2022

# Using the Fraction of Missing Information (FMI) in Selecting Auxiliary Variables to Impute Missingness in Confirmatory Factor Analysis (CFA)

Dareen Taha Alzahrani
*University of Denver*

# Using the Fraction of Missing Information (FMI) in Selecting Auxiliary Variables to Impute Missingness in Confirmatory Factor Analysis (CFA)


## Abstract

This study aimed to investigate the effectiveness of using the fraction of missing information (FMI) to select auxiliary variables in imputing missing data in confirmatory factor analysis (CFA). This was done by conducting two studies (a simulation study and an empirical study). A Monte Carlo simulation technique was used to compare the performance and the effect of the restrictive strategy based on FMI and the inclusive strategy on parameter estimate bias and parameter estimate efficiency. The missing data mechanisms, missing data proportion, correlation strength between the analysis variables and auxiliary variables, and the inclusive and restrictive strategies were assessed in the simulation study for their impact on three dependent variables: bias, mean squared error (MSE), and confidence interval coverage of parameters. In addition, the difference between the inclusive and restrictive strategies was examined using empirical data where missing data were designed with two levels of missingness (15% and 30%) and two missingness mechanisms to assess their impact on parameter estimate bias, gain in efficiency, and power. In the simulation study, factorial ANOVAs were conducted to assess the design factors and their interactions' effects. The results indicated that the design factors had no impact on study. The two strategies showed no impact on parameter estimate bias for the empirical data. Yet, the restrictive strategy based on the FMI outperformed the inclusive strategy in terms of gains in efficiency and power. Thus, there is an initial support of using the FMI to evaluate the auxiliary variables.



## Document Type

Dissertation

## Degree Name

Ph.D.

## Department

Quantitative Research Methods

## First Advisor

Duan Zhang

## Second Advisor

Kathy Green

## Third Advisor

Jeanine Coleman

## Keywords

Confirmatory factor analysis, Fraction of missing information, Missing data

## Subject Categories

Other Statistics and Probability | Physical Sciences and Mathematics | Statistical Methodology | Statistics and Probability

# Using the Fraction of Missing Information (FMI) in Selecting Auxiliary Variables to Impute Missingness in Confirmatory Factor Analysis (CFA)

---

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

---

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Dareen Alzahrani

June 2022

Advisor: Dr. Duan Zhang

Author: Dareen Alzahrani
Title: Using the Fraction of Missing Information (FMI) in Selecting Auxiliary Variables to Impute Missingness in Confirmatory Factor Analysis (CFA)
Advisor: Dr. Duan Zhang
Degree Date: June 2022

## Abstract

This study aimed to investigate the effectiveness of using the fraction of missing information (FMI) to select auxiliary variables in imputing missing data in confirmatory factor analysis (CFA). This was done by conducting two studies (a simulation study and an empirical study). A Monte Carlo simulation technique was used to compare the performance and the effect of the restrictive strategy based on FMI and the inclusive strategy on parameter estimate bias and parameter estimate efficiency. The missing data mechanisms, missing data proportion, correlation strength between the analysis variables and auxiliary variables, and the inclusive and restrictive strategies were assessed in the simulation study for their impact on three dependent variables: bias, mean squared error (MSE), and confidence interval coverage of parameters. In addition, the difference between the inclusive and restrictive strategies was examined using empirical data where missing data were designed with two levels of missingness (15% and 30%) and two missingness mechanisms to assess their impact on parameter estimate bias, gain in efficiency, and power. In the simulation study, factorial ANOVAs were conducted to assess the design factors and their interactions' effects. The results indicated that the design factors had no impact on study. The two strategies showed no impact on parameter estimate bias for the empirical data. Yet, the restrictive strategy based on the FMI outperformed the inclusive

strategy in terms of gains in efficiency and power. Thus, there is an initial support of using the FMI to evaluate the auxiliary variables.

**Acknowledgments**

First and foremost, I would like to thank Allah for giving me the strength and the knowledge to pursue my study abroad. Without his support and mercy, this achievement would not have been possible.

Actively pursuing any Ph.D. involves the entire family and advisors, and they were all my cheerleaders. I should begin with an acknowledgment of my parents. Thank you for encouraging me always to chase my dream with passion, no matter what kind of struggles I encounter. I am also deeply grateful to my lovely husband, Mohsen. Thank you for giving me relentless support and love during the completion of this journey. I must also acknowledge my children, Loreen and Baraa, who grew up watching their mother spend long hours studying and who learned first-hand the value of education and the need for persistence. I would like to express my indebtedness to all my siblings who always believed in me, even when I had my doubts.

Special thanks go to my advisor, Dr. Duan Zhang, for reviewing endless drafts of this dissertation and offering invaluable suggestions to improve the quality of my research and writing. I also want to express my heartfelt thanks to Dr. Kathy Green, not only for her continuous support throughout my Ph.D. years but also for providing truly warm welcomes and considerable encouragement that meant a lot to me as an international student. I am also grateful to Dr. Jeanine Coleman for her excellent teaching. It was an honor to be her student.

## Table of Contents

Chapter One: Introduction ..... 1

Chapter Two: Literature Review ..... 6
- Missing Data Mechanism ..... 6
- Missing Data Treatment Techniques ..... 7
- Auxiliary Variable Definition ..... 9
- The Importance of Auxiliary Variables ..... 10
- Strategies of including auxiliary variables into an imputation model ..... 14
- The Fraction of Missing Information (FMI) ..... 19
- Obtaining the FMI from FIML ..... 22
- Interpretation of the FMI ..... 23
- The Use of FMI in the Literature ..... 24
- Factors Affecting FMI ..... 25
- Using FMI for Auxiliary Variables Selection in Missing Data Imputation ..... 26
- Research Questions ..... 29
- The Study's Significance ..... 30

Chapter Three: Method ..... 33
- Simulation Study ..... 33
  - Research Design ..... 33
  - Conditions ..... 34
  - Data Generation ..... 36
  - Generating Missing Values ..... 36
  - Evaluation Criteria ..... 39
- Empirical Study ..... 41
  - Empirical Data ..... 41
  - Data Manipulation ..... 42
  - Generating MAR ..... 45
  - Analytic Strategy ..... 46
  - Study Outcomes ..... 46

Chapter Four: Results ..... 48
- The Simulation Study's Results ..... 48
  - Selection Procedure ..... 48

Parameter Estimate Bias ..... 50
Mean Squared Error ..... 60
Confidence Interval Coverage (CIC) ..... 70
FMI Properties ..... 78
The Empirical Study's Results ..... 81
The Process of Selecting Auxiliary Variables ..... 81
Inclusive and Restrictive Strategies ..... 81
Distinguishing Type A and B Auxiliary Variables ..... 82
Bias ..... 82
Efficiency ..... 86

Chapter Five: Discussion ..... 90
The Main Findings ..... 90

References ..... 101
Appendix A ..... 110
Appendix B ..... 112
Appendix C ..... 114

## List of Tables

Chapter Three: Method ..... 32
Table 1 ..... 44
Table 2 ..... 45

Chapter Four: Results ..... 48
Table 3 ..... 52
Table 4 ..... 53
Table 5 ..... 54
Table 6 ..... 55
Table 7 ..... 56
Table 8 ..... 57
Table 9 ..... 58
Table 10 ..... 59
Table 11 ..... 62
Table 12 ..... 63
Table 13 ..... 64
Table 14 ..... 65
Table 15 ..... 66
Table 16 ..... 67
Table 17 ..... 68
Table 18 ..... 69
Table 19 ..... 71
Table 20 ..... 72
Table 21 ..... 75
Table 22 ..... 77
Table 23 ..... 79
Table 24 ..... 80
Table 25 ..... 84
Table 26 ..... 73
Table 27 ..... 75
Table 28 ..... 75
Table 29 ..... 79
Table 30 ..... 80
Table 31 ..... 83
Table 32 ..... 84
Table 33 ..... 85

## Chapter One: Introduction

It is very common that researchers in the social and behavioral sciences collect item-level data using questionnaires, where they face the problem of missing data in the analysis phase of their studies. Missing data on the items are inevitable, where responders may refuse to answer sensitive items or accidentally skip items. The prevalence of missing data encouraged methodologists to propose and improve statistical methods to handle the missing data in a way that will not preclude researchers from obtaining valid inferences. Several methods have been developed to deal with missing data that can be classified as traditional and modern methods (Enders, 2010), where the modern techniques such as full information maximum likelihood (FIML) and multiple imputation (MI) outperformed traditional methods like listwise deletion or mean imputation when data are missing at random (MAR) or missing not at random (MNAR) based on simulation studies' results (Enders, 2001; Enders & Bandalos, 2001; Madley-Dowd et al., 2019).

In addition, the literature on missing data suggests adding auxiliary variables to the imputation model can increase the power and reduce the estimation bias of the model parameters (Collins et al. 2001; Enders, 2010). Most of the studies recommend the inclusive strategy of using auxiliary variables, where the researchers add all possible

covariates to the imputation model, as these variables will improve the estimation, especially if they are highly correlated with the outcome (Collins et al., 2001; Raykov &West, 2016; Yoo, 2009), even when these variables are incomplete (Eners, 2008; Wang & Deng, 2016). On the other hand, some studies found that auxiliary variables did not positively impact the imputation model (Hardt et al., 2012; Mustillo, 2012). Yuan and Savalei (2012) found that auxiliary variables might increase the standard error (SE) with a small sample size, high missing data proportion, and non-normal distribution of the auxiliary variables. Also, it was found that the impact of adding auxiliary variables to the imputation model could be harmful to the precision of the model estimate by increasing the SE and the fraction of missing information (FMI) when the auxiliary variables are incomplete (Madley-Dowd et al., 2019).

Thus, the inclusive strategy might not be the solution to overcome missingness in some situations, and it can increase the bias in the model (Hardt et al., 2012). However, Collins et al. (2001) justified using the inclusive strategy over the restrictive strategy, which includes a selected subset of auxiliary variables in the imputation model, as including all available auxiliary variables will decrease the chance of omitting any cause of missingness.

This study examined the effect of using the FMI to evaluate the effectiveness of auxiliary variables, as it is assumed that the FMI may help researchers to select the optimal auxiliary variables and avoid the problems of convergence, difficulty in implementations, and omitting the auxiliary variables that may add information to the

model. The FMI can inform researchers about the amount of information that can be returned by including auxiliary variables, so researchers can use the restrictive strategy to select and include the auxiliary variable that reduces the FMI.

Based on previous works that compared the inclusive and restrictive strategies, we can infer the recommendation of using the inclusive strategy as a way to enhance the plausibility of meeting the MAR assumption. This recommendation comes from the concern that using the restrictive strategy might lead to excluding an auxiliary variable that is related to the missingness, which in turn leads to non-ignorable missingness. This study proposes using the FMI as a tool to evaluate the candidate auxiliary variables. It was found that the FMI can be used as an indicator for the missing mechanism; specifically, if the FMI is greater than the missingness rate, a missing data mechanism may be nonignorable (Nishimura et al., 2016). The researcher is aware of only two studies that examined the use of the FMI to select auxiliary variables (Andridge et al., 2015; Madley-Dowd et al., 2019); however, Andridge et al. (2015) applied the method using empirical data where the results cannot speak to the impact of using this method on the bias. While Madley-Dowd et al. (2019) examined the proposed method using a simulation design, they did not test this method with a measurement model and condition like the incomplete auxiliary variables that might affect the FMI performance. Thus, this study examined the case of having incomplete auxiliary variables, which reflects a condition that researchers often face. This study evaluated using the FMI as a tool to select the auxiliary variables and examined the performance of the FMI under different conditions.

Due to the growing application of CFA by researchers in social sciences (Guo et al., 2009; Jackson et al., 2009; Martens, 2005; Raykov et al., 1991), this study utilized this model. In a review of articles published in the *Personality and Social Psychology Bulletin* during the years 1996, 1998, and 2000, 12% of the articles used CFA. A similar trend has been noticed in social work research where Guo et al. (2009) found that the CFA is the most common type of structural equation modeling (SEM) that has been used in the top-ranked social work journals published from 2001 to 2007.

Giving that quantitative research in the social sciences relies on using scales to measure latent constructs, it is important to examine the use of the FMI to include the auxiliary variables in the imputation process of the CFA model.

By conducting this study, the researcher aims to contribute to the literature of missing data by extending the previous research on the identification strategies of effective auxiliary variables. The results of this simulation study can help guide applied researchers in modeling CFA with an incomplete dataset by identifying and selecting useful auxiliary variables among a set of candidate auxiliary variables, which in turn could help to enhance the plausibility of the MAR assumption.

This study examined the effectiveness of using the FMI to select auxiliary variables in imputing missing data in the CFA model through a Monte Carlo simulation. This was done by comparing the effect of the restrictive strategy based on FMI and inclusive strategy on parameter estimate bias and parameter estimate efficiency. In addition, this study answered the research question of the influence of missing data mechanisms, missing data proportion, correlation strength between the analysis variables

and auxiliary variables, and the use of the restrictive strategy and the inclusive strategy on the parameter estimate bias and parameter estimate efficiency

## Chapter Two: Literature Review

### Missing Data Mechanism

There are two terms in the missing data literature that are often used interchangeably, which are missing data patterns and missing data mechanisms. It is important to distinguish between these two terms as they have different meanings (Enders, 2010). While the missing data pattern represents the location of the missing data, the missing data mechanism refers to the relationship between the data and missingness and explains the reason for missingness (Enders, 2010). The focus of the literature is more on the missing data mechanisms as the pattern of the missing data is no longer important since the modern missing data methods maximum likelihood (ML) and multiple imputation (MI) are well suited for any pattern (Enders, 2010).

Missing data mechanisms are the foundation for Rubin's (1976) missing data theory. Rubin (1976) classified missing data into three categories based on the probability of missing data relying on the measured variables in the data set. The process of Missing Completely at Random (MCAR) happens when the missingness is independent of the observed and the missing value in the data (Enders, 2010). This mechanism has a restricted assumption as it assumes the missingness is completely unrelated to the measured variables, which rarely happens unless the missingness is planned (Little & Rhemtulla, 2013). Missing at Random (MAR) exists when there is a systematic relationship between the probability of missing data and one or more measured variables.

Missing not at random (MNAR) occurs when the probability of missingness depends on the missing value (Enders, 2010). The difference between the mechanisms is based on how the distribution of missingness is related to the data values.

**Missing Data Treatment Techniques**

Several methods have been developed to deal with missing data that can be classified as traditional and modern methods (Enders, 2010). Traditional methods include mean substitution, listwise deletion, pairwise deletion, and single imputation methods. Modern missing data methods include full information maximum likelihood (FIML) and multiple imputation (MI). Reviewing the practice of handling missing data in empirical studies showed that the listwise method is the most common technique that is used in psychological journals using the CFA model (Jackson et al., 2009), epidemiological journals using multi-item instruments (Eekhout et al., 2012), prevention research (Lang & Little, 2016), cluster randomized trials (Fiero et al., 2016), and in medical journals using randomized controlled trials (Wood et al., 2004). While it is common to use complete case analysis, simulation studies revealed that modern techniques such as FIML and MI outperformed the traditional methods when data were MAR or MNAR (Enders, 2001; Enders & Bandalos, 2001; Madley-Dowd et al., 2019). A more detailed review of the modern methods is provided in the next section.

**Multiple Imputation**. MI was proposed by Rubin (1978), which generates multiple independent imputed data sets and then conducts inferences by averaging across them. This process takes into account the uncertainty of parameter estimates that is caused by missing data (Enders, 2010). Compared to FIML and traditional methods, MI

is more complicated as it involves multiple steps. Appling MI is done through three phases: an imputation phase, an analysis phase, and a pool phase. In the first phase, multiple imputed datasets are generated with missing values filled in. Then, the analysis step is used to fit the hypothesized model to each imputed dataset. In the final step, the pooling step, results across imputed data sets are combined to produce the results (Enders, 2010).

**The Full Information Maximum Likelihood.** In structural equation modeling (SEM), FIML is a recommended method to handle missing data. This method was known to produce unbiased and efficient parameter estimates under the assumption of MAR and normality (Enders & Bandalos,2001)

FIML was designed to use individual likelihood functions by maximizing the sum of the log of "case wise" likelihood functions (Enders, 2010). This was expressed as:

$$logLi = -\frac{Ki}{2} log\,(2\pi) - \frac{1}{2} log\,|\sum i| - \frac{1}{2}\,(Yi - \mu i)\,T\sum -1\,(Yi - \mu i) \quad (1)$$

where $K_i$ represents the number of complete data points for case $i$, $Y_i$ denotes the data for the case, the mean vector $\mu_i$ and the covariance matrix $\Sigma_i$ represent parameters unique to each case (Enders, 2010). Thus, FIML allows the dimensions of the mean vector and covariance matrix to be person-specific. The term direct-maximum likelihood is often used in the literature to describe the FIML as it is considered as a direct approach to handling missing data. In other words, with FIML, the missing values are not imputed, instead, FIML estimates the model parameters and the associated standard errors directly from the observed data of each individual case.

It is clear from the description above the difference between MI and FIML in handling missing data. While MI allows for different imputation and analysis models, the FIML, which is considered a model-based approach, handles the missing data within a single iterative step. This gives MI an advantage and flexibility when the researcher includes auxiliary variables in the imputation model. Despite these differences, both methods produced the same results under the condition that both had the same imputation and analysis models when the number of imputations was sufficiently large (Collins et al., 2001; Graham et al., 2007). In addition, the assumption of missingness ignorability was required for both methods.

In the context of structural equation modeling, SEM, researchers had tended to prefer the FIML method to handle missing data (e.g., Enders, 2010; Enders & Bandalos,2001; Raykov, 2005; Savalei & Bentler, 2009). It seems that this preference was based on convenience as the FIML is available in most SEM software (e.g., *Amos, LISREL, and Mplus*) rather than a theoretical reason (Savalei & Rhemtulla, 2012). Since the goal of this study is to handle missing data using auxiliary variables in the CFA model, the FIML method will be the focus of this study as it represents a challenge in incorporating a large number of auxiliary variables, and as it is available in most SEM programs.

**Auxiliary Variable Definition**

As has been discussed, methodologists recommended using modern methods in handling missing data to reduce estimation bias and increase power under less restrictive assumptions compared with traditional approaches (Allison, 2003; Collin et al.,2001;

Enders, 2010). The effectiveness of these modern methods depended on meeting the MAR assumption, which means that all variables that are related to the missingness or variables with missing data should be included in the process of imputing the missing data (Allison, 2003; Collin et al.,2001; Enders, 2010). These kinds of variables are called auxiliary variables that are generally not of direct interest to researchers, but they are used to keep the assumptions about ignorability of the missing data plausible.

**The Importance of Auxiliary Variables**

To maximize the benefit of using modern methods, methodologists recommend using auxiliary variables that are related to the missing data but not of substantive interest for the main analysis. MI and FIML allow for including auxiliary variables in different ways to help meet the MAR assumption, which reduces the bias of the estimates. A common recommendation is to include the available auxiliary variables in the imputation process. Despite the recommendation, little work has been done in evaluating and assessing the benefits of the candidate auxiliary variables.

In most cases, with incomplete datasets, researchers had some extra variables that might not be of direct interest to the research analysis, but they can provide indirect information about the likely values of missing data. Using this information could effectively recover missing data to the degree of the association between these auxiliary variables and missingness or incomplete variables (Collins et al., 2001; Enders, 2010).

Collins et al., (2001) classified auxiliary variables into three categories: A type A auxiliary variable is correlated with the incomplete variable and missingness, type B is the variable that correlates with the incomplete variable but not the missingness, and type

C is the variable that correlates with neither the incomplete variable nor the missingness. They found that including type A auxiliary variables reduced bias, and both type A and B improved the efficiency of the estimated parameter.

In line with Collins et al. (2001) results, Howard et al. (2015) found that including auxiliary variables that were related to missingness using MI reduced the bias compared with omitting these variables from the imputation model.

Enders (2008) has extended this proof with incomplete auxiliary variables, while in practice, it is expected that auxiliary variables themselves would have missing data. Using a FIML based model, Enders (2008) conducted a simulation study to examine the impact of using incomplete auxiliary variables under different conditions, including a missing data mechanism for the auxiliary variable (MCAR and MNAR), moderate and strong associations between the auxiliary variable and the model variables, and different missing data rates for the auxiliary variable (25% and 50%). The complete auxiliary variables outperformed the incomplete ones, and the inclusion of an auxiliary variable that was highly correlated with model variables improved the parameter estimation. In his study, the proportion and the mechanism of the missing data of the auxiliary variables showed little impact on the bias. However, the missing data pattern was the factor that most influenced bias in the regression model parameters. For instance, the proportion of observations missing for both the auxiliary and dependent variables was the cause of the most observed bias in parameter estimation; specifically, when about 15% of the cases were missing for both the auxiliary and dependent variables, the most extreme bias values were observed. Thus, Enders (2008) recommended checking missing data patterns when

including incomplete auxiliary variables in the imputation model. Overall, this study recommends including the auxiliary variable even if it has a substantial proportion of missing data.

Wang and Deng (2016) drew the same conclusion after expanding Enders's (2008) study by testing the impact of different missing mechanisms for both the outcome and the auxiliary variables under the FIML using the CFA model. They varied the conditions of the simulation study to include different sample size (100, 200, 500, and 1000), missingness rates (5%, 10 %, 15 %, and 20 %), missingness mechanism combinations for both the outcome and the auxiliary variables (MCAR-MCAR, MCAR-MAR, MCAR-MNAR, MAR-MCAR, MAR-MAR, and MAR-MNAR ), number of auxiliary variables (1, 3, and 5), and the magnitude of the association between auxiliary variables and the outcome (low, moderate, and high). They found that including auxiliary variables, even when they were incomplete, improved the parameter estimates in most cases. Based on the results of this study, researchers recommended including auxiliary variables even if they have low correlations with variables of interest in the model.

Another simulation study that explored the role of auxiliary variables in CFA using MI was conducted by Yoo (2009). The sample size (200, 500), missingness rates (10%, 20%), missingness mechanism combinations (MCAR–MCAR, MCAR–MAR, MAR–MAR, MCAR– MNAR, MAR–MNAR, and MNAR–MNAR), missingness types (linear or convex), and the absence or presence of the auxiliary variables were used to examine their impacts on convergence failure, bias, standard error, and confidence interval coverage of parameters. In this simulation study, the researcher used highly

correlated auxiliary variables that ranged between 0.48 and 0.72, type A auxiliary variables that were associated with incomplete variables and the indicators of missingness, and type B auxiliary variable that were associated only with the incomplete variable.

The results showed that MI performed very well under ignorable missingness regardless of the type of missing data, missingness combination, and strategy for including auxiliary variables. In addition, MI showed robust results with the nonignorable linear type of missingness but not with the nonignorable convex type, which resulted in unacceptable bias. However, using the inclusive strategy with the nonignorable convex type improved the estimation and solved the problem. The difference in the performance of the inclusive and restrictive strategies depended on the magnitude of the correlation between auxiliary and model variables as she found that the inclusive strategy outperformed the restrictive strategy using variables with correlations between 0.48 and 0.72. Thus, Yoo (2009) recommended including the auxiliary variables in the imputation model; especially, when the correlation between auxiliary variables and model variables is relatively high (e.g., $\geq 0.48$).

Despite the effectiveness of using auxiliary variables that were established in the previous simulation studies, some studies questioned the usefulness of auxiliary variables. In a simulation study, Mustillo (2012) argued that the benefits of auxiliary variables appear to be minimal. She tested the impact of using different types of auxiliary variables, three levels of missing data (10%, 20%, and 30%), and two missingness

mechanisms (MCAR, MAR). She came to the conclusion that including any type of auxiliary variables did not appreciably impact the coefficient bias or efficiency.

However, Mustillo (2012) acknowledged the advantage of including auxiliary variables in increasing the power. In other words, she mentioned that including type A and B variables in the MAR models reduced the SE from 0.0240 to 0.0226, which led to an increase in the sample size from 1988 to 2242. Thus, even when the auxiliary variables appear to be ineffective in improving efficiency, they can increase the power.

**Strategies of including auxiliary variables in an imputation model**

As concluded from prior research, auxiliary variables can play important roles in the imputation process by decreasing bias and increasing power which improves the efficiency of the estimation process (Enders, 2010; Rhemtulla & Hancock, 2016). Generally, in the imputation literature, the approaches taken to choose auxiliary variables can be divided into two categories: the inclusive strategy and the restrictive strategy. Methodologists have generally recommended the inclusive strategy, which encourages the generous use of all available auxiliary variables from a dataset rather than the restrictive strategy, that includes a selected subset of auxiliary variables (Collins et al., 2001; Enders, 2010). The rationale for recommending the inclusive strategy is that including all possible covariates in the imputation model reduces the chance of omitting an important cause of missingness (Collins et al., 2001). In addition, Collins et al. (2001) pointed out that when including too many auxiliary variables, the worst results are neutral.

On the other hand, from a practical view, some researchers recommended the restrictive strategy as including too many auxiliary variables can introduce implementation difficulties with ML estimation, especially with SEM (Enders, 2010). Since ML handles missing data problems simultaneous with model-fitting, incorporating a large number of auxiliary variables can be difficult to implement using a saturated correlates approach (Enders, 2010). Thus, using a few auxiliary variables can be adequate to satisfy the MAR assumption assuming that any potential cause of missing data is explained by the selected auxiliary variables.

Even though Enders (2010) mentioned the superiority of the inclusive strategy over the restrictive strategy, he acknowledged the difficulty of applying this method and the consequences of including too many auxiliary variables. Taking into account the benefits of using auxiliary variables, it is important to consider how best to select influential auxiliary variables. Enders (2010) noted that the literature review could be a good source to provide researchers with ideas for important auxiliary variables to include. Many researchers also proposed alternative methods to select auxiliary variables. There is no rule of thumb for selecting a particular set of auxiliary variables. However, there are many efforts from methodologists in proposing and testing methods for choosing a restrictive set of auxiliary variables. The first recommended approach is to look at the correlation of auxiliary variables with the research variables and select auxiliary variables with high correlations. There is no rule of thumb for the value of the correlation coefficient, but Graham (2009) recommended adding auxiliary variables that correlated

with research variables at $r \geq 0.50$, and Collins et al. (2001) suggested using auxiliary variables that correlated about $r \geq 0.40$ with the variables to be imputed.

One method that can be used is comparing mean differences of the complete and incomplete data groups among potential auxiliary variables. If this comparison results in a significant difference between the two groups for a given variable, this means that this variable should be included as an auxiliary variable. This can be tested using independent *t*-tests after creating complete and incomplete data groups for each variable. However, selecting auxiliary variables based on mean differences ignores the covariance information, which may prevent researchers from effectively implementing the restrictive auxiliary variable strategy (Enders, 2010). Another criticism of this method is that using hypothesis testing to identify effective auxiliary variables may not be helpful as it will be sensitive to the sample size (Raykov & Marcoulides, 2014).

In responding to this criticism, Raykov and Marcoulides (2014) proposed using latent variable modeling to obtain point and interval estimations of mean and variance differences on potential auxiliary variables across the groups of cases with complete data and with missing data on an outcome of interest. They mentioned that this approach could be useful to find measures on which cases with missing data on a variable of interest are different from those with complete data on the outcome. Applying this method requires a large sample size, normal data, and it should be used along with an appropriate method for estimating the correlations between the candidate auxiliary variables and the outcome in a joint effort to detect effective auxiliary variables. One drawback of Raykov and Marcoulides's (2014) study is that they only describe the

application of the proposed method without showing the effect of using this method on the outcome estimation. It would be advantageous to perform additional comparisons between the proposed method to select auxiliary variables and an inclusive strategy that includes all available auxiliary variables to see the effect of the proposed method.

Another study by Raykov and West (2016) in the area of identifying useful auxiliary variables proposed a method to evaluate the potential auxiliary variables by estimating the correlation between the outcome and the auxiliary variables and the correlation between the auxiliary variables and the outcome's residual. They stated that effective auxiliary variables could be the covariates that show notable point estimates of their correlations with the incomplete outcome after accounting for its relationships with independent variables. Raykov and West (2016) suggested using this proposed method in tandem with the group difference examination approach that was described in Raykov and Marcoulides's (2014) study. Both procedures can be considered complementary to each other in identifying effective auxiliary variables (Raykov & West, 2016).

They demonstrated the proposed method using real data where they estimated the bivariate and semi-partial correlations between six auxiliary variables and the outcomes using latent variable modeling. Based on the results, they included four auxiliary variables that had acceptable bivariate and semi-partial correlations with the outcome. They claimed that by using this imputed model, they were able to use the whole dataset in a prediction model that had three significant predictors. The prediction model explained 74% of the outcome variance. However, stating that all the predictor variables in the model contributed significantly in explaining the outcome variance is not sufficient

to conclude that using the proposed method is beneficial without comparing the results with the unimputed model or a model that includes all available auxiliary variables.

A study by Howard et al. (2015) proposed the idea of using principal components analysis (PCA) to reduce the number of the included auxiliary variables. Conducting a simulation study, the researchers compared the performance of the inclusive strategy and the PCA strategy under different conditions. Their study examined the impact of different magnitudes of correlations (six levels), eight rates of missing data ranging from 10% to 80% in increments of 10%, missing data mechanism (linear and nonlinear MAR), and 39 sample sizes ranging from 50 to 1,000 on parameter bias, root mean squared error, and confidence interval coverage. Consistent with previous studies, the results showed bias estimates when the missingness was linear and auxiliary variables were not included in the model, which resulted in MNAR by omitting variables that have a linear relationship with missingness. In line with Collins et al. (2001), this study found that including auxiliary variables when the missingness was nonlinear removed the parameter bias with no need to include the interaction variable. In addition, the PCA approach exhibited comparable performance to the inclusive strategy across a linear and nonlinear cause of missingness, which indicated the effectiveness of PCA in reducing the number of auxiliary variables without increasing bias. However, it is important to mention that PCA requires complete data, which means researchers need to take a further step by imputing incomplete auxiliary variables before using PCA. In this study, researchers imposed 10% MCAR on the auxiliary variables to address this problem. However, the missingness rate and mechanism of auxiliary variables can be more complicated in real data. Also, in

applying this method to empirical data, it was not clear how many principal components should be included in the imputation model.

**The Fraction of Missing Information (FMI)**

The concept of information, in statistics, means the amount of information available for inference about parameters (Rhemtulla & Hancock, 2016; Savalei & Rhemtulla, 2012). The inverse relationship between the information available and the size of the standard error of a specific parameter helps to understand the concept of missing information. The less information we have to estimate a parameter, the less we know about this parameter, which results in a larger standard error (Rhemtulla & Hancock, 2016; Savalei & Rhemtulla, 2012).

This concept of missing information was introduced by Orchard and Woodbury (1972) where they demonstrated that the available information from an incomplete dataset is equal to complete information minus missing information. Thus, they introduced the fraction of missing information as the ratio of missing information to complete information. Orchard and Woodbury (1972) discussed the impact of missing data on sampling variability in the context of maximum likelihood, which can be explained as large maximum likelihood standard errors because of inflated variance from missing data. Therefore, the missed information can influence the efficiency of parameters by increasing their standard errors (Rhemtulla & Hancock, 2016; Savalei & Rhemtulla, 2012).

As the missing information principle is grounded in likelihood theory, we will follow Savalei and Rhemtulla's (2012) illustration of the missing information principle in the context of ML.

The likelihood of the complete data can be expressed as:

$$L\ (\theta|\ X) = L_1\ (\theta|Y)\ f_2\ (Z|Y,\theta) \qquad (2)$$

where *X*, *Y*, and *Z* represent the complete, observed, and missing data for *n* cases, respectively. From the previous equation, the likelihood of the complete data can be defined as the likelihood of observed data times the conditional density of the missing data given observed data (Savalei & Rhemtulla, 2012).

Based on the previous equation, we can partition the derivative of the log-likelihood as:

$$\frac{\partial \log L\ (\theta|X)}{\partial \theta'} = \frac{\partial \log L1(\theta|Y)}{\partial \theta'} + \frac{\partial \log f2(\mathrm{Z}|\mathrm{Y},\theta)}{\partial \theta'} \qquad (3)$$

Defining the information matrix as the covariance matrix of the score vector (Rao, 2002, as cited in Savalei & Rhemtulla, 2012 ), and based on Equation 2, the information about the parameters $\theta$ that would be available from complete data can be expressed as:

$$\mathrm{J_X} = \mathrm{cov}\left(\frac{\partial \log L\ (\theta|X)}{\partial \theta\prime}\right) \qquad (4)$$

The information about the parameters $\theta$ that would be available from the observed data is the covariance of

$$\mathrm{J_Y} = \mathrm{cov}\left(\frac{\partial \log L1(\theta|Y)}{\partial \theta\prime}\right) \qquad (5)$$

The information that would be available from missing data about $\theta$ is the covariance of

$$J_{X|Y} = J_{Z|Y} = \text{cov}\left(\frac{\partial \log f2(Z|Y,\theta)}{\partial \theta'}\right) \quad (6)$$

Thus, Orchard and Woodbury's (1972) missing information principle can be expressed as:

$$J_X = J_Y + J_{X|Y} \quad (7)$$

If the data are complete, we can estimate the ML parameter $\hat{\theta}_{ML}$ by maximizing the complete data likelihood function $L(\theta|X)$ assuming a normal distribution for $\hat{\theta}_{ML}$ with a covariance matrix equal to the inverse of the complete data information matrix.

$$\alpha \text{ cov}\left(\sqrt{n}\,\hat{\theta}_{ML}\right) = J_X^{-1} \quad (8)$$

When there are missing data, the $\hat{\theta}_{FIML}$ can be estimated by maximizing the incomplete data likelihood function $L_1(\theta|Y)$ with the condition that the missingness is either MCAR or MAR, assuming a normal distribution for $\hat{\theta}_{FIML}$ with a covariance matrix equal to the inverse of the incomplete data information matrix.

$$\alpha \text{ cov}\left(\sqrt{n}\,\hat{\theta}_{FIML}\right) = J_Y^{-1} \quad (9)$$

Based on Orchard and Woodbury's (1972) missing information principle, we can infer that $J_X^{-1} \leq J_Y^{-1}$ because the diagonal of $J_Y^{-1}$ are expected to be larger than $J_X^{-1}$ due to the standard error when there is missing data. The influence of missing data can be seen in the greater variability of $\hat{\theta}_{FIML}$ comparing with $\hat{\theta}_{ML}$ (Savalei & Rhemtulla, 2012).

Since the diagonal of $J_X^{-1}$ and $J_Y^{-1}$ indicates the parameter estimates' variances of the complete and incomplete data, respectively, the fraction of missing information for a given parameter $\theta_j$ can be stated as:

$$\lambda_j = 1 - \frac{\{J_X^{-1}\}_{jj}}{\{J_Y^{-1}\}_{jj}} = 1 - \frac{SE_{j,c}^2}{SE_{j,o}^2} \quad (10)$$

Where $SE_{j,c}^2$ is the standard error of $\hat{\theta}_{ML}$ based on complete data, and $SE_{j,o}^2$ is the standard error of the FIML estimate $\hat{\theta}_{FIML}$ based on incomplete data. From this equation, we can infer that the FMI quantifies the amount of a parameter's information that is lost due to missingness.

**Obtaining the FMI from FIML**

The traditional way to estimate the FMI requires using MI. However, with the preference of using the FIML method among SEM researchers, Savalei and Rhemtulla (2012) demonstrate how to obtain an estimate of the FMI from FIML. They noted that the FMI estimate from FIML is superior to the obtained estimate from MI when the number of imputations is small. They described four steps that can be applied in many SEM software packages (e.g., *Mplus, R, and EQS*) to estimate the FMI for each parameter. It requires running the SEM program twice; first fitting the model using FIML to the original incomplete data to obtain the standard error for each parameter of interest. Then, running the same model using FIML to the model-implied means and covariance matrix and vector of means obtained from the previous step as input into the complete data ML routine. In the next step, the FMI for each parameter can be estimated as one minus the ratio of the corresponding squared standard errors from each output.

**Interpretation of the FMI**

Savalei and Rhemtulla (2012) suggested many approaches to interpret the FMI as it relates to different statistical concepts. One approach is to relate it to the relative efficiency, which measures the amount of information loss due to missingness and relates negatively to the FMI. In other words, relative efficiency can be computed as the ratio of the sampling variances of the complete data estimates to the incomplete data estimates (Rhemtulla et al., 2014), which can be expressed as:

$$\frac{SE_{j,c}^2}{SE_{j,o}^2} \tag{11}$$

Recalling the equation for estimating the FMI

$$\lambda_j = 1 - \frac{\{J_X^{-1}\}_{jj}}{\{J_Y^{-1}\}_{jj}} = 1 - \frac{SE_{j,c}^2}{SE_{j,o}^2} \tag{12}$$

We can see how FMI and relative efficiency are related as the FMI is one minus the relative efficiency. Thus, the FMI can be interpreted as the loss of efficiency in the estimation of a particular parameter (Savalei & Rhemtulla, 2012).

Additionally, the FMI could be interpreted as the loss of statistical power caused by missing data (Savalei & Rhemtulla, 2012). The effective sample size that would have achieved the same efficiency for a parameter with complete data can be defined as:

$$N_j^* = N(1 - \lambda_j) \tag{13}$$

The more information available to estimate a parameter, the smaller its standard error, and thus the greater the statistical power of a significance test on that parameter (Rhemtulla & Hancock, 2016). The lower the FMI, the less information lost, and the higher the quality of estimation (Enders, 2010; Savalei & Rhemtulla, 2012).

**The Use of FMI in the Literature**

The fraction of missing data has been used for different reasons; it was used as a tool for evaluating survey quality during data collection to assess the risk of non-response, so actions can be taken based on the FMI estimate (Wagner, 2010; Wagner, 2012). For instance, based on a pre-specified FMI, researchers monitored the change that happened over time for the FMI. When the FMI estimate was high, researchers could intervene to obtain additional respondents (Wagner, 2010). Another study was conducted by Andridge and Little (2011) used the FMI to evaluate the risk of non-response once the data collection was completed.

In addition, the FMI was widely used in MI to determine the number of imputations to achieve reasonable relative efficiency (Bodner 2008; Harel, 2007; Rubin, 1987; Von Hippel, 2020). It was suggested by Rubin (1987) that the number of imputations (2-10) is sufficient in most cases when researchers use MI. However, Bodner (2008) questioned this guideline and emphasized that researchers should consider the FMI to determine the needed number of the imputations. With a higher FMI, more imputations were necessary for stable estimations of the *p* values, confidence interval half-widths, and estimated FMI (Bodner, 2008).

In general, it is recommended that researchers and practitioners report the FMI when they deal with missingness as an index of the estimation accuracy and a diagnostic tool for the impact of missingness (Enders, 2010; Lang & Little, 2016; Rhemtulla & Hancock, 2016; Savalei & Rhemtulla, 2012).

**Factors Affecting FMI**

Before we discuss the factors that affect the FMI, we should distinguish between the FMI and the proportion of missing data. The proportion of missing data does not reflect the influence of missingness on the accuracy of parameter estimates. However, the amount of missing information can be an informative diagnostic tool that communicates how much the estimation of a parameter was affected by missingness (Enders, 2010; Orchard & Woodbury, 1972; Rhemtulla & Hancock, 2016; Savalei & Rhemtulla, 2012). On the other hand, the amount and pattern of missingness can affect the information available to estimate parameters; specifically, the missing information is influenced by which variables are missing (Rhemtulla & Hancock, 2016). The rates of univariate and pairwise missingness can lead to a higher FMI and standard error (Madley-Dowd et al., 2019). The missingness mechanism can influence the FMI as the FMI will be bounded by the largest rates of missingness when the mechanism is MCAR, while it can be higher than the missingness rate when the mechanism is MAR (Savalei & Rhemtulla, 2012). The intercorrelation among the variables in the model is another factor that can affect the FMI (Andridge & Thompson, 2015). If we have highly correlated variables, the missingness in one of them will not affect the missing information in the data when including the other variables. Since the FMI depends on the specific set of variables used in the model, including extra variables that are highly correlated with the missingness can improve estimation and result in a lower FMI (Enders, 2010). Thus, it is recommended to add auxiliary variables that correlate highly with the imputed variable. Additionally, if we have multiple auxiliary variables, the correlation among them could affect the FMI. For

example, adding two auxiliary variables that are highly collinear into an imputation model might not decrease the FMI even if they are highly correlated with the imputed variable (Andridge & Thompson, 2015).

**Using FMI for Auxiliary Variable Selection in Missing Data Imputation**

As mentioned above, most of the research suggests using the inclusive strategy for including auxiliary variables in the imputation models. However, it is not feasible to include all variables in practice as they may introduce problematic situations such as multicollinearity (Mustillo, 2012), or implementation difficulty, especially with FIML, which requires incorporating the auxiliary variables with the saturated correlate model (Enders, 2010). In other words, including a large number of auxiliary variables in the MI analysis is easier because the auxiliary variables will play a role only in the imputation phase, and there will be no need to include them in the analysis phase. Thus, MI can handle a larger number of auxiliary variables than a maximum likelihood analysis (Enders, 2010). Additionally, in some cases, including all the available covariates may lead to an increase in computation time or in failure to converge (White et al., 2011), especially with a small sample size or longitudinal design when the number of covariates approaches the number of cases (Hardt et al., 2012). Therefore, Enders (2010) suggested limiting the number of included auxiliary variables to use the most useful variables correlated with incomplete analysis variables. This indicates the importance of selecting informative auxiliary variables to help recover the missed information. It is assumed that the FMI can be a useful tool to select the most informative auxiliary variables that can decrease the fraction of missing information by recovering the lost information.

In a study investigating the FMI use to select potential auxiliary variables, Andridge et al. (2015) proposed using proxy pattern□mixture (PPM) to obtain a maximum likelihood estimate of the FMI. Then, they selected the auxiliary variables based on the FMI. The goal of using the PPM to estimate the FMI is that PPM helped to avoid the instability of imputation-based estimates and helped estimate the FMI under both MAR and MNAR assumptions. The use of PPM reduced a set of auxiliary variables to a single "proxy" variable which then was used for imputation under either a bivariate normal model or a bivariate gamma model.

The variable selection procedure to select the best auxiliary variables can be done using the forward selection procedure (Andridge & Thompson, 2015). They used a forward selection procedure to include and evaluate auxiliary variables, where the FMI was estimated for each auxiliary variable one at a time. The variable that produced the smallest FMI was then selected into the imputation model. The other auxiliary variables were entered one at a time, and the variable producing the largest decrease in FMI was selected. In other words, they added auxiliary variables to the imputation model, and they monitored how this changed the FMI. When the extra variables did not improve the imputation model, they removed them. This procedure continued until changes in the FMI were not large enough to be worth adding additional auxiliary variables.

They administrated the use of the proposed method in empirical data. They found that adding auxiliary variables to the imputation model decreases the FMI in most cases, except for variables with low associations with the outcome and strong associations with

missingness. They utilized a real data set, so they did not address the performance of auxiliary variables on the estimate's accuracy.

In a recent study, Madley-Dowd et al. (2019) compared the use of MI and complete case analysis (CCA). They examined the utility of the FMI as a guide to select the most beneficial auxiliary variables. They generated a normally distributed dataset that consisted of an outcome, Y, an independent variable, X, and 11 auxiliary variables, Z, with different magnitudes of association between Y and Z. All the variables were only correlated with the outcome, and the outcome was the only variable with missingness. Missingness proportions were manipulated at different levels (0%, 5%, 10%, 20%, 40%, 60%, 80%, and 90%).

The simulation study results showed an association between the increase of the empirical SE and the FMI of any given proportion of missing data, especially at high missingness proportions. Adding auxiliary variables to the model resulted in a decrease in the FMI value and SE. This indicated the gain in efficiency by including auxiliary variables in the imputation model. In addition, it was noticed that at different proportions of missing data but similar FMI values, the SEs were approximately the same, which demonstrated that the FMI is a good measure of the estimate's precision.

Moreover, they used empirical data to demonstrate the comparison between MI and CCA and the application of FMI. The results showed that using MI reduced bias and improved efficiency, and including auxiliary variables improved the estimation regardless of the proportion of missing data. However, the empirical example results indicated that the impact of adding additional auxiliary variables to the imputation model was not

consistent. For example, introducing an auxiliary variable to the imputation model that did not reduce the FMI can negatively affect the precision of model estimates. This was noticed with the incomplete auxiliary variable. The researchers suggested including auxiliary variables based on the FMI to ensure that these variables are adding information to the model.

However, a closer look at the study design reveals that the simulation design's conditions did not reflect real empirical data where auxiliary variables are expected to correlate to each other and be incomplete too.

The researcher is aware of only the previous two studies that examined FMI use to select auxiliary variables (Andridge et al., 2015; Madley-Dowd et al., 2019). However, each one suffers from certain weaknesses. Thus, this study extended previous studies by examining FMI use in selecting incomplete auxiliary variables as well as comparing the effect of the restrictive strategy based on the FMI and inclusive strategy on the parameter estimate using FIML methods in the CFA context.

**Research Questions**

Two research questions were addressed in this study.

1. Is including a smaller set of auxiliary variables based on FMI (restrictive strategy) as beneficial as including all possible auxiliary variables (inclusive strategy) for the parameter estimate bias and parameter estimate efficiency in CFA?
2. How are the parameter estimate bias and parameter estimate efficiency influenced by the missing data mechanism, missing data proportion,

correlation strength between the analysis variables and auxiliary variables, and the used strategy of including auxiliary variables (restrictive and inclusive)?

**The Study's Significance**

Based on previous work, we can classify the approaches taken to choose auxiliary variables into two categories: the inclusive strategy and the restrictive strategy. Previous simulation studies that compared the inclusive and restrictive strategies' effectiveness recommended using the inclusive strategy to enhance the plausibility of meeting the MAR assumption (Collins et al., 2001; Yoo, 2009). Practically, using a few auxiliary variables can adequately satisfy the MAR assumption, assuming that the selected auxiliary variables explain any potential cause of missing data.

This study proposes using the FMI to evaluate the candidate auxiliary variables to include the most effective auxiliary variables in the imputation model as the FMI can help identify a covariate that recovers some of the missed information. Researchers found that the missing data mechanism was nonignorable when the FMI was greater than the missingness rate (Nishimura et al., 2016). Therefore, the FMI can be used to indicate the missing data mechanism.

It is assumed that using the restrictive strategy based on the FMI would help overcome methodologists' concerns about omitting an auxiliary variable related to missingness (Collins et al., 2001; Yoo, 2009). The researcher is aware of only two studies that examined the FMI use to select auxiliary variables (Andridge et al., 2015; Madley-Dowd et al., 2019). However, Andridge et al. (2015) applied the method using empirical

data where the results cannot speak to the impact of using this method on bias estimation. While Madley-Dowd et al. (2019) examined the proposed method using a simulation design, the simulation design's conditions did not reflect real empirical data where variables are expected to correlate to each other and be incomplete. Thus, this study aims to expand previous works in different ways.

First, the aforementioned studies used auxiliary variables that were completely observed. However, researchers should expect to deal with incomplete auxiliary variables in practice. Therefore, this study simulated the uninvestigated situations where all auxiliary variables are incomplete.

Second, lacking in the previously mentioned studies is generalization to other models such as CFA, as both studies used regression analysis. Therefore, this study applied FMI to select auxiliary variables for the imputation in a basic two-factor measurement model using FIML as an imputation method.

Third, even though Andridge et al. (2015) and Madley-Dowd et al. (2019) examined FMI use to select auxiliary variables, they did not compare the performance of the inclusive strategy and the restrictive strategy that select covariates based on the FMI. Based on empirical data, Andridge et al. (2015) used the PPM to reduce a set of auxiliary variables to a single "proxy" variable used for the imputation.

On the other hand, Madley-Dowd et al. (2019) designed a simulation study focused on comparing the CCA and the MI. They applied FMI use in choosing the auxiliary variables. Thus, this study compared the inclusive strategy that includes all the auxiliary variables into the imputation model and the restrictive strategy that selects auxiliary

variables that reduce the FMI. In addition, Madley-Dowd et al.'s (2019) simulation study only manipulated the correlation between auxiliary variables and the outcome. Still, they set the correlation among auxiliary variables to be zero. This study reflected the real situations where auxiliary variables are expected to be correlated.

This study aimed to contribute to the literature by extending the previous research on selecting informative auxiliary variables. Since this study employs CFA, which is considered a sub-model of SEM, its results can help applied researchers who use this model in their analyses. As quantitative research usually employs instruments to collect data, researchers are supposed to present evidence for the instrument's psychometric properties. Validity and reliability are the core psychometric properties that should be examined and reported for any measure in the study. Construct validity, in particular, is used to establish evidence about the degree to which the instrument measures what it is designed to measure by examining the measure's structure and the relationship between the factors and the indicators. The CFA model is one of the methods used to examine construct validity. Thus, it is not surprising to find common use of this model in the field of social science (Guo et al., 2009; Jackson et al., 2009; Martens, 2005; Raykov et al., 1991), as it plays an important role in developing and establishing validity evidence of the developed instruments. As a sub-model of SEM, CFA can be used even in more complex models and serve as the measurement model in the structural model. Given this model's importance and common use, this study can help researchers deal with the inevitable problem of having missingness on items. Particularly, it is hoped that this study can provide researchers with a more effective strategy to identify and select auxiliary variables to include in the imputation mod.

## Chapter Three: Method

The FMI use to select the auxiliary variables was assessed using a Monte Carlo simulation study and an existing empirical dataset.

### Simulation Study

#### *Research Design*

Following Yoo (2009), the analysis model reflected two correlated factors and six indicators. Specifically, three indicators were used per latent variable, with each indicator loading on only one factor. This model is similar to the population model used by Enders (2008), Enders and Peugh (2004), Wang and Deng (2016), and Savalei and Bentler (2009). The correlation between the two latent variables was 0.40, and the factor variance was fixed at 1 (Enders & Peugh, 2004; Yoo, 2009). The factor loadings were 0.70, and the error variances were 0.51 as defined by 1 minus the squared factor loadings (Enders, 2008). Factor loadings ($\leq .7$) are commonly seen in studies measuring concrete constructs such as cognitive abilities, reasoning abilities, or attitudes (McNeish et al., 2017). The current study's sample size was fixed at 500, as it is common to see this sample size in SEM simulation studies (e.g., Enders, 2008; Wang & Deng, 2016; Yoo, 2009; Yuan & Savalei, 2014).

This analysis model was chosen as it shares features with many models in the existing CFA simulation literature, and its two-factor, six-indicator model overlaps partially or completely with models appearing in many notable CFA simulation studies (e.g., Enders, 2008; Enders & Peugh, 2004; Savalei & Bentler, 2009; Yoo, 2009).

Besides, a CFA model is generally considered the fundamental step in a more complex structural equation model (Kline, 2016).

The number of the generated auxiliary variables in this study were ten auxiliary variables with a mean of 0 and a standard deviation of 1. The correlation among auxiliary variables was set at 0.4. Following Collins et al. (2001) and Yoo (2009), type A and type B auxiliary variables were generated. While type A auxiliary variables are variables that associate with incomplete variables and missingness indicators, type B auxiliary variables are variables that associate only with the incomplete variables (Collins et al., 2001).

### *Conditions*

The design factors chosen for this study are the ones that have been identified as important in affecting the FMI's performance. The levels of each condition were selected to reflect realistic situations.

**Correlation Strength.** As has been stated in previous studies, the magnitude of the association between the auxiliary variables and analysis variables can influence the imputation process (Collins et al., 2001; Enders, 2008; Enders, 2010; Enders & Peugh, 2004; Yoo, 2009) and the FMI performance (Andridge & Thompson, 2015; Savalei & Rhemtulla, 2012). Methodologists recommended using auxiliary variables that correlate

about $r \geq 0.40$ with the variables to be imputed (Collins et al., 2001; Enders, 2010; Yoo, 2009). However, since simulation studies should be designed to imitate real data studies (Burton et al., 2006), the manipulated levels of the correlations between the auxiliary variables and the analysis variables were determined based on two steps.

First, the researcher generated two independent datasets; one of these datasets had ten variables that can be considered auxiliary variables while the other represented a measurement model of two latent factors and six observed items. The correlations among auxiliary variables were manipulated at different levels (0.3, 0.4, 05, and 0.8) to observe any pattern of correlations between auxiliary variables and factor indicators. The result of estimating correlation matrices of 250 datasets is that the two datasets were correlated at a very low level that never exceeded 0.177, regardless of the magnitude of the correlation among auxiliary variables.

In the second step, multiple publicly accessible large-scale datasets were consulted to examine the actual and expected relationship between scale items and covariates. Based on these datasets, most auxiliary variables correlated with the items at a low level, few covariates correlated moderately with some items (0.4-0.54), and the correlations among auxiliary variables varied between 0.2-0.5.

Based on the previous exploratory steps, the correlations between the auxiliary variables and the analysis variables were manipulated at two levels: low (0.3) and moderate (0.6), which the researcher judged to be realistic.

**Missing Data Proportion.** This study investigated two levels of missing data (15% and 30%) (Savalei & Bentler, 2009; Yoo, 2009). The same missing portion was imposed

on each simulation study's indicators and auxiliary variables. The motivation for imposing missingness on the auxiliary variables is based on the idea that real data will have incomplete auxiliary variables (Enders, 2008; Hardt et al., 2012; Yoo, 2009) as well as the results of Madley-Dowd et al.'s (2019) study that showed an increase in the FMI with adding an auxiliary variable that had 38% missingness.

**Missing Data Mechanisms.** The MCAR and MAR were used in this study. The choice to examine these two missing data mechanisms is influenced by Savalei and Rhemtulla (2012), as they indicate that the FMI can be affected by missing mechanisms. Since the FIML assumes that data are MCAR or MAR, the MNAR mechanism was not examined in this study.

### *Data Generation*

The Mplus program 8.6 (Muthén & Muthén, 2017) was used to generate the data, which consisted of 32 conditions with 100 replications for each condition. Then, the R program (R Core Team, 2021) was utilized to run the CFA model using the Lavaan package (Rosseel, 2012). Finally, SPSS (IBM Corp., 2021) was used to analyze the results of the 32 conditions. The following sections describe the analysis steps taken in this study in more detail.

### *Generating Missing Values*

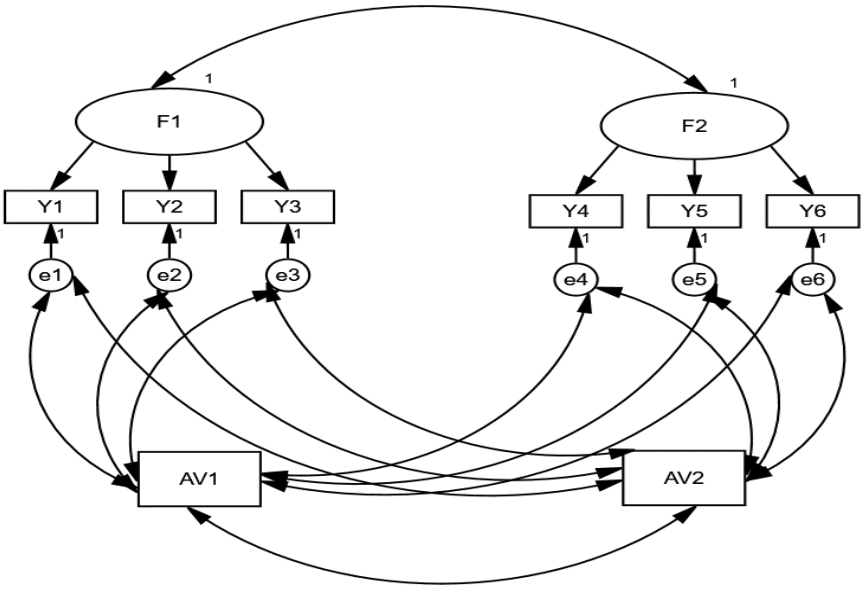
Mplus was used to facilitate missing data generation for both missing data mechanisms (MCAR and MAR). The missing values were created for both the items and auxiliary variables. To create MCAR, Mplus allows users to specify the proportion of data

missing for each variable in the model using the PATMISS option. For the MAR mechanism, logistic regression models were used to generate incomplete data on the items and the auxiliary variables based on values of the first auxiliary variable (X1). Thus, X1 was considered the type A auxiliary variable, and the rest of the auxiliary variables represented type B auxiliary variables.

Once the full dataset was generated, R was used to apply the CFA model using both strategies of including auxiliary variables (inclusive and restrictive based on FMI). To run the CFA model using the auxiliary variables. The Lavaan package was used. Lavaan allows the saturated correlates model to incorporate auxiliary variables into the FIML estimation routine (Graham, 2003). This model was implemented according to the following rules: (a) auxiliary variables must be correlated with the exogenous manifest variables, (b) auxiliary variables must be correlated with residuals of all predicted manifest variables, and (c) auxiliary variables must be correlated with one another. However, since this study utilized the CFA model with two latent factors, the first rule does not apply to this analysis (see Figure 1).

**Figure 1**

*The Saturated Correlates Model*

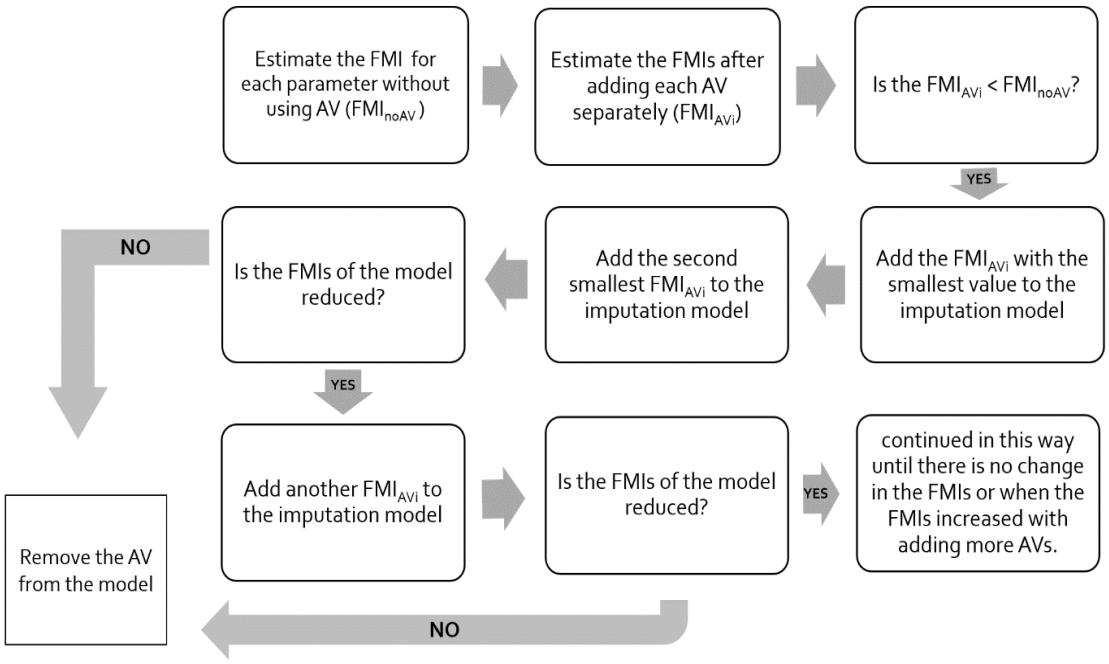



The variable selection procedure to select the best auxiliary variables was done using the forward selection procedure (Andridge & Thompson, 2015). In the beginning, the FMI was estimated for each parameter by fitting the CFA model without adding any auxiliary variable ($FMI_{no\ AV}$). This model served as the baseline model to which subsequent models could be compared. Then, the FMI was estimated using each auxiliary variable one at a time ($FMI_{AVi}$). Then, the variable that produces the smallest FMI was selected to be entered into the imputation model. All the auxiliary variables were entered one at a time, and the procedure continued until no further decrease was observed in the FMI (Figure 2). Then, each simulation's results (i.e., parameter estimates and standard errors) were saved into an Excel sheet to be analyzed. Considering that two strategies included auxiliary

variables (inclusive and restrictive based on the FMI), 64 Excel files were imported into SPSS for descriptive and inferential analyses.

**Figure 2**

*The Forward Selection Procedure*




### *Evaluation Criteria*

**Parameter Bias**. The current study's main outcome of interest was parameter bias. The average parameter estimate from each simulation condition is compared to the true population parameter resulting in an estimate of raw bias. To make this result more comparable with previous research, raw bias was reported as a percentage relative to the true parameter value (Collins et al., 2001; Enders & Bandalos, 2001; Yoo, 2009), which can be expressed as:

$$\% \text{ Bias} = \frac{\theta\, est - \theta 0}{\theta 0} * 100 \quad (14)$$

Where the numerator represents the raw bias, which is the difference between the average parameter estimate across replications within a design cell ($\theta_{est}$) and the population value ($\theta_0$). According to Muthen et al. (1987), bias smaller than 10% to 15% is considered acceptable in most SEM contexts.

**Mean Squared Error (MSE).** MSE represents the average squared difference between a parameter estimate and the true population value. It can be decomposed into the sum of the squared bias and the variance of the estimate. Thus, when a parameter estimate is unbiased, MSE quantifies the sampling variance, or efficiency of an estimate. For a biased outcome, the measure serves to quantify the overall accuracy of an estimate, combining bias and sampling variance.

**Confidence Interval Coverage (CIC).** CIC can be computed as the percentage of replications in a design cell that leads to 95% confidence intervals containing the population value. Following Collins et al. (2001), a coverage value below 90% is considered problematic, indicating an inflated Type I error rate. Optimally, a parameter estimate is expected to be covered 95% of the time, which is 95 times out of the 100 replications in this study. Therefore, the confidence intervals of each parameter were obtained with each replication's results, and frequency analyses were applied to track the number of replications in which the true value is included in the 95% confidence interval for each parameter.

In summary, a Monte Carlo simulation was run to compare the inclusive strategy and the restrictive strategy based on FMI. Specifically, there were 32 combinations

(2x2x2x2x2): two missing data mechanisms (MCAR, MAR) for the items and the auxiliary variables, two missing proportions (15%, 30%) for the items and the auxiliary variables, two levels of the magnitude of the associations between the auxiliary variable and the model variables (low and moderate). For each cell, there were 100 replications. These factors' impacts were examined in terms of bias, MSE, and confidence interval coverage of parameters. Analysis of variance (ANOVA) was conducted to determine the design factors' impact on the dependent variables.

## Empirical Study

In addition to the simulation work, the FMI use to select the auxiliary variables was investigated in an empirical data example to demonstrate their actual data performances.

### *Empirical Data*

The dataset was based on a sample of the Midlife in the United States study (MIDUS) wave three collected in 2013. The data include 3,291 participants and hundreds of variables such as questions about cognitive functioning, economic recession experiences, optimism, coping, stressful life events, caregiving, and variables about demographic information. The data are publicly available from the Inter-university Consortium for Political and Social Research at https://www.icpsr.umich.edu/web/ICPSR/studies/36346 and https://www.icpsr.umich.edu/web/ICPSR/studies/37095.

The Brief Test of Adult Cognition by Telephone (BTACT), administered in MIDUS wave three, was used for the CFA model in this study.

**The Brief Test of Adult Cognition by Telephone (BTACT).** This battery consists of seven tests. The first latent construct, Executive Function, is measured by five tasks:

Stop and Go Switch Task (SGST), 30 Seconds and Counting Task (30-SACT), Number Series, Category Verbal Fluency, and Backward Digit Span. The second latent construct, Episodic Memory, is measured by two tasks: word list recall immediate and word list recall delayed. The measurement model's cognitive function framework was adapted from Lachman et al. (2014), where the seven tasks were the indicators in the measurement model. This battery has been used to measure cognitive function in many studies (e.g., Bhattacharyya et al., 2020; Charles et al., 2020; Hartanto et al., 2020) and to measure Executive Function (Roiland et al., 2015; Lin et al., 2014).

***Data Manipulation***

Given that the missingness rate in the MIDUS dataset was very low (less than 5%), the missing data were removed, and the complete data (N=2673) was used as the population data. Then, different missingness rates were imposed on the scale items (15%-30%), and two missing mechanisms (MCAR & MAR) were generated using the missMethods package in R (Rockel, 2020). Then, Mplus was used to fit the CFA model and estimate each parameter's FMI.

**The Auxiliary Variables and Correlation.** Since the simulation study manipulated the auxiliary variables' correlation with the items at two levels (low and moderate) and they were all at the same level in each condition, it was interesting to use auxiliary variables that vary in their correlation with the imputed items. The potential variables in the data set of MIDUS wave three that can be used as auxiliary variables were age and nine variables about the Personality in Intellectual Contexts (PIC). The correlation between age and the BTACT items ranged between .12 and .4 (Table 1), and the magnitude

of the correlation between personality variables and the BTACT items ranged between .01 and .3 (Table 2). These ten variables represented the low and moderate levels of the correlation magnitude with the items. To include a covariate that can correlate in a higher level with the imputed items, seven items of the BTACT scales from wave two were used as auxiliary variables. The correlation between wave two items and wave three items ranged between .12 and .8 (Table 1). This makes the total number of the used auxiliary variables 17 variables. One of these variables was completely observed, and the rest of the variables had missing data.

**Table 1**

*Correlations Between the Imputed Items, BTACT Wave 2 Items, and Age.*

| Variables | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y1 | 1 | | | | | | | | | | | | | | |
| Y2 | .8 | 1 | | | | | | | | | | | | | |
| Y3 | .32 | .31 | 1 | | | | | | | | | | | | |
| Y4 | .29 | .24 | .21 | 1 | | | | | | | | | | | |
| Y5 | .28 | .23 | .35 | .39 | 1 | | | | | | | | | | |
| Y6 | .28 | .25 | .34 | .43 | .55 | 1 | | | | | | | | | |
| Y7 | -.19 | -.14 | -.15 | -.26 | -.24 | -.37 | 1 | | | | | | | | |
| X1 | .47 | .45 | .24 | .24 | .23 | .23 | -.13 | 1 | | | | | | | |
| X2 | .46 | .51 | .24 | .23 | .21 | .23 | -.12 | .77 | 1 | | | | | | |
| X3 | .26 | .24 | .45 | .17 | .31 | .3 | -.13 | .32 | .31 | 1 | | | | | |
| X4 | .22 | .19 | .19 | .65 | .36 | .37 | -.17 | .24 | .22 | .19 | 1 | | | | |
| X5 | .22 | .18 | .32 | .33 | .66 | .46 | -.23 | .25 | .22 | .36 | .35 | 1 | | | |
| X6 | .22 | .19 | .32 | .38 | .51 | .84 | -.31 | .23 | .22 | .32 | .38 | .45 | 1 | | |
| X7 | -.18 | -.15 | -.21 | -.27 | -.32 | -.45 | .39 | -.18 | -.16 | -.19 | -.28 | -.29 | -.47 | 1 | |
| Age | -.31 | -.31 | -.19 | -.33 | -.31 | -.41 | .12 | -.19 | -.22 | -.1 | -.23 | -.18 | -.36 | .24 | 1 |

*Note*. Y1-Y7= the imputed variables of the BTACT from wave 3, X1-X10 = auxiliary variables from the BTACT scale of wave 2.

**Table 2**

*Correlations Between the Imputed Items and PIC Variables.*

| Variables | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y1 | 1 | | | | | | | | | | | | | | | |
| Y2 | .8 | 1 | | | | | | | | | | | | | | |
| Y3 | .32 | .31 | 1 | | | | | | | | | | | | | |
| Y4 | .29 | .24 | .21 | 1 | | | | | | | | | | | | |
| Y5 | .28 | .23 | .35 | .39 | 1 | | | | | | | | | | | |
| Y6 | .28 | .25 | .34 | .43 | .55 | 1 | | | | | | | | | | |
| Y7 | -.19 | -.14 | -.15 | -.26 | -.24 | -.37 | 1 | | | | | | | | | |
| P1 | -.13 | -.11 | -.11 | -.13 | -.14 | -.16 | .11 | 1 | | | | | | | | |
| P2 | .09 | .08 | .06 | .08 | .05 | .02 | -.01 | .02 | 1 | | | | | | | |
| P3 | .14 | .14 | .19 | .18 | .31 | .25 | -.15 | -.14 | .13 | 1 | | | | | | |
| P4 | .11 | .1 | .08 | .11 | .14 | .11 | -.09 | -.17 | .4 | .32 | 1 | | | | | |
| P5 | -.04 | -.06 | -.05 | -.03 | -.02 | -.07 | .06 | .17 | -.07 | .001 | -.07 | 1 | | | | |
| P6 | .07 | .06 | .07 | .04 | .04 | .004 | -.02 | -.06 | .61 | .11 | .45 | -.01 | 1 | | | |
| P7 | .16 | .13 | .16 | .19 | .27 | .24 | -.17 | -.2 | .19 | .36 | .39 | -.05 | .22 | 1 | | |
| P8 | .13 | .13 | .1 | .1 | .1 | .1 | -.06 | -.13 | .36 | .18 | .52 | -.1 | .44 | .32 | 1 | |
| P9 | .18 | .2 | .13 | .12 | .1 | .09 | -.1 | -.16 | .33 | .2 | .37 | -.16 | .36 | .31 | .38 | 1 |

*Note*.Y1-Y7= the imputed variables of the BTACT, P1-P9= auxiliary variables from the PIC.

***Generating MAR***

For MCAR, 15% and 30% of the responses in each item were removed randomly using the missMethods package (Rockel, 2020). For the MAR, the missingness on items

was imposed based on age. In other words, age represents type A auxiliary variables where missingness likely increases with higher value of age. To validate the generated missing data, SPSS was used to determine the missingness rates and the missing mechanisms of each data file. Appendix C shows the missing counts and rates for each item across conditions. To check the missingness mechanisms, dummy variables were created for missing indicators, and differences between the two groups in the auxiliary variables were explored. In MCAR conditions, some items showed differences in some auxiliary variables, however, these differences were not practically significant as the effect sizes were *d*s <.2. Thus, the MCAR condition was assumed with paying attention to the impact of including and removing these auxiliary variables in the analysis. For MAR conditions, the missingness in all the items related to age with effect sizes *d*s ≥.3. Besides, the missingness in some items showed differences in some auxiliary variables but with effect sizes of ds <.2.

### *Analytic Strategy*

Similar to the simulation study, the forward selection procedure was applied to select the auxiliary variables and form an imputation model.

### *Study Outcomes*

***Bias.*** Using the complete data as the population parameters, the percentage bias was calculated for each parameter (factor loadings, factor correlation, and error variances) (Table 26). The threshold range that states bias smaller than 10% is acceptable was used in this analysis (Muthen et al., 1987).

***Efficiency***. Relative efficiencies have been used to compare estimators in simulation studies (Enders & Bandalos, 2001; Savalei & Bentler, 2009). Following Savalei and Bentler (2009), relative efficiency was used to determine the gain in efficiency due to including auxiliary variables, where the variance of the parameter estimates based on the model that did not use auxiliary variables was used over the variance of the same parameter based on the model that used one of the strategies (inclusive – restrictive).

***Power***. The FMI was used to show the differences between the two strategies regarding the loss of statistical power caused by missing data.

The equation for the effective sample size was

$$N_j^* = N\,(1 - \lambda_j) \qquad ((13)$$

This indicates the sample size that would have achieved the same efficiency for a parameter with complete d

# Chapter Four: Results

## The Simulation Study's Results

### *Selection Procedure*

In selecting the included auxiliary variables based on the FMI, the researcher faced the difficulty of choosing the auxiliary variable as the FMI estimated values for all the parameters did not differ notably between auxiliary variables, even with type A auxiliary variables. As mentioned in the third chapter, this simulation study's design included two types of auxiliary variables: a type A auxiliary variable, correlated with the incomplete variable and missingness, and a type B variable that correlates with the incomplete variable but not the missingness. Based on the research design, the first auxiliary variable was set to be type A, and the rest of the auxiliary variables were type B. Including a type A auxiliary is important as it changes the assumption of the missingness to the ignorable situation. In the conditions where the items' missing mechanism is MAR, including a type A auxiliary variable changed the non-ignorable situations to ignorable. Thus, it was expected that the type A auxiliary variable would reduce the FMI more than the type B auxiliary variable. However, only in six conditions out of 16, the type A auxiliary variable showed more reduction in the FMI of the factor correlation and error variances than the type B auxiliary variables. Yet, the difference between type A and type B auxiliary variables in reducing the FMI was very small.

To demonstrate this, a condition with the missingness rate of 15% at both items and auxiliary variables and items with MAR mechanism and auxiliary variables with

MCAR mechanism is used as an example. Compared to the FMI of the model that did not include auxiliary variables, adding the type A auxiliary variable decreased the factor correlation's FMI to .006, and the factor loadings and the error variances FMIs decreased about 0.005. The effect of using the type B auxiliary variable showed that the factor correlation's reduction in the FMI was 0.0058, the reduction in the FMIs of the factor loadings was 0.005, and the error variances were 0.004. This example showed how small the difference was between the impact of the auxiliary variables on the FMI.

In the subsequent steps, when more auxiliary variables were added into the imputation model across the board, the FMI decreased in all parameters. Thus, using the restrictive strategy, which includes auxiliary variables based on the FMI, suggested that the imputation model should include the ten auxiliary variables. Because of the above reasons, the results indicated that in 27 conditions, there was no difference between the inclusive strategy and the restrictive strategy based on the FMI since both suggested including all the auxiliary variables. In four conditions, the restrictive strategy based on FMI suggested including eight auxiliary variables, and only in one condition, the imputation model included six auxiliary variables.

The following section will show the parameter estimate bias results, MSE, and CIC.

***Parameter Estimate Bias***

In line with Enders and Bandalos (2001) and Yoo (2009), the percentage bias was calculated in reference to the true population parameter for factor loadings, factor correlation, and error variances (see equation 14). Muthen et al. (1987) suggested that bias smaller than 10% to 15% is acceptable in most SEM contexts. Thus, this threshold range will be referenced in the following discussion.

Factorial analysis of variance (ANOVA) was conducted to test the design factors' impact (missing data rate, missing data mechanism, the magnitude of the correlation, and the strategy of including the auxiliary variables) on the parameter estimate bias.

Only the factors and interactions that were identified as statistically significant ($p \leq 0.05$) and with large enough effect sizes ($\eta^2 > 0.01$) (Cohen, 1988) are presented and discussed with additional details. The rationale is that when the effect size is very small, even if the p-value is significant, the difference is often of limited to no practical significance. As suggested by Cohen (1988), a rule of thumb was employed to assess $\eta^2$ effect size: small when $\eta^2 > 0.01$, medium when $\eta^2 \geq 0.06$, and large when $\eta^2 \geq 0.14$.

A total of 13 factorial ANOVA models were conducted to test the impact of the five factors on the parameter estimate bias of six-factor loadings, six error variances, and the factor correlation.

Tables 3-6 show the factor loading parameters' mean parameter estimate bias percentage by the missing data rate, missing data mechanism, the correlation's magnitude,

and the strategy of including auxiliary variables. As seen in the tables, all the cells yielded no biased estimates.

Based on ANOVA results, no interaction or main effect yielded statistically significant results with partial eta square of value that exceeded Cohen's (1988) medium effect size criteria (.06). The mean percentage of bias for the factor loading Y1 showed a statistically significant main effect ($p$ = .02) for the missingness rate of the items with a very small effect size ($\eta^2$ = .001). This pattern of the small effect size was noticed in other parameters where the effect sizes never exceeded .004.

Results for the factor correlation parameter were similar to those of the factor loadings and error variances. As before, no interaction or main effect yielded statistically significant results with a partial eta square value larger than (.01). Consistent with previous results, all the cells showed no biased estimates (Tables 3- 6).

Finally, results from the error variances were quite similar to those reported for the factor loadings. Again, no interaction or main effect yielded statistically significant results with partial eta square value that exceeded (.01). Consistent with previous results, all the cells yielded no bias (Tables 7-10). However, there was some variability among the mean percentage error variances bias compared to the factor correlation and factor loadings. This could be due to the sensitivity of error variance to the missingness compared to the factor correlation and factor loadings, as the FMIs' results showed that error variances indicated the higher FMIs compared to the factor correlation and factor loadings.

**Table 3**

*Mean Percentage for Factor Loadings and Factor Correlation Bias by Factors Design for the Moderate Magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | Inclusive | .56 | -.92 | 1.91 | .47 | -1.09 | -.28 | .77 |
| | | | Restrictive | .46 | -.92 | 1.98 | .48 | -1.11 | -.31 | .78 |
| | | MAR | Inclusive | .62 | -.97 | 1.93 | .49 | -1.16 | -.29 | .72 |
| | | | Restrictive | .67 | -.96 | 1.88 | .52 | -1.05 | -.32 | .78 |
| | MAR | MCAR | Inclusive | .59 | -1.21 | 1.53 | .66 | -1.33 | -.52 | .79 |
| | | | Restrictive | .56 | -1.15 | 1.52 | .71 | -1.23 | -.54 | .83 |
| | | MAR | Inclusive | .6 | -1.24 | 1.58 | .6 | -1.38 | -.58 | .71 |
| | | | Restrictive | .58 | -1.5 | 1.55 | .66 | -1.48 | -.49 | .71 |
| 30% | MCAR | MCAR | Inclusive | .55 | -.85 | 1.87 | .46 | -1.07 | -.31 | .77 |
| | | | Restrictive | .56 | -.86 | 1.82 | .46 | -.93 | -.41 | .83 |
| | | MAR | Inclusive | .53 | -.91 | 1.9 | .43 | -1 | -.38 | .74 |
| | | | Restrictive | .56 | -.88 | 1.85 | .42 | -.87 | -.46 | .77 |
| | MAR | MCAR | Inclusive | .56 | -1.09 | 1.51 | .58 | -1.29 | -.57 | .71 |
| | | | Restrictive | .48 | -1.06 | 1.52 | .59 | -1.31 | -.47 | .69 |
| | | MAR | Inclusive | .54 | -1.17 | 1.36 | .49 | -1.3 | -.68 | .47 |
| | | | Restrictive | .56 | -1.17 | 1.36 | .55 | -1.23 | -.73 | .37 |

*Note:* AV= auxiliary variable

**Table 4**

*Mean Percentage for Factor Loadings and Factor Correlation Bias by Factors Design for the Moderate Magnitude and the 30% of the Item Missingness Rate.*

| AV Missing Rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | Inclusive | .96 | -.22 | 1.35 | 1.03 | -1.33 | -1.17 | .14 |
| | | | Restrictive | .81 | -.2 | 1.41 | 1.01 | -1.09 | -1.37 | .26 |
| | | MAR | Inclusive | 1.04 | -18 | 1.25 | 1.09 | -1.29 | -1.21 | .25 |
| | | | Restrictive | .94 | -.17 | 1.30 | 1.08 | -1.08 | -1.39 | .31 |
| | MAR | MCAR | Inclusive | 1.2 | -1.22 | 1.27 | 1.11 | -1.45 | -1.06 | .29 |
| | | | Restrictive | 1.21 | -1.25 | 1.24 | 1.14 | -1.29 | -1.27 | .59 |
| | | MAR | Inclusive | 1.32 | -1.30 | 1.29 | 1.23 | -1.58 | -1.24 | .16 |
| | | | Restrictive | 1.35 | -1.37 | 1.24 | 1.24 | -1.44 | -1.44 | .16 |
| 30% | MCAR | MCAR | Inclusive | 1.03 | -.08 | 1.31 | .97 | -1.21 | -1.15 | .05 |
| | | | Restrictive | 1.07 | -.17 | 1.35 | 1.01 | -.1.12 | -1.26 | .21 |
| | | MAR | Inclusive | .96 | -.04 | 1.22 | .9 | -1.12 | -1.24 | .07 |
| | | | Restrictive | .94 | -.04 | 1.23 | .97 | -.92 | -1.46 | .17 |
| | MAR | MCAR | Inclusive | 1.26 | -1.22 | 1.32 | 1.06 | -1.38 | -1.18 | .27 |
| | | | Restrictive | 1.27 | -1.23 | 1.31 | 1.1 | -1.23 | -1.41 | .49 |
| | | MAR | Inclusive | 1.15 | -1.4 | 1.13 | .83 | -1.52 | -1.35 | -.18 |
| | | | Restrictive | 1.16 | -1.35 | 1.07 | .86 | -1.39 | -1.56 | .003 |

*Note*: AV= auxiliary variable.

**Table 5**

*Mean Percentage for Factor Loadings and Factor Correlation Bias by Factors Design for the Low Magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | Inclusive | .49 | -.96 | 1.98 | .39 | -1.1 | -.36 | 1.02 |
| | | | Restrictive | .5 | -.96 | 1.98 | .38 | -1.11 | -.35 | 1.03 |
| | | MAR | Inclusive | .47 | -.97 | 1.98 | .42 | -1.12 | -.37 | .98 |
| | | | Restrictive | .44 | -.98 | 1.97 | .45 | -1.1 | -.38 | .95 |
| | MAR | MCAR | Inclusive | .75 | -.98 | 1.66 | .55 | -.97 | -.44 | .96 |
| | | | Restrictive | .76 | -1.01 | 1.63 | .56 | -.95 | -.41 | .87 |
| | | MAR | Inclusive | .72 | -.97 | 1.66 | .56 | -1 | -.42 | .86 |
| | | | Restrictive | .7 | -.99 | 1.63 | .55 | -.98 | -.41 | .82 |
| 30% | MCAR | MCAR | Inclusive | .49 | -.93 | 1.95 | .4 | -1.13 | -.41 | .94 |
| | | | Restrictive | .46 | -.94 | 1.94 | .44 | -1.1 | -.42 | .9 |
| | | MAR | Inclusive | .51 | -.96 | 1.95 | .39 | -1.11 | -.41 | 1.01 |
| | | | Restrictive | .53 | -.96 | 1.96 | .39 | -1.1 | -.4 | .99 |
| | MAR | MCAR | Inclusive | .7 | -.93 | 1.64 | .54 | -.99 | -41 | .87 |
| | | | Restrictive | .67 | -.9 | 1.67 | .56 | -1 | -.43 | .84 |
| | | MAR | Inclusive | .74 | -.95 | 1.6 | .56 | -.95 | -.45 | .95 |
| | | | Restrictive | .75 | -.98 | 1.59 | .56 | -.96 | -.45 | .83 |

*Note:* AV= auxiliary variable.

**Table 6**

*Mean Percentage for Factor Loadings and Factor Correlation Bias by Factors Design for the Low Magnitude and the 30% of the Item Missingness Rate.*

| AV Missing Rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | Inclusive | .86 | -.1 | 1.55 | 1.25 | -1.43 | -1.73 | .27 |
| | | | Restrictive | .82 | -.08 | 1.55 | 1.27 | -1.36 | -1.78 | .16 |
| | | MAR | Inclusive | .81 | -.12 | 1.6 | 1.28 | -1.46 | -1.72 | .22 |
| | | | Restrictive | .87 | -.1 | 1.62 | 1.24 | -1.45 | -1.75 | .22 |
| | MAR | MCAR | Inclusive | 1.17 | -1.21 | 1.72 | 1.31 | -1.08 | -1.43 | 1.35 |
| | | | Restrictive | 1.18 | -1.19 | 1.76 | 1.3 | -1.12 | -1.42 | 1.32 |
| | | MAR | Inclusive | 1.08 | -1.25 | 1.74 | 1.32 | -1.15 | -1.41 | 1.36 |
| | | | Restrictive | 1.07 | -1.24 | 1.76 | 1.36 | -1.11 | -1.42 | 1.41 |
| 30% | MCAR | MCAR | Inclusive | .89 | -.13 | 1.53 | 1.2 | -1.36 | -1.7 | -.01 |
| | | | Restrictive | .89 | -.17 | 1.6 | 1.21 | -1.34 | -1.65 | .12 |
| | | MAR | Inclusive | .94 | -.15 | 1.57 | 1.16 | -1.31 | -1.66 | .09 |
| | | | Restrictive | .86 | -.11 | 1.55 | 1.23 | -1.26 | -1.71 | .04 |
| | MAR | MCAR | Inclusive | 1.16 | -1.25 | 1.7 | 1.23 | -1.07 | -1.46 | 1.24 |
| | | | Restrictive | 1.14 | -1.1 | 1.74 | 1.27 | -1.1 | -1.44 | 1.28 |
| | | MAR | Inclusive | 1.22 | -1.28 | 1.7 | 1.24 | -1.06 | -1.45 | 1.27 |
| | | | Restrictive | 1.19 | -1.29 | 1.73 | 1.24 | 1.08 | -1.47 | 1.23 |

*Note:* AV= auxiliary variable.

**Table 7**

*Mean Percentage for Error Variance Bias by Factors Design for the Moderate Magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15% | MCAR | MCAR | Inclusive | -.09 | 1.96 | -3.57 | -2.31 | .79 | -1.46 |
| | | | Restrictive | .07 | 1.95 | -3.72 | -2.41 | .83 | -1.35 |
| | | MAR | Inclusive | -.21 | 2.08 | -3.6 | -2.32 | .83 | -1.5 |
| | | | Restrictive | -.19 | 2.1 | -3.58 | -2.37 | .63 | -1.48 |
| | MAR | MCAR | Inclusive | -.5 | 2.42 | -3.04 | -2.25 | 1.13 | -.95 |
| | | | Restrictive | -.42 | 2.34 | -3 | -2.3 | .95 | -.93 |
| | | MAR | Inclusive | -.61 | 2.47 | -3.07 | -2.17 | 1.1 | -.96 |
| | | | Restrictive | -.6 | 2.34 | -3.04 | -2.21 | 1.21 | -.95 |
| 30% | MCAR | MCAR | Inclusive | -.19 | 1.89 | -3.42 | -2.25 | .8 | -1.46 |
| | | | Restrictive | -.15 | 1.8 | -3.35 | -2.3 | .61 | -1.27 |
| | | MAR | Inclusive | -.16 | 2.01 | -3.41 | -2.17 | .71 | -1.35 |
| | | | Restrictive | -.17 | 1.89 | -3.35 | -2.19 | .49 | -1.18 |
| | MAR | MCAR | Inclusive | -.57 | 2.25 | -2.83 | -2.11 | 1.06 | -.92 |
| | | | Restrictive | -.49 | 2.25 | -2.79 | -2.19 | .99 | -.96 |
| | | MAR | Inclusive | -.64 | 2.34 | -2.68 | -2.07 | .98 | -.83 |
| | | | Restrictive | -.6 | 2.35 | -2.71 | -2.16 | .86 | -.79 |

**Table 8**

*Mean Percentage for Error Variance Bias by Factors Design for the Moderate Magnitude and the 30% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15% | MCAR | MCAR | Inclusive | -1.31 | 1.05 | -2.2 | -2.22 | 1.44 | -.55 |
| | | | Restrictive | -1.17 | 1.02 | -2.25 | -2.26 | 1 | -.3 |
| | | MAR | Inclusive | -1.35 | .97 | -1.97 | -2.3 | 1.39 | -.52 |
| | | | Restrictive | -1.29 | .95 | -2.01 | -2.26 | .97 | -.3 |
| | MAR | MCAR | Inclusive | -1.76 | 2.25 | -3.11 | -2.52 | 1.44 | -.9 |
| | | | Restrictive | -1.72 | 2.25 | -2.99 | -2.63 | 1.08 | -.54 |
| | | MAR | Inclusive | -1.96 | 2.25 | -3.08 | -2.81 | 1.5 | -.7 |
| | | | Restrictive | -2 | 2.32 | -2.98 | -2.83 | 1.24 | -.59 |
| 30% | MCAR | MCAR | Inclusive | -1.44 | .86 | -1.94 | -1.94 | 1.35 | -.71 |
| | | | Restrictive | -1.45 | .95 | -2.02 | -2.02 | 1.04 | -.54 |
| | | MAR | Inclusive | -1.46 | .76 | -1.69 | -1.94 | 1.28 | -.47 |
| | | | Restrictive | -1.47 | .63 | -1.63 | -2.06 | .9 | -.09 |
| | MAR | MCAR | Inclusive | -2 | 2.24 | -3 | -2.34 | 1.36 | -.77 |
| | | | Restrictive | -1.96 | 2.2 | -2.9 | -2.46 | 1.03 | -.38 |
| | | MAR | Inclusive | -2.05 | 2.31 | -2.81 | -2.31 | 1.36 | -.72 |
| | | | Restrictive | -2.06 | 2.22 | -2.7 | -2.47 | 1.01 | -.35 |

**Table 9**

*Mean Percentage for Error Variance Bias by Factors Design for the Low Magnitude and the 15% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15% | MCAR | MCAR | Inclusive | .08 | 1.95 | -3.59 | -2.45 | .53 | -1.2 |
| | | | Restrictive | .1 | 1.92 | -3.6 | -2.43 | .53 | -1.23 |
| | | MAR | Inclusive | .1 | 1.94 | -3.61 | -2.48 | .54 | -1.21 |
| | | | FMI | .12 | 1.92 | -3.59 | -2.51 | .52 | -1.19 |
| | MAR | MCAR | Inclusive | -.53 | 2.17 | -2.62 | -2.3 | .91 | -.79 |
| | | | Restrictive | -.54 | 2.12 | -2.63 | -2.34 | .87 | -.82 |
| | | MAR | Inclusive | -.51 | 2.14 | -2.66 | -2.32 | .93 | -.82 |
| | | | Restrictive | -.5 | 2.1 | -2.65 | -2.32 | .9 | -.84 |
| 30% | MCAR | MCAR | Inclusive | .08 | 1.91 | -3.59 | -2.49 | .53 | -1.15 |
| | | | Restrictive | .1 | 1.87 | -3.56 | -2.53 | .5 | -1.12 |
| | | MAR | Inclusive | .11 | 1.92 | -3.56 | -2.45 | .53 | -1.18 |
| | | | Restrictive | .11 | 1.9 | -3.58 | -2.45 | .52 | -1.19 |
| | MAR | MCAR | Inclusive | -.49 | 2.08 | -2.62 | -2.29 | .9 | -.82 |
| | | | Restrictive | -.43 | 2.02 | -2.63 | -2.33 | .9 | -.82 |
| | | MAR | Inclusive | -51 | 2.09 | -2.56 | -2.3 | .88 | -.8 |
| | | | Restrictive | -.54 | 2.06 | -2.58 | -2.36 | .88 | -.83 |

**Table 10**

*Mean Percentage for Error Variance Bias by Factors Design for the Low Magnitude and the 30% of the Item Missingness Rate.*

| AV missingness rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Bias % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15% | MCAR | MCAR | Inclusive | -.95 | .79 | -2.21 | -2.86 | 1.13 | .2 |
| | | | Restrictive | -.9 | .7 | -2.22 | -2.9 | 1.04 | .28 |
| | | MAR | Inclusive | -.88 | .78 | -2.32 | -2.88 | 1.17 | .16 |
| | | | Restrictive | -.92 | .76 | -2.36 | -2.83 | 1.15 | .2 |
| | MAR | MCAR | Inclusive | -1.68 | 2.68 | -3.01 | -3.15 | .87 | -.42 |
| | | | Restrictive | -1.7 | 2.66 | -3.07 | -3.2 | .89 | -.4 |
| | | MAR | Inclusive | -1.6 | 2.7 | -3.09 | -3.15 | .89 | -.46 |
| | | | Restrictive | -1.6 | 2.65 | -3.11 | -3.21 | .84 | -.44 |
| 30% | MCAR | MCAR | Inclusive | -.97 | .83 | -2.27 | -2.74 | 1.02 | .1 |
| | | | Restrictive | -.99 | .87 | -2.41 | -2.82 | 1.08 | .06 |
| | | MAR | Inclusive | -.98 | .83 | -2.29 | -2.7 | .99 | .06 |
| | | | Restrictive | -.95 | 74 | -2.24 | -2.83 | .96 | .16 |
| | MAR | MCAR | Inclusive | -1.73 | 2.73 | -3.06 | -3.08 | .82 | -.46 |
| | | | Restrictive | -1.69 | 2.43 | -3.1 | -3.13 | .81 | -.38 |
| | | MAR | Inclusive | -1.79 | 2.75 | -3.04 | -3.05 | .83 | -.44 |
| | | | Restrictive | -1.78 | 2.75 | -3.1 | -3.11 | .83 | -.39 |

*Note:* AV= auxiliary variable.

***Mean Squared Error***

Tables 11-18 show the MSE's results for each parameter under the different missingness rates for both items and auxiliary variables, the massing mechanism for items and auxiliary variables, the magnitude of the correlation, and the strategy for including auxiliary variables. The SPSS was used to conduct ANOVA analysis to quantify the design factors' effects and these factors' interactions on parameter MSE. To meet the assumption of normality, logMSE was used for all parameters. The homogeneity of variance assumption was violated ($p < 0.01$) for some parameters, but analysis of variance is robust concerning violation of homogeneity of variance with a balanced design.

Based on ANOVA results, no interaction or main effect yielded statistically significant results with partial eta square of value that exceeded Cohen (1988) medium effect size criteria ($\eta^2$=.06). The items' missingness rate shows a statistically significant main effect on the MSE for six parameters. For the factor loading Y2, the items' missingness rate has a main effect $F(1,6336) = 232.47, p < .001, \eta^2 = .04$ with a lower MSE for the missingness rate of 15% (mean = .001, SD < 0.001) than that for the missingness rate of 30% (mean = .002, SD < 0.001). For the factor loading Y3, the items' missingness rate has a main effect $F$ (1, 6336) = 106.703, $p < .001$, $\eta^2 = .02$ with a lower MSE for the missingness rate of 15% (mean = .001, SD < 0.001) than that for the missingness rate of 30% (mean = .002, SD < 0.001).

The same pattern was found for the error variance of Y1, the items' missingness rate has a main effect $F$ (1, 6336) = 131.071, $p < .001$, $\eta^2 = .02$ with a lower MSE for the missingness rate of 15% (mean = .001, SD < 0.001) than that for the missingness rate of

30% (mean = .002, SD < 0.001). The error variance of Y2 shows significant main effect for the items' missingness rate $F$ (1, 6336) = 108.646, $p$ < .001, $\eta^2$ = .02 with a lower MSE for the missingness rate of 15% (mean = .001, SD < 0.001) than that for the missingness rate of 30% (mean = .002, SD < 0.001).

Similarly, the error variance of Y3 shows significant main effect for the items' missingness rate $F$ (1, 6336) = 99.647, $p$ < .001, $\eta^2$ = .02 with a lower MSE for the missingness rate of 15% (mean = .001, SD < 0.001) than that for the missingness rate of 30% (mean = .002, SD < 0.001).

For the factor correlation $F$ (1, 6336) = 94.958, $p$ < .001, $\eta^2$ = .02 with a lower MSE for the missingness rate of 15% (mean = .001, SD < 0.001) than that for the missingness rate of 30% (mean = .002, SD < 0.001). All the other main effects and interaction had uninterpretable effect sizes that contributed less than or equal to 1% of the total variance in MSE.

**Table 11**

*Mean Squared Error for Factor Loadings and Factor Correlation by Factors Design for the Moderate Magnitude and the 15% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
|---|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0013 | .0012 | .001 | .001 | .001 | .001 | .0013 |
| | | | Restrictive | .0013 | .0012 | .001 | .001 | .0011 | .001 | .0013 |
| | | MAR | Inclusive | .0012 | .0012 | .001 | .001 | .001 | .0011 | .0013 |
| | | | Restrictive | .0012 | .0012 | .001 | .001 | .001 | .001 | .0013 |
| | MAR | MCAR | Inclusive | .0013 | .0011 | .001 | .001 | .001 | .001 | .0013 |
| | | | Restrictive | .0013 | .0011 | .001 | .001 | .001 | .001 | .0013 |
| | | MAR | Inclusive | .0013 | .0011 | .001 | .001 | .001 | .001 | .0013 |
| | | | Restrictive | .0013 | .0011 | .001 | .001 | .001 | .001 | .0014 |
| 30% | MCAR | MCAR | Inclusive | .0012 | .0011 | .001 | .001 | .0011 | .001 | .0013 |
| | | | Restrictive | .0012 | .0011 | .001 | .001 | .0011 | .001 | .0013 |
| | | MAR | Inclusive | .0012 | .0012 | .001 | .001 | .0011 | .0011 | .0013 |
| | | | Restrictive | .0012 | .0012 | .001 | .001 | .0011 | .0011 | .0013 |
| | MAR | MCAR | Inclusive | .0012 | .0011 | .001 | .001 | .001 | .001 | .0013 |
| | | | Restrictive | .0013 | .0011 | .001 | .001 | .001 | .001 | .0013 |
| | | MAR | Inclusive | .0012 | .0011 | .001 | .001 | .0011 | .001 | .0013 |
| | | | Restrictive | .0013 | .0011 | .001 | .001 | .0011 | .001 | .0014 |

*Note:* AV= auxiliary variable.

**Table 12**

Mean Squared Error for Factor Loadings and Factor Correlation by Factors Design for the Moderate Magnitude *and the 30% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
|---|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0016 | .0017 | .0015 | .0015 | .0011 | .0013 | .0017 |
| | | | Restrictive | .0016 | .0017 | .0014 | .0015 | .0011 | .0013 | .0017 |
| | | MAR | Inclusive | .0015 | .0016 | .0014 | .0015 | .0011 | .0013 | .0017 |
| | | | Restrictive | .0016 | .0016 | .0014 | .0015 | .0011 | .0013 | .0017 |
| | MAR | MCAR | Inclusive | .0016 | .0019 | .0015 | .0013 | .0012 | .0014 | .0018 |
| | | | Restrictive | .0017 | .0018 | .0015 | .0012 | .0012 | .0014 | .0019 |
| | | MAR | Inclusive | .0016 | .0019 | .0015 | .0013 | .0012 | .0014 | .0018 |
| | | | Restrictive | .0017 | .0019 | .0015 | .0013 | .0013 | .0014 | .0018 |
| 30% | MCAR | MCAR | Inclusive | .0014 | .0016 | .0014 | .0015 | .0012 | .0013 | .0018 |
| | | | Restrictive | .0015 | .0017 | .0015 | .0015 | .0012 | .0012 | .0018 |
| | | MAR | Inclusive | .0015 | .0017 | .0014 | .0015 | .0012 | .0013 | .0018 |
| | | | Restrictive | .0016 | .0016 | .0014 | .0015 | .0012 | .0013 | .0018 |
| | MAR | MCAR | Inclusive | .0016 | .0018 | .0015 | .0013 | .0013 | .0014 | .0019 |
| | | | Restrictive | .0017 | .0018 | .0015 | .0013 | .0012 | .0014 | .0019 |
| | | MAR | Inclusive | .0016 | .0019 | .0015 | .0013 | .0013 | .0014 | .0019 |
| | | | Restrictive | .0016 | .0019 | .0015 | .0013 | .0013 | .0014 | .0019 |

*Note*: AV= auxiliary variable.

**Table 13**

Mean Squared Error for Factor Loadings and Factor Correlation by Factors Design for the Low Magnitude *and the 15% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
|---|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | | | Restrictive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | | MAR | Inclusive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | | | Restrictive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | MAR | MCAR | Inclusive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |
| | | | Restrictive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |
| | | MAR | Inclusive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |
| | | | Restrictive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |
| 30% | MCAR | MCAR | Inclusive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | | | Restrictive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | | MAR | Inclusive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | | | Restrictive | .0013 | .0012 | .0011 | .0012 | .0011 | .0012 | .0014 |
| | MAR | MCAR | Inclusive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |
| | | | Restrictive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |
| | | MAR | Inclusive | .0013 | .0011 | .0012 | .0011 | .0011 | .0011 | .0014 |
| | | | Restrictive | .0014 | .0011 | .0012 | .0011 | .0011 | .001 | .0014 |

*Note:* AV= auxiliary variable.

**Table 14**

Mean Squared Error for Factor Loadings and Factor Correlation by Factors Design for the Low Magnitude *and the 30% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
|---|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | | | Restrictive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | | MAR | Inclusive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | | | Restrictive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | MAR | MCAR | Inclusive | .0017 | .002 | .0018 | .0015 | .0014 | .0016 | .0021 |
| | | | Restrictive | .0018 | .002 | .0018 | .0015 | .0013 | .0017 | .0021 |
| | | MAR | Inclusive | .0018 | .002 | .0017 | .0015 | .0014 | .0016 | .002 |
| | | | Restrictive | .0018 | .002 | .0018 | .0015 | .0014 | .0016 | .002 |
| 30% | MCAR | MCAR | Inclusive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | | | Restrictive | .0017 | .0019 | .0016 | .0016 | .0012 | .0015 | .002 |
| | | MAR | Inclusive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | | | Restrictive | .0017 | .0019 | .0016 | .0016 | .0013 | .0015 | .002 |
| | MAR | MCAR | Inclusive | .0018 | .002 | .0018 | .0015 | .0014 | .0016 | .0021 |
| | | | Restrictive | .0018 | .002 | .0018 | .0015 | .0013 | .0016 | .0021 |
| | | MAR | Inclusive | .0017 | .002 | .0017 | .0015 | .0014 | .0016 | .0021 |
| | | | Restrictive | .0017 | .002 | .0017 | .0015 | .0014 | .0016 | .0021 |

*Note:* AV= auxiliary variable.

**Table 15**

Mean Squared Error for Error Variances by factors design for the moderate magnitude *and the 15% of the Item Missingness Rate.*

| AV missingness rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | e1 | e2 | e3 | e4 | e5 | e6 |
|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0017 | .0015 | .0013 | .0015 | .0014 | .0016 |
| | | | Restrictive | .0017 | .0015 | .0013 | .0015 | .0014 | .0015 |
| | | MAR | Inclusive | .0017 | .0015 | .0013 | .0015 | .0014 | .0016 |
| | | | Restrictive | .0017 | .0015 | .0014 | .0015 | .0014 | .0016 |
| | MAR | MCAR | Inclusive | .0017 | .0013 | .0014 | .0013 | .0013 | .0015 |
| | | | Restrictive | .0018 | .0013 | .0014 | .0013 | .0013 | .0015 |
| | | MAR | Inclusive | .0017 | .0013 | .0014 | .0013 | .0013 | .0015 |
| | | | Restrictive | .0017 | .0012 | .0014 | .0015 | .0015 | .0015 |
| 30% | MCAR | MCAR | Inclusive | .0016 | .0015 | .0013 | .0015 | .0015 | .0016 |
| | | | Restrictive | .0016 | .0015 | .0013 | .0015 | .0015 | .0016 |
| | | MAR | Inclusive | .0017 | .0015 | .0014 | .0014 | .0015 | .0016 |
| | | | Restrictive | .0017 | .0015 | .0013 | .0014 | .0015 | .0016 |
| | MAR | MCAR | Inclusive | .0017 | .0013 | .0014 | .0013 | .0013 | .0015 |
| | | | Restrictive | .0017 | .0013 | .0015 | .0013 | .0013 | .0015 |
| | | MAR | Inclusive | .0017 | .0013 | .0014 | .0013 | .0013 | .0015 |
| | | | Restrictive | .0017 | .0013 | .0015 | .0013 | .0013 | .0015 |

**Table 16**

*Mean Squared Error for Error Variances by factors design for the moderate magnitude and the 30% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | e1 | e2 | e3 | e4 | e5 | e6 |
|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0023 | .002 | .0017 | .0019 | .0016 | .0019 |
| | | | Restrictive | .0023 | .0019 | .0017 | .0019 | .0017 | .0019 |
| | | MAR | Inclusive | .0023 | .0019 | .0017 | .0019 | .0017 | .0019 |
| | | | Restrictive | .0023 | .0019 | .0016 | .0019 | .0017 | .0019 |
| | MAR | MCAR | Inclusive | .0026 | .0019 | .002 | .0017 | .0018 | .0022 |
| | | | Restrictive | .0025 | .0018 | .002 | .0016 | .0018 | .0022 |
| | | MAR | Inclusive | .0026 | .0019 | .002 | .0017 | .0018 | .0022 |
| | | | Restrictive | .0026 | .0019 | .002 | .0017 | .0018 | .0022 |
| 30% | MCAR | MCAR | Inclusive | .0023 | .0021 | .0017 | .002 | .0019 | .0019 |
| | | | Restrictive | .0023 | .002 | .0018 | .002 | .0018 | .0018 |
| | | MAR | Inclusive | .0023 | .002 | .0017 | .0019 | .0017 | .0018 |
| | | | Restrictive | .0022 | .002 | .0016 | .0018 | .0018 | .0019 |
| | MAR | MCAR | Inclusive | .0027 | .0019 | .002 | .0016 | .0019 | .0022 |
| | | | Restrictive | .0026 | .0019 | .002 | .0016 | .0019 | .0023 |
| | | MAR | Inclusive | .0026 | .002 | .0021 | .0016 | .0019 | .0022 |
| | | | Restrictive | .0025 | .0019 | .002 | .0016 | .0019 | .0022 |

**Table 17**

*Mean Squared Error for Error Variances by factors design for the Low magnitude and the 15% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | e1 | e2 | e3 | e4 | e5 | e6 |
|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0017 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | | | Restrictive | .0018 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | | MAR | Inclusive | .0017 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | | | Restrictive | .0017 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | MAR | MCAR | Inclusive | .0018 | .0014 | .0015 | .0013 | .0014 | .0015 |
| | | | Restrictive | .0019 | .0014 | .0015 | .0014 | .0014 | .0015 |
| | | MAR | Inclusive | .0018 | .0014 | .0015 | .0013 | .0014 | .0015 |
| | | | Restrictive | .0018 | .0014 | .0015 | .0013 | .0013 | .0015 |
| 30% | MCAR | MCAR | Inclusive | .0017 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | | | Restrictive | .0017 | .0015 | .0014 | .0015 | .0015 | .0016 |
| | | MAR | Inclusive | .0017 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | | | Restrictive | .0017 | .0015 | .0014 | .0015 | .0015 | .0017 |
| | MAR | MCAR | Inclusive | .0018 | .0014 | .0015 | .0013 | .0014 | .0015 |
| | | | Restrictive | .0018 | .0014 | .0015 | .0013 | .0014 | .0015 |
| | | MAR | Inclusive | .0018 | .0014 | .0015 | .0013 | .0014 | .0015 |
| | | | Restrictive | .0018 | .0014 | .0015 | .0013 | .0014 | .0016 |

**Table 18**

*Mean Squared Error for Error Variances by factors design for the Low magnitude and the 30% of the Item Missingness Rate.*

| AV missing rate | Item Missing Mechanism | AV Missing Mechanism | Strategy | e1 | e2 | e3 | e4 | e5 | e6 |
|---|---|---|---|---|---|---|---|---|---|
| 15% | MCAR | MCAR | Inclusive | .0025 | .002 | .0018 | .0022 | .0018 | .0021 |
| | | | Restrictive | .0025 | .0021 | .0018 | .0023 | .0018 | .0021 |
| | | MAR | Inclusive | .0025 | .002 | .0019 | .0022 | .0018 | .0021 |
| | | | Restrictive | .0025 | .0021 | .0019 | .0022 | .0018 | .0021 |
| | MAR | MCAR | Inclusive | .0026 | .0021 | .0023 | .0021 | .0019 | .0026 |
| | | | Restrictive | .0026 | .0021 | .0023 | .0021 | .0019 | .0026 |
| | | MAR | Inclusive | .0025 | .0021 | .0023 | .002 | .0019 | .0026 |
| | | | Restrictive | .0026 | .0021 | .0023 | .002 | .0019 | .0026 |
| 30% | MCAR | MCAR | Inclusive | .0024 | .0021 | .0018 | .0022 | .0019 | .0021 |
| | | | Restrictive | .0024 | .0021 | .0019 | .0022 | .0019 | .0021 |
| | | MAR | Inclusive | .0024 | .0021 | .0018 | .0022 | .0018 | .0021 |
| | | | Restrictive | .0024 | .0021 | .0018 | .0022 | .0019 | .0021 |
| | MAR | MCAR | Inclusive | .0026 | .0022 | .0023 | .0021 | .0019 | .0025 |
| | | | Restrictive | .0026 | .0021 | .0023 | .002 | .0019 | .0025 |
| | | MAR | Inclusive | .0026 | .0022 | .0023 | .002 | .0019 | .0026 |
| | | | Restrictive | .0026 | .0022 | .0023 | .002 | .0019 | .0026 |

### *Confidence Interval Coverage (CIC)*

CIC's coverage rate was computed as the percentage of times that the 95% confidence intervals of the parameter estimates contain the true parameter values. Tables 19-22 give the 95% confidence interval coverage rates of the 13 parameters by the design factors.

The coverage rates for inclusive and restrictive strategies based on FMI were generally well above the 90% mark across all conditions for all factor loadings and factor correlation.

In terms of the error variances, all parameters performed well under the two strategies and across the conditions, except for the error variances of the Y6 item, which showed problematic coverage (CIC < .90). There was no exact pattern for the low CIC in this parameter. Still, the problematic coverages were produced in some conditions of the low correlation between the items and auxiliary variables, which could be due to sampling variability (Table 22)

**Table 19**

*Confidence Interval Coverage for Factor loadings and factor correlation by factors design for the moderate magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | IS | 92 | 96 | 96 | 96 | 95 | 97 | 94 |
| | | | RS | 93 | 97 | 97 | 97 | 95 | 98 | 95 |
| | | MAR | IS | 94 | 96 | 96 | 97 | 95 | 97 | 94 |
| | | | RS | 94 | 96 | 96 | 96 | 95 | 98 | 94 |
| | MAR | MCAR | IS | 94 | 98 | 96 | 95 | 98 | 97 | 96 |
| | | | RS | 94 | 97 | 96 | 96 | 99 | 98 | 94 |
| | | MAR | IS | 95 | 98 | 96 | 94 | 97 | 97 | 95 |
| | | | RS | 94 | 99 | 97 | 96 | 98 | 98 | 94 |
| 30% | MCAR | MCAR | IS | 94 | 97 | 95 | 97 | 95 | 97 | 94 |
| | | | RS | 94 | 97 | 95 | 98 | 96 | 97 | 94 |
| | | MAR | IS | 94 | 96 | 95 | 96 | 95 | 97 | 94 |
| | | | RS | 94 | 96 | 95 | 96 | 95 | 97 | 93 |
| | MAR | MCAR | IS | 95 | 98 | 96 | 96 | 97 | 98 | 94 |
| | | | RS | 96 | 98 | 96 | 95 | 97 | 98 | 94 |
| | | MAR | IS | 96 | 99 | 96 | 93 | 98 | 98 | 95 |
| | | | RS | 96 | 98 | 96 | 94 | 97 | 97 | 93 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

**Table 20**

*Confidence Interval Coverage for Factor loadings and factor correlation by factors design for the moderate magnitude and the 30% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | IS | 94 | 92 | 95 | 95 | 98 | 98 | 95 |
| | | | RS | 94 | 94 | 95 | 95 | 98 | 99 | 95 |
| | | MAR | IS | 94 | 93 | 95 | 95 | 97 | 98 | 95 |
| | | | RS | 95 | 95 | 96 | 95 | 99 | 99 | 96 |
| | MAR | MCAR | IS | 93 | 90 | 96 | 98 | 98 | 95 | 96 |
| | | | RS | 93 | 92 | 96 | 98 | 98 | 95 | 95 |
| | | MAR | IS | 93 | 93 | 96 | 97 | 98 | 94 | 96 |
| | | | RS | 93 | 92 | 95 | 99 | 98 | 93 | 96 |
| 30% | MCAR | MCAR | IS | 97 | 94 | 95 | 96 | 99 | 97 | 95 |
| | | | RS | 97 | 96 | 95 | 96 | 99 | 99 | 95 |
| | | MAR | IS | 96 | 95 | 96 | 96 | 97 | 97 | 95 |
| | | | RS | 98 | 93 | 96 | 95 | 97 | 98 | 96 |
| | MAR | MCAR | IS | 93 | 92 | 95 | 97 | 97 | 97 | 95 |
| | | | RS | 93 | 93 | 95 | 97 | 98 | 95 | 92 |
| | | MAR | IS | 94 | 93 | 95 | 96 | 98 | 97 | 95 |
| | | | RS | 94 | 92 | 94 | 97 | 98 | 95 | 94 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

**Table 21**

*Confidence Interval Coverage for Factor loadings and factor correlation by factors design for the low magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | IS | 95 | 97 | 95 | 97 | 95 | 95 | 96 |
| | | | RS | 95 | 97 | 95 | 97 | 95 | 96 | 96 |
| | | MAR | IS | 95 | 97 | 95 | 97 | 95 | 93 | 94 |
| | | | RS | 95 | 99 | 95 | 97 | 95 | 94 | 94 |
| | MAR | MCAR | IS | 94 | 99 | 96 | 95 | 98 | 98 | 93 |
| | | | RS | 94 | 99 | 96 | 95 | 98 | 98 | 93 |
| | | MAR | IS | 94 | 98 | 96 | 95 | 98 | 97 | 93 |
| | | | RS | 94 | 98 | 96 | 95 | 95 | 97 | 93 |
| 30% | MCAR | MCAR | IS | 95 | 97 | 96 | 97 | 95 | 94 | 94 |
| | | | RS | 95 | 97 | 96 | 97 | 95 | 95 | 95 |
| | | MAR | IS | 95 | 97 | 96 | 97 | 95 | 94 | 93 |
| | | | RS | 95 | 97 | 95 | 97 | 95 | 95 | 94 |
| | MAR | MCAR | IS | 94 | 98 | 96 | 95 | 98 | 98 | 93 |
| | | | RS | 94 | 99 | 96 | 95 | 98 | 98 | 93 |
| | | MAR | IS | 94 | 99 | 95 | 95 | 98 | 98 | 93 |
| | | | RS | 95 | 98 | 96 | 95 | 98 | 97 | 93 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

**Table 22**

*Confidence Interval Coverage for Factor loadings and factor correlation by factors design for the low magnitude and the 30% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | F |
| 15% | MCAR | MCAR | IS | 96 | 92 | 96 | 96 | 98 | 96 | 92 |
| | | | RS | 95 | 92 | 97 | 96 | 98 | 96 | 93 |
| | | MAR | IS | 96 | 92 | 97 | 96 | 98 | 96 | 93 |
| | | | RS | 95 | 92 | 95 | 96 | 98 | 96 | 93 |
| | MAR | MCAR | IS | 94 | 91 | 96 | 95 | 97 | 91 | 92 |
| | | | RS | 95 | 91 | 96 | 95 | 98 | 91 | 93 |
| | | MAR | IS | 95 | 91 | 97 | 95 | 97 | 91 | 93 |
| | | | RS | 94 | 91 | 97 | 95 | 97 | 91 | 93 |
| 30% | MCAR | MCAR | IS | 95 | 93 | 97 | 96 | 97 | 96 | 92 |
| | | | RS | 95 | 93 | 98 | 96 | 98 | 96 | 93 |
| | | MAR | IS | 95 | 92 | 97 | 95 | 97 | 96 | 92 |
| | | | RS | 95 | 92 | 98 | 95 | 98 | 97 | 92 |
| | MAR | MCAR | IS | 94 | 90 | 97 | 94 | 97 | 91 | 92 |
| | | | RS | 94 | 92 | 97 | 94 | 98 | 92 | 93 |
| | | MAR | IS | 95 | 90 | 97 | 94 | 97 | 91 | 93 |
| | | | RS | 95 | 90 | 97 | 94 | 98 | 93 | 93 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

**Table 23**

*Confidence Interval Coverage for error variances by factors design for the moderate magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15% | MCAR | MCAR | IS | 92 | 94 | 94 | 93 | 95 | 93 |
| | | | RS | 93 | 95 | 94 | 93 | 95 | 93 |
| | | MAR | IS | 92 | 94 | 94 | 93 | 95 | 92 |
| | | | RS | 91 | 95 | 94 | 93 | 96 | 90 |
| | MAR | MCAR | IS | 92 | 96 | 94 | 95 | 95 | 94 |
| | | | RS | 93 | 96 | 95 | 95 | 98 | 95 |
| | | MAR | IS | 92 | 98 | 94 | 95 | 94 | 93 |
| | | | RS | 93 | 97 | 94 | 95 | 94 | 93 |
| 30% | MCAR | MCAR | IS | 94 | 93 | 94 | 93 | 94 | 91 |
| | | | RS | 93 | 93 | 94 | 93 | 94 | 92 |
| | | MAR | IS | 92 | 94 | 9 | 94 | 93 | 91 |
| | | | RS | 91 | 93 | 94 | 93 | 93 | 91 |
| | MAR | MCAR | IS | 94 | 97 | 95 | 95 | 94 | 96 |
| | | | RS | 95 | 97 | 94 | 9 | 93 | 96 |
| | | MAR | IS | 93 | 97 | 94 | 95 | 95 | 96 |
| | | | RS | 93 | 97 | 95 | 95 | 93 | 95 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

**Table 24**

*Confidence Interval Coverage for error variances by factors design for the moderate magnitude and the 30% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15% | MCAR | MCAR | IS | 93 | 93 | 96 | 95 | 97 | 93 |
| | | | RS | 94 | 93 | 96 | 96 | 97 | 93 |
| | | MAR | IS | 93 | 94 | 96 | 95 | 98 | 95 |
| | | | RS | 94 | 95 | 96 | 96 | 97 | 93 |
| | MAR | MCAR | IS | 92 | 98 | 93 | 96 | 98 | 91 |
| | | | RS | 94 | 98 | 92 | 96 | 98 | 91 |
| | | MAR | IS | 94 | 98 | 93 | 96 | 99 | 91 |
| | | | RS | 94 | 98 | 93 | 96 | 100 | 92 |
| 30% | MCAR | MCAR | IS | 93 | 93 | 96 | 95 | 97 | 93 |
| | | | RS | 93 | 95 | 96 | 95 | 97 | 92 |
| | | MAR | IS | 93 | 93 | 97 | 96 | 97 | 93 |
| | | | RS | 94 | 94 | 97 | 96 | 97 | 93 |
| | MAR | MCAR | IS | 93 | 97 | 93 | 96 | 98 | 91 |
| | | | RS | 93 | 97 | 92 | 96 | 98 | 90 |
| | | MAR | IS | 93 | 97 | 93 | 95 | 97 | 92 |
| | | | RS | 92 | 97 | 93 | 96 | 97 | 92 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

**Table 25**

*Confidence Interval Coverage for error variances by factors design for the low magnitude and the 15% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15 | MCAR | MCAR | IS | 94 | 96 | 93 | 94 | 96 | **88** |
| | | | RS | 93 | 96 | 93 | 94 | 96 | **88** |
| | | MAR | IS | 93 | 96 | 93 | 94 | 96 | **88** |
| | | | RS | 92 | 96 | 93 | 94 | 95 | **88** |
| | MAR | MCAR | IS | 92 | 97 | 95 | 97 | 95 | 93 |
| | | | RS | 93 | 96 | 95 | 97 | 95 | 92 |
| | | MAR | IS | 93 | 97 | 94 | 97 | 95 | 93 |
| | | | RS | 93 | 96 | 94 | 97 | 95 | 93 |
| 30 | MCAR | MCAR | IS | 94 | 95 | 93 | 93 | 95 | 90 |
| | | | RS | 94 | 95 | 93 | 93 | 94 | 91 |
| | | MAR | IS | 93 | 95 | 93 | 97 | 94 | 91 |
| | | | RS | 92 | 96 | 93 | 97 | 94 | 91 |
| | MAR | MCAR | IS | 93 | 97 | 94 | 97 | 95 | 93 |
| | | | RS | 93 | 97 | 94 | 97 | 95 | 93 |
| | | MAR | IS | 93 | 97 | 95 | 97 | 94 | 93 |
| | | | RS | 93 | 97 | 95 | 97 | 94 | 92 |

*Note*. Bolded cells indicate that the coverage rate is less than 90%.

**Table 26**

*Confidence Interval Coverage for error variances by factors design for the low magnitude and the 30% of the Item Missingness Rate.*

| AV Missing Rate | Item Mechanism | AV Missing Mechanism | Strategy | CIC % | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | e1 | e2 | e3 | e4 | e5 | e6 |
| 15 | MCAR | MCAR | IS | 96 | 95 | 98 | 98 | 97 | 91 |
| | | | RS | 95 | 96 | 98 | 98 | 97 | 92 |
| | | MAR | IS | 96 | 96 | 98 | 98 | 97 | 91 |
| | | | RS | 96 | 96 | 98 | 98 | 97 | 92 |
| | MAR | MCAR | IS | 93 | 93 | 94 | 95 | 98 | **89** |
| | | | RS | 93 | 94 | 94 | 96 | 98 | **89** |
| | | MAR | IS | 93 | 94 | 95 | 96 | 97 | **89** |
| | | | RS | 91 | 94 | 95 | 96 | 98 | **89** |
| 30 | MCAR | MCAR | IS | 95 | 96 | 97 | 98 | 97 | 92 |
| | | | RS | 95 | 95 | 97 | 98 | 97 | 91 |
| | | MAR | IS | 96 | 96 | 97 | 98 | 97 | 92 |
| | | | RS | 95 | 97 | 97 | 98 | 97 | 92 |
| | MAR | MCAR | IS | 92 | 94 | 95 | 96 | 97 | **89** |
| | | | RS | 93 | 93 | 93 | 96 | 98 | **89** |
| | | MAR | IS | 93 | 93 | 95 | 96 | 98 | **89** |
| | | | RS | 93 | 93 | 94 | 96 | 97 | **89** |

*Note*. Bolded cells indicate that the coverage rate is less than 90%.

### *FMI Properties*

Since few articles talk about FMI properties, it is important to discuss the FMI properties observed in this study. To understand the effect of using auxiliary variables on the FMI, the FMI properties without using auxiliary variables for each parameter are reported first.

**FMI Properties Without Using Auxiliary Variables**. Analyzing the model without using any auxiliary variable showed that each parameter's FMI increased as the items' missingness rate increased (Table 27).

**FMI Properties Using Auxiliary Variables**. For the model with auxiliary variables, the FMI of the inclusive strategy is reported as both strategies showed similar results. The findings showed that adding auxiliary variables to the model decreased the FMI values for all parameters, conditioning on the magnitude of the correlation between auxiliary variables and items (Table 28). In other words, the decrease in the FMI was clear when the magnitude of the correlation between auxiliary variables and items was moderate and became very notable with adding more auxiliary variables to the model. This finding is not surprising, as the higher the correlations between auxiliary variables and items are, the more information the auxiliary variables will add to the model imputation.

Another pattern observed in the results is related to the empirical SE. There was an association between each parameter's SE and the FMI. Specifically, the SE and the FMI increased as the items' missingness rates increased. Using auxiliary variables that correlated with items at a moderate level (.6), the SE and the FMI decreased.

**Table 21**

*FMI Parameters Without Adding Auxiliary Variables (AVs)*

| Item Missing Mechanism | Item Missing Rate | Factor Correlation | Facto Loading Y1 | Error Variance Y1 |
|---|---|---|---|---|
| MCAR | 15% | .14 | .21 | .27 |
| | 30% | .31 | .42 | .49 |
| MAR | 15% | .16 | .23 | .28 |
| | 30% | .33 | .43 | .5 |

**Table 22**

*FMI Parameter with the Inclusive Strategy*

| Magnitude | Item Missing Mechanism | Item Missing rate | Factor Correlation | Factor Loading Y1 | Error Variance Y1 |
|---|---|---|---|---|---|
| Moderate | MCAR | 15% | .12 | .17 | .23 |
| | | 30% | .24 | .36 | .45 |
| | MAR | 15% | .13 | .19 | .25 |
| | | 30% | .26 | .37 | .45 |
| Low | MCAR | 15% | .14 | .21 | .28 |
| | | 30% | .31 | .42 | .49 |
| | MAR | 15% | .16 | .22 | .28 |
| | | 30% | .32 | .42 | .5 |

## The Empirical Study's Results

### *The Process of Selecting Auxiliary Variables*

In this empirical example, selecting auxiliary variables based on their FMI was much easier across the conditions as the FMIs differed among the auxiliary variables. As was expected, the highly correlated auxiliary variables (wave 2 BTACT items) showed lower FMIs in most parameters compared to lower correlated auxiliary variables (personality variables). The difference between FMIs becomes clearer between the imputation models. For example, in some conditions, using wave two BTACT items as the imputation model showed the lowest FMIs in most parameters compared to the model that included all the personality variables.

### *Inclusive and Restrictive Strategies*

In the simulation study, there were no differences between the inclusive and restrictive strategies as the FMI suggested, including all auxiliary variables into the imputation model. As mentioned previously, the restrictive strategy showed lower FMIs only in a few conditions. In the empirical study, however, and using auxiliary variables that varied in their correlations with the imputed items, the restrictive strategy in all conditions showed lower FMIs in most parameters (13 out of 15) compared to inclusive strategy. The restrictive strategy included seven variables in the conditions of MCAR 15%, MCAR 30%, and MAR 30%, whereas five variables were included in the condition of MAR 15%.

### *Distinguishing Type A and B Auxiliary Variables*

In this study, age was used as a type A auxiliary variable. The parameters' FMIs after including age in the imputation model showed lower FMIs compared to personality variables. Yet, some wave two BTACT items outperformed age in decreasing the FMIs. In the selection process of including auxiliary variables, the covariates that produced lower FMI were chosen to form an imputation model. In the condition of MAR 30%, selecting wave two BTACT items as the auxiliary variables showed the lowest FMIs in most parameters. However, after adding age into this imputation model, the FMI increased in 13 parameters out of 15, and the FMIs after adding age were identical to the FMIs after using age separately. The same was observed in the condition of MAR 15%. Thus, FMI could not distinguish the type A auxiliary variable among the other variables.

### *Bias*

In the conditions of MCAR with a 15% missingness rate, the inclusive and restrictive strategies showed similar results as all the parameters produced ignorable percentage bias except for the error variance of item 1. For MCAR with missingness of 30%, the inclusive strategy produces three biased estimates (the estimates of the factor loading of item three and the error variances of items 1 and 2). However, the restrictive strategy showed two biased estimates (the error variances of items 1 and 2). In this condition, the restrictive strategy resulted in lower bias levels and lower biased estimates than the inclusive one.

For MAR with missingness of 15%, both strategies showed similar results as all the parameters produced ignorable percentage bias except for the error variance of item 1. In the MAR with missingness of 30%, the error variance estimates of item 1 showed a bias higher than 10%. Neither the inclusive strategy nor the restrictive strategy could solve the bias problem. However, the restrictive strategy resulted in lower bias levels than the inclusive one in nine parameters.

In general, both strategies showed a similar pattern when the missingness rate was 15%, yet, with a higher missingness rate, the restrictive strategy resulted in lower bias levels and/or lower biased estimates than the inclusive one (Tables 29-30).

**Table 23**

*Mean Percentage for Factor Loadings and Factor Correlation Bias by Factor Design.*

| Missing Mechanism | Missing Rate | Strategy | Bias % | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | F |
| MCAR | 15% | Inclusive | -1.33 | 1.38 | 2.74 | -2.97 | 1.01 | -1.81 | 0 | .22 |
| | | Restrictive | -1.46 | 1.56 | 2.02 | -1.9 | 1.29 | -1.47 | -1.4 | -.22 |
| | 30% | Inclusive | -2.98 | 4.91 | **10.1** | -. 65 | -1.75 | -1.83 | -4.22 | 3.85 |
| | | Restrictive | -3.29 | 4.96 | 6.92 | -.38 | -.36 | -1.29 | 0 | 1.34 |
| MAR | 15% | Inclusive | -1.06 | -0.75 | -3.03 | -0.41 | -0.55 | -1.06 | 4.92 | 1.56 |
| | | Restrictive | -1.64 | 1.56 | -3.03 | -.65 | -.55 | -.13 | -4.22 | .44 |
| | 30% | Inclusive | -2.62 | 1.96 | 3.75 | -2.17 | -.27 | -.12 | 4.92 | 2.69 |
| | | Restrictive | -1.64 | 1.6 | 2.88 | -.5 | -2.21 | -1.31 | 6.33 | 1.34 |

*Note*. Bolded cells indicate that the mean percentage bias is higher than 10%.

**Table 30**

*Mean Percentage for Error Variance Bias by Factors Design.*

| Missing Mechanism | Missing Rate | Strategy | Bias % | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | e1 | e2 | e3 | e4 | e5 | e6 | e7 |
| MCAR | 15% | Inclusive | **28.32** | -4.84 | -.47 | -1.64 | -.23 | -1.52 | -5.31 |
| | | Restrictive | **30.32** | -5.27 | -.11 | 1.14 | -.78 | -.46 | -4.25 |
| | 30% | Inclusive | **75.93** | **-16.81** | -7.08 | -.29 | 3.79 | 5.37 | 1.06 |
| | | Restrictive | **81.45** | **-17.2** | -5.43 | -.69 | 1.1 | 4.53 | 0 |
| MAR | 15% | Inclusive | **24.6** | -4.53 | .05 | 1.86 | .63 | 3.62 | -4.25 |
| | | Restrictive | **36.84** | -7.13 | .47 | 1 | 1.18 | .03 | -4.25 |
| | 30% | Inclusive | **55.13** | -1.59 | -2.47 | 4.03 | 1.02 | 5.09 | 0 |
| | | Restrictive | **40.6** | .47 | -1.71 | 1.68 | 2.92 | 1.35 | 0 |

*Note*. Bolded cells indicate that the mean percentage bias is higher than 10%.

### *Efficiency*

Shifting toward the impact of using inclusive and restrictive strategies on efficiency and power, the relative gain in efficiency and the loss of statistical power caused by missing data were estimated. A specific parameter's FMI value can be interpreted as the loss of efficiency in the estimation of that parameter (Savalei & Rhemtulla, 2012). Thus, the restrictive strategy, which produced lower FMI values in most parameters (13 out of 15) across conditions, indicated that the loss of efficiency in estimating these parameters due to missing data was lower with using the restrictive strategy than with the inclusive strategy.

Moreover, the relative gain in efficiency results showed that the restrictive strategy improved efficiency relative to the absence of the auxiliary variables in most parameters, with more improvement with a higher missingness rate (Tables 31-32). However, across conditions, the inclusive strategy showed no improvement as most parameters were as effective as the model with no auxiliary variables. Some parameters showed a loss of efficiency when the mechanism was MAR (Tables 31-32).

In terms of power, the FMI could be interpreted as the loss of statistical power due to missing data as it can be used to estimate the effective sample size. This indicates the sample size that would have achieved the same efficiency for a parameter with complete data. To illustrate this point, let's take the condition of MAR with 30% of missingness as an example. In Table 31, the results of the effective sample size for all parameters are based on three imputation models: model without using auxiliary variables, model using the inclusive strategy, and model based on the restrictive strategy. The model with no auxiliary

variables can be used as the baseline to compare inclusive and restrictive strategies' impact on efficiency and power. For example, under the model that did not include auxiliary variables, the factor loading of the sixth item is based on the effective sample size of $N^* = 2673(1-0.401) = 1601$, which means that its variability is as high as it would have been had it been based on a complete data set with only 1601 cases instead of 2673. This reflects the loss of power as the sample size decreased from 2673 to 1601. The inclusive strategy showed identical results to the model without any auxiliary variables for the same parameter (Table 33).

On the other hand, using the restrictive strategy resulted in an effective sample size of 1981, which means that the included auxiliary variables increased the sample size by 380 compared to the inclusive model and the model with no auxiliary variables. This indicates that while using the restrictive strategy based on the FMI could lead to gains in efficiency and power, the inclusive strategy that adds all available auxiliary variables could cross out the benefit of some of these auxiliary variables, and it could result in the same impact of not using any auxiliary variables. The improvement in efficiency and power by using the restrictive strategy was observed in most parameters, becoming larger with a higher rate of missingness.

**Table 24**

*Gain in Efficiency in Factor Loadings and Factor Correlation Bias by Factor Design.*

| Missing Mechanism | Missing Rate | Strategy | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | F |
|---|---|---|---|---|---|---|---|---|---|---|
| MCAR | 15% | Inclusive | 1 | 1 | 1 | 1 | 1 | 1.01 | 1 | 1 |
| | | Restrictive | 1.02 | 1 | 1 | 1.06 | 1.06 | 1.1 | 1.31 | 1 |
| | 30% | Inclusive | 1 | 1 | 1 | 1.01 | 1 | 1.01 | 1 | 1 |
| | | Restrictive | 1.03 | 1.03 | 1.11 | 1.18 | 1.17 | 1.28 | 1 | 1 |
| MAR | 15% | Inclusive | 1 | 1 | 1 | 1.01 | 1 | 1.01 | 1 | 1 |
| | | Restrictive | 1.04 | 1 | 1 | 1.08 | 1.06 | 1.12 | 1 | 1 |
| | 30% | Inclusive | 1 | 1 | 1.05 | 1 | 1 | 1 | 1 | 1 |
| | | Restrictive | 1.03 | 1.06 | 1.11 | 1.18 | 1.17 | 1.27 | 1 | 1.09 |

**Table 32**

*Gain in Efficiency in Error Variance by Factors Design.*

| Missing Mechanism | Missing Rate | Strategy | e1 | e2 | e3 | e4 | e5 | e6 | e7 |
|---|---|---|---|---|---|---|---|---|---|
| MCAR | 15% | Inclusive | 1 | 1.009 | 1 | 1.002 | 1 | 1.004 | 1 |
| | | Restrictive | 1.03 | 1.02 | 1 | 1.02 | 1.08 | 1.07 | 1 |
| | 30% | Inclusive | 1.009 | 1.02 | 1 | 1.006 | 1.03 | 1.009 | 1 |
| | | Restrictive | 1.04 | 1.04 | 1.07 | 1.12 | 1.17 | 1.24 | 1 |
| MAR | 15% | Inclusive | **0.98** | **0.99** | 1 | 1 | 1 | 1.003 | 1 |
| | | Restrictive | 1.04 | 1 | 1 | 1.07 | 1.08 | 1.12 | 1 |
| | 30% | Inclusive | 1 | 1 | 1 | **0.99** | 1 | 1.001 | 1 |
| | | Restrictive | 1.02 | 1.06 | 1.03 | 1.12 | 1.13 | 1.28 | 1 |

*Note*. Bolded cells indicate the loss of efficiency.

**Table 33**

*The Effective Sample Size*

| Parameters | Model with no AV | IS | RS |
|---|---|---|---|
| Factor loading 1 | 1581.657 | 1581.657 | 1631.47 |
| Factor loading 2 | 1642.069 | 1642.069 | 1734.563 |
| Factor loading 3 | 1581.657 | 1665.997 | 1757.268 |
| Factor loading 4 | 1677.557 | 1677.557 | 1973.067 |
| Factor loading 5 | 1688.858 | 1688.858 | 1982.063 |
| Factor loading 6 | 1601.217 | 1601.217 | 1981.569 |
| Factor loading 7 | 1617 | 1617 | 1617 |
| Factor correlation | 1856.25 | 1856.25 | 2021.172 |
| Error variance 1 | 1347.295 | 1347.295 | 1434.043 |
| Error variance 2 | 1376.297 | 1376.297 | 1448.821 |
| Error variance 3 | 1683.673 | 1683.673 | 1738.424 |
| Error variance 4 | 1646.899 | 1647.793 | 1811.195 |
| Error variance 5 | 1519.023 | 1519.023 | 1724.771 |
| Error variance 6 | 1352.936 | 1352.645 | 1647.283 |
| Error variance 7 | 2673 | 2673 | 2673 |

*Note:* AV= auxiliary variable, IS= inclusive strategy, RS= restrictive strategy.

# Chapter Five: Discussion

In this chapter, a summary of the main findings, the integration of this study's results with the literature, this study's limitations, and recommendations for researchers and future studies are provided.

## The Main Findings

***Bias***. One of the outcomes of interest in this study was bias, which is the difference between the expected value of the parameter and the true or population parameter. Based on the simulation study results and the empirical study, the two strategies did not differ in parameter estimate bias. In the simulation study, both strategies were almost identical as the restrictive strategy that selected auxiliary variables based on the FMI suggested including all the auxiliary variables in most conditions. As mentioned previously, the FMI showed small differences when including different auxiliary variables. In other words, after adding each auxiliary variable separately, each parameter's FMI was estimated, and the resulting FMIs displayed small differences with the change in auxiliary variables. This small difference might be due to the similarity between the auxiliary variables regarding the magnitude of the correlation with the items. To clarify, since all auxiliary variables correlated with items at the same level, the FMI did not show notable differences between them in the information gained. Therefore, it is not surprising that the performances of both

strategies were almost the same in all the study's outcomes since the inclusive and restrictive strategies were very similar across the board. Generating ten auxiliary variables that vary in their correlations with the items might show different results. However, varying the correlations of the ten variables and manipulating the other study factors would complicate the generation process. Thus, the empirical data in the second phase of this study was used to overcome this limitation.

In addition, running the model without using any auxiliary variables resulted in no parameter estimate bias across conditions. This indicates that the FIML performs well in handling missing data assuming MAR. Thus, the auxiliary variables had little to contribute to reducing bias. This result is consistent with past studies, such as Savalei and Bentler (2009). They found no bias in parameter estimations across conditions, even without using any auxiliary variables with missingness rates of 15% and 30% and with low levels of correlation between items and auxiliary variables (.1 & .3). Another related work that used the CFA model and similar results is Yoo's (2009) study. She examined both linear and nonlinear types of missingness under MCAR, MAR, and MNAR with missingness rates of 10% and 20% using the MI techniques. The findings showed that restrictive and inclusive strategies produced ignorable bias levels in parameter estimates regardless of missingness rate and sample size when the missingness type was MCAR or linear-MAR. The bias in estimation was observed when the mechanism was MNAR or when the type of missingness was nonlinear, which was not covered in the current study. Enders (2008) mentioned that omitting auxiliary variables produced bias compared to the model that included them. Yet, these levels of bias are still considered ignorable under the threshold range that the current

study followed (Muthen et al., 1987). For example, in the structural model, the parameter resulting from regressing the latent variable Y on X was .553 when the auxiliary variable was omitted, while the population parameter was .6. By applying the threshold range that the current study followed (Muthen et al., 1987), we can infer that this level of bias is ignorable. The current study's finding is consistent with previous literature, which indicated that the modern approaches in handling missing data performed well in estimating parameters with up to 30% missingness rates with random missingness. Thus, auxiliary variables have little to add to the imputation process in terms of bias.

The empirical example did not show differences in the parameter estimation bias between the two strategies. Across all conditions, both produced the same amount of parameter bias. The biased parameter was the error variance of the first item, which has the highest factor loading among the other items (.96). In one condition, the second item's error variances produced unacceptable bias levels, which also had the second-highest factor loading (.8). Likewise, Yoo's (2009) study reported that the error variance of the items with high factor loadings showed biased estimates when the conditions were non-ignorable (MNAR) or nonlinear. However, Yoo (2009) examined two low missingness rates (10% & 20%), so we might observe biased error variance estimates with ignorable linear conditions similar to the current study when the factor loadings (≥.95) with missingness rate ≥ 15%, or when the factor loadings (≥.8) with missingness rate higher than 20%. This finding might indicate that the error variance associated with higher factor loadings seems to have been more easily disrupted by missingness than those with lower factor loadings (Yoo, 2009). More research is needed on the impact of missingness on parameter estimates

when the factor loading's magnitude is high. It is worth noting that the restrictive strategy produced fewer biased estimates or lower bias levels in most parameters when the missingness rate was 30%.

***Efficiency.*** The accuracy and precision of parameter estimates were assessed in this study. As expected, the simulation study did not find differences between the two strategies as both were almost identical in the number of the included auxiliary variables. However, the empirical data distinguished between the inclusive strategy and the restrictive strategy. The results showed that the restrictive strategy improved the efficiency of parameter estimation compared to the absence of auxiliary variables and more than the improvement of the inclusive strategy in terms of efficiency, as reflected by FMI. This can be observed in the results of the gain efficiency due to the inclusion of auxiliary variables. The restrictive strategy resulted in a more efficient parameter estimate relative to the absence of the auxiliary variables. In contrast, the inclusive strategy resulted in parameter estimates that were as efficient as the model with no auxiliary variables, and in some cases, it showed losses of efficiency.

As the restrictive strategy produced lower FMI values in most parameters (13 out of 15) across conditions, the loss of efficiency in estimating these parameters due to missing data was lower than the inclusive strategy. This could imply the loss of statistical power due to missing data. The restrictive strategy could increase power whereas the inclusive strategy showed no improvement in power. It indicates that selecting auxiliary variables based on the FMI can help choose variables that can recapture some of the lost information, which will improve power. The inclusive strategy with all available auxiliary

variables could complicate the model by adding noise to the model. Thus, the "expected asymptotic gains in efficiency may be canceled out by variability due to fitting a larger number of parameters" (Savalei & Bentler, 2009, pp. 488-489). The empirical data example provides initial support for the restrictive strategy, suggesting that it can perform better than the inclusive strategy in power and efficiency. Future research would be needed to examine the findings from this empirical study.

***Type A Auxiliary Variables.*** One important finding that is worth noting is related to the type A auxiliary variable which is the variable that associates with incomplete variables and missingness indicators. In the current study, including the type A auxiliary variable showed no difference in bias compared to omitting it. Collins et al. (2001) found that omitting type A auxiliary variables introduced bias to the parameter estimations when the missingness rate was 50%, and the magnitude of the correlation was very high ($r = .9$). Another condition affected by omitting the type A auxiliary variable was the condition with the nonlinear missingness (Collins et al., 2001; Yoo, 2009). In the current study that focused on linear-MAR with low and moderate missingness rates and correlation magnitudes, omitting type A did not produce bias. This indicates that the type A auxiliary variable will effectively handle linear missingness when the missingness rate and the magnitude of the correlation with the imputed variable are very high or when the missingness is nonlinear (Collins et al., 2001; Yoo, 2009).

**Recommendations**

Given the current study's results and the simulation studies that utilized the CFA model to examine the impact of missing data on parameters estimate bias (Enders, 2008;

Enders & Peugh, 2004; Savalei & Bentler, 2009; Yoo, 2009), a few conclusions can be drawn within the limit of the common examined conditions in these studies (missingness rate up to 30%, assuming MCAR or linear-MAR). Since these studies found no differences in bias between auxiliary variables' absence and presence (Enders, 2008; Enders & Peugh, 2004; Savalei & Bentler, 2009), and no differences in bias with using auxiliary variables that associated with imputed items at a low level (Enders & Peugh, 2004; Savalei & Bentler, 2009) or even high level (Enders, 2008; Yoo, 2009), and no differences between the inclusive and restrictive strategies (Yoo, 2009), the modern approaches in handling missing data should perform well in parameter estimation. One exception might be applied to items with high factor loadings (.8-.9) that might produce biased estimates.

Omitting the cause of missingness (the type A auxiliary variable) was not detrimental in terms of bias when the mechanism was linear-MAR (Enders, 2008; Enders & Peugh, 2004; Savalei & Bentler, 2009; Yoo, 2009). However, omitting the type A auxiliary variable could cause bias in parameter estimation when the missingness is nonlinear (Yoo, 2009).

That being said, with a missingness rate of up to 30% and assuming linear-MAR, researchers and practitioners should use modern techniques (MI-FIML) to handle missing data as they showed ignorable parameter estimation bias, even without the use of auxiliary variables.

Although using auxiliary variables showed no impact on estimation bias, researchers should consider the efficiency and power improvement that auxiliary variables can add (Mustillo, 2014). The empirical study demonstrated the impact of adding auxiliary

variables in efficiency and power. Nevertheless, this improvement was consistent only with using the restrictive strategy. In line with Mustillo (2012), the inclusive strategy led to inconsistent results in gaining power and decreasing SE. Thus, the best way to maximize gain and minimize loss is to select auxiliary variables based on the FMI. The current study restricts the common recommendation for including all available variables (Collins et al., 2001) to include only variables that reduced the FMI. Therefore, it is important to evaluate the candidate auxiliary variables and know whether they are likely to be beneficial. The recommended way to evaluate the auxiliary variable's usefulness is by estimating the FMI, indicating how much information we can gain by adding this variable to the model. Consistent with Madley-Dowd et al. (2019), the observed positive association between the FMI and SE shows that FMI can be a useful indicator of the gain in efficiency and power.

**Limitations and Directions for Future Research**

While the current simulation study has included several common design factors, such as the correlation magnitude between auxiliary variables and items, missingness proportion, and mechanism of missingness, other absent factors may also have substantial impacts on the performance of the FMI. For example, the effect of using auxiliary variables with different levels of correlations (low, moderate, and high) has been observed in the empirical example where the FMI shows different estimates, which helped to form an imputation model using the restrictive strategy that increased power and efficiency compared to the inclusive strategy. However, the results cannot be generalized due to the lack of replications in the empirical study. Future studies should examine the selection of auxiliary variables with different levels of correlation magnitude based on the FMI.

In addition, there are multiple other factors worth investigating that have not been included in this study, such as factor loading magnitude that showed different results in the parameter estimation bias when the magnitude was strong. Thus, it might be useful to include factor loading magnitude as a controlled level in another study. Moreover, because this study only used MCAR and linear-MAR, future studies should include nonlinear-MAR and MNAR and compare the FMI performance with the different types of missingness.

In this simulation study, the sample size was fixed at 500 which is considered a large sample size. It may be that the design factors showed no impact on the study's outcomes since the sample size was large. Yoo (2009) found that a sample size of 200 with a missingness rate of 20% tended to result in larger standard error estimates. Thus, with a smaller sample size (100-200), the missing mechanism, missingness rate, and magnitude of the correlation could have more impact on bias and efficacy. Additionally, while it is possible that the benefits seen in this study on efficiency and power were small, this could be due to the efficient large sample size (500). Hence, it is plausible that more benefits in efficiency and power will be observed with a smaller sample size.

This study utilized CFA model which is the basic model and first step in building most types of SEM models. It will be interesting to know if the results of this study can be generalized to other SEM models. Given the consistency of the study's results with previous works (Andridge et al., 2015; Madley-Dowd et al., 2019) on applying the use of FMI in the selection of auxiliary variables, the use of this technique can be used in more statistical analysis such as logistic regression, canonical-correlation, and discriminate analysis. Madley-Dowd et al. (2019) found that FMI is an effective tool in selecting

auxiliary variables to impute the linear regression analysis. In this study, the empirical study showed initial support for using this technic to impute CFA, which can suggest the applicability of the FMI in evaluating auxiliary variables to impute other statistical models.

In conclusion, findings from the current study may not generalize to situations where missingness is substantially higher than 30% or when the missing mechanism is nonlinear. Thus, further studies may extend the study's scope by examining additional factors and/or using alternative levels of the design factors.

**Reflection**

In the following paragraphs, I would like to share the personal experiences and struggles that I endured during the dissertation process.

Simulation studies are considered experimental studies where the researcher manipulates some factors to examine their impacts on some outcomes. In the process of designing my simulation study, I tried to focus on some factors found in the literature of missing data to influence the bias and efficiency of parameter estimates. This study examined using the FMI to select and evaluate the beneficial auxiliary variables among a large number of candidate covariates. Some factors that were of interest are missingness rate and missing mechanisms in both items and auxiliary variables, the magnitude of the correlation between items and auxiliary variables, the variability of the correlation magnitude between items and auxiliary variables in the same condition, the magnitude of factor lodgings of the items, number of candidate auxiliary variables, and sample size. Varying all previous factors in one study will broaden the scope of the study and will

introduce complexity in generation data with different levels of previous factors. Thus, some factors were fixed in this study (sample size, the variability of the correlation magnitude between items and auxiliary variables in the same matrix, magnitude of factor lodgings of the items, number of candidate auxiliary variables). To choose the levels of the manipulated factors, the researcher considered levels of factors that reflect real-world conditions commonly faced by researchers. For example, the magnitudes of the correlation between items and auxiliary variables were chosen to be moderate and low as consulting many actual public data showed this range of magnitudes.

The results of the simulation study were disappointing as there was no difference between the inclusive and restrictive strategies. The lack of variability of the correlation between items and auxiliary variables in the same condition was expected to be the reason for the absence of the difference between the two strategies.

In order to overcome the simulation study limitations, in the second study, I used empirical data where I tried to examine some of the uncovered factors. For example, 17 covariates were included as auxiliary variables that represented low, moderate, and high correlation with the imputed items. As expected, the variability of the correlation between the imputed items and auxiliary variables showed differences between the two strategies in terms of efficiency and power.

In simulation studies where many factors could affect the results and are hard to be all manipulated, it would be a good idea to build the simulation study based on actual data. In other words, the actual data can be treated as the population where the parameters from the existing data set represent the population parameter. This way researcher will

not need to manipulate all variables, and the design will reflect more real-world conditions.

## References


Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology (1965), 112*(4), 545-557.

Andridge, R. R., & Little, R. J. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics*, *27*(2), 153.

Andridge, R., & Thompson, K. J. (2015). Using the fraction of missing information to identify auxiliary variables for imputation procedures via proxy pattern□mixture models. *International Statistical Review, 83*(3), 472-492.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. *Advanced Structural Equation Modeling: Issues and Techniques*, *243*, 277.

Bhattacharyya, K., Meng, H., Hueluer, G., & Hyer, K. (2020). Movement therapy and cognitive function in middle-aged and older adults: A 10-year study. *Innovation in Aging*, *4*(Suppl 1), 364-365.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling, 15*(4), 651-675.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279-4292.

Charles, S. T., Mogle, J., Piazza, J. R., Karlamangla, A., & Almeida, D. M. (2020). Going the distance: The diurnal range of cortisol and its association with cognitive and physiological functioning. *Psychoneuroendocrinology*, *112*, 104516.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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

Eekhout, I., de Boer, R. M., Twisk, J. W., de Vet, H. C., & Heymans, M. W. (2012). Missing data: a systematic review of how they are reported and handled. *Epidemiology*, *23*(5), 729-732.

Enders, C. (2008). A Note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 15*(3), 434-448.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

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*(3), 430-457.

Enders, C. K., & Peugh, J. L. (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling*, *11*(1), 1-19.

Fiero, M. H., Huang, S., Oren, E., & Bell, M. L. (2016). Statistical analysis and handling of missing data in cluster randomized trials: A systematic review. *Trials*, *17*(1), 1-10.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, *10*(1), 80-100.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549-576.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206-213.

Guo, B., Perron, B. E., & Gillespie, D. F. (2009). A systematic review of structural equation modelling in social work research. *British Journal of Social Work*, *39*(8), 1556-1574.

Hardt, J., Herke, M., & Leonhart, R. (2012). Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. *BMC Medical Research Methodology*, *12*(1), 1-13.

Harel, O. (2007). Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology, 4*(1), 75-89.

Hartanto, A., Yong, J. C., Toh, W. X., Lee, S. T., Tng, G. Y., & Tov, W. (2020). Cognitive, social, emotional, and subjective health benefits of computer use in

adults: A 9-year longitudinal study from the Midlife in the United States (MIDUS). *Computers in Human Behavior*, *104*, 106179.

Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research*, *50*(3), 285-299.

IBM Corp. (2017). *IBM SPSS Statistics for Windows*. Armonk, NY: IBM Corp. Retrieved from https://hadoop.apache.org

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6.

Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods*, *51*(5), 2337-2355.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Lachman, M. E., Agrigoroaei, S., Tun, P. A., & Weaver, S. L. (2014). Monitoring cognitive functioning: Psychometric properties of the Brief Test of Adult Cognition by Telephone. *Assessment*, *21*(4), 404-417.

Lang, K., & Little, T. (2016). Principled Missing Data Treatments. *Prevention Science, 19*(3), 284-294.

Lin, F., Heffner, K., Mapstone, M., Chen, D. G. D., & Porsteisson, A. (2014). Frequency of mentally stimulating activities modifies the relationship between

cardiovascular reactivity and executive function in old age. *The American Journal of Geriatric Psychiatry*, *22*(11), 1210-1221.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199-204.

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, *110*, 63-73.

Martens, M. P. (2005). The use of structural equation modeling in counseling psychology research. *The Counseling Psychologist*, *33*(3), 269-298.

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, *100*(1), 43-52.

Mustillo, S. (2012). The effects of auxiliary variables on coefficient bias and efficiency in multiple imputation. *Sociological Methods & Research*, *41*(2), 335-361.

Muthen, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431-462.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8)*. Los Angeles, CA: Authors.

Nishimura, R., Wagner, J., & Elliott, M. (2016). Alternative indicators for the risk of non□response bias: a simulation study. *International Statistical Review*, *84*(1), 43-62.

Orchard, T., & Woodbury, M. A. (1972). *A missing information principle: theory and applications*. Paper presented at the Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525-556.

Raykov, T. (2005). Analysis of longitudinal data with missing values via covariance structure modeling using full-information maximum likelihood. *Structural Equation Modeling., 12*, 331.

Raykov, T., & Marcoulides, G. A. (2014). Identifying useful auxiliary variables for incomplete data analyses: A note on a group difference examination approach. *Educational and Psychological Measurement*, *74*(3), 537-550.

Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in Psychology and Aging: Some proposed guidelines. *Psychology and Aging*, *6*(4), 499.

Raykov, T., & West, B. T. (2016). On enhancing plausibility of the missing at random assumption in incomplete data analyses via evaluation of response-auxiliary variable correlations. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 45-53.

Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. *Educational Psychologist*, *51*(3-4), 305-316.

Rhemtulla, M., Jia, F., Wu, W., & Little, T.D. (2014). Planned missing designs to optimize the efficiency of latent growth parameter estimates. *International Journal of Behavioral Development, 38*(5), 423-434.

Rockel, T. (2020). *Package 'missMethods'*. https://cran.r project.org/web/packages/missMethods/missMethods.pdf

Roiland, R. A., Lin, F., Phelan, C., & Chapman, B. P. (2015). Stress regulation as a link between executive function and pre-frailty in older adults. *The journal of Nutrition, Health & Aging*, *19*(8), 828-838.

Rosseel, Y. (2012). lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. https://www.jstatsoft.org/v48/i02/

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592.

Rubin, D. (1978). Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. *Survey Research Methods., 1*, 20.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys / Donald B. Rubin.* (Wiley series in probability and mathematical statistics. Applied probability and statistics). New York: Wiley.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social psychology bulletin*, *28*(12), 1629-1646.

Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling, 16*(3), 477-497.

Savalei, V., & Rhemtulla, M. (2012). On obtaining estimates of the fraction of missing information from full information maximum likelihood. *Structural Equation Modeling, 19*(3), 477-494.

Von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research, 49*(3), 699-718.

Wang, M. C., & Deng, Q. (2016). The mechanism of auxiliary variables in full information maximum likelihood–based structural equation models with missing data. *Acta Psychologica Sinica*, *48*(11), 1489-1498.

Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opin. Quart.*, *74*, 223–243.

Wagner, J. (2012). A comparison of alternative indicators for the risk of nonresponse bias. *Public Opin. Quart.*, 76, 555–575.

White, I. R., Royston, P. & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399.

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*, *1*(4), 368-376.

Yoo, J. E. (2009). The effect of auxiliary variables and multiple imputation on parameter estimation in confirmatory factor analysis. *Educational and Psychological Measurement, 69*(6), 929-947.

Yuan, K., & Savalei, V. (2014). Consistency, bias and efficiency of the normal-distribution-based MLE: The role of auxiliary variables. *Journal of Multivariate Analysis, 124*, 353.

## Appendix A

Mplus syntax to generate the simulation data

```
TITLE:    CFA MCAR with 15%, and r is Moderate

MONTECARLO:
          NAMES = y1-y6 x1-x10;
          NOBSERVATIONS = 500;
          NREPS = 100;
          SEED = 4533;
          PATMISS = y1(.15) y2(.15) y3(.15) y4(.15) y5(.15) y6(.15)
          x1(.15) x2(.15) x3(.15) x4(.15) x5(.15) x6(.15)
          x7(.15) x8(.15) x9(.15) x10(.15);

PATPROBS = 1;
   REPSAVE = ALL;
   SAVE = MCAR-H.rep*.dat;

ANALYSIS: TYPE = BASIC;

MODEL POPULATION:
          [x1-x10@0];
          x1-x10*1;
          f1 BY y1-y3*0.7;
   f2 BY y4-y6*0.7;
```

```
        f1@1;
    f2@1;
        y1-y6*.51;
    f1 with f2*.4;
    y1-y6 with x1-x10@.42;
    x1-x10 with x1-x10@.4;

OUTPUT:         TECH1  TECH9;
```

## Appendix B

R code for Running CFA

```
library(semTools)
library(lavaan)



mcar <- read.csv("dataforR.csv", na.strings = "999")


#set up model
  mcar.fmi <- 'F1 =~ y1 + y2 + y3
F2 =~ y4 + y5 + y6'


  fit.mcarfmi <- cfa.auxiliary(mcar.fmi, data = mcar,std.lv=TRUE, missing = "fiml",
estimator = "ml", information ="observed",
                          aux = ("x1"))


  #save model standard errors
  SE.step1 <- parameterEstimates(fit.mcarfmi)$se
  #get model-implied covariance matrix and means
  cov.cfa <- fitted.values(fit.mcarfmi)$cov
  means.cfa <- fitted.values(fit.mcarfmi)$mean
  #run the model using model-implied cov. matrix and means as input
  step2.FMI <- cfa(mcar.fmi, sample.cov = cov.cfa, sample.mean =
                     means.cfa, sample.nobs = 500, std.lv = TRUE,
                   meanstructure = TRUE, information = "observed")
```

```
#get standard errors
SE.step2 <- parameterEstimates(step2.FMI)$se
#compute vector of fraction of missing information estimates
temp_store_params$FMI <- 1-(SE.step2^2/SE.step1^2)
```

## Appendix C

**Table 1**

*Missing Counts and Missing Rates for each Condition*

| Condition | Missing Counts /Missing Rate % | | | | | | |
|---|---|---|---|---|---|---|---|
| | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 |
| MCAR 15% | 401/15 | 401/15 | 401/15 | 401/15 | 401/15 | 401/15 | 401/15 |
| MCAR 30% | 802/30 | 802/30 | 802/30 | 802/30 | 802/30 | 802/30 | 802/30 |
| MAR 15% | 401/15 | 401/15 | 401/15 | 401/15 | 401/15 | 401/15 | 401/15 |
| MAR 30% | 802/30 | 802/30 | 802/30 | 802/30 | 802/30 | 802/30 | 802/30 |

**Table 2**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y1.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.006 | .045 | .020 | 1 | .887 | .994 |
| | x2 | .018 | .039 | .202 | 1 | .653 | 1.018 |
| | x3 | -.097 | .048 | 4.155 | 1 | .042 | .907 |
| | x4 | .021 | .011 | 3.481 | 1 | .062 | 1.021 |
| | x5 | -.047 | .049 | .900 | 1 | .343 | .954 |
| | x6 | .011 | .007 | 2.154 | 1 | .142 | 1.011 |
| | x7 | .118 | .336 | .122 | 1 | .726 | 1.125 |
| | age | .003 | .006 | .193 | 1 | .660 | 1.003 |
| | p1 | -.057 | .055 | 1.081 | 1 | .299 | .944 |
| | p2 | -.004 | .047 | .006 | 1 | .939 | .996 |
| | p3 | .067 | .049 | 1.836 | 1 | .175 | 1.069 |
| | p4 | .000 | .042 | .000 | 1 | .995 | 1.000 |
| | p5 | .033 | .042 | .604 | 1 | .437 | 1.033 |
| | p6 | -.032 | .050 | .392 | 1 | .531 | .969 |
| | p7 | -.005 | .042 | .014 | 1 | .905 | .995 |
| | p8 | .037 | .041 | .830 | 1 | .362 | 1.038 |
| | p9 | -.030 | .043 | .504 | 1 | .478 | .970 |
| | Constant | -2.544 | .841 | 9.158 | 1 | .002 | .079 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 3**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y2.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.018 | .046 | .146 | 1 | .702 | .983 |
| | x2 | .033 | .040 | .689 | 1 | .406 | 1.033 |
| | x3 | .030 | .048 | .400 | 1 | .527 | 1.031 |
| | x4 | .027 | .011 | 5.640 | 1 | .018 | 1.027 |
| | x5 | -.077 | .050 | 2.391 | 1 | .122 | .926 |
| | x6 | -.007 | .007 | .963 | 1 | .326 | .993 |
| | x7 | -.015 | .336 | .002 | 1 | .963 | .985 |
| | age | .008 | .006 | 1.444 | 1 | .230 | 1.008 |
| | p1 | .047 | .049 | .923 | 1 | .337 | 1.048 |
| | p2 | .035 | .047 | .571 | 1 | .450 | 1.036 |
| | p3 | .043 | .048 | .825 | 1 | .364 | 1.044 |
| | p4 | .025 | .043 | .323 | 1 | .570 | 1.025 |
| | p5 | .008 | .042 | .038 | 1 | .845 | 1.008 |
| | p6 | .005 | .050 | .011 | 1 | .918 | 1.005 |
| | p7 | -.048 | .042 | 1.317 | 1 | .251 | .953 |
| | p8 | .047 | .041 | 1.285 | 1 | .257 | 1.048 |
| | p9 | -.038 | .043 | .748 | 1 | .387 | .963 |
| | Constant | -2.819 | .840 | 11.259 | 1 | <.001 | .060 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 4**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y3.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.016 | .045 | .126 | 1 | .723 | .984 |
| | x2 | .007 | .039 | .030 | 1 | .862 | 1.007 |
| | x3 | .059 | .047 | 1.597 | 1 | .206 | 1.061 |
| | x4 | -.008 | .011 | .489 | 1 | .484 | .992 |
| | x5 | -.089 | .049 | 3.339 | 1 | .068 | .915 |
| | x6 | .005 | .007 | .414 | 1 | .520 | 1.005 |
| | x7 | .197 | .321 | .375 | 1 | .540 | 1.217 |
| | age | .006 | .006 | 1.065 | 1 | .302 | 1.006 |
| | p1 | -.091 | .055 | 2.721 | 1 | .099 | .913 |
| | p2 | .017 | .046 | .128 | 1 | .721 | 1.017 |
| | p3 | .033 | .046 | .517 | 1 | .472 | 1.034 |
| | p4 | .032 | .042 | .593 | 1 | .441 | 1.033 |
| | p5 | .015 | .042 | .129 | 1 | .719 | 1.015 |
| | p6 | -.023 | .050 | .217 | 1 | .642 | .977 |
| | p7 | .012 | .042 | .078 | 1 | .780 | 1.012 |
| | p8 | -.021 | .040 | .283 | 1 | .595 | .979 |
| | p9 | -.022 | .042 | .265 | 1 | .607 | .979 |
| | Constant | -2.446 | .819 | 8.926 | 1 | .003 | .087 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 5**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y4.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.032 | .046 | .483 | 1 | .487 | .969 |
| | x2 | .041 | .039 | 1.107 | 1 | .293 | 1.042 |
| | x3 | .032 | .047 | .463 | 1 | .496 | 1.033 |
| | x4 | .000 | .011 | .001 | 1 | .980 | 1.000 |
| | x5 | .019 | .049 | .147 | 1 | .701 | 1.019 |
| | x6 | -.011 | .007 | 2.096 | 1 | .148 | .989 |
| | x7 | -.077 | .339 | .052 | 1 | .820 | .926 |
| | age | .008 | .006 | 1.706 | 1 | .192 | 1.008 |
| | p1 | -.124 | .060 | 4.311 | 1 | .038 | .883 |
| | p2 | -.057 | .047 | 1.457 | 1 | .227 | .945 |
| | p3 | .133 | .054 | 6.108 | 1 | .013 | 1.143 |
| | p4 | -.009 | .043 | .043 | 1 | .836 | .991 |
| | p5 | .015 | .042 | .128 | 1 | .721 | 1.015 |
| | p6 | .048 | .050 | .909 | 1 | .340 | 1.049 |
| | p7 | .027 | .044 | .382 | 1 | .536 | 1.028 |
| | p8 | .066 | .040 | 2.695 | 1 | .101 | 1.068 |
| | p9 | -.035 | .043 | .650 | 1 | .420 | .966 |
| | Constant | -2.809 | .855 | 10.792 | 1 | .001 | .060 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 6**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y5.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .012 | .046 | .064 | 1 | .801 | 1.012 |
| | x2 | .021 | .039 | .285 | 1 | .593 | 1.021 |
| | x3 | .044 | .047 | .852 | 1 | .356 | 1.045 |
| | x4 | .022 | .011 | 3.724 | 1 | .054 | 1.022 |
| | x5 | -.078 | .049 | 2.470 | 1 | .116 | .925 |
| | x6 | -.003 | .007 | .225 | 1 | .635 | .997 |
| | x7 | -.197 | .340 | .338 | 1 | .561 | .821 |
| | age | .006 | .006 | .761 | 1 | .383 | 1.006 |
| | p1 | -.039 | .054 | .518 | 1 | .472 | .962 |
| | p2 | -.043 | .048 | .797 | 1 | .372 | .958 |
| | p3 | .070 | .048 | 2.180 | 1 | .140 | 1.073 |
| | p4 | .058 | .042 | 1.885 | 1 | .170 | 1.060 |
| | p5 | -.019 | .043 | .196 | 1 | .658 | .981 |
| | p6 | .058 | .051 | 1.335 | 1 | .248 | 1.060 |
| | p7 | -.118 | .040 | 8.588 | 1 | .003 | .889 |
| | p8 | -.087 | .042 | 4.309 | 1 | .038 | .917 |
| | p9 | .014 | .043 | .108 | 1 | .743 | 1.014 |
| | Constant | -2.201 | .841 | 6.844 | 1 | .009 | .111 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 7**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y6.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .043 | .045 | .925 | 1 | .336 | 1.044 |
| | x2 | -.059 | .039 | 2.252 | 1 | .133 | .943 |
| | x3 | -.064 | .048 | 1.796 | 1 | .180 | .938 |
| | x4 | -.007 | .011 | .387 | 1 | .534 | .993 |
| | x5 | -.003 | .049 | .003 | 1 | .959 | .997 |
| | x6 | .011 | .007 | 2.333 | 1 | .127 | 1.011 |
| | x7 | -.030 | .340 | .008 | 1 | .929 | .970 |
| | age | -.007 | .006 | 1.320 | 1 | .251 | .993 |
| | p1 | .084 | .048 | 3.039 | 1 | .081 | 1.088 |
| | p2 | -.019 | .047 | .164 | 1 | .686 | .981 |
| | p3 | .004 | .046 | .006 | 1 | .939 | 1.004 |
| | p4 | .000 | .043 | .000 | 1 | .992 | 1.000 |
| | p5 | .010 | .043 | .053 | 1 | .818 | 1.010 |
| | p6 | .014 | .050 | .079 | 1 | .779 | 1.014 |
| | p7 | -.040 | .042 | .929 | 1 | .335 | .961 |
| | p8 | .032 | .041 | .616 | 1 | .433 | 1.032 |
| | p9 | .039 | .044 | .784 | 1 | .376 | 1.039 |
| | Constant | -1.518 | .832 | 3.327 | 1 | .068 | .219 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 8**

*Checking the Missing Mechanisms (MCAR -15%) for variable Y7.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .006 | .046 | .015 | 1 | .904 | 1.006 |
| | x2 | -.079 | .040 | 3.951 | 1 | .047 | .924 |
| | x3 | .004 | .048 | .006 | 1 | .938 | 1.004 |
| | x4 | .004 | .011 | .138 | 1 | .710 | 1.004 |
| | x5 | .083 | .049 | 2.792 | 1 | .095 | 1.086 |
| | x6 | -.006 | .007 | .653 | 1 | .419 | .994 |
| | x7 | -.062 | .339 | .034 | 1 | .854 | .939 |
| | age | -.007 | .006 | 1.397 | 1 | .237 | .993 |
| | p1 | -.087 | .055 | 2.454 | 1 | .117 | .917 |
| | p2 | .065 | .047 | 1.903 | 1 | .168 | 1.067 |
| | p3 | .064 | .048 | 1.777 | 1 | .183 | 1.066 |
| | p4 | -.020 | .042 | .237 | 1 | .626 | .980 |
| | p5 | .036 | .043 | .698 | 1 | .404 | 1.036 |
| | p6 | .010 | .051 | .037 | 1 | .848 | 1.010 |
| | p7 | -.053 | .041 | 1.672 | 1 | .196 | .948 |
| | p8 | -.120 | .042 | 8.065 | 1 | .005 | .887 |
| | p9 | .042 | .043 | .951 | 1 | .329 | 1.043 |
| | Constant | -.964 | .839 | 1.320 | 1 | .251 | .381 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 9**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y1.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.071 | .036 | 3.865 | 1 | .049 | .932 |
| | x2 | .064 | .031 | 4.311 | 1 | .038 | 1.066 |
| | x3 | -.019 | .037 | .273 | 1 | .601 | .981 |
| | x4 | .030 | .009 | 11.803 | 1 | <.001 | 1.031 |
| | x5 | -.030 | .038 | .617 | 1 | .432 | .970 |
| | x6 | .004 | .006 | .602 | 1 | .438 | 1.004 |
| | x7 | .183 | .262 | .486 | 1 | .486 | 1.201 |
| | age | .002 | .005 | .220 | 1 | .639 | 1.002 |
| | p1 | -.005 | .041 | .013 | 1 | .909 | .995 |
| | p2 | .058 | .037 | 2.534 | 1 | .111 | 1.060 |
| | p3 | .085 | .038 | 5.025 | 1 | .025 | 1.089 |
| | p4 | -.007 | .033 | .045 | 1 | .832 | .993 |
| | p5 | .037 | .033 | 1.228 | 1 | .268 | 1.037 |
| | p6 | .003 | .039 | .007 | 1 | .934 | 1.003 |
| | p7 | -.006 | .033 | .036 | 1 | .850 | .994 |
| | p8 | .011 | .032 | .126 | 1 | .723 | 1.011 |
| | p9 | -.054 | .034 | 2.610 | 1 | .106 | .947 |
| | Constant | -2.129 | .656 | 10.518 | 1 | .001 | .119 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 10**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y2.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.025 | .035 | .484 | 1 | .487 | .976 |
| | x2 | .013 | .030 | .191 | 1 | .662 | 1.013 |
| | x3 | .052 | .037 | 2.001 | 1 | .157 | 1.053 |
| | x4 | .007 | .009 | .666 | 1 | .414 | 1.007 |
| | x5 | -.047 | .038 | 1.487 | 1 | .223 | .955 |
| | x6 | -.004 | .006 | .626 | 1 | .429 | .996 |
| | x7 | -.160 | .263 | .369 | 1 | .543 | .852 |
| | age | .000 | .005 | .007 | 1 | .935 | 1.000 |
| | p1 | -.050 | .041 | 1.447 | 1 | .229 | .951 |
| | p2 | -.053 | .037 | 2.090 | 1 | .148 | .948 |
| | p3 | .075 | .038 | 4.019 | 1 | .045 | 1.078 |
| | p4 | .048 | .033 | 2.118 | 1 | .146 | 1.049 |
| | p5 | .044 | .033 | 1.802 | 1 | .179 | 1.045 |
| | p6 | .018 | .039 | .209 | 1 | .648 | 1.018 |
| | p7 | .008 | .033 | .054 | 1 | .816 | 1.008 |
| | p8 | .000 | .032 | .000 | 1 | .998 | 1.000 |
| | p9 | -.018 | .033 | .298 | 1 | .585 | .982 |
| | Constant | -1.285 | .650 | 3.916 | 1 | .048 | .277 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 11**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y3.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .021 | .035 | .345 | 1 | .557 | 1.021 |
| | x2 | -.018 | .031 | .331 | 1 | .565 | .983 |
| | x3 | -.090 | .037 | 5.805 | 1 | .016 | .914 |
| | x4 | .005 | .009 | .315 | 1 | .574 | 1.005 |
| | x5 | .003 | .038 | .008 | 1 | .928 | 1.003 |
| | x6 | .005 | .006 | .843 | 1 | .358 | 1.005 |
| | x7 | .057 | .261 | .047 | 1 | .828 | 1.058 |
| | age | -.002 | .005 | .128 | 1 | .720 | .998 |
| | p1 | .050 | .039 | 1.579 | 1 | .209 | 1.051 |
| | p2 | -.017 | .037 | .210 | 1 | .646 | .983 |
| | p3 | .001 | .036 | .001 | 1 | .980 | 1.001 |
| | p4 | .015 | .033 | .206 | 1 | .650 | 1.015 |
| | p5 | .026 | .033 | .630 | 1 | .427 | 1.027 |
| | p6 | .028 | .039 | .516 | 1 | .473 | 1.029 |
| | p7 | -.023 | .033 | .511 | 1 | .475 | .977 |
| | p8 | -.022 | .032 | .472 | 1 | .492 | .978 |
| | p9 | .025 | .034 | .536 | 1 | .464 | 1.025 |
| | Constant | -.930 | .646 | 2.070 | 1 | .150 | .395 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 12**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y4.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.026 | .036 | .520 | 1 | .471 | .975 |
| | x2 | -.020 | .031 | .443 | 1 | .506 | .980 |
| | x3 | -.054 | .037 | 2.118 | 1 | .146 | .947 |
| | x4 | .012 | .009 | 2.000 | 1 | .157 | 1.013 |
| | x5 | .099 | .038 | 6.630 | 1 | .010 | 1.104 |
| | x6 | -.009 | .006 | 2.281 | 1 | .131 | .991 |
| | x7 | -.051 | .260 | .039 | 1 | .844 | .950 |
| | age | -.003 | .005 | .349 | 1 | .555 | .997 |
| | p1 | -.080 | .042 | 3.520 | 1 | .061 | .924 |
| | p2 | .007 | .037 | .038 | 1 | .845 | 1.007 |
| | p3 | -.017 | .035 | .223 | 1 | .637 | .983 |
| | p4 | .014 | .033 | .189 | 1 | .664 | 1.014 |
| | p5 | -.001 | .034 | .001 | 1 | .977 | .999 |
| | p6 | .061 | .039 | 2.369 | 1 | .124 | 1.062 |
| | p7 | -.040 | .033 | 1.521 | 1 | .217 | .961 |
| | p8 | -.085 | .032 | 6.941 | 1 | .008 | .918 |
| | p9 | .005 | .033 | .018 | 1 | .892 | 1.005 |
| | Constant | .227 | .644 | .124 | 1 | .724 | 1.255 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 14**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y5.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.035 | .035 | 1.003 | 1 | .317 | .965 |
| | x2 | .005 | .030 | .031 | 1 | .860 | 1.005 |
| | x3 | .001 | .036 | .001 | 1 | .981 | 1.001 |
| | x4 | -.011 | .009 | 1.545 | 1 | .214 | .989 |
| | x5 | .013 | .038 | .118 | 1 | .731 | 1.013 |
| | x6 | .004 | .006 | .409 | 1 | .523 | 1.004 |
| | x7 | -.087 | .260 | .112 | 1 | .738 | .917 |
| | age | -.004 | .005 | .803 | 1 | .370 | .996 |
| | p1 | .076 | .039 | 3.774 | 1 | .052 | 1.078 |
| | p2 | -.021 | .036 | .348 | 1 | .556 | .979 |
| | p3 | -.004 | .035 | .010 | 1 | .919 | .996 |
| | p4 | -.033 | .033 | 1.005 | 1 | .316 | .968 |
| | p5 | -.018 | .033 | .285 | 1 | .593 | .982 |
| | p6 | -.004 | .039 | .010 | 1 | .919 | .996 |
| | p7 | .026 | .033 | .615 | 1 | .433 | 1.026 |
| | p8 | .038 | .031 | 1.460 | 1 | .227 | 1.039 |
| | p9 | .051 | .033 | 2.374 | 1 | .123 | 1.053 |
| | Constant | -.545 | .640 | .725 | 1 | .394 | .580 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 15**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y6.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .009 | .035 | .061 | 1 | .805 | 1.009 |
| | x2 | -.010 | .030 | .102 | 1 | .749 | .990 |
| | x3 | .020 | .037 | .292 | 1 | .589 | 1.020 |
| | x4 | -.006 | .009 | .391 | 1 | .532 | .994 |
| | x5 | -.029 | .038 | .558 | 1 | .455 | .972 |
| | x6 | .001 | .006 | .013 | 1 | .909 | 1.001 |
| | x7 | -.167 | .264 | .398 | 1 | .528 | .847 |
| | age | -.004 | .005 | .570 | 1 | .450 | .996 |
| | p1 | .023 | .040 | .327 | 1 | .568 | 1.023 |
| | p2 | .039 | .036 | 1.121 | 1 | .290 | 1.039 |
| | p3 | .010 | .036 | .076 | 1 | .782 | 1.010 |
| | p4 | -.052 | .033 | 2.452 | 1 | .117 | .950 |
| | p5 | -.003 | .033 | .006 | 1 | .938 | .997 |
| | p6 | .023 | .039 | .332 | 1 | .564 | 1.023 |
| | p7 | .050 | .033 | 2.224 | 1 | .136 | 1.051 |
| | p8 | -.014 | .032 | .188 | 1 | .664 | .986 |
| | p9 | -.017 | .033 | .255 | 1 | .613 | .983 |
| | Constant | -.594 | .646 | .845 | 1 | .358 | .552 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 16**

*Checking the Missing Mechanisms (MCAR -30%) for variable Y7.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.026 | .035 | .547 | 1 | .460 | .974 |
| | x2 | .032 | .031 | 1.065 | 1 | .302 | 1.032 |
| | x3 | .014 | .037 | .151 | 1 | .697 | 1.014 |
| | x4 | .015 | .009 | 3.067 | 1 | .080 | 1.016 |
| | x5 | -.012 | .038 | .098 | 1 | .755 | .988 |
| | x6 | -.010 | .006 | 3.271 | 1 | .071 | .990 |
| | x7 | -.165 | .262 | .397 | 1 | .529 | .848 |
| | age | .003 | .005 | .515 | 1 | .473 | 1.003 |
| | p1 | .019 | .040 | .225 | 1 | .635 | 1.019 |
| | p2 | .036 | .037 | .969 | 1 | .325 | 1.037 |
| | p3 | -.045 | .035 | 1.687 | 1 | .194 | .956 |
| | p4 | -.049 | .033 | 2.237 | 1 | .135 | .952 |
| | p5 | .026 | .033 | .604 | 1 | .437 | 1.026 |
| | p6 | -.063 | .040 | 2.542 | 1 | .111 | .939 |
| | p7 | .083 | .034 | 6.053 | 1 | .014 | 1.086 |
| | p8 | .018 | .032 | .303 | 1 | .582 | 1.018 |
| | p9 | -.022 | .033 | .440 | 1 | .507 | .978 |
| | Constant | -.689 | .642 | 1.153 | 1 | .283 | .502 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 17**

*Checking the Missing Mechanisms (MAR -15%) for variable Y1.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .020 | .041 | .233 | 1 | .629 | 1.020 |
| | x2 | -.019 | .035 | .285 | 1 | .593 | .981 |
| | x3 | -.028 | .043 | .422 | 1 | .516 | .972 |
| | x4 | -.004 | .011 | .155 | 1 | .694 | .996 |
| | x5 | .002 | .044 | .002 | 1 | .965 | 1.002 |
| | x6 | -.012 | .006 | 3.612 | 1 | .057 | .988 |
| | x7 | .000 | .000 | .376 | 1 | .540 | 1.000 |
| | age | .029 | .006 | 26.009 | 1 | <.001 | 1.029 |
| | p1 | .000 | .000 | .932 | 1 | .334 | 1.000 |
| | p2 | .000 | .000 | 1.439 | 1 | .230 | 1.000 |
| | p3 | .000 | .000 | .730 | 1 | .393 | 1.000 |
| | p4 | .000 | .000 | .026 | 1 | .871 | 1.000 |
| | p5 | .000 | .000 | .095 | 1 | .759 | 1.000 |
| | p6 | .000 | .000 | 1.258 | 1 | .262 | 1.000 |
| | p7 | .000 | .000 | .722 | 1 | .396 | 1.000 |
| | p8 | .000 | .000 | .332 | 1 | .564 | 1.000 |
| | p9 | .000 | .000 | .977 | 1 | .323 | 1.000 |
| | Constant | -3.000 | .558 | 28.917 | 1 | <.001 | .050 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 18**

*Checking the Missing Mechanisms (MAR -15%) for variable Y2.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .003 | .041 | .006 | 1 | .939 | 1.003 |
| | x2 | -.019 | .036 | .275 | 1 | .600 | .982 |
| | x3 | -.032 | .041 | .620 | 1 | .431 | .968 |
| | x4 | .004 | .010 | .182 | 1 | .670 | 1.004 |
| | x5 | .001 | .005 | .031 | 1 | .861 | 1.001 |
| | x6 | .000 | .001 | .077 | 1 | .781 | 1.000 |
| | x7 | .000 | .000 | .761 | 1 | .383 | 1.000 |
| | age | .025 | .005 | 20.782 | 1 | <.001 | 1.025 |
| | p1 | .000 | .000 | 3.582 | 1 | .058 | 1.000 |
| | p2 | .000 | .000 | 2.063 | 1 | .151 | 1.000 |
| | p3 | .000 | .000 | .205 | 1 | .651 | 1.000 |
| | p4 | .000 | .001 | .122 | 1 | .727 | 1.000 |
| | p5 | .001 | .001 | .539 | 1 | .463 | 1.001 |
| | p6 | .000 | .000 | .387 | 1 | .534 | 1.000 |
| | p7 | .000 | .000 | .451 | 1 | .502 | 1.000 |
| | p8 | .000 | .001 | .071 | 1 | .790 | 1.000 |
| | p9 | .000 | .001 | .218 | 1 | .641 | 1.000 |
| | Constant | -3.226 | .509 | 40.152 | 1 | <.001 | .040 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 19**

*Checking the Missing Mechanisms (MAR -15%) for variable Y3.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.073 | .041 | 3.102 | 1 | .078 | .930 |
| | x2 | .055 | .035 | 2.378 | 1 | .123 | 1.056 |
| | x3 | -.006 | .041 | .024 | 1 | .878 | .994 |
| | x4 | .001 | .010 | .014 | 1 | .906 | 1.001 |
| | x5 | .001 | .001 | .211 | 1 | .646 | 1.001 |
| | x6 | .000 | .001 | .095 | 1 | .757 | 1.000 |
| | x7 | .000 | .000 | .375 | 1 | .541 | 1.000 |
| | age | .028 | .005 | 27.606 | 1 | <.001 | 1.029 |
| | p1 | .000 | .000 | 3.696 | 1 | .055 | 1.000 |
| | p2 | .000 | .000 | .066 | 1 | .797 | 1.000 |
| | p3 | .000 | .000 | .356 | 1 | .551 | 1.000 |
| | p4 | .000 | .000 | 1.572 | 1 | .210 | 1.000 |
| | p5 | .000 | .000 | .284 | 1 | .594 | 1.000 |
| | p6 | .000 | .000 | .287 | 1 | .592 | 1.000 |
| | p7 | .000 | .000 | .678 | 1 | .410 | 1.000 |
| | p8 | .000 | .000 | .008 | 1 | .930 | 1.000 |
| | p9 | .000 | .000 | 3.789 | 1 | .052 | 1.000 |
| | Constant | -3.304 | .506 | 42.695 | 1 | <.001 | .037 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 20**

*Checking the Missing Mechanisms (MAR -15%) for variable Y4.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.008 | .041 | .038 | 1 | .845 | .992 |
| | x2 | -.004 | .035 | .011 | 1 | .916 | .996 |
| | x3 | .042 | .042 | 1.040 | 1 | .308 | 1.043 |
| | x4 | .011 | .010 | 1.195 | 1 | .274 | 1.011 |
| | x5 | .000 | .002 | .044 | 1 | .834 | 1.000 |
| | x6 | -.016 | .006 | 6.947 | 1 | .008 | .984 |
| | x7 | .000 | .000 | .395 | 1 | .530 | 1.000 |
| | age | .026 | .006 | 21.348 | 1 | <.001 | 1.026 |
| | p1 | .000 | .000 | .009 | 1 | .924 | 1.000 |
| | p2 | -.001 | .002 | .112 | 1 | .738 | .999 |
| | p3 | .000 | .000 | .605 | 1 | .437 | 1.000 |
| | p4 | .003 | .005 | .295 | 1 | .587 | 1.003 |
| | p5 | .000 | .002 | .076 | 1 | .783 | 1.000 |
| | p6 | .000 | .000 | 1.393 | 1 | .238 | 1.000 |
| | p7 | -.001 | .002 | .082 | 1 | .774 | .999 |
| | p8 | .000 | .002 | .028 | 1 | .867 | 1.000 |
| | p9 | -.001 | .002 | .073 | 1 | .787 | .999 |
| | Constant | -3.178 | .554 | 32.948 | 1 | <.001 | .042 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 21**

*Checking the Missing Mechanisms (MAR -15%) for variable Y5.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .079 | .041 | 3.729 | 1 | .053 | 1.082 |
| | x2 | -.046 | .035 | 1.770 | 1 | .183 | .955 |
| | x3 | -.034 | .042 | .645 | 1 | .422 | .967 |
| | x4 | -.006 | .010 | .311 | 1 | .577 | .994 |
| | x5 | -.021 | .043 | .253 | 1 | .615 | .979 |
| | x6 | .000 | .002 | .041 | 1 | .839 | 1.000 |
| | x7 | .000 | .000 | 1.530 | 1 | .216 | 1.000 |
| | age | .031 | .005 | 31.313 | 1 | <.001 | 1.031 |
| | p1 | -.035 | .027 | 1.634 | 1 | .201 | .966 |
| | p2 | .000 | .000 | .049 | 1 | .825 | 1.000 |
| | p3 | .000 | .000 | .184 | 1 | .668 | 1.000 |
| | p4 | -.061 | .036 | 2.994 | 1 | .084 | .940 |
| | p5 | .000 | .000 | .058 | 1 | .810 | 1.000 |
| | p6 | .000 | .000 | 2.703 | 1 | .100 | 1.000 |
| | p7 | .000 | .000 | .207 | 1 | .649 | 1.000 |
| | p8 | .096 | .036 | 7.150 | 1 | .007 | 1.101 |
| | p9 | .000 | .000 | .173 | 1 | .677 | 1.000 |
| | Constant | -3.691 | .514 | 51.565 | 1 | <.001 | .025 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 22**

*Checking the Missing Mechanisms (MAR -15%) for variable Y6.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.028 | .041 | .447 | 1 | .504 | .973 |
| | x2 | .045 | .035 | 1.625 | 1 | .202 | 1.046 |
| | x3 | .071 | .043 | 2.803 | 1 | .094 | 1.074 |
| | x4 | .007 | .010 | .445 | 1 | .505 | 1.007 |
| | x5 | -.035 | .044 | .641 | 1 | .423 | .965 |
| | x6 | -.004 | .006 | .447 | 1 | .504 | .996 |
| | x7 | .000 | .000 | 1.646 | 1 | .200 | 1.000 |
| | age | .030 | .006 | 27.511 | 1 | <.001 | 1.030 |
| | p1 | -.001 | .002 | .124 | 1 | .725 | .999 |
| | p2 | .000 | .000 | .106 | 1 | .745 | 1.000 |
| | p3 | .000 | .000 | 1.920 | 1 | .166 | 1.000 |
| | p4 | .000 | .000 | 1.435 | 1 | .231 | 1.000 |
| | p5 | -.001 | .007 | .014 | 1 | .906 | .999 |
| | p6 | .000 | .000 | .049 | 1 | .825 | 1.000 |
| | p7 | .000 | .000 | 3.325 | 1 | .068 | 1.000 |
| | p8 | .001 | .007 | .034 | 1 | .855 | 1.001 |
| | p9 | .000 | .001 | .142 | 1 | .706 | 1.000 |
| | Constant | -3.972 | .564 | 49.617 | 1 | <.001 | .019 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 23**

*Checking the Missing Mechanisms (MAR -30%) for variable Y1.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.023 | .036 | .423 | 1 | .515 | .977 |
| | x2 | -.014 | .031 | .193 | 1 | .661 | .986 |
| | x3 | -.015 | .038 | .149 | 1 | .699 | .986 |
| | x4 | .005 | .009 | .310 | 1 | .578 | 1.005 |
| | x5 | .006 | .039 | .021 | 1 | .883 | 1.006 |
| | x6 | .001 | .006 | .033 | 1 | .856 | 1.001 |
| | x7 | .113 | .262 | .186 | 1 | .667 | 1.119 |
| | age | .036 | .005 | 50.101 | 1 | <.001 | 1.036 |
| | p1 | .017 | .040 | .186 | 1 | .666 | 1.018 |
| | p2 | .014 | .037 | .141 | 1 | .708 | 1.014 |
| | p3 | -.028 | .036 | .622 | 1 | .430 | .972 |
| | p4 | -.021 | .034 | .402 | 1 | .526 | .979 |
| | p5 | .019 | .034 | .335 | 1 | .563 | 1.020 |
| | p6 | -.018 | .040 | .210 | 1 | .647 | .982 |
| | p7 | .070 | .034 | 4.191 | 1 | .041 | 1.073 |
| | p8 | -.030 | .033 | .830 | 1 | .362 | .971 |
| | p9 | .033 | .034 | .945 | 1 | .331 | 1.034 |
| | Constant | -3.412 | .665 | 26.352 | 1 | <.001 | .033 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 24**

*Checking the Missing Mechanisms (MAR -30%) for variable Y2.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .032 | .036 | .818 | 1 | .366 | 1.033 |
| | x2 | -.026 | .031 | .716 | 1 | .397 | .974 |
| | x3 | .011 | .037 | .090 | 1 | .764 | 1.011 |
| | x4 | .003 | .009 | .089 | 1 | .765 | 1.003 |
| | x5 | .012 | .039 | .097 | 1 | .756 | 1.012 |
| | x6 | .007 | .006 | 1.399 | 1 | .237 | 1.007 |
| | x7 | -.129 | .266 | .236 | 1 | .627 | .879 |
| | age | .040 | .005 | 62.855 | 1 | <.001 | 1.041 |
| | p1 | .102 | .039 | 6.783 | 1 | .009 | 1.107 |
| | p2 | -.017 | .037 | .201 | 1 | .654 | .983 |
| | p3 | .006 | .036 | .028 | 1 | .867 | 1.006 |
| | p4 | .044 | .034 | 1.685 | 1 | .194 | 1.044 |
| | p5 | -.011 | .034 | .117 | 1 | .732 | .989 |
| | p6 | .020 | .040 | .257 | 1 | .612 | 1.020 |
| | p7 | -.032 | .033 | .938 | 1 | .333 | .968 |
| | p8 | -.012 | .032 | .134 | 1 | .714 | .988 |
| | p9 | .012 | .034 | .120 | 1 | .729 | 1.012 |
| | Constant | -4.065 | .669 | 36.976 | 1 | <.001 | .017 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 25**

*Checking the Missing Mechanisms (MAR -30%) for variable Y3.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .042 | .036 | 1.343 | 1 | .246 | 1.042 |
| | x2 | -.019 | .031 | .380 | 1 | .537 | .981 |
| | x3 | .004 | .038 | .013 | 1 | .909 | 1.004 |
| | x4 | -.022 | .009 | 5.696 | 1 | .017 | .978 |
| | x5 | .020 | .039 | .259 | 1 | .611 | 1.020 |
| | x6 | .004 | .006 | .521 | 1 | .470 | 1.004 |
| | x7 | .470 | .258 | 3.329 | 1 | .068 | 1.600 |
| | age | .036 | .005 | 50.044 | 1 | <.001 | 1.036 |
| | p1 | -.009 | .040 | .044 | 1 | .833 | .992 |
| | p2 | .006 | .037 | .027 | 1 | .870 | 1.006 |
| | p3 | .024 | .036 | .433 | 1 | .510 | 1.024 |
| | p4 | -.022 | .034 | .423 | 1 | .516 | .978 |
| | p5 | .040 | .033 | 1.471 | 1 | .225 | 1.041 |
| | p6 | .014 | .040 | .117 | 1 | .732 | 1.014 |
| | p7 | -.010 | .033 | .088 | 1 | .766 | .990 |
| | p8 | .013 | .033 | .160 | 1 | .689 | 1.013 |
| | p9 | .008 | .034 | .051 | 1 | .821 | 1.008 |
| | Constant | -3.912 | .664 | 34.733 | 1 | <.001 | .020 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 25**

*Checking the Missing Mechanisms (MAR -30%) for variable Y4.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .023 | .036 | .403 | 1 | .526 | 1.023 |
| | x2 | .002 | .031 | .003 | 1 | .959 | 1.002 |
| | x3 | -.013 | .037 | .114 | 1 | .736 | .987 |
| | x4 | .005 | .009 | .292 | 1 | .589 | 1.005 |
| | x5 | -.028 | .039 | .530 | 1 | .467 | .972 |
| | x6 | .004 | .006 | .503 | 1 | .478 | 1.004 |
| | x7 | -.328 | .269 | 1.486 | 1 | .223 | .721 |
| | age | .043 | .005 | 69.765 | 1 | <.001 | 1.043 |
| | p1 | .049 | .040 | 1.523 | 1 | .217 | 1.051 |
| | p2 | -.092 | .038 | 6.000 | 1 | .014 | .912 |
| | p3 | .044 | .037 | 1.415 | 1 | .234 | 1.045 |
| | p4 | -.021 | .033 | .391 | 1 | .532 | .979 |
| | p5 | -.033 | .034 | .961 | 1 | .327 | .967 |
| | p6 | .026 | .040 | .434 | 1 | .510 | 1.027 |
| | p7 | .014 | .034 | .171 | 1 | .679 | 1.014 |
| | p8 | .040 | .032 | 1.519 | 1 | .218 | 1.041 |
| | p9 | .019 | .034 | .326 | 1 | .568 | 1.019 |
| | Constant | -3.832 | .670 | 32.673 | 1 | <.001 | .022 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 26**

*Checking the Missing Mechanisms (MAR -30%) for variable Y5.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.008 | .036 | .056 | 1 | .813 | .992 |
| | x2 | -.002 | .031 | .004 | 1 | .947 | .998 |
| | x3 | -.025 | .038 | .436 | 1 | .509 | .975 |
| | x4 | .000 | .009 | .002 | 1 | .965 | 1.000 |
| | x5 | -.036 | .039 | .853 | 1 | .356 | .965 |
| | x6 | .012 | .006 | 4.016 | 1 | .045 | 1.012 |
| | x7 | .121 | .260 | .216 | 1 | .642 | 1.128 |
| | age | .040 | .005 | 61.387 | 1 | <.001 | 1.040 |
| | p1 | .042 | .040 | 1.122 | 1 | .289 | 1.043 |
| | p2 | -.099 | .038 | 6.753 | 1 | .009 | .906 |
| | p3 | .006 | .036 | .024 | 1 | .878 | 1.006 |
| | p4 | -.032 | .033 | .914 | 1 | .339 | .968 |
| | p5 | -.012 | .034 | .121 | 1 | .728 | .988 |
| | p6 | .062 | .040 | 2.355 | 1 | .125 | 1.064 |
| | p7 | -.006 | .033 | .032 | 1 | .858 | .994 |
| | p8 | .022 | .033 | .439 | 1 | .508 | 1.022 |
| | p9 | .005 | .034 | .019 | 1 | .892 | 1.005 |
| | Constant | -3.605 | .662 | 29.629 | 1 | <.001 | .027 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 27**

*Checking the Missing Mechanisms (MAR -30%) for variable Y6.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | -.009 | .036 | .060 | 1 | .806 | .991 |
| | x2 | -.006 | .031 | .040 | 1 | .841 | .994 |
| | x3 | -.009 | .038 | .059 | 1 | .808 | .991 |
| | x4 | -.007 | .009 | .566 | 1 | .452 | .993 |
| | x5 | .030 | .039 | .581 | 1 | .446 | 1.030 |
| | x6 | .008 | .006 | 1.659 | 1 | .198 | 1.008 |
| | x7 | -.473 | .272 | 3.022 | 1 | .082 | .623 |
| | age | .050 | .005 | 94.896 | 1 | <.001 | 1.052 |
| | p1 | -.084 | .043 | 3.874 | 1 | .049 | .919 |
| | p2 | -.027 | .038 | .504 | 1 | .478 | .973 |
| | p3 | -.042 | .036 | 1.389 | 1 | .239 | .959 |
| | p4 | -.033 | .034 | .946 | 1 | .331 | .968 |
| | p5 | .021 | .034 | .375 | 1 | .540 | 1.021 |
| | p6 | .009 | .040 | .045 | 1 | .832 | 1.009 |
| | p7 | .013 | .034 | .147 | 1 | .702 | 1.013 |
| | p8 | -.008 | .033 | .060 | 1 | .806 | .992 |
| | p9 | -.001 | .034 | .002 | 1 | .965 | .999 |
| | Constant | -3.167 | .668 | 22.461 | 1 | <.001 | .042 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.

**Table 28**

*Checking the Missing Mechanisms (MAR -30%) for variable Y7.*

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | x1 | .015 | .036 | .173 | 1 | .677 | 1.015 |
| | x2 | .012 | .031 | .161 | 1 | .688 | 1.012 |
| | x3 | -.019 | .038 | .260 | 1 | .610 | .981 |
| | x4 | -.017 | .009 | 3.631 | 1 | .057 | .983 |
| | x5 | .049 | .039 | 1.588 | 1 | .208 | 1.050 |
| | x6 | .007 | .006 | 1.324 | 1 | .250 | 1.007 |
| | x7 | .149 | .261 | .324 | 1 | .569 | 1.160 |
| | age | .036 | .005 | 50.271 | 1 | <.001 | 1.036 |
| | p1 | .104 | .039 | 6.899 | 1 | .009 | 1.109 |
| | p2 | -.059 | .037 | 2.493 | 1 | .114 | .943 |
| | p3 | -.022 | .036 | .383 | 1 | .536 | .978 |
| | p4 | .054 | .034 | 2.619 | 1 | .106 | 1.056 |
| | p5 | -.029 | .034 | .753 | 1 | .386 | .971 |
| | p6 | .031 | .040 | .612 | 1 | .434 | 1.031 |
| | p7 | -.027 | .033 | .661 | 1 | .416 | .973 |
| | p8 | -.027 | .032 | .715 | 1 | .398 | .973 |
| | p9 | .064 | .034 | 3.466 | 1 | .063 | 1.066 |
| | Constant | -3.665 | .663 | 30.527 | 1 | <.001 | .026 |

a. Variable(s) entered on step 1: x1, x2, x3, x4, x5, x6, x7, age, p1, p2, p3, p4, p5, p6, p7, p8, p9.