higher power was. Percent missing and missing mechanism were unrelated to power.

Table 5. *Logistic Regression Predicting Power for PMI*

Predictor	**β**	**Wald χ^2**	**Sig.**	**Odds Ratio**
Sample Size	0.024	4577.103	0.000	1.025
Percent Missing	-0.270	0.275	0.600	0.764
Difference in Factor Loadings	-0.414	19.256	0.000	0.661
Number of Items	0.011	85.225	0.000	1.011
Missing Mechanism	0.011	0.117	0.732	1.011

A test of the full model for MI versus a model with intercept only was statistically significant, $\chi^2(5) = 11908.340$, $p < 0.001$. The model was able to correctly classify power for significant MI estimates 82.7% of the time and non-significant MI estimates 56.7% of the time for an overall percentage of 76.3%.

Table 6 shows logistic regression coefficients, Wald tests, and odds ratios for each of the predictors. Employing a 0.05 criterion of statistical significance, sample size, difference in factor loadings, and number of items had significant partial effects. As sample size increased, power also increased. The greater the different in factor loadings the more power decreased and the more items on the scale the higher power was.

Table 6. *Logistic Regression Predicting Power for MI*

Predictor	**β**	**Wald χ^2**	**Sig.**	**Odds Ratio**
Sample Size	0.025	5013.448	0.000	1.026
Percent Missing	0.005	0.000	0.993	1.005
Difference in Factor Loadings	-0.509	29.648	0.000	0.601
Number of Items	0.008	55.492	0.000	1.009
Missing Mechanism	-0.004	0.020	0.888	0.996

Table 7 shows power for each of the conditions for both PMI and MI computed with a factorial ANOVA. Power was greatest for a sample size of 200 for both PMI and MI. The least amount of power was seen in conditions with a small sample size, more missing data, a larger difference in factor loadings, and a smaller number of items. Power for both PMI and MI was approximately the same for all conditions.

Table 7. *Power for All Conditions Using PMI and MI*

Conditions					PMI	MI
Sample Size	*Percent Missing*	*Difference in Factor Loadings*	*Number of Items*	*Missing Mechanism*		
50	2%	0	8	MAR	0.544	0.514
				MCAR	0.542	0.518
			36	MAR	0.580	0.536
				MCAR	0.584	0.540
		0.2	8	MAR	0.506	0.466
				MCAR	0.502	0.446
			36	MAR	0.574	0.520
				MCAR	0.578	0.524
		0.4	8	MAR	0.502	0.482
				MCAR	0.498	0.472
			36	MAR	0.554	0.490
				MCAR	0.558	0.504
	5%	0	8	MAR	0.546	0.516
				MCAR	0.526	0.516
			36	MAR	0.574	0.522
				MCAR	0.582	0.537
		0.2	8	MAR	0.498	0.464
				MCAR	0.516	0.470
			36	MAR	0.568	0.510
				MCAR	0.574	0.521
		0.4	8	MAR	0.488	0.466
				MCAR	0.496	0.462
			36	MAR	0.554	0.532
				MCAR	0.557	0.492
	10%	0	8	MAR	0.528	0.522

Conditions					**PMI**	**MI**
Sample Size	*Percent Missing*	*Difference in Factor Loadings*	*Number of Items*	*Missing Mechanism*		
				MCAR	0.526	0.512
			36	MAR	0.574	0.540
				MCAR	0.686	0.600
		0.2	8	MAR	0.504	0.458
				MCAR	0.486	0.454
			36	MAR	NA	NA
				MCAR	0.526	0.526
		0.4	8	MAR	0.480	0.466
				MCAR	0.494	0.460
			36	MAR	NA	NA
				MCAR	0.614	0.500
200	2%	0	8	MAR	0.976	0.982
				MCAR	0.976	0.976
			36	MAR	0.988	0.992
				MCAR	0.988	0.990
		0.2	8	MAR	0.974	0.970
				MCAR	0.972	0.970
			36	MAR	0.988	0.992
				MCAR	0.986	0.990
		0.4	8	MAR	0.962	0.952
				MCAR	0.958	0.954
			36	MAR	0.986	0.988
				MCAR	0.988	0.990
	5%	0	8	MAR	0.978	0.978
				MCAR	0.974	0.972
			36	MAR	0.990	0.992

Conditions					PMI	MI
Sample Size	*Percent Missing*	*Difference in Factor Loadings*	*Number of Items*	*Missing Mechanism*		
				MCAR	0.988	0.992
		0.2	8	MAR	0.972	0.968
				MCAR	0.970	0.968
			36	MAR	0.988	0.992
				MCAR	0.986	0.990
		0.4	8	MAR	0.962	0.956
				MCAR	0.958	0.958
			36	MAR	0.986	0.986
				MCAR	0.988	0.988
	10%	0	8	MAR	0.972	0.974
				MCAR	0.978	0.978
			36	MAR	0.990	0.988
				MCAR	0.990	0.993
		0.2	8	MAR	0.966	0.962
				MCAR	0.978	0.972
			36	MAR	0.990	0.990
				MCAR	0.985	0.985
		0.4	8	MAR	0.958	0.950
				MCAR	0.956	0.950
			36	MAR	0.986	0.988
				MCAR	0.988	0.983

* NA indicates this combination of variables was not observed in the data

CHAPTER IV: DISCUSSION

This study was conducted to investigate the effects of MI and PMI for handling missing data for multiple-item questionnaires. Prior to conducting analyses, it was hypothesized that when data were MAR, MI would perform better than PMI with a higher percentage of missing data (Hypothesis 1). Additionally, it was thought that when data was MCAR with unequal factor loadings, MI would outperform PMI with a higher percent missing, smaller sample size, and a small number of items (Hypothesis 3). Neither hypothesis 1 nor hypothesis 3 was supported as PMI slightly outperformed MI in a variety of conditions. However, there is mild support for hypothesis 2, when data is MCAR with equal factor loadings MI and PMI will have the same results regardless of percent missing or number of items, as MI and PMI performed about the same in all conditions, including MCAR with equal factor loadings.

Theoretical Implications

Previous research has shown PMI to be comparable to MI when dealing with item-level missing data, aligning with the results of this study (Savalei & Rhemtulla, 2017; Schafer & Graham, 2002). Schafer and Graham (2002) reasoned that PMI may be a reasonable alternative to MI because the bias in the scales tends to decrease as the scales become more correlated with each other. Additionally, if the items to be averaged can form a single, well-defined domain with a difference in factor loadings of not more than 0.20, and the reliability of the scale is high ($\alpha > 0.70$) Schafer and Graham (2002) and Graham (2012) believe PMI to be a reasonable alternative to MI.

This study found PMI to be surprisingly robust, especially for cases with a large number of variables and small samples in which MI would not converge. Coinciding with the results of this study, Newman (2014) recommends the use of PMI for construct-level analyses due to the complicated nature of using MI on item-level missing data; as experienced in this study, MI does not always converge when there is a large number of variables. Additionally, Graham (2009) recommended that the number of variables should be kept small when sample sizes are small to ensure MI converges. Furthermore, Roth, Switzer, and Switzer (1999) support the use of PMI when estimating results from unidimensional scales as PMI utilizes all of the available data and acknowledges differences across people by using different items to create the imputed means; PMI also yielded the least biased regression coefficients when compared to other deletion and single imputation techniques (Roth, Switzer, & Switzer, 1999).

Based on the results of this study, practitioners would benefit from using PMI over MI when the number of variables is large relative to sample size. Another advantage of PMI over MI is that PMI is relatively easy to conduct and does not require sophisticated statistical software. However, if software is available and the number of variables is small, MI might provide more accurate results as MI produced less parameter bias than PMI in this study when the sample size was 50.

Limitations and Future Directions

The present research sought to provide recommendations for the use of MI and PMI when data are missing on multiple-item questionnaires. However, due to the lack of support for the hypotheses, not many recommendations can be made. Additionally, various problems occurred when simulating MI data, and a good amount of cases had to be eliminated thereby reducing sample size.

This study investigated a relatively simple model with equal item means. Therefore, future research should consider more complex models, with means differing across items as this is more realistic and likely will affect how PMI performs because PMI assumes equal item means. Additionally, the model only tested a correlation which does not consider the mean structure of the variables.

This study did not consider analytic methods of handling item missing data when scale scores are ofinterest, however, Savalei and Rhemtulla (2017) provided a way to use FIML in these cases although it currently may not be practical for practitioners. Because analytic missing data methods such as FIML were not considered; the inclusion of other methods could expand upon these findings and offer more insight into how best to handle missing data.

Conclusions

This research sought to provide recommendations for the use of PMI and MI under varying conditions of sample size, factor loadings, percent missing data, type of missing data, and scale length. Additionally, this study aimed to provide recommendations for the use of MI or PMI when investigating the relationship among scale scores for multi-item questionnaires that measure a single construct. Results showed mixed findings; PMI seemed to outperform MI in the scenarios tested in this study due to nonconvergence, however MI may outperform PMI when the number of variables is small relative to sample size as MI produced the least amount of parameter bias. There were no effects for mean squared error, and power was greatest for PMI for the vast majority of conditions. Therefore, practitioners would benefit from using PMI, unless dealing with data with a small number of variables relative to sample size.

References

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods & Research*, *28*, 301-309.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*, 5-37. doi: 10.1016/j.jsp.2009.10.001.

Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal, 1*, 287-316. doi:10.1080/10705519409539983.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*, 107–112.

Collins, L.M., Schafer, J.L. & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 330-351. doi: 10.1037//1082-989X.6.4.330.

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology, 125*, 175-191. doi: 10.1080/00221309809595542.

Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology, 71*, 500-507. doi:10.1037/0021-9010.71.3.500

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, *8*, 322. doi: 10.1037/1082-989X.8.3.322.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430-457.

Fichman, M., & Cummings, J. M. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods, 6*, 282-308.

Graham, J. W. (2012). *Missing data: statistics for social and behavioral sciences.* New York, NY: Springer.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, *60*, 549-576. doi: 10.1146/annurev.psych.58.110405.08553.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197-218. doi: 10.1207/s15327906mbr3102_3.

Howell, D.C. (2015a, July 13). Treatment of missing data- part 1. Retrieved from https://www.uvm.edu/~dhowell/StatPages/Missing_Data/Missing.html.

Howell, D.C. (2015b, June 26). Treatment of missing data- part 2. Retrieved from https://www.uvm.edu/~dhowell/StatPages/Missing_Data/Missing-Part-Two.html

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. (2nd ed.) Hoboken, NJ: John Wiley & Sons.

Littvay, L. (2009). Questionnaire design considerations with planned missing data. *Review of Psychology*, *16*, 103-113.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods, 17,* 372-411. doi: 10.1177/1094428114548590

Newman, D. A. (2009). Missing data techniques and low response rates: The role of systematic nonresponse parameters. In C. E. Lance, R. J. Vandenberg, C. E. Lance, R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 7-36). New York, NY, US: Routledge/Taylor & Francis Group.

Parent, M. C. (2013). Handling item-level missing data: Simpler is just as good. *The Counseling Psychologist, 41*, 568-600. doi: 0011000012445176.

Pornprasertmanit, S., Miller, P., & Schoemann, A. (2016). *simsem: SIMulated structural equation modeling. R package version 0.5-10.* http://CRAN.R-project.org/package=simsem

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roth, P.L., Switzer III, F. S., Swtizer, 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*, 581-592.

Savalei, V., & Rhemtulla, M. (2017). Normal theory two-stage ML estimator when data are missing at the item level. *Journal of Educational and Behavioral Statistics*, *20*, 1-27. doi: 10.3102/1076998617694880.

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

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*, 545-571. doi:10.1207/s15327906mbr3304_5

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177. doi:10.1037/1082-989X.7.2.147

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*, 1-10. doi:10.1037/a0018082

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*, 317-329. doi:10.1037/1082-989X.6.4.317

Spector, P. E. (1985). Measurement of human service staff satisfaction: Development of the job satisfaction survey. *American Journal of Community Psychology, 13*, 693. Retrieved from http://search.proquest.com.jproxy.lib.ecu.edu/docview/1295894482?accountid=10639

Van Buuren, S., Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1-67. http://www.jstatsoft.org/v45/i03/.

Wayman, J. C. (2003, April). Multiple imputation for missing data: What is it and how can I use it. In *Annual Meeting of the American Educational Research Association, Chicago, IL* (pp. 2-16).

West, S. G. (2001). New approaches to missing data in psychological research: Introduction to the special section. *Psychological Methods*, *6*, 315-316. doi:10.1037/1082-989X.6.4.315

Appendix A: Between-Subjects Effects for Parameter Bias

Effects	***F***	**η_p^2**
Sample Size	$F(1, 33276) = 2.013, p = 0.156$	.000
Percent Missing	$F(2, 33276) = 0.266, p = 0.766$	.000
Factor Loadings	$F(2, 33276) = 25.673, p < .001$	.002
Number of Items	$F(1, 33276) = 108.073, p < .001$	.003
Missing Mechanism	$F(1, 33276) = 0.880, p < .001$	.000
Sample Size * Percent Missing	$F(2, 33276) = 0.660, p = 0.517$	.000
Sample Size * Factor Loadings	$F(2, 33276) = 1.016, p = 0.362$	.000
Sample Size * Number of Items	$F(1, 33276) = 0.488, p = 0.485$	.000
Sample Size * Missing Mechanism	$F(1, 33276) = 1.855, p = 0.170$	.000
Percent Missing * Factor Loadings	$F(4, 33276) = 0.784, p = 0.535$	.000
Percent Missing * Number of Items	$F(2, 33276) = 0.794, p = 0.452$	.000
Percent Missing * Missing Mechanism	$F(2, 33276) = 1.278, p = 0.279$	.000
Factor Loadings * Number of Items	$F(2, 33276) = 1.832, p = 0.160$	.000
Factor Loadings * Missing Mechanism	$F(2, 33276) = 0.153, p = 0.858$	.000
Number of Items * Missing Mechanism	$F(1, 33276) = 1.171, p = 0.279$	.000
Sample Size * Percent Missing * Factor Loadings	$F(4, 33276) = 0.595, p = 0.666$	.000
Sample Size * Percent Missing * Number of Items	$F(2, 33276) = 0.625, p = 0.535$	.000
Sample Size * Percent Missing * Missing Mechanism	$F(2, 33276) = 1.832, p = 0.160$	.000
Sample Size * Factor Loadings * Number of Items	$F(2, 33276) = 1.427, p = 0.240$	.000
Sample Size * Factor Loadings * Missing Mechanism	$F(2, 33276) = 0.040, p = 0.961$	.000
Sample Size * Number of Items * Missing Mechanism	$F(1, 33276) = 2.390, p = 0.122$	.000
Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 0.630, p = 0.641$	.000
Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 0.095, p = 0.984$	.000

Effects	***F***	η_p^2
Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 1.116$, $p = 0.328$	.000
Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.099$, $p = 0.906$	.000
Sample Size * Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 0.725$, $p = 0.574$	.000
Sample Size * Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 0.048$, $p = 0.996$	.000
Sample Size * Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 1.913$, $p = 0.148$	.000
Sample Size * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.031$, $p = 0.970$	.000
Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(4, 33276) = 0.063$, $p = 0.993$	.000
Sample Size * Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.027$, $p = 0.974$	.000

Appendix B: Between-Subjects Effects for Mean Squared Error

Effects	*F*	η_p^2
Sample Size	$F(1, 33276) = 89956.028, p < .001$	.730
Percent Missing	$F(2, 33276) = 1.938, p = 0.144$	.000
Factor Loadings	$F(2, 33276) = 36.092, p < .001$	.002
Number of Items	$F(1, 33276) = 60.685, p < .001$	.002
Missing Mechanism	$F(1, 33276) = 11.42, p < .001$	.000
Sample Size * Percent Missing	$F(2, 33276) = 1.153, p = 0.316$	.000
Sample Size * Factor Loadings	$F(2, 33276) = 23.070, p < .001$	.001
Sample Size * Number of Items	$F(1, 33276) = 23.432, p < .001$	.001
Sample Size * Missing Mechanism	$F(1, 33276) = 4.553, p = 0.033$	.000
Percent Missing * Factor Loadings	$F(4, 33276) = 12.113, p < .001$	.001
Percent Missing * Number of Items	$F(2, 33276) = 1.182, p = 0.307$	.000
Percent Missing * Missing Mechanism	$F(2, 33276) = 7.773, p < .001$	.000
Factor Loadings * Number of Items	$F(2, 33276) = 5.338, p = 0.005$	.000
Factor Loadings * Missing Mechanism	$F(2, 33276) = 0.686, p = 0.504$	.000
Number of Items * Missing Mechanism	$F(1, 33276) = 9.608, p = 0.002$	.000
Sample Size * Percent Missing * Factor Loadings	$F(4, 33276) = 11.301, p < .001$	.001
Sample Size * Percent Missing * Number of Items	$F(2, 33276) = 2.270, p = 0.103$	.000
Sample Size * Percent Missing * Missing Mechanism	$F(2, 33276) = 4.695, p = 0.009$	.000
Sample Size * Factor Loadings * Number of Items	$F(2, 33276) = 17.318, p < .001$	.001
Sample Size * Factor Loadings * Missing Mechanism	$F(2, 33276) = 1.692, p = 0.184$	.000
Sample Size * Number of Items * Missing Mechanism	$F(1, 33276) = 8.720, p = 0.003$	.000
Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 11.910, p < .001$	.001
Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 0.520, p = 0.721$	.000

Effects	***F***	**η_p^2**
Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 7.202$, $p = 0.001$	.000
Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 1.369$, $p = 0.254$	.000
Sample Size * Percent Missing * Factor Loadings * Number of Items	$F(4, 33276) = 15.346$, $p < .001$	.002
Sample Size * Percent Missing * Factor Loadings * Missing Mechanism	$F(4, 33276) = 0.673$, $p = 0.611$	.000
Sample Size * Percent Missing * Number of Items * Missing Mechanism	$F(2, 33276) = 7.641$, $p < .001$	.000
Sample Size * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.192$, $p = 0.826$	.000
Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(4, 33276) = 0.457$, $p = 0.767$	.000
Sample Size * Percent Missing * Factor Loadings * Number of Items * Missing Mechanism	$F(2, 33276) = 0.100$, $p = 0.905$	.000