A Comparison of Imputation Strategies for Ordinal Missing Data on Likert Scale Variables

Wei Wu   Fan Jia

University of Kansas

Craig Enders

Arizona State University

Abstract

This article compares a variety of imputation strategies for ordinal missing data on Likert scale variables (number of categories = 2, 3, 5, or 7) in recovering reliability coefficients, mean scale scores, and regression coefficients of predicting one scale score from another. The examined strategies include imputing using normal data models with naïve rounding/without rounding, using latent variable models, and using categorical data models such as discriminant analysis and binary logistic regression (for dichotomous data only), multinomial and proportional odds logistic regression (for polytomous data only). The result suggests that both the normal model approach without rounding and the latent variable model approach perform well for either dichotomous or polytomous data regardless of sample size, missing data proportion, and asymmetry of item distributions. The discriminant analysis approach also performs well for dichotomous data. Naïvely rounding normal imputations or using logistic regression models to impute ordinal data are not recommended as they can potentially lead to substantial bias in all or some of the parameters.

A Comparison of Imputation Strategies for Ordinal Missing Data on Likert Scale Variables

Ordinal categorical data such as those measured using Likert scales (e.g., 3 or 5 point Likert scales) are very common in social and behavioral sciences. When there are missing data on the ordinal variables due to nonresponse or attrition, the missing data need to be handled appropriately so that accurate parameter estimates and statistical inference on population parameters can be obtained. One of the popular ways to handle missing data is to impute them (i.e., replace the missing data by some plausible values). However, most of the missing data literature in the social and behavioral sciences has focused on imputation methods for normally distributed data. Methods for categorical variables have not been systematically investigated. In particular, we focus on imputation because recent software advances provide researchers with more than one option for handling discrete response scales.

The appropriate methods to handle missing data are contingent on the cause of missingness (missing data mechanism). Rubin (1976) outlined three missing data mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR indicates a scenario where the probability of missing a score is not related to any observed variables or the missing value. MAR indicates a scenario where the probability of missing a score is dependent on an observed variable but not the missing value itself. Once the observed variable is controlled for, missingness has no association with the unobserved scores on the incomplete variable. Missing not at random (MNAR) indicates a scenario where the probability of missing a score is dependent on the missing value itself even with observed variables controlled.

When the missing data mechanism is MCAR or MAR (referred to as ignorable missingness), multiple imputation (MI) is an appropriate method to impute the missing data

(Rubin, 1987; Little, & Rubin, 2002; Sinharay, Stern, & Russell, 2001). MI generates multiple

copies of independent imputed datasets. To account for uncertainty in estimation due to

missingness, missing data are predicted from observed data based on a different set of parameter

values in each imputed dataset (Schafer & Olsen, 1998; Rubin, 1987). After imputation, the

target analysis (e.g. a regression analysis) is applied to each imputed dataset. The point and

standard error estimates from the analyses are then pooled to generate the final results (Rubin,

1987).

Three strategies are generally used to impute the missing ordinal data (described in more

details below): treat them as normally distributed and apply well-established normal data

imputation models (e.g., linear regression models and saturated models[1]), generate imputations

using a true categorical data model (e.g., logistic imputation models), or generate imputations

based on a latent variable model. Given the existence of multiple strategies, it is beneficial for

researchers to know which strategy or strategies are appropriate given how the ordinal data are to

be analyzed following imputation. The current study aims to address this question in a

commonly encountered context where the ordinal variables are items in Likert scales and will be

aggregated (sum or average) to scale scores for further analysis. In this context, MI has

advantages over other popular missing data techniques such as full information maximum

likelihood (FIML). MI can impute missing data at the item level but FIML relies on the analysis

model in which the item scores are not directly included (Gottschall, West, & Enders, 2012). As

a result, the observed information in the dataset is fully utilized by MI but not by FIML. In

addition, if the missingness on one item is determined by another, MI would minimize the bias

---

[1] In saturated models, all non-redundant elements in the mean vector and covariance matrix are
freely estimated.

due to the missingness by borrowing information from the item that predicts the missing data. Note that in practice, researchers often deal with missing item scores within a scale by simply averaging over the remaining items. This is not recommended because it will cause biased parameter estimates for analyses on scale scores and biased reliability measures (Enders, 2003; Enders, 2010; p. 51).

The rest of the article is organized as follows. We start with an introduction of the commonly used algorithms in MI followed by the strategies to impute missing ordinal data. We then review the past research on the available strategies. After the literature review, a simulation study is conducted to evaluate the strategies in the target context. An empirical example is also provided to illustrate the strategies. This article is concluded by discussing the implications and limitations of the simulation study as well as potential directions for future research.

Imputation Algorithms

Multiple algorithms have been developed to generate multiply-imputed data. The most popular one is the Markov chain Monte Carlo (MCMC; Schafer, 1997) algorithm. Briefly speaking, MCMC facilitates MI by drawing values of imputation parameters (e.g., mean vector and covariance matrix if the imputation model is saturated or intercept and regression coefficients if the imputation model is a logistic regression model) from their posterior distribution (s) through an iterative procedure. Note that we use the MCMC algorithm as a general term to refer to a family of algorithms commonly used in MI such as the data augmentation algorithm and Gibbs sampler. In other words, these algorithms are just special cases of a MCMC algorithm.

MCMC allows missing data on multiple variables to be imputed from either a joint distribution or a series of univariate distributions. The joint distribution approach imputes

missing data on multiple variables simultaneously, which is particularly suitable for normal data. In contrast, the univariate distribution approach imputes missing data on a variable-by-variable basis. The univariate approach is also called chained equation or sequential regression equation imputation because the variables are imputed separately and sequentially at each iteration. It is also referred to as fully conditional imputation because prediction of missing data on one variable is conditional on the current values of all of the other variables at a specific iteration (Raghunathan, Lepkowski, van Hoewyk, & Solenberger, 2001; van Buuren, 2007; 2012). One major advantage of the univariate approach is that the imputation model can be easily tailored to the nature of each variable. Thus it is very flexible in accommodating mixed types of missing data (e.g., continuous, categorical, and semi-continuous). For instance, one variable can be imputed using a categorical data model while another can be imputed using a normal data model. In the following, we illustrate the MCMC algorithm with joint and chained equation imputation in more details as it will help readers understand how imputation is actually done for the strategies considered in the article.

Joint distribution approach. For joint imputation, MCMC starts with initial values on the imputation parameters and then iterate between two steps. Schafer (1997) describes the two steps as posterior step (P step) and imputation step (I step). Let $\mathbf{y}_{obs}$ and $\mathbf{y}_{miss}$ represent observed and missing data on a vector of variables respectively, the two steps at a specific iteration, say tth iteration, can be described as follows.

P step: Draw a random set of imputation parameter values ($\boldsymbol{\theta}^{(t)}$) from the posterior distribution of the imputation parameters ($\boldsymbol{\theta}$) which typically contains the elements in a mean vector and covariance matrix, e.g., when variables are normally distributed.

$$\boldsymbol{\theta}^{(t)} \sim p(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{y}_{miss}^{(t-1)}) \tag{1}$$

Here $\mathbf{y}_{miss}^{(t-1)}$ represents predicted missing data from the (t-1)th iteration.

I step: Predict missing data from the observed data conditional on $\boldsymbol{\theta}^{(t)}$. This is equivalent to drawing random values from the predictive probability distribution of $\mathbf{y}_{miss}$. This distribution is typically multivariate normal.

$$\mathbf{y}_{miss}^{(t)} \sim p(\mathbf{y}_{miss} \mid \mathbf{y}_{obs}, \boldsymbol{\theta}^{(t)}) \tag{2}$$

The predicted missing data are then carried to the P step of the next iteration to update the posterior distribution of imputation parameters**.**

Univariate distribution approach. In the univariate distribution approach, the two steps described in the joint distribution approach apply to every variable to be imputed. Suppose there are p variables to be imputed in the order of $y_1, ..., y_p$, each MCMC iteration would contain $2p$ steps. Note that the order is decided arbitrarily in this illustration example. In practice, it can be determined based on the amount of missing data on each variable. The two steps for y_1 at iteration t can be described as follows.

P Step: Draw a random set of parameter values ($\boldsymbol{\theta}_1^{(t)}$) from the posterior distribution of the imputation parameters conditional on the values of the other variables (imputed or complete) from the previous iteration t-1.

$$\boldsymbol{\theta}_1^{(t)} \sim p(\boldsymbol{\theta}_1 \mid y_{1(obs)}, y_2^{(t-1)}, ..., y_p^{(t-1)}) \tag{3}$$

I Step: Impute missing data on y_1 conditional on all of the other variables and $\boldsymbol{\theta}_1^{(t)}$. Again, this is equivalent to drawing random values from the predictive probability distribution of $y_{1(\text{miss})}^{(t)}$.

$$y_{1(miss)}^{(t)} \sim p(y_{1(miss)} \mid y_{1(obs)}, y_2^{(t-1)}, ..., y_p^{(t-1)}, \boldsymbol{\theta}_1^{(t)}) \tag{4}$$

The imputed values $y_{1(\text{miss})}^{(t)}$ are then used to update the posterior distribution of imputation parameters for the next variable y_2 at iteration t ($\boldsymbol{\theta}_2^{(t)}$), as shown below.

$$\boldsymbol{\theta}_2^{(t)} \sim p(\boldsymbol{\theta}_2 \mid y_{2(obs)}, y_1^{(t)}, y_3^{(t-1)}, ..., y_p^{(t-1)}) \tag{5}$$

The missing data on y_2 are then predicted from a random draw from the predictive probability distribution of $y_{2(\text{miss})}^{(t)}$ conditional on $\boldsymbol{\theta}_2^{(t)}$ and the current values of the other variables.

$$y_{2(miss)}^{(t)} \sim p(y_{2(miss)} \mid y_{2(obs)}, y_1^{(t)}, y_3^{(t-1)}, ..., y_p^{(t-1)}, \boldsymbol{\theta}_2^{(t)}) \tag{6}$$

The two steps are repeated for the rest of the variables in the iteration. As mentioned above, each variable can be imputed by a different model. When a variable y is normal, $\boldsymbol{\theta}$ may contain linear regression coefficients or elements in mean vector and covariance matrix, when y is dichotomous, $\boldsymbol{\theta}$ may contain logistic regression coefficients, when y is ordinal, $\boldsymbol{\theta}$ may contain ordinal logistic regression coefficients, etc. Correspondingly, the predictive probability distribution of y_{miss} is tailored to the scale; normal data are drawn from a normal distribution,

binary data are drawn from a binomial distribution, and polytomous data are drawn from a

cumulative logistic distribution, etc.

The MCMC algorithm with joint or chained equation imputation typically cycles through

many iterations. The algorithm converges when the posterior distribution(s) of the imputation

parameters stabilizes. After convergence, the imputed data sets can be saved. However, they will

be saved at specified intervals but not every iteration to avoid the autocorrelations among the

imputation parameters at adjacent iterations.

Most of the strategies considered in the study are implemented using an MCMC

algorithm. Another algorithm which is relevant to the current study is the so-called EMB

algorithm. This algorithm combines expectation-maximization (EM) algorithm with

bootstrapping to impute missing data (Honaker, King, & Blackwell, 2011). Specially, EMB

applies bootstrapping to the sample data and then uses the EM algorithm to obtain the maximum

likelihood estimate of the mean vector and covariance matrix of the variables in an imputation

model for each bootstrapped sample. The EM estimate from each bootstrapped sample is then

treated as a random draw of the imputation parameters and used to impute the missing data. For

more information on the EM algorithm and bootstrapping, researchers can refer to Dempster,

Laird, and Rubin (1977) and Efron (1979) respectively.

EMB assumes multivariate normality which is suitable for joint imputation of normal

data. In this regard, it is theoretically equivalent to the MCMC algorithm and runs much faster

than MCMC for two reasons: 1) Convergence of the EM algorithm is more straightforward than

convergence of the MCMC algorithm. It does not require the usual convergence diagnostics in

MCMC; 2) Unlike MCMC, no autocorrelation exits among the imputation parameter values

from different bootstrap samples. As a result, every bootstrap sample is used to generate the

imputed datasets.

Imputation Strategies

Having illustrated the predominant imputation algorithms, we now turn to the strategies

that can be used to predict missing ordinal data. As mentioned above, we focus on three types of

strategies. The first is to treat the ordinal missing data as normally distributed and use the models

that assume normality to impute the missing data. Because it is easy to define the joint

distribution of variables with normal distributions, missing data are usually imputed jointly with

either an MCMC or EMB algorithm although the imputation can be also done sequentially with

an MCMC algorithm. The normal model strategy is available in almost all of the MI packages

and has been used most widely by researchers. The resulting imputed values are continuous and

may be directly used in analysis or rounded to discrete metrics.

The second is to impute ordinal missing data using models designed for categorical data

such as linear discriminant analysis, and logistic regression models. For the categorical data

models, imputation is usually done using the MCMC algorithm with chained equation

imputation. The imputed values from these approaches preserve the original metric of the ordinal

variables. In what follows, we briefly describe the major characteristics of the categorical data

models.

We only consider the discriminant analysis model for dichotomous data although it can

be also used for polytomous data. Discriminant analysis assumes that the predictors of group

membership follow a multivariate normal distribution with means that vary across categories but

a covariance matrix that is constant over categories (often pooled over categories, see Shafer,

1997). For each imputation, group means and pooled covariance matrix are drawn from their

posterior distributions and used to derive a linear combination of the predictors (referred to as

discriminant function) which maximizes the group difference. The discriminant function scores

are then used along with prior probabilities of the categories (usually the sample proportions of

each category) to compute the posterior probabilities a participant belonging to each category

(van Buuren & Groothuis-Oudshoorn, 2011). The missing values are imputed based on the

posterior probabilities typically in a way such as a participant is assigned to the category with the

highest posterior probability. This imputation model is currently available in MI packages such

as SAS PROC MI, MICE in R, and IVEware.

In terms of logistic regression models, there are different types of logistic regression

models depending on the number of categories (K) in the ordinal data. For dichotomous data (i.e.,

$K = 2$), there are binary logistic regression models which link predictors linearly with the log

odds (logit) of the probability of falling into a category versus the other. The parameters in a

binary logistic regression model include the intercept and regression coefficients associated with

each of the predictors. For polytomous data ($K \geq 3$), multinomial and proportional odds logistic

regression models may be used. The former is designed for nominal data but can be also used for

ordinal data. Suppose there are K categories, a multinomial regression model contains K-1

equations linking predictors linearly with the log odds of falling into each category versus

category K. The intercept and regression coefficients associated with the predictors are allowed

to be different across the equations. A proportional odds regression model also contains K-1

equations. These equations link predictors linearly with the log-odds of falling into or below

each category versus falling above it (cumulative logit). However, unlike the multinomial model,

the effects of the predictors (i.e., regression coefficients) are assumed to be constant across

equations (referred to as proportional odds assumption). That is the influence of the predictors is

assumed to be the same across all adjacent categories of the ordinal variable (e.g., a one-unit increase in a predictor has the same influence on the probability of belonging to categories 1 and 2, categories 2 and 3, etc.).

In theory, the multinomial and proportional odds logistic regression models have their own pros and cons. The proportional odds model takes the order of response categories into account while the multinomial model does not. In addition, it is more parsimonious (contains fewer parameters) than the multinomial model. On the other hand, the multinomial model is not restricted to the proportional odds assumption. Thus when this assumption is violated, it may be preferred to the proportional odds model. More information on the logistic regression models can be found in Agresti (2007). The binary and proportional odds logistic regression models for imputation are available in SAS PROC MI, MICE in R, and IVEware. The multinomial logistic regression imputation model is available in MICE in R, and IVEware.

The third type of strategy is to assume that there is a continuous latent variable underlying each observed ordinal variable (Asparouhov, & Muthén, 2010a; McKelvey & Zavoina, 1975). The underlying latent variables are typically assumed to follow a multivariate normal distribution. The imputed values are first imputed at the latent variable level using a normal data model and then discretized based on estimated thresholds (see more details below). This latent variable model is a formulation of a cumulative/ordinal probit model and is currently available in Mplus (Asparouhov, & Muthén, 2010a).

Let $\mathbf{y}^*$ represent the vector of latent variables underlying a vector of categorical variables. The following equation shows the default imputation model (a saturated model) for $\mathbf{y}^*$ in Mplus (Asparouhov, & Muthén, 2010a; 2010b).

$$\mathbf{y}^* \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Theta}) \tag{7}$$

MVN stands for multivariate normal distribution. $\boldsymbol{\mu}$ is a vector of latent means which are fixed to 0 for identification purpose because each latent mean is redundant with one of the threshold parameters for the corresponding ordinal variable. Fixing the means to be 0 allows all of the thresholds to be freely estimated[2]. $\boldsymbol{\Theta}$ is the covariance matrix of the latent variables. The diagonal elements of $\boldsymbol{\Theta}$ are fixed at 1 to set the scale and the off diagonal elements are freely estimated. Specially, for complete cases, the $\mathbf{y}^*$ values follow a truncated normal distribution, such that they are bounded by the appropriate threshold parameters. For incomplete cases, the $\mathbf{y}^*$ values are unbounded since we are not able to condition on the discrete scores.

Once the imputed values are obtained for $\mathbf{y}^*$, they are discretized using the thresholds as follows (Asparouhov, & Muthén, 2010a; 2010b).

$$y = \begin{cases} 1 \text{ if } -\infty < y^* \leq \tau_{1j} \\ k \text{ if } \tau_{(k-1)} < y^* \leq \tau_k \text{ and } 2 \leq k \leq K-1 \\ K \text{ if } \tau_{(K-1)} < y^* < +\infty \end{cases} . \tag{8}$$

y represents an ordinal variable. k indicates category, $k = 1, ..., K$. y^* is the latent variable underlying y. $\tau_1, ... \tau_{(K-1)}$ represent the K -1 thresholds, respectively. To illustrate, suppose y has 3 categories. Two threshold parameters (τ_1 and τ_2) are estimated which separate the normal distribution into three segments as shown below. $y = 1$ if y^* is less than τ_1, $y = 2$ if y^* exceeds τ_1 but not τ_2, and $y = 3$ if y^* exceeds τ_2.

---

[2] An alternative way to identify the model is to fix one of the threshold parameters to be 0 and freely estimate the latent means.

$$y = \begin{cases} 1 \text{ if } -\infty < y^* \leq \tau_1 \\ 2 \text{ if } \tau_1 < y^* \leq \tau_2 \\ 3 \text{ if } \tau_2 < y^* \leq +\infty \end{cases} \tag{9}$$

Because missing data are imputed at the latent variable level and the latent variables are assumed to follow a multivariate normal distribution, it is straightforward to establish the joint distribution even when there is a mixture of ordinal variables and continuous variables[3] (Asparouhov, & Muthén, 2010a). Thus the missing data are usually imputed jointly with this strategy. The imputation process involves a few additional steps that are not present in the joint model described in Equations 1 and 2, e.g., drawing threshold parameters from their posterior distributions and categorize latent normal imputations using the threshold parameters. Interested researchers can refer to Asparouhov and Muthén (2010b) and Cowles (1996) for more details on the imputation process.

Literature Review

The past research on imputation strategies for ordinal data has been focused on three issues: 1) the robustness of the normal model approach without rounding to discontinuity and nonnormality due to ordinal data, 2) the rounding strategies for the resulting continuous imputed data from the normal model approach, and 3) the comparison of the normal model approach to categorical model approaches.

With regard to the first issue, many of the previous studies suggested that with a moderate proportion of missing data ($\leq 30\%$), the normal model approach without rounding was able to accurately reproduce the mean of polytomous data (Ake, 2005; Horton, Lipsitz, & Parzen,

---

[3] In this case, the joint distribution of the continuous observed variables and latent variables underlying the ordinal variables is used for imputation. This joint distribution is often assumed to be multivariate normal.

2003), a linear regression coefficient of predicting a continuous outcome variable from a binary variable with missing data (Allison, 2005), Cronbach's alpha for likert scales and Pearson correlations among ordinal items with 3, 5 or 7 categories for MCAR and MAR data (see Enders, 2003; Leite & Beretvas, 2004). However, with a larger proportion of missing data, the normal model approach without rounding can potentially result in biased parameter estimates. For example, Leite and Beretvas (2004) found that with 50% missing data, the normal model approach led to underestimated correlations among the ordinal items.

Many rounding strategies have been proposed to discretize the continuous imputed data from the normal model approach. The simplest strategy is to round the imputed values to the nearest integer (naïve rounding). However, naive rounding is generally not recommended for estimation of proportions and regression coefficients when the ordinal variables are predictors (see Ake, 2005; Allison, 2005; Horton et al., 2003; Lee, Galati, Simpson, & Carlin, 2012). Finch and Margraf (2008) also reported substantial bias in parameter estimates due to naïve rounding in item response theory analysis. However, naïve rounding seemed acceptable for estimation of parameters such as odds ratio in logistic regression (Bernaards, Belin, & Schafer, 2007). Alternative rounding strategies (e.g., adaptive rounding, two-stage calibration rounding, and indicator based rounding strategies) have been also developed in the past decades (see Demirtas, 2010; Lee et al., 2012; Yucel, He, & Zaslavsky, 2008). However, these rounding strategies are generally cumbersome to implement. In addition, none of them have emerged as superior to the other under all analytical scenarios examined in the past research. For example, Lee et al. (2012) found that all of the rounding strategies resulted in biased estimates of the proportions of ordinal variables with 4 categories. Based on the review, the current study will only consider naïve rounding for its simplicity.

Limited research has compared the normal model approach to a subset of the categorical model approaches mentioned above. Using simulated data of $N = 500$, Allison (2005) found that the binary logistic regression and discriminant analysis approaches were comparable to the normal model approach without rounding to impute MCAR data on a binary variable, and were superior to the normal approach without rounding to impute MAR data on the binary variable in estimating mean (proportion) of the binary variable especially when the distribution of the binary variable is severely asymmetric (proportion $\leq .2$). However, the normal model approach without rounding performed as well as the logistic regression and discriminant analysis in correctly recovering the linear regression coefficient when the binary variable with missing data was used to predict a continuous variable. Finch (2010) compared the performance of the normal model approach with naïve rounding and multinomial logistic regression approach to impute 5 ordinal items with 5 categories. He examined the effect of imputation strategy on the regression coefficients in an ordinal regression analysis where one of the ordinal variables is predicted by all of the rest. He varied sample size ($N = 200, 500,$ or 1000) and missing data mechanism (MCAR or MAR) and set the percent of missing data on the 5 variables to 25%. He found that the multinomial logistic regression approach yielded more bias in the point and standard error estimates of the logistic regression coefficients than the normal model approach with naïve rounding. However, the biases tended to decrease as N increased.

Purpose of Study

Our literature review suggests that the past research on the imputation strategies has been conducted in a variety of analytical contexts (e.g., linear and logistic regression models, coefficient alpha, and estimation of variable means). However, little research has been done to evaluate the imputation strategies in a commonly encountered scenario were the ordinal variables

are to be aggregated to scale scores for further analysis. In addition, some imputation strategies

have only become available or accessible to researchers in the recent years. They have not been

included in the previous literature. For example, the past literature has much less compared the

proportional odds logistic regression[4] and the latent variable approach.

This article is thus designed to compare a large set of currently available imputation

strategies for Likert scale variables using Monte Carlo simulation. We restrict our attention to

scale scores in the article, as this is an important starting point for understanding the behavior of

categorical imputation models and is overlooked in the previous research. Specifically, we

examine the extent to which the imputation strategies will influence the estimation of reliability

coefficients (Cronbach's alpha), mean scale scores, as well as regression coefficients of

predicting one scale score from another. The examined strategies include the normal model

approach without rounding and with naïve rounding, the latent variable approach, the binary

logistic regression and the linear discriminant analysis (only for dichotomous data), and the

multinominal and proportional odds logistic regression approaches (only for polytomous data).

<div align="center">

Simulation Study
</div>

Data Generation

In the simulation study, we generated 14 variables in total. 12 of them were ordinal (A_1 to

A_6, and B_1 to B_6) and 2 of them were continuous (*aux*1 and *aux*2). The 12 ordinal variables

belong to two scales A and B, each containing six items. For the purpose of generating data, we

assumed that there was a continuous variable underlying each ordinal variable. We first

---

[4] Finch (2010) examined a stochastic regression imputation method in which ordinal missing data
are predicted by a proportional odds logistic regression model (see Sulis & Porcu, 2008). He
found that this approach provided comparable performance to the normal model approach with
naïve rounding. However, the stochastic regression imputation is not directly comparable to the
imputation done through an MCMC algorithm.

generated 12 standard normal variables using a two factor model with each factor indicated by

six variables. The correlation between the two factors was set to .3 (a medium effect). The

loadings of the variables on the corresponding factor were all set to .7. On the continuous metric,

this model would yield within-scale item correlations of .49, and between-scale item correlations

of .15. This model would also yield Conbach's alphas equal to or above .7 (a commonly used

threshold for acceptable internal reliability) for each scale on the discrete metric (Nunnally,

1978). The ordinal variables were then created by discretizing the continuous variables using

thresholds. With K categories, K-1 thresholds need to be specified (more details on the thresholds

used to generate the ordinal variables follow). After generating the ordinal variables, two scale

scores (X_A and X_B) are computed by summing their corresponding item scores.

Missing data generation. We imposed missing data on half of the ordinal items in each

scale (A_1 - A_3, and B_1 - B_3). The two continuous variables (*aux*1 and *aux*2) are used to generate

missing data. *aux*1 and *aux*2 had a correlation of .5 with each other and with each of the two

scale scores. The missingness on $A_1 - A_3$ is determined by aux1. We rank ordered aux_1 from the

smallest to the largest and computed the probability of being missing based on the rank order.

For example, for each case, the probability of having missing data on A_1 is computed as $1 -$ the

rank order of the value on aux_1/N. This probability is then compared to a random number drawn

from a uniform distribution. If the probability is higher than the random number, the observation

is treated as missing. This procedure is continued until the desired percentage (e.g., 30%) of

missing data is achieved. The missing data on the B items are imposed in a similar fashion

except that the missingness is determined by the rank of the sum of aux_1 and aux_2.

The missing data generated in this way are MAR. We only considered MAR data in the

simulation considering 1) pure MCAR is rare in practice and 2) the method that works for MAR

data usually also works for MCAR data. To account for the MAR missingness, *aux*1 and *aux*2

are included in the imputation models as auxiliary variables.

Design Factors

The following factors are manipulated in the simulation study:

1) Number of categories in the ordinal items ($K = 2, 3, 5,$ or 7).

2) Sample size ($N = 125$ or 500). These Ns fall in the range of the typical sample sizes in

social and behavioral science research. For ease of presentation, we arbitrarily call $N = 125$ a

small sample size and $N = 500$ a large sample size.

3) Proportion of missing data on each of the six incomplete items (30% or 50%). These

missing data proportions are moderate or large proportions examined in the previous literature.

We selected relatively large proportions to reveal the difference among the imputation strategies.

These missing data rates are not uncommon in some studies such as longitudinal studies

(Gustavson, von Soest, Karevold, and Røysamb , 2012).

4) Asymmetry of thresholds used to discretize continuous distributions (symmetric,

moderately asymmetric, or severely asymmetric). Because the distributions of the ordinal items

are not necessarily symmetric in practice, we manipulated the asymmetry of thresholds at the

three levels to introduce different levels of asymmetry in the item distributions. Table 1 shows

the thresholds used to generate the levels of asymmetry for ordinal variables with different Ks.

These thresholds are borrowed from Rhemtulla, Brosseau-Liard, and Savalei (2012). To

visualize the resulting item distributions, the theoretical frequency distributions for the ordinal

data with $N = 100$ are presented in Figure 1.

In sum, there are 48 conditions for each imputation strategy. Each condition was

replicated 1000 times. To remove any possible simulation error in comparison of the strategies,

the same set of simulated data is used across imputation strategies. Since we have maximum

50% missing data on some of the variables, 50 imputed datasets are obtained for each replication

in accordance with White, Royston, and Wood's (2011) suggestion.

[Table 1 is about here]

[Figure 1 is about here]

Parameters to be Evaluated

In the current study, we examine how imputation strategy influences the estimation of

three parameters: Cronbach's alpha coefficient (a measure of internal consistency) of Scale B,

mean scale score of Scale B, and linear regression coefficient of predicting X_B from X_A. Note that

we reported the Cronbach's alpha or mean scale score for Scale B only because the items in the

Scales A and B are generated in the same fashion thus the results for the two scales are

comparable. The mean scale score and regression coefficient are scale-level quantities. The alpha

coefficient is an item-level statistic which is calculated based on averaged Pearson correlation

coefficients across imputations following typical practice.

$$\alpha = (p \times \bar{r}) / [1 + (p-1) \times \bar{r}] \tag{10}$$

where p is the number of items and $\bar{r}$ is the average correlation.

The population values for the three parameters for ordinal data with different numbers of

categories and shapes of item distributions are presented in Table 2. The population values are

computed based on a single large simulated data set ($N = 500,000$) of complete ordinal data

because the categorization process prevents us from analytically determining the parameter

values.

<div align="center">[Table 2 is about here]</div>

Evaluation Criteria

Bias in point estimate. Following Collins, Schafer, and Kam (2001), we used %

standardized bias to quantify the bias in a specific parameter (θ).

$$\% \text{ standardized bias} = 100\% \times \frac{(\sum_{i=1}^{q} \hat{\theta}_i \big/ q) - \theta_0}{s_\theta} \tag{11}$$

where $\hat{\theta}_i$ is the parameter estimate for the ith replication. q is the number of replications. θ_0 is the

population parameter value. s_θ is the standard deviation of the parameter estimates across

replications (i.e., empirical standard error). Collins et al. (2001) suggested that % standardized

bias greater than 40% (i.e., more than 40% of a standard error) represents estimation bias that is

of practical significance.

Mean square error (MSE). For each parameter, *MSE* is the averaged squared

discrepancy between the parameter estimate and the population parameter value.

$$MSE = \sum_{i=1}^{q} (\hat{\theta}_i - \theta_0)^2 / q \tag{12}$$

MSE is a measure of overall accuracy (the extent to which sample estimates deviate from the

population parameter value) which is a combination of both bias and variance of estimation. A

strategy with a smaller *MSE* is more accurate thus is preferred. To facilitate comparison of *MSE*s

among multiple strategies, a ratio of *MSE* was computed for each strategy relative to the normal

model approach without rounding given that it is the strategy that have enjoyed the most

widespread use by researchers. An *MSE* ratio larger than one indicates that a specific strategy is

less accurate than the normal model approach without rounding.

95% CI coverage. We also reported the 95% CI coverage for the regression slope and

the mean scale score. The 95% CI coverage for each parameter is calculated as the proportion of

the 95% CIs that cover the population value across replications. The upper and lower limits of

the CI for the regression slope or mean scale score are computed based on the corresponding

standard error estimate. The upper and lower limits of the CI for Cronbach alpha are computed

as follows (Fan, & Thompson, 2001):

$$CI_{upper} = 1 - [(1 - \hat{\alpha}) \times F_{(\gamma/2), df_1, df_2}]$$

$$CI_{lower} = 1 - [(1 - \hat{\alpha}) \times F_{(1-\gamma/2), df_1, df_2}]$$

where $\hat{\alpha}$ is the sample estimate of the alpha coefficient. γ is the significance level ($\gamma = .05$ for a

95% CI). $F_{(\gamma/2), df_1, df_2}$ and $F_{(1-\gamma/2), df_1, df_2}$ represent percentiles $\gamma/2$ and $1 - \gamma/2$ respectively in an F

distribution with $df_1 = (N - 1)$ and $df_2 = (N - 1)(p - 1)$, where N indicates sample size and p

indicates number of items, respectively.

Data Analysis

A repeated measures ANOVA was conducted to examine the effect of the design factors

on standardized biases for each of the four types of ordinal data and each of the three parameters

(12 ANOVAs in total). We did not run the analyses for MSE or CI coverage because these

outcomes violate the normality assumption of ANOVAs. In addition, since a larger bias tended

to be associated with a larger *MSE* and lower CI coverage (see the result section), the ANOVAs can also give us some sense about how the design factors affected the two outcomes.

Among the design factors, sample size, missing data proportions, and level of asymmetry are between-subject factors. The type of strategy is a within-subject factor because the same set of simulated data was used across imputation strategies. We included all possible main and interaction effects in the ANOVA models and computed an eta square (η^2) for each effect as the effect size measure. For a between-subject effect, η^2 is computed as the ratio of the sum of squares explained by the effect and the total between-subject sum of squares. For a within-subject effect, η^2 is computed as the ratio of the sum of squares explained by the effect by the total within-subject sum of squares. To simplify presentation, we only reported the effects with a $\eta^2 > .05$.

Software

R, Mplus, and SAS were used to perform the tasks in the simulation study. The normal model approach was implemented using Amelia II in R (Honaker, King, & Blackwell, 2011) which uses the EMB algorithm for imputation. The latent variable model approach is implemented in Mplus 7 (Muthén & Muthén, 1998-2012) through the MCMC with the joint imputation. Mplus automatically monitors convergence of the MCMC algorithm using the potential scale reduction convergence criterion (Asparouhov & Muthén, 2010b; Gelman & Rubin, 1992). After convergence, the imputed datasets were saved every 100 iterations. The discriminant analysis and logistic regression approaches were implemented using MICE in R (van Buuren & Groothuis-Oudshoorn, 2011) through the MCMC algorithm with chained equations imputation. 5 burn-in iterations (the iterations occur before convergence) were specified following the general recommendation of the MICE developers (van Buuren &

Groothuis-Oudshoorn, 2011). To ensure that MCMC converged appropriately with 5 burn-in iterations, we checked the trace plots of the parameter estimates across MCMC iterations. We also compared the result from 5 burn-in iterations to that from 20 burn-in iterations for one condition (50% missing data, $K = 7$, and the item distributions were severely asymmetric). We did not see any substantial difference between the two sets of result. R was used to compute the alpha coefficient, the mean scale score, and the regression coefficient of predicting X_B from X_A. PROC GLM in SAS 9.2 was used to perform the repeated measures ANOVAs.

Results

The substantial effects ($\eta^2 > .05$) from the repeated measures ANOVAs are summarized in Table 3. The standardized bias, MSE, and CI coverage for each method under each combination of N and distribution asymmetry were summarized in Tables 4 - 7 for $K = 2, 3, 5$, and 7, respectively. As shown in Table 3, missing data proportion did not have a substantial influence on standardized bias for any of the parameters, controlling for the other factors. The MSE and CI coverage results also showed similar patterns between the two missing data proportions (30% and 50%). Thus, for simplicity, we present the average outcomes across the two proportions in Tables 4 - 7. In what follows, we describe the results for the dichotomous data first followed by those for the polytomous data.

[Table 3 is about here]

Dichotomous Data

For dichotomous data, the standardized bias was substantially influenced by imputation strategy, N, and distribution asymmetry for any of the parameters (see Table 3). Specifically, imputation strategy interacted with distribution asymmetry (η^2 ranged from .09 to .29) or N (η^2 ranged from .08 to .12) to affect standardize bias, indicating that distribution asymmetry or N

effect differed across imputation strategies. Figure 2 displays the plot of % standardized biases

against imputation strategies for each N and level of distribution asymmetry for regression slope.

The same patterns were seen in the two interaction effects for the other parameters. As shown in

Figure 2, the logistic regression approach and naïve rounding were more sensitive to N than the

other approaches. N influenced the two approaches differently. Increase in N improved the

performance of the logistic regression approach while worsened the performance of naïve

rounding. In terms of the effect of distribution asymmetry, all of the approaches were influenced

by distribution asymmetry to some extent; however, the logistic regression approach was most

sensitive to distribution asymmetry. As shown in Figure 2, the bias due to the logistic regression

approach increased dramatically as distribution asymmetry increased.

[Figure 2 is about here]

Taking a further look at the magnitude of the standardized biases, MSEs, and CI coverage

probabilities under each combination of N and distribution asymmetry, it is obvious that the

normal model approach without rounding, the latent variable approach, and the discriminant

analysis approach not only were least sensitive to the design factors, but also yielded more

accurate estimates (|standardized bias| < 40%), lower MSEs, and better CI coverage rates for the

parameters under examination than naïve rounding and the logistic regression approach (see

Table 4).

In contrast, the logistic regression approach yielded severely underestimated alpha

coefficient (highest standardized bias = -124%, highest MSE ratio = 2.83, lowest CI coverage =

55%), overestimated mean scale score (highest standardized bias = 206%, highest MSE ratio =

4.83, lowest CI coverage = 58%) with asymmetric distributions, and underestimated regression

slope with $N = 125$ and severe asymmetry (standardized bias = -61%). Naïve rounding, although

it produced acceptable estimates for regression slope, yielded substantially underestimated alpha

coefficient (highest standardized bias = -102%, highest *MSE* ratio = 2.19, lowest CI coverage =

74%) regardless of *N* and item distributions, and overestimated mean scale score (highest

standardized bias = 66%, highest *MSE* ratio = 1.37, lowest CI coverage = 89%) with *N* = 500 and

asymmetric distributions.

[Table 4 is about here]

Polytomous Data

For polytomous data, a two-way interaction effect was found between imputation strategy

and distribution asymmetry (η^2 ranged from .09 to .47) for many of the conditions. The two-way

interaction effects for the regression slope under *K* = 3, 5 or 7 were plotted in Figure 3. The same

interaction effects for the other parameters share a similar pattern as those shown in Figure 3. As

shown in Figure 3, only naïve rounding and the two logistic regression approaches (multinomial

and proportional odds) were influenced by distribution asymmetry. In general, larger biases were

associated with asymmetric distributions. However, the interaction effect exhibited different

patterns between *K* = 3 and *K* = 5 or 7. Specifically, with *K* = 3, whether the distributions were

moderately or severely asymmetric made significant difference. The biases associated with

moderate asymmetry were substantially lower than those associated with severe asymmetry.

However, with *K* = 5 or 7, the two levels of asymmetry led to very similar results for the two

logistic regression approaches.

[Figure 3 is about here]

A two-way interaction effect was also found between imputation strategy and *N*

(η^2 ranged from .08 to .12) for many of the conditions (see Table 3). Representative interaction

effects for $K = 3$, 5 or 7 were plotted in Figure 4. Again, naïve rounding and the two logistic regression approaches were more influenced by N than the others. Specifically, the proportional odds approach yielded more bias with a smaller N regardless of the number of categories. However, the N effect for the multinomial approach differed between $K = 3$ and $K = 5$ or 7. With $K = 3$, more bias was associated with a smaller N, while with $K = 5$ or 7, more bias was associated with a larger N. This result is likely due to the fact that the multinomial approach ignores the order information in the response categories. The bias due to this problem became only noticeable with $K > 3$ and was exacerbated as N increased. Naive rounding seemed to be influenced by N only when K is small. Similar to the result for dichotomous data, the bias due to naïve rounding increased as N increased.

[Figure 4 is about here]

Furthermore, a three-way interaction among strategy, N, and distribution asymmetry was found for the regression slope for ordinal data with $K = 5$ categories ($\eta^2 = .09$), and the mean scale score for ordinal data with any number of categories (η^2 ranged from .10 to 14). The three-way interaction effects suggested that N and distribution asymmetry interacted with each other to affect the outcomes in these cases and this interaction effect differed across imputation strategies. Representative three-way interaction effects were plotted in Figure 5. As shown in Figure 5, the interaction effect seemed only to occur with naïve rounding and the two logistic regression approaches. For these approaches, N affected the bias most when the item distributions were severely asymmetric within each K. For the proportional odds logistic regression approach, the largest bias occurred when severe asymmetry was combined with a small N. However, for the

multinomial regression approach and naïve rounding, the largest bias occurred when severe asymmetry was combined with a large *N*.

[Figure 5 is about here]

Taking a further look at the magnitude of the standardized biases, along with the MSEs and CI coverage probabilities, associated with the examined methods for polytomous data (see Tables 5 - 7). It is clear that the latent variable model approach and the normal model approach without rounding outperformed the other approaches regardless of the parameter and number of categories. Between the two approaches the latent variable approach performed slightly better than the normal model approach by producing smaller standardized biases and *MSE*s when the distributions were severely asymmetric.

[Tables 5 to 7 are about here]

Naïve rounding produced acceptable estimates for the regression slope under all conditions with only one exception: *K* = 3, the item distributions were severely asymmetric, and *N* = 500. However, it yielded underestimated alpha coefficients (highest standardized bias = -83%, highest *MSE* ratio = 1.64, lowest CI coverage = 73%), and overestimated mean scale scores (highest standardized bias = 118%, highest *MSE* ratio = 2.15, lowest CI coverage = 79%) when the distributions were severely asymmetric and *N* was large.

The two logistic regression models performed poorly under most of the conditions. They yielded substantial biases in the parameter estimates, large *MSE* ratios, and unacceptable CI coverage probabilities. In general, they led to underestimated alpha coefficients (highest standardized bias = -312%, highest *MSE* ratio = 29.86, lowest CI coverage = 1%), underestimated regression slopes (highest standardized bias = -148%, highest *MSE* ratio = 3.08,

lowest CI coverage = 62%), and overestimated mean scale scores (highest standardized bias = 471%, highest *MSE* ratio = 11.37, lowest CI coverage = 3%).

The result also suggests that the performance of naïve rounding and the two logistic regression approaches tended to be influenced by number of categories (*K*). As *K* increased, the logistic regression approaches became less and less likely to yield acceptable results for the parameters. For instance, with *K* = 3, the multinomial approach was acceptable for all three parameters when *N* = 500 and the item distributions were NOT severely asymmetric. However, with *K* = 5, it yielded acceptable result for only the scale level parameters and the item distributions need to be symmetric. Once *K* increased to 7, it failed under all conditions. Turning to the proportional odds logistic regression model, it was able to produce acceptable result for the scale level parameters when item distributions were NOT severely asymmetric with *K* = 3 categories. With more categories (5 or 7), however, it only performed well with symmetric distributions and required a larger *N* with 7 categories than with 5 categories. For naïve rounding, the direction is opposite. As *K* increased, naïve rounding became more and more likely to yield acceptable results for the parameters especially with a small *N*.

Another interesting observation is that the standardized bias in the alpha coefficient was generally greater than that in the scale-level parameters (mean scale score and regression slope) when the item distributions were not severely asymmetric, implying that the item-level parameters (correlations in this case) may be more impacted by imputation strategy than the scale level parameters. This is not surprising as the scale-level parameters probably absorbed this bias in a way that yields more acceptable parameter estimates.

Empirical Example

An empirical example is provided to illustrate the examined strategies for polytomous data. The data for the example is from a study conducted by Salyers et al. (2014) to examine the effect of burnout of staffs working in community health on quality of care in mental healthcare. The participants are 113 employees at a Community Mental Health Center. Staff burnout and quality of care were measured using a 7-point and 5-point Likert scales respectively. Each scale contains a few subscales. For simplicity, we selected one subscale from each scale for analysis. Specifically, we selected the personal accomplishment subscale (8 items, measures feelings of competence and successful achievement in one's work) from the burnout scale and the client-centered care subscale (11 items, measures the degree to which a clinician work with clients around their identified goals) from the quality of care scale.

In the original dataset, the missing data on the client-centered care items ranged from 2.7% - 27.1% (average = 14.6%) and the missing data on the personal accomplishment items ranged from 1.8% to 4.4% in the original dataset. To approximate the amount of missing data used in the simulation study (the average missing data proportion was 15% for the conditions with 30% missing data on half of the items), we imposed additional 20% missing data on half of the personal accomplishment items in a way such that the missingness was determined by gender (MAR missing). With the additional missing data, the missing data on the personal accomplishment items ranged from 1.8% to 24.8% (average = 13.0%).

Since the data in the two scales have 5 or 7 categories, we applied the strategies for ordinal data with more than 2 categories (i.e., the normal model approach without rounding and with naïve rounding, the latent variable, the multinomial regression, and the proportional odds regression model approaches). The software packages used for the strategies are the same as those used in the simulation study (the R and Mplus codes are included in the Appendices A and

B respectively). 50 imputations were obtained for each strategy. After imputation, a scale score

was computed for each subscale by summing the corresponding item scores. In parallel to the

parameters considered in the original study and the simulation study, we computed the

Cronbach's alpha and mean scale score for each subscale, and the linear regression coefficient of

predicting client-centered care from personal accomplishment at the scale score level. It is

expected that higher sense of personal accomplishment is associated with higher quality of

client-centered care.

The result is reported in Table 8. Consistent with the simulation study, the latent

variable model approach and the normal model approach without rounding produced almost

identical result on all of the parameters. Naïve rounding performed comparably to the normal

model approach without rounding in this case due to small N and relatively large number of

categories in the two scales. The multinominal and proportional odds logistic regression

approaches yielded different estimates on the parameters especially the regression coefficient

than the other approaches.

[Table 8 is about here]

However, there were a few results from the empirical example that were different from

those from the simulation study. One is that the two logistic regression approaches produced

larger regression coefficients than those from the other approaches, while the simulation study

suggests that they tended to underestimate the regression coefficient. This result is probably due

to the fact that most of the items in the example had negatively skewed distributions (the average

skewness of the personal achievement items was -1.04 and the average skewness of the client-

centered care items was -0.91) while only positively skewed distributions were generated in the

simulation study (see Figure 1). This implies that the direction of biases would be influenced by the direction of skewness.

The other is that the imputation strategies did not seem different as much in the estimation of the alpha-coefficient or mean scale score as in the simulation study. A possible explanation is that we reported standardized biases in the simulation study but raw estimates in the empirical example given that the population values and the empirical standard errors were unknown in the empirical example. Small difference in the raw metric might translate to a large difference in the standardized metric if the empirical standard error is small. For example, the standard error of the mean client-centered care scale score from the analysis was .77 for the proportional odds model approach. If this standard error is used to standardize the difference in the mean scale score between the latent variable and proportional odds approaches, then the 1.29 unit difference in the raw metric (see Table 8) will turned out to be 1.68 (i.e., 168%) units difference in the standardized metric, which would be deemed substantial based on the threshold of 40%.

Conclusion and Discussion

In the current article, a large scale simulation study is conducted to compare the performance of a variety of imputation strategies for ordinal missing data in recovering reliability coefficients, mean scale scores, and regression coefficients of predicting one scale score from another. Based on the result from the simulation study, we recommend the latent variable approach, the normal model approach without rounding for either dichotomous or polytomous data, and the discriminant analysis approach for dichotomous data. These approaches were least influenced by the design factors and yielded reliable and consistent result

across all conditions examined in the study. These approaches yielded not only the lowest biases, but also the best MSEs and CI coverage rates in general.

We generally don't recommend naïve rounding for ordinal data for the following reasons: a) It does not offer a better performance than the recommended approaches, although it seems acceptable for the estimation of the regression slope, and b) It could potentially result in substantial bias in alpha coefficients and mean scale scores especially when N is large and item distributions are asymmetric.

We do not recommend the logistic regression approaches (i.e., binary, multinomial, and proportional odds) for either dichotomous or polytomous data either. These approaches, although designed specifically for categorical data, yielded poor estimates, MSEs, and CI coverage probabilities for the parameters considered in the study under many of the conditions. They were also highly sensitive to N, asymmetry of item distributions, and number of categories. This sensitivity is probably related to the fact that the logistic regression models are highly influenced by sparseness of the ordinal data. To reduce the sparseness, a large N is usually required. This N requirement will also increase when the distributions get more asymmetric, the number of categories increases, or the number of variables increases in the imputation model. The reason Allison (2005) found that the binary logistic regression model approach yielded satisfactory results with $N = 500$ is probably that there is only one binary variable in his imputation model. In our study, there are 12 ordinal variables in the imputation model and the variables can have more than 2 categories. In this case, $N = 500$ was apparently insufficient. In real world studies, imputation models could get more complicated than ours. Thus, although the logistic regression model approaches may be theoretically sound, their real word applicability may be highly

compromised by the sample size requirement. This problem is also noted in Schafer (1997; p. 239).

Limitations and Future Directions

Given the scope of the study, we could not cover every analytical context that is of interest to researchers. For instance, it would be interesting to know how the strategies would perform if an ordinal variable with missing data was used to predict a continuous variable in a linear regression analysis or if the ordinal variable was the dependent variable in an ordinal logistic regression analysis. In the latter case, the normal model approach without rounding would not be applicable because the imputed values have to be discrete. As reviewed previously, Finch (2010) found that the multinomial regression model approach was inferior to the normal approach with naïve rounding in the estimation of the ordinal logistic regression coefficients. However, it is unclear how the proportional odds logistic regression and the latent variable approaches would perform in this context. A study that considers these analytical contexts is currently being conducted.

In addition, we only considered parametric models for imputation in the current study. Recent research has shown that ordinal missing data may also be imputed using nonparametric models such as random forest models (Stekhoven & Bühlmann, 2012). The random forest model is a machine learning technique which has many desirable features. First, it can be applied to a variety of categorical data types (including ordinal data) and can be used when there is a mix of categorical and continuous data (Stekhoven & Bühlmann, 2012). Second, it does not require any distributional assumption. Third, it can accommodate nonlinear relationships such as interactions among variables (Doove, van Buuren, & Dusseldorp, 2014; Shah, Bartlett, Carpenter, Nicholas, & Hemingway, 2014). This method has been implemented in R packages MICE and *missForest*

(Doove, van Buuren, & Dusseldorp, 2014; Stekhoven & Bühlmann, 2012). It would be interesting to compare this nonparametric approach to the parametric model strategies.

One might argue that the reason the latent variable approach performed well in the current study is that it matches with the data generation mechanism: the ordinal data are generated from normally distributed latent variables. Thus it would be interesting to know how this approach would perform if the latent distributions are in fact nonnormal or if the ordinal data are generated from truly categorical distributions such as cumulative logistic distributions. We briefly investigated the former case by generating ordinal data from moderately nonnormal latent distributions (skewness = 2.25, kurtosis = 7) for a subset of the conditions. We found that the conclusion remained the same, implying that the shape of the latent variable distributions might not matter as much as the shape of the observed item distributions. We are not clear on how the imputation strategies examined in the article would perform under the latter case especially whether the logistic regression approaches would perform better if they conform to the data generation mechanism. This may be an avenue for future research.

Another limitation is that we computed the alpha coefficient based on Pearson correlations. Although this is a common practice, it is more appropriate to compute the alpha coefficient based on polychoric correlations if the items cannot be assumed to be continuous (Zumbo, Gadermann, & Zeisser, 2007). Future research needs to be conducted to examine whether the conclusion of the simulation study can be generalized to alpha coefficients computed based on polychoric correlations.

Furthermore, the items were simulated to share the same level of distribution asymmetry. In practice, the shape of item distribution is likely to vary across items. As a preliminary investigation, we mixed items with symmetric, moderately asymmetric, or severely asymmetric

distributions in one imputation model for a subset of the conditions examined in the study. The results suggest that the influence of the items with asymmetric distributions seemed to be offset by the items with symmetric distributions to a certain degree. However, even with mixed distributions, the latent variable approach and the normal model approach without rounding still outperformed the other approaches. In addition, as mentioned above, we only considered positively skewed distributions. Thus the result from the simulation study particularly the direction of the biases may not be directly generalized to negatively skewed distributions. As shown in the empirical example, with negatively skewed distributions, the logistic regression approaches led to overestimated instead of underestimated regression coefficients. This issue needs to be further investigated.

Finally, we only examined Ns up to 500. Since the performance of the binary and proportional odds logistic regression seemed improved as N increased, it would be interesting to investigate whether a larger N would enable them to deliver a satisfactory result. As a preliminary investigation, we examined $N = 1000$ for the imputation strategies for ordinal data with 3 categories. We found that $N = 1000$ did improve the performance of the proportional odds logistic regression approach. However, the standardized biases in the parameters from the approach were still above the threshold (40%) when the distributions were severely asymmetric. Another interesting observation is that as N increased to 1000, the normal model approach without rounding started to yield unacceptable standardized biases ($> 40\%$) in the alpha coefficient when the item distributions were severely asymmetric. In contrast, the results on all of the parameters from the latent variable approach still stayed below threshold, implying that the latent variable approach may outperform the normal model approach without rounding with a

large *N*. Based on these preliminary results, a future study is warranted to investigate the large

sample properties of the imputation strategies.

In sum, we compared multiple imputation strategies for ordinal missing data in the

context where the incomplete ordinal variables are to be aggregated to scale scores. This

comparison resulted in the recommendation of the normal model approach without rounding and

the latent variable model approach. In practice, researchers are often faced with datasets

containing missing data on a wide range of variable types. For example, continuous and nominal

missing data may coexist with ordinal missing data. It is important to note that the two

recommended approaches are still applicable in this case if the nominal missing data are imputed

as sets of dummy variables. Allison (2002, p. 40) suggested that missing dummy codes can be

imputed under a normal model and proposed a simple process to recreate the dummy codes

based on the continuous imputed values. Nominal missing data can also be imputed as a single

variable using categorical models (e.g., multinomial logistic regression model) or nonparametric

models (e.g., random forest). To our knowledge, these imputation models cannot be

implemented along with the latent variable approach in the existing software packages. However,

they can be implemented along with the normal model approach through chained equation

imputation (e.g., MICE), with ordinal and continuous missing data imputed under a normal

model. A study is currently being conducted to compare these different strategies to impute

nominal missing data (Lang & Wu, 2015)."

References

Agresti, A. (2007). *An introduction to categorical data analysis*. Hoboken: NJ: Wiley.

Ake, C. F. (2005). Rounding after multiple imputation with non-binary categorical covariates. Paper presented at the annual meeting of the SAS Users Group International, Philadelphia, PA.

Allison, P. D. (2002). *Missing data.* Thousand Oaks, CA: Sage Publications. doi: 10.4135/9780857020994.n4

Allison, P. D. (2005). Imputation of categorical variables with PROC MI. *SUGI 30 Proceedings, 113–30, 1–14*. Retrieved from http://www2.sas.com/proceedings/sugi30/113-30.pdf.

Asparouhov, T., & Muthén, B. (2010a). *Multiple imputation with Mplus*. Retrieved from http://statmodel2.com/download/Imputations7.pdf.

Asparouhov, T., & Muthén, B. (2010b). *Bayesian analysis using Mplus: Technical implementation.* Retrieved from http://www.statmodel.com/download/Bayes3.pdf.

Bernaards, C. A., Belin, T. R., & Schafer, J. L. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine, 26*(6)*, 1368-1382.

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

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing, 6*(2), 101-111.

Demirtas, H. (2010). A distance-based rounding strategy for post-imputation ordinal data. *Journal of Applied Statistics, 37*(3), 489-500.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1-38.

Doove, L., van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis, 72*, 92-104.

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". *The Annals of Statistics,7*(1): 1–26.

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*(3), 322-337.

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

Fan, X., & Thompson, B. (2001). Confidence Intervals for Effect Sizes Confidence Intervals about Score Reliability Coefficients, Please: An EPM Guidelines Editorial. *Educational and Psychological Measurement, 61*(4), 517-531.

Finch, W. H., & Margraf, M. (2008). Imputation of categorical missing data: A Comparison of multivariate normal and multinomial methods. Retrieved from www.mwsug.org/proceedings/2008/stats/MWSUG-2008-S05.pdf

Finch, W. H. (2010). Imputation methods for missing categorical questionnaire data: A comparison of approaches. *Journal of Data Science, 8,* 361-378.

Gelman, A., & Rubin, D. B. (1992). "Inference from Iterative Simulation using Multiple Sequences". *Statistical Science*, 7, p. 457–511.

Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A Comparison of Item-Level and Scale-Level Multiple Imputation for Questionnaire Batteries. *Multivariate Behavioral Research, 47*(1), 1–25.

Gustavson, K., von Soest, T., Karevold, E., & Røysamb, E. (2012). Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health, 12*(1), 918.

Honaker, J., & King, G., & Blackwell, M. (2011). Amelia II: A Program for Missing Data. *Journal of Statistical Software, 45*, 1–47.

Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *The American Statistician, 57*(4), 229–232.

Lang, K. M., & Wu, W. (2015). *What to do with incomplete nominal variables? A comparison of multiple imputation strategies.* Manuscript in preparation.

Lee, K. J., Galati, J. C., Simpson, J. A., & Carlin, J. B. (2012). Comparison of methods for imputing ordinal data using multivariate normal imputation: a case study of nonlinear effects in a large cohort study. *Statistics in Medicine, 31*, 4164–4174.

Leite, W., & Beretvas, S. N. (2004). The performance of multiple imputation for Likert-type items with missing data. *Journal of Modern Applied Statistical Methods, 9*(1), 64–74.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data.* Hoboken (NJ): Wiley-Interscience.

McKelvey, R., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level variables. *Journal of Mathematical Sociology, 4,* 103-120.

Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén. Retrieved from

http://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r6_

web.pdf.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A

multivariate technique for multiply imputing missing values using a sequence of

regression models. *Survey Methodology, 27,* 85-96.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be

treated as continuous? A comparison of robust continuous and categorical SEM

estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354 - 373.

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

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

Salyers, M. P., Fukui, S., Rollins, A. L., Firmin, R., Gearhart, T., Noll, J. P., Williams, S., &

Davis, C. J. (2014). Burnout and self-reported quality of care in community mental health.

Administration and Policy in Mental Health and Mental Health Services Research, 1-9.

doi: 10.1007/s10488-014-0544-6

Schafer, J., & Olsen, M. (1998). Multiple imputation for multivariate missing-data problems: A

data analyst's perspective. *Multivariate Behavioral Research, 33*(4), 545-571.

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

Hall/CRC.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison

of Random Forest and Parametric Imputation Models for Imputing Missing Data Using

MICE: A CALIBER Study. *American journal of epidemiology, 179*(6), 764-774.

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

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112-118.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A caliber study. *American Journal of Epidemiology, 179(6),* 764-774.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16,* 219–242.

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman & Hall.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3).

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

Yucel, R. M., He, Y., & Zaslavsky, A. M. (2008). Using calibration to improve rounding in imputation. *The American Statistician, 62*(2)*,* 125-129.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21-29.

Table 1

Thresholds Used to Generate the Ordinal Data [a]

# of categories	Thresholds as z scores					
	Symmetric					
2	.00					
3	-.83	.83				
5	-1.50	-.50	.50	1.50		
7	-1.79	-1.07	-.36	.36	1.07	1.79
	Moderately asymmetric					
2	.36					
3	-.50	.76				
5	-.70	.39	1.16	2.05		
7	-1.43	-.43	.38	.94	1.44	2.54
	Severely asymmetric					
2	1.04					
3	.58	1.13				
5	.05	.44	.84	1.34		
7	-.25	.13	.47	.81	1.18	1.64

[a] The thresholds are borrowed from Rhemtulla, et al. (2012).

Table 2

Population Values of the Parameters Considered in the Study

	Alpha	Regression Slope	Mean Scale Score
2 categories			
Symmetric	.74	.22	9.00
Moderately asymmetric	.74	.21	8.16
Extremely asymmetric	.69	.18	6.89
3 categories			
Symmetric	.79	.24	12.00
Moderately asymmetric	.80	.24	11.49
Extremely asymmetric	.76	.21	8.46
5 categories			
Symmetric	.83	.25	18.00
Moderately asymmetric	.82	.25	13.50
Extremely asymmetric	.80	.23	12.60
7 categories			
Symmetric	.84	.25	24.00
Moderately asymmetric	.83	.25	19.18
Extremely asymmetric	.82	.24	16.47

Table 3

η^2 s from the Repeated Measures ANOVAs

Effect	2 categories	3 categories	5 categories	7 categories
Alpha				
strategy	.67	.44	.72	.75
strategy× N	.12	.14	.08	.11
strategy×asymmetry	.09	.13		
Regression Slope				
Strategy	.44	.21	.44	.57
strategy× N	.09			.09
strategy×asymmetry	.20	.34	.19	.09
strategy× N ×asymmetry			.09	
Mean Scale Score				
Asymmetry		.19	.21	.23
Strategy	.47	.21	.38	.35
strategy× N	.08	.07	.06	.07
strategy×asymmetry	.29	.47	.31	.33
strategy× N ×asymmetry		.14	.11	.10

Note. The values in the table are percentages. All of the main and interaction effects are

included in the ANOVA models. However, only the effects with a η^2 > .05 are displayed

in this table. N = sample size; mprop = missing data proportion; asymmetry = distribution

asymmetry; strategy = type of imputation strategy;

Table 4

The Simulation Result for Ordinal Data with K = 2 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125				symmetric					
Normal	-23%	1.00	89%	-1%	1.00	90%	1%	1.00	95%
Naïve rounding	-57%	1.32	87%	-8%	0.93	91%	15%	0.89	95%
Logistic regression	-54%	1.22	88%	-6%	0.93	91%	10%	0.69	97%
Discriminant	-12%	1.03	87%	-1%	1.01	90%	1%	0.93	93%
Latent	-22%	0.97	89%	6%	1.01	90%	-9%	0.93	95%
				moderately asymmetric					
Normal	-25%	1.00	88%	2%	1.00	90%	-6%	1.00	95%
Naïve rounding	-57%	1.34	85%	-11%	0.95	91%	27%	0.93	96%
Logistic regression	-77%	1.68	83%	-34%	1.06	89%	72%	1.10	93%
Discriminant	-16%	1.03	87%	3%	0.96	90%	2%	0.91	94%
Latent	-25%	1.00	88%	4%	0.99	90%	-2%	0.89	95%
				severely asymmetric					
Normal	-34%	1.00	80%	8%	1.00	88%	-18%	1.00	94%
Naïve rounding	-42%	1.10	78%	6%	0.93	89%	26%	1.01	94%
Logistic regression	-124%	2.83	55%	-61%	1.29	83%	206%	4.83	58%
Discriminant	-8%	0.91	77%	18%	1.00	88%	0%	0.85	94%
Latent	-33%	1.06	79%	4%	1.04	87%	8%	0.88	94%
N = 500				symmetric					
Normal	-8%	1.00	89%	2%	1.00	92%	2%	1.00	95%
Naïve rounding	-90%	1.95	78%	-17%	0.94	92%	29%	0.96	94%
Logistic regression	-31%	1.10	88%	0%	0.96	92%	4%	0.87	96%
Discriminant	24%	1.18	83%	2%	0.99	92%	-7%	0.94	94%
Latent	-3%	0.98	89%	9%	0.99	91%	-9%	0.95	95%
				moderately asymmetric					
Normal	-23%	1.00	90%	8%	1.00	91%	-12%	1.00	95%
Naïve rounding	-102%	2.19	74%	-30%	1.02	92%	57%	1.15	92%
Logistic regression	-58%	1.36	85%	-18%	0.98	92%	37%	0.98	94%
Discriminant	0%	1.06	87%	4%	0.95	92%	-5%	0.92	94%
Latent	-15%	0.99	89%	6%	0.97	92%	-2%	0.91	96%
				severely asymmetric					
Normal	-40%	1.00	80%	12%	1.00	88%	-33%	1.00	94%
Naïve rounding	-76%	1.58	70%	-24%	0.99	88%	66%	1.31	89%
Logistic regression	-81%	1.62	69%	-32%	1.05	86%	92%	1.57	87%
Discriminant	17%	1.00	75%	9%	0.97	88%	-8%	0.81	95%
Latent	-19%	0.93	81%	7%	0.99	87%	7%	0.83	95%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the *MSE* ratios >1.2, and the CI coverage rates < 90% are underscored.

Table 5
Simulation Result for Ordinal Data with K = 3 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-25%	1.00	91%	3%	1.00	92%	2%	1.00	95%
Naïve rounding	-55%	1.28	89%	-2%	.95	93%	9%	.95	95%
Multinomial	-64%	1.87	82%	41%	1.26	87%	-67%	.99	95%
Proportional odds	-87%	2.01	82%	-8%	.94	92%	7%	.72	97%
Latent	-22%	.96	91%	7%	.99	92%	-6%	.98	95%
				moderately asymmetric					
Normal	-25%	1.00	89%	1%	1.00	91%	1%	1.00	96%
Naïve rounding	-49%	1.22	89%	-6%	.96	92%	17%	.94	95%
Multinomial	-64%	1.78	83%	-3%	1.09	90%	9%	.64	98%
Proportional odds	-82%	1.90	82%	-22%	1.02	91%	37%	.79	96%
Latent	-25%	.99	90%	3%	.98	92%	-3%	.95	95%
				severely asymmetric					
Normal	-36%	1.00	85%	5%	1.00	89%	-12%	1.00	95%
Naïve rounding	-50%	1.21	82%	-15%	1.00	90%	53%	1.16	92%
Multinomial	-168%	6.01	34%	-92%	1.94	75%	350%	9.55	10%
Proportional odds	-155%	4.82	43%	-90%	1.88	75%	299%	6.95	31%
Latent	-34%	1.02	84%	0%	.99	89%	11%	.91	95%
N = 500					symmetric				
Normal	-6%	1.00	92%	5%	1.00	92%	1%	1.00	94%
Naïve rounding	-75%	1.64	85%	-5%	.95	92%	12%	.97	94%
Multinomial	-18%	1.11	91%	18%	1.07	91%	-34%	1.04	92%
Proportional odds	-45%	1.28	89%	-1%	.99	92%	1%	.91	94%
Latent	-5%	1.00	92%	10%	.99	92%	-9%	.99	94%
				moderately asymmetric					
Normal	-13%	1.00	92%	5%	1.00	93%	0%	1.00	94%
Naïve rounding	-72%	1.58	86%	-11%	.97	93%	29%	.99	94%
Multinomial	-29%	1.16	91%	-3%	1.01	93%	2%	.88	94%
Proportional odds	-49%	1.28	90%	-10%	.98	93%	15%	.91	94%
Latent	-13%	1.02	92%	7%	.98	93%	-5%	.96	94%
				severely asymmetric					
Normal	-46%	1.00	83%	12%	1.00	90%	-23%	1.00	93%
Naïve rounding	-83%	1.59	73%	-45%	1.19	87%	118%	2.15	79%
Multinomial	-93%	1.78	69%	-55%	1.28	86%	121%	2.20	76%
Proportional odds	-91%	1.73	70%	-51%	1.23	87%	105%	1.85	81%
Latent	-26%	.91	84%	6%	.97	90%	6%	.88	93%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the
MSE ratios >1.2, and the CI coverage rates < 90% are underscored.

Table 6

The Simulation Result for Ordinal Data with K = 5 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-23%	1.00	91%	2%	1.00	92%	1%	1.00	96%
Naïve rounding	-37%	1.11	90%	-1%	.98	92%	4%	.97	96%
Multinomial	-163%	6.61	45%	-21%	1.06	91%	11%	1.25	92%
Proportional odds	-190%	8.36	38%	-30%	1.02	91%	-17%	0.88	95%
Latent	-26%	1.01	90%	4%	.99	92%	-8%	.99	96%
				moderately asymmetric					
Normal	-31%	1.00	90%	5%	1.00	92%	-5%	1.00	95%
Naïve rounding	-45%	1.15	89%	-4%	.97	92%	16%	.96	95%
Multinomial	-162%	7.09	41%	-40%	1.30	87%	127%	2.35	76%
Proportional odds	-236%	16.95	16%	-117%	2.43	70%	195%	5.80	43%
Latent	-35%	1.04	91%	2%	.97	91%	0%	.96	95%
				severely asymmetric					
Normal	-34%	1.00	87%	3%	1.00	91%	-9%	1.00	96%
Naïve rounding	-43%	1.14	87%	-15%	1.01	91%	42%	1.05	93%
Multinomial	-135%	4.88	51%	-56%	1.52	83%	194%	3.40	57%
Ordinal	-239%	14.35	16%	-128%	2.62	67%	403%	8.71	11%
Latent	-35%	1.04	88%	-1%	.97	91%	6%	.92	95%
N = 500					symmetric				
Normal	-6%	1.00	92%	5%	1.00	92%	2%	1.00	94%
Naïve rounding	-39%	1.17	91%	0%	.98	93%	7%	0.98	95%
Multinomial	-133%	5.60	55%	56%	1.34	88%	-102%	2.04	84%
Proportional odds	-108%	2.68	74%	-9%	.98	93%	5%	0.88	95%
Latent	-15%	1.03	92%	3%	.98	93%	0%	0.99	94%
				moderately asymmetric					
Normal	-33%	1.00	91%	8%	1.00	92%	-8%	1.00	95%
Naïve rounding	-65%	1.37	87%	-12%	.99	93%	33%	1.04	93%
Multinomial	-212%	16.02	16%	-119%	2.87	69%	171%	5.04	56%
Proportional odds	-156%	4.47	53%	-88%	1.76	82%	123%	2.32	76%
Latent	-35%	1.04	90%	0%	.97	93%	8%	.98	94%
				severely asymmetric					
normal	-43%	1.00	87%	10%	1.00	92%	-16%	1.00	95%
Naïve rounding	-66%	1.33	82%	-36%	1.13	91%	89%	1.60	86%
Multinomial	-166%	5.57	42%	-116%	2.70	70%	262%	6.91	39%
Proportional odds	-120%	2.47	65%	-80%	1.65	83%	139%	2.54	72%
Latent	-32%	.96	87%	1%	.97	92%	9%	.93	95%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the
   MSE ratios > 1.2, and the CI coverage rates < 90% are underscored.

Table 7

The Simulation Result for Ordinal Data with K =7 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-25%	1.00	90%	1%	1.00	92%	-1%	1.00	96%
Naïve rounding	-32%	1.05	90%	0%	.99	92%	1%	.98	96%
Multinomial	-168%	6.36	44%	-23%	1.09	89%	-47%	2.25	79%
Proportional odds	-298%	24.64	11%	-60%	1.11	90%	-65%	1.59	85%
Latent	-34%	1.06	90%	2%	.99	91%	-5%	.98	96%
					moderately asymmetric				
Normal	-32%	1.00	90%	3%	1.00	92%	-3%	1.00	96%
Naïve rounding	-39%	1.07	89%	-1%	0.98	92%	6%	.97	96%
Multinomial	-159%	6.13	45%	-32%	1.21	89%	73%	1.51	87%
Proportional odds	-290%	27.74	7%	-132%	2.40	70%	154%	3.65	60%
Latent	-36%	1.04	89%	1%	0.98	92%	0%	.96	96%
					severely asymmetric				
Normal	-33%	1.00	88%	4%	1.00	91%	-8%	1.00	95%
Naïve rounding	-40%	1.10	87%	-11%	1.00	92%	34%	1.01	94%
Multinomial	-139%	4.94	52%	-36%	1.32	86%	147%	2.43	71%
Proportional odds	-312%	28.28	1%	-148%	3.05	62%	471%	11.37	3%
Latent	-39%	1.10	88%	-2%	.97	91%	10%	.94	95%
N = 500					symmetric				
Normal	-6%	1.00	92%	4%	1.00	93%	0%	1.00	94%
Naïve rounding	-24%	1.07	92%	2%	.99	93%	4%	.99	94%
Multinomial	-278%	26.51	7%	-45%	1.37	87%	6%	1.43	90%
Proportional odds	-162%	5.20	53%	-16%	1.02	93%	3%	.90	94%
Latent	-10%	1.02	92%	6%	0.99	93%	-7%	1.00	94%
					moderately asymmetric				
Normal	-34%	1.00	90%	6%	1.00	92%	-5%	1.00	95%
Naïve rounding	-52%	1.17	88%	-3%	.98	93%	13%	0.98	95%
Multinomial	-296%	29.86	2%	-133%	3.08	66%	163%	4.56	60%
Proportional odds	-237%	14.56	17%	-114%	2.30	76%	132%	2.72	76%
Latent	-31%	.99	90%	3%	.97	93%	2%	.97	94%
					severely asymmetric				
Normal	-42%	1.00	88%	9%	1.00	91%	-12%	1.00	95%
Naïve rounding	-59%	1.24	85%	-28%	1.08	91%	73%	1.38	89%
Multinomial	-201%	10.52	20%	-115%	2.85	68%	281%	8.06	21%
Proportional odds	-168%	4.28	47%	-112%	2.32	74%	195%	4.01	53%
Latent	-30%	.94	89%	3%	.97	92%	3%	0.94	95%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the
MSE ratios >1.2, and the CI coverage rates < 90% are underscored.

Table 4
The Simulation Result for Ordinal Items with 2 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-23%	1.00	89%	-1%	1.00	90%	1%	1.00	95%
Naïve rounding	-57%	1.32	87%	-8%	0.93	91%	15%	0.89	95%
Logistic regression	-54%	1.22	88%	-6%	0.93	91%	10%	0.69	97%
Discriminant	-12%	1.03	87%	-1%	1.01	90%	1%	0.93	93%
Latent	-22%	0.97	89%	6%	1.01	90%	-9%	0.93	95%
				moderately asymmetric					
Normal	-25%	1.00	88%	2%	1.00	90%	-6%	1.00	95%
Naïve rounding	-57%	1.34	85%	-11%	0.95	91%	27%	0.93	96%
Logistic regression	-77%	1.68	83%	-34%	1.06	89%	72%	1.10	93%
Discriminant	-16%	1.03	87%	3%	0.96	90%	2%	0.91	94%
Latent	-25%	1.00	88%	4%	0.99	90%	-2%	0.89	95%
				severely asymmetric					
Normal	-34%	1.00	80%	8%	1.00	88%	-18%	1.00	94%
Naïve rounding	-42%	1.10	78%	6%	0.93	89%	26%	1.01	94%
Logistic regression	-124%	2.83	55%	-61%	1.29	83%	206%	4.83	58%
Discriminant	-8%	0.91	77%	18%	1.00	88%	0%	0.85	94%
Latent	-33%	1.06	79%	4%	1.04	87%	8%	0.88	94%
N = 500					symmetric				
Normal	-8%	1.00	89%	2%	1.00	92%	2%	1.00	95%
Naïve rounding	-90%	1.95	78%	-17%	0.94	92%	29%	0.96	94%
Logistic regression	-31%	1.10	88%	0%	0.96	92%	4%	0.87	96%
Discriminant	24%	1.18	83%	2%	0.99	92%	-7%	0.94	94%
Latent	-3%	0.98	89%	9%	0.99	91%	-9%	0.95	95%
				moderately asymmetric					
Normal	-23%	1.00	90%	8%	1.00	91%	-12%	1.00	95%
Naïve rounding	-102%	2.19	74%	-30%	1.02	92%	57%	1.15	92%
Logistic regression	-58%	1.36	85%	-18%	0.98	92%	37%	0.98	94%
Discriminant	0%	1.06	87%	4%	0.95	92%	-5%	0.92	94%
Latent	-15%	0.99	89%	6%	0.97	92%	-2%	0.91	96%
				severely asymmetric					
Normal	-40%	1.00	80%	12%	1.00	88%	-33%	1.00	94%
Naïve rounding	-76%	1.58	70%	-24%	0.99	88%	66%	1.31	89%
Logistic regression	-81%	1.62	69%	-32%	1.05	86%	92%	1.57	87%
Discriminant	17%	1.00	75%	9%	0.97	88%	-8%	0.81	95%
Latent	-19%	0.93	81%	7%	0.99	87%	7%	0.83	95%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the *MSE* ratios >1.2, and the CI coverage rates < 90% are underscored.

Table 5

Simulation Result for Ordinal Data with 3 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-25%	1.00	91%	3%	1.00	92%	2%	1.00	95%
Naïve rounding	-55%	1.28	89%	-2%	.95	93%	9%	.95	95%
Multinomial	-64%	1.87	82%	41%	1.26	87%	-67%	.99	95%
Proportional odds	-87%	2.01	82%	-8%	.94	92%	7%	.72	97%
Latent	-22%	.96	91%	7%	.99	92%	-6%	.98	95%
					moderately asymmetric				
Normal	-25%	1.00	89%	1%	1.00	91%	1%	1.00	96%
Naïve rounding	-49%	1.22	89%	-6%	.96	92%	17%	.94	95%
Multinomial	-64%	1.78	83%	-3%	1.09	90%	9%	.64	98%
Proportional odds	-82%	1.90	82%	-22%	1.02	91%	37%	.79	96%
Latent	-25%	.99	90%	3%	.98	92%	-3%	.95	95%
					severely asymmetric				
Normal	-36%	1.00	85%	5%	1.00	89%	-12%	1.00	95%
Naïve rounding	-50%	1.21	82%	-15%	1.00	90%	53%	1.16	92%
Multinomial	-168%	6.01	34%	-92%	1.94	75%	350%	9.55	10%
Proportional odds	-155%	4.82	43%	-90%	1.88	75%	299%	6.95	31%
Latent	-34%	1.02	84%	0%	.99	89%	11%	.91	95%
N = 500					symmetric				
Normal	-6%	1.00	92%	5%	1.00	92%	1%	1.00	94%
Naïve rounding	-75%	1.64	85%	-5%	.95	92%	12%	.97	94%
Multinomial	-18%	1.11	91%	18%	1.07	91%	-34%	1.04	92%
Proportional odds	-45%	1.28	89%	-1%	.99	92%	1%	.91	94%
Latent	-5%	1.00	92%	10%	.99	92%	-9%	.99	94%
					moderately asymmetric				
Normal	-13%	1.00	92%	5%	1.00	93%	0%	1.00	94%
Naïve rounding	-72%	1.58	86%	-11%	.97	93%	29%	.99	94%
Multinomial	-29%	1.16	91%	-3%	1.01	93%	2%	.88	94%
Proportional odds	-49%	1.28	90%	-10%	.98	93%	15%	.91	94%
Latent	-13%	1.02	92%	7%	.98	93%	-5%	.96	94%
					severely asymmetric				
Normal	-46%	1.00	83%	12%	1.00	90%	-23%	1.00	93%
Naïve rounding	-83%	1.59	73%	-45%	1.19	87%	118%	2.15	79%
Multinomial	-93%	1.78	69%	-55%	1.28	86%	121%	2.20	76%
Proportional odds	-91%	1.73	70%	-51%	1.23	87%	105%	1.85	81%
Latent	-26%	.91	84%	6%	.97	90%	6%	.88	93%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the *MSE* ratios > 1.2, and the CI coverage rates < 90% are underscored.

Table 6

The Simulation Result for Ordinal Data with 5 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-23%	1.00	91%	2%	1.00	92%	1%	1.00	96%
Naïve rounding	-37%	1.11	90%	-1%	.98	92%	4%	.97	96%
Multinomial	-163%	6.61	45%	-21%	1.06	91%	11%	1.25	92%
Proportional odds	-190%	8.36	38%	-30%	1.02	91%	-17%	0.88	95%
Latent	-26%	1.01	90%	4%	.99	92%	-8%	.99	96%
					moderately asymmetric				
Normal	-31%	1.00	90%	5%	1.00	92%	-5%	1.00	95%
Naïve rounding	-45%	1.15	89%	-4%	.97	92%	16%	.96	95%
Multinomial	-162%	7.09	41%	-40%	1.30	87%	127%	2.35	76%
Proportional odds	-236%	16.95	16%	-117%	2.43	70%	195%	5.80	43%
Latent	-35%	1.04	91%	2%	.97	91%	0%	.96	95%
					severely asymmetric				
Normal	-34%	1.00	87%	3%	1.00	91%	-9%	1.00	96%
Naïve rounding	-43%	1.14	87%	-15%	1.01	91%	42%	1.05	93%
Multinomial	-135%	4.88	51%	-56%	1.52	83%	194%	3.40	57%
Ordinal	-239%	14.35	16%	-128%	2.62	67%	403%	8.71	11%
Latent	-35%	1.04	88%	-1%	.97	91%	6%	.92	95%
N = 500					symmetric				
Normal	-6%	1.00	92%	5%	1.00	92%	2%	1.00	94%
Naïve rounding	-39%	1.17	91%	0%	.98	93%	7%	0.98	95%
Multinomial	-133%	5.60	55%	56%	1.34	88%	-102%	2.04	84%
Proportional odds	-108%	2.68	74%	-9%	.98	93%	5%	0.88	95%
Latent	-15%	1.03	92%	3%	.98	93%	0%	0.99	94%
					moderately asymmetric				
Normal	-33%	1.00	91%	8%	1.00	92%	-8%	1.00	95%
Naïve rounding	-65%	1.37	87%	-12%	.99	93%	33%	1.04	93%
Multinomial	-212%	16.02	16%	-119%	2.87	69%	171%	5.04	56%
Proportional odds	-156%	4.47	53%	-88%	1.76	82%	123%	2.32	76%
Latent	-35%	1.04	90%	0%	.97	93%	8%	.98	94%
					severely asymmetric				
normal	-43%	1.00	87%	10%	1.00	92%	-16%	1.00	95%
Naïve rounding	-66%	1.33	82%	-36%	1.13	91%	89%	1.60	86%
Multinomial	-166%	5.57	42%	-116%	2.70	70%	262%	6.91	39%
Proportional odds	-120%	2.47	65%	-80%	1.65	83%	139%	2.54	72%
Latent	-32%	.96	87%	1%	.97	92%	9%	.93	95%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the *MSE* ratios > 1.2, and the CI coverage rates < 90% are underscored.

Table 7
The Simulation Result for Ordinal Data with 7 Categories

	Alpha			Regression slope			Mean scale score		
	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage	Bias	*MSE* Ratio	CI Coverage
N = 125					symmetric				
Normal	-25%	1.00	90%	1%	1.00	92%	-1%	1.00	96%
Naïve rounding	-32%	1.05	90%	0%	.99	92%	1%	.98	96%
Multinomial	-168%	6.36	44%	-23%	1.09	89%	-47%	2.25	79%
Proportional odds	-298%	24.64	11%	-60%	1.11	90%	-65%	1.59	85%
Latent	-34%	1.06	90%	2%	.99	91%	-5%	.98	96%
					moderately asymmetric				
Normal	-32%	1.00	90%	3%	1.00	92%	-3%	1.00	96%
Naïve rounding	-39%	1.07	89%	-1%	0.98	92%	6%	.97	96%
Multinomial	-159%	6.13	45%	-32%	1.21	89%	73%	1.51	87%
Proportional odds	-290%	27.74	7%	-132%	2.40	70%	154%	3.65	60%
Latent	-36%	1.04	89%	1%	0.98	92%	0%	.96	96%
					severely asymmetric				
Normal	-33%	1.00	88%	4%	1.00	91%	-8%	1.00	95%
Naïve rounding	-40%	1.10	87%	-11%	1.00	92%	34%	1.01	94%
Multinomial	-139%	4.94	52%	-36%	1.32	86%	147%	2.43	71%
Proportional odds	-312%	28.28	1%	-148%	3.05	62%	471%	11.37	3%
Latent	-39%	1.10	88%	-2%	.97	91%	10%	.94	95%
N = 500					symmetric				
Normal	-6%	1.00	92%	4%	1.00	93%	0%	1.00	94%
Naïve rounding	-24%	1.07	92%	2%	.99	93%	4%	.99	94%
Multinomial	-278%	26.51	7%	-45%	1.37	87%	6%	1.43	90%
Proportional odds	-162%	5.20	53%	-16%	1.02	93%	3%	.90	94%
Latent	-10%	1.02	92%	6%	0.99	93%	-7%	1.00	94%
					moderately asymmetric				
Normal	-34%	1.00	90%	6%	1.00	92%	-5%	1.00	95%
Naïve rounding	-52%	1.17	88%	-3%	.98	93%	13%	0.98	95%
Multinomial	-296%	29.86	2%	-133%	3.08	66%	163%	4.56	60%
Proportional odds	-237%	14.56	17%	-114%	2.30	76%	132%	2.72	76%
Latent	-31%	.99	90%	3%	.97	93%	2%	.97	94%
					severely asymmetric				
Normal	-42%	1.00	88%	9%	1.00	91%	-12%	1.00	95%
Naïve rounding	-59%	1.24	85%	-28%	1.08	91%	73%	1.38	89%
Multinomial	-201%	10.52	20%	-115%	2.85	68%	281%	8.06	21%
Proportional odds	-168%	4.28	47%	-112%	2.32	74%	195%	4.01	53%
Latent	-30%	.94	89%	3%	.97	92%	3%	0.94	95%

Note. The biases reported in the table are percent standardized biases. The biases > 40%, the
MSE ratios >1.2, and the CI coverage rates < 90% are underscored.

Table 8

The Result from the Empirical Example

Strategy	Alpha Scale 1	Alpha Scale 2	Regression Slope	Mean Scale Score (Scale 1)	Mean Scale Score (Scale 2)
Normal	.78 [0.71 – 0.84]	0.87 [0.84 – 0.91]	0.59 [0.43 - 0.76]	43.66 [42.54 - 44.78]	45.89 [44.62 - 47.16]
Naïve rounding	.78 [0.71 – 0.84]	0.87 [0.84 – 0.91]	0.60 [0.43 - 0.76]	43.60 [42.49 - 44.70]	45.87 [44.64 - 47.10]
Multinomial	.76 [0.68 – 0.82]	0.89 [0.86 – 0.92]	0.71 [0.53 - 0.88]	43.36 [42.24 - 44.48]	45.32 [44.04 - 46.61]
Proportional odds	.73 [0.64 – 0.80]	0.90 [0.87 – 0.92]	0.72 [0.52 - 0.92]	42.98 [41.80- 44.15]	44.53 [43.03 - 46.04]
Latent	.78 [0.72 – 0.84]	0.88 [0.85 – 0.91]	0.59 [0.45 - 0.79]	43.67 [42.55 - 44.78]	45.82 [44.57 - 47.06]

Note. The values in the brackets are 95% CI limits.

Scale 1 = personal accomplishment subscale; Scale 2 = client-centered care subscale.
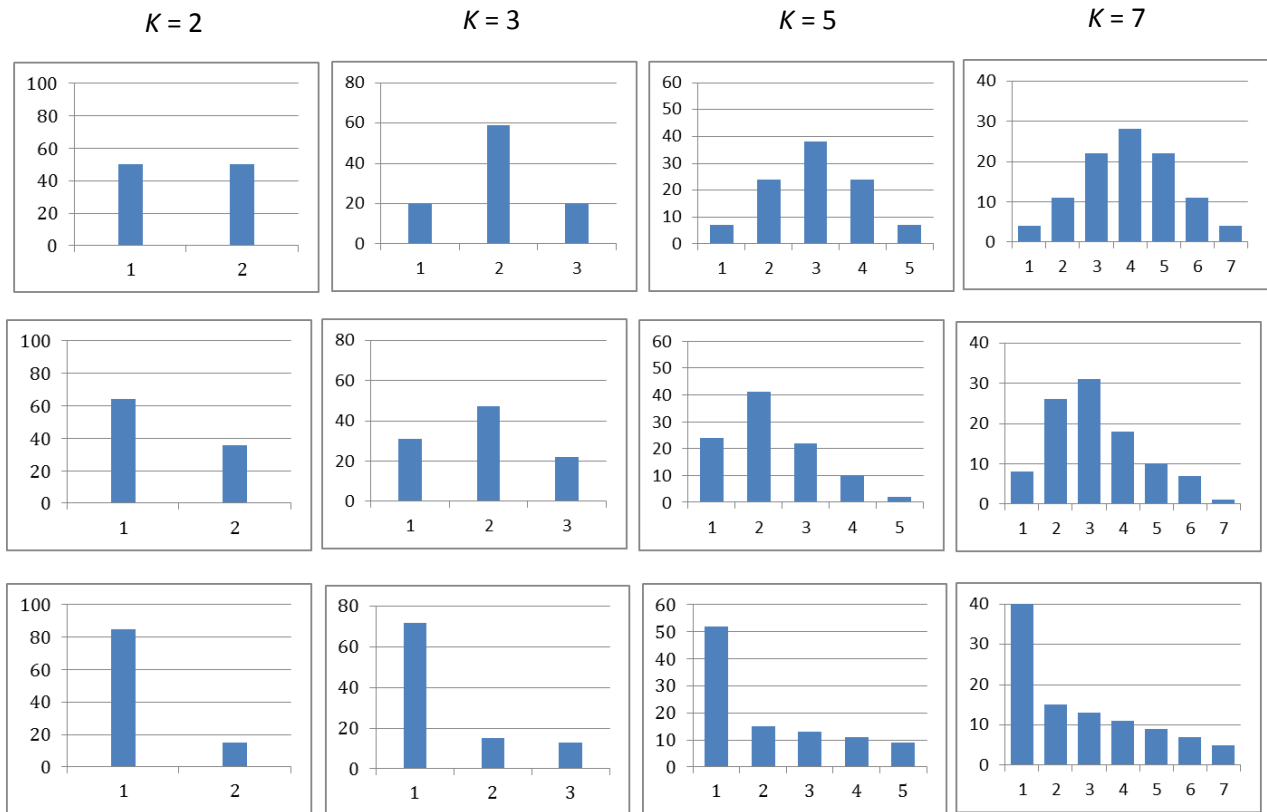
Figure 1. Theoretical frequency distributions of ordinal data. From top to bottom: symmetric,

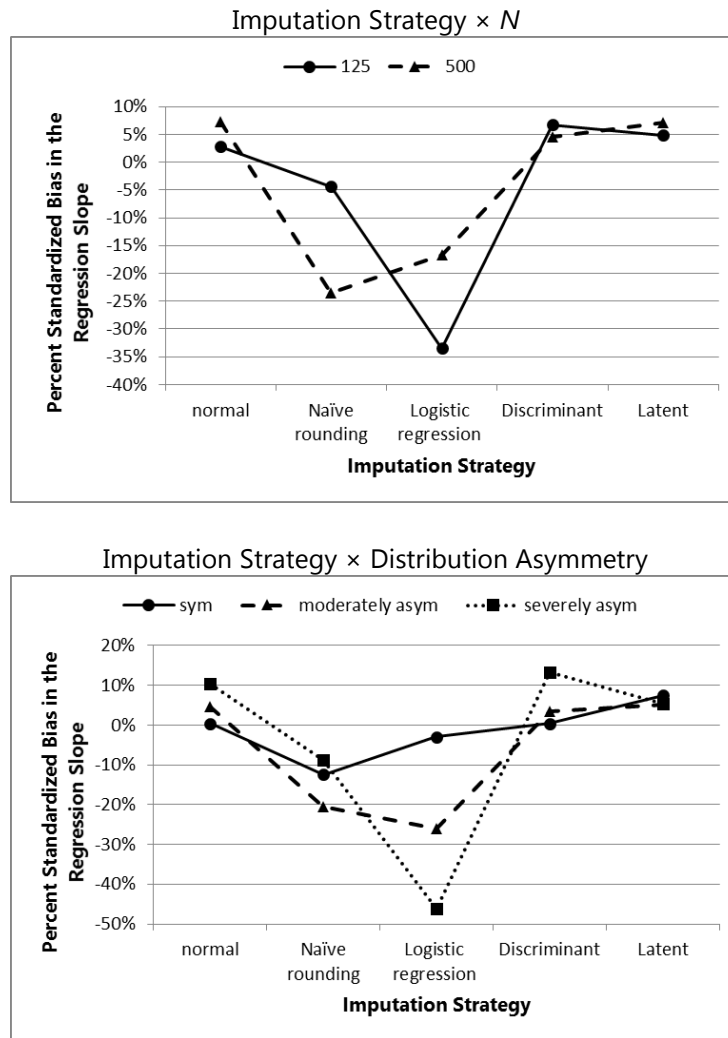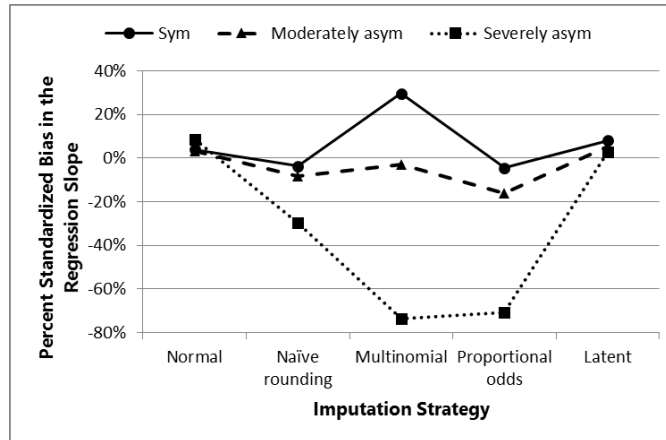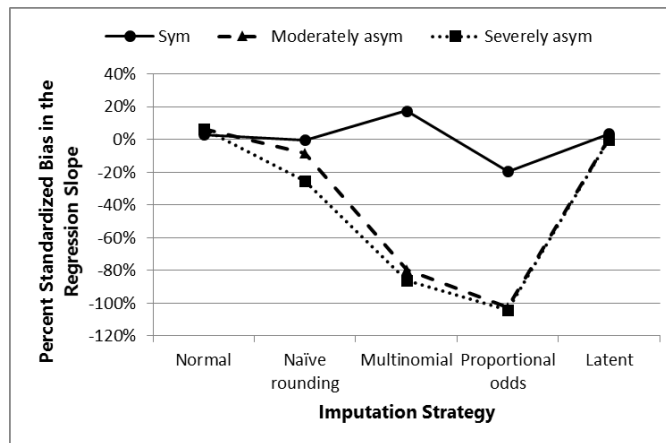moderately asymmetric, and severely asymmetric.

Figure 2. The two-way interaction effects between imputation strategy and *N* and between imputation strategy and distribution asymmetry for the regression slope for dichotomous data.
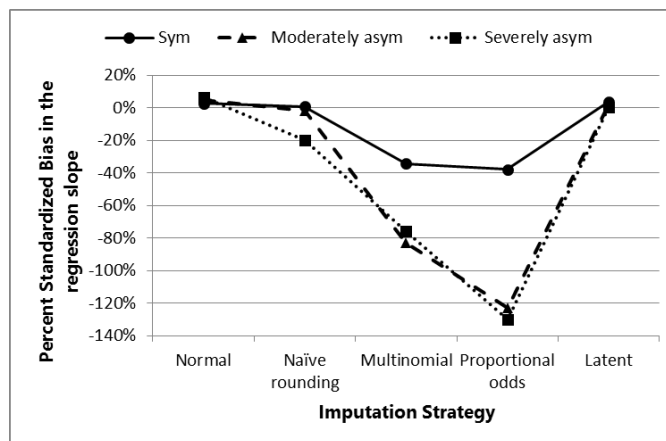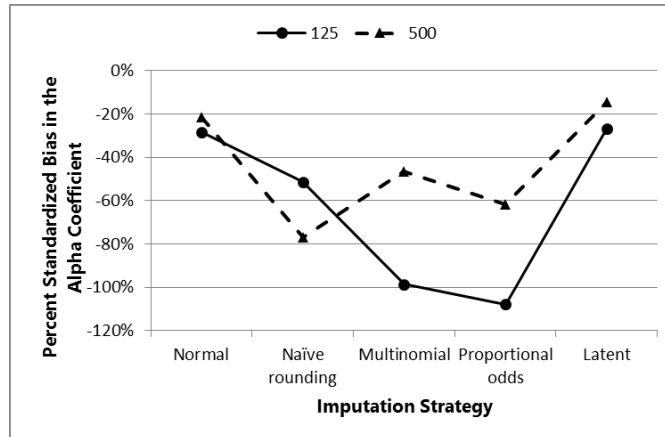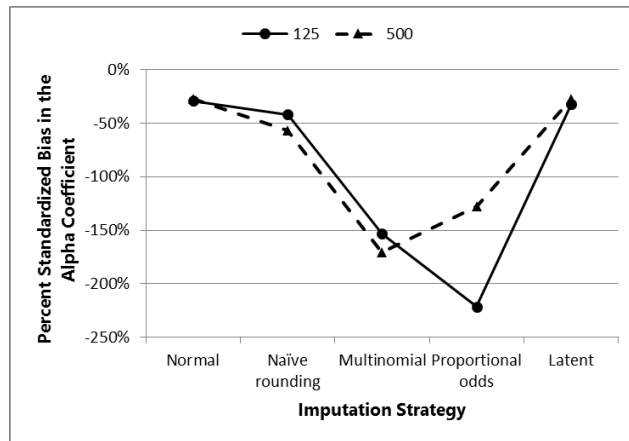
$K = 3$

$K = 5$

$K = 7$

Figure 3. Representative two-way interaction effects between imputation strategy and distribution asymmetry for polytomous data.
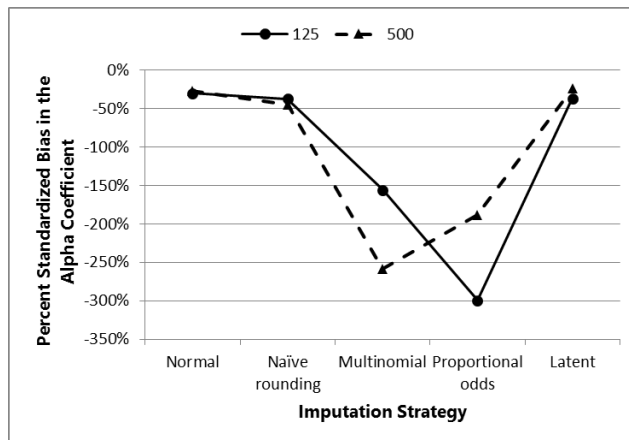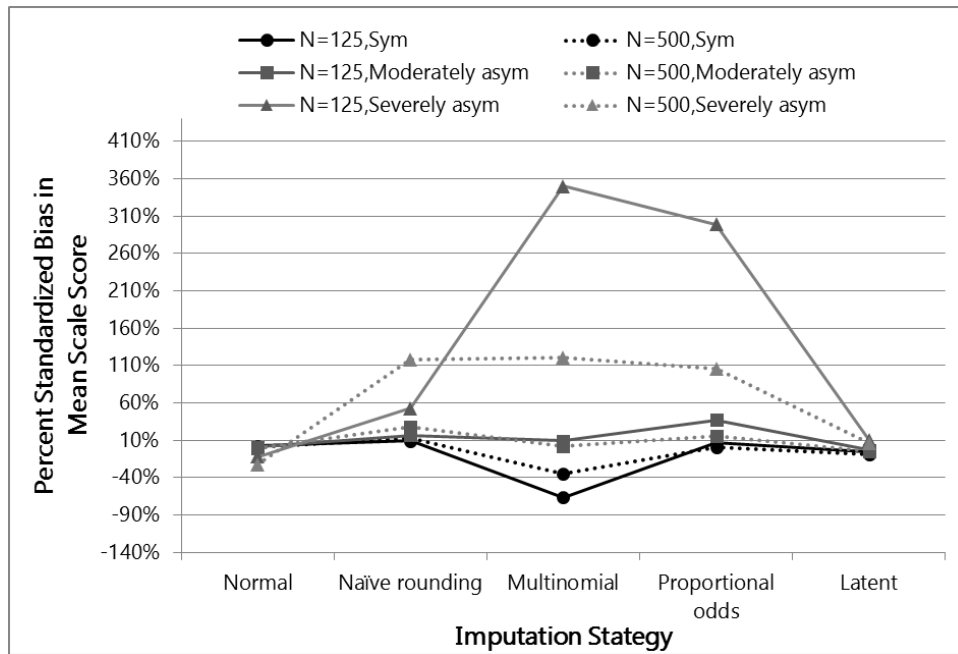
$K = 3$

$K = 5$

$K = 7$

Figure 4. Representative two-way interaction effects between imputation strategy and *N* for polytomous data.
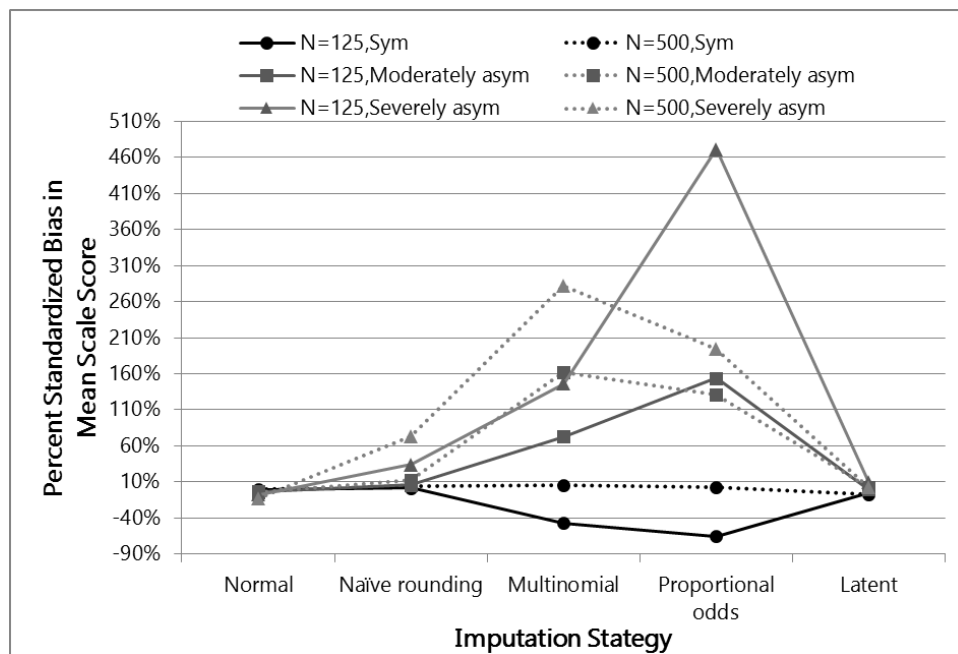
$K = 3$

$K = 7$

Figure 5. Representative three-way interaction effects among imputation strategy, distribution asymmetry, and *N* for polytomous data.

Appendix A

R code for the normal data model and logistic regression model approaches

for the empirical example.

```
############# Multiple Imputation using Amelia/mice
##      install.packages("Amelia")
##      install.packages("mice")
##      install.packages("psych")
##      install.packages("lavaan")
library(Amelia)
library(mice)

## impMeth==1 observed continuous (Amelia)
## impMeth==2 naive rounding (Amelia)
## impMeth==3 multinomial logistic (mice)
## impMeth==4 proportional odds (mice)

datMiss <- read.table("qoc_for_R.dat", header=T)
impMeth <- 1
nImp <- 50


if (impMeth==1){
        ##impute
        imp <- amelia(datMiss, m=nImp, p2s=0, empri =.01*nrow(datMiss))
        ##extract imputed data
        datImpd <- list()
        for (i in 1:nImp){
                datImpd[[i]] <- imp$imputations[[i]]
        }

} else if (impMeth==2){
        ##impute
        imp <- amelia(datMiss, m=nImp, p2s=0, empri =.01*nrow(datMiss))
        ##extract imputed data
        b1 <- seq(1.5, (7-0.5), 1)
        b2 <- seq(1.5, (5-0.5), 1)
        datImpd <- list()
        for (i in 1:nImp){
                datImpd[[i]] <- imp$imputations[[i]]
                for (j in 2:9) {
                        datImpd[[i]][,j] <- cut(datImpd[[i]][,j], breaks=c((min(datImpd[[i]][,j])-999), b1,
(max(datImpd[[i]][,j])+999)),
                        include.lowest = TRUE, labels=FALSE)
                }
                for (j in 10:20)
```

```
                              datImpd[[i]][,j] <- cut(datImpd[[i]][,j], breaks=c((min(datImpd[[i]][,j])-999), b2,
    (max(datImpd[[i]][,j])+999)),
                              include.lowest = TRUE, labels=FALSE)
                    }

    } else if (impMeth==3){
          ##impute
          datMiss[,1:20] <- lapply(datMiss[,1:20] , as.factor)
          imp <- mice(datMiss, meth=c("logreg", rep("polyreg",19)), m=nImp ,print=FALSE)
          ##extract imputed data
          imputations <- complete(imp, "long")
          datImpd <- list()
          for (i in 1:nImp){
                    datImpd[[i]] <- imputations[imputations$.imp==i, -1:-2]
                    nnd1 <- lapply(datImpd[[i]], as.character)
                    nnd2 <- lapply(nnd1, as.numeric)
                    datImpd[[i]] <- do.call("cbind", nnd2)
          }

    } else if (impMeth==4){
          ##impute
          datMiss[,1:20] <- lapply(datMiss[,1:20] , as.ordered)
          imp <- mice(datMiss, meth=c("logreg", rep("polr",19)), m=nImp ,print=FALSE)
          ##extract imputed data
          imputations <- complete(imp, "long")
          datImpd <- list()
          for (i in 1:nImp){
                    datImpd[[i]] <- imputations[imputations$.imp==i, -1:-2]
                    nnd1 <- lapply(datImpd[[i]], as.character)
                    nnd2 <- lapply(nnd1, as.numeric)
                    datImpd[[i]] <- do.call("cbind", nnd2)
          }
    }

    for (m in 1:nImp){
    write.table(datImpd[[m]], file=paste0("impMeth",impMeth ,"_", m, ".dat"), row.names=F, col.names=F)
    }
```

Appendix B

Mplus code for the latent variable approach for the empirical example.

TITLE: multiple imputation
DATA: FILE IS qoc_for_mplus.dat;
VARIABLE: NAMES = gender a1-a8 b1-b11;
USEVARIABLES = gender a1-a8 b1-b11;
MISSING = ALL (-999);
DATA IMPUTATION:
IMPUTE = gender (c) a1-a8 (c) b1-b11 (c);
NDATASETS = 50;
SAVE = impMethMplus_*.dat;
ANALYSIS: TYPE = BASIC;