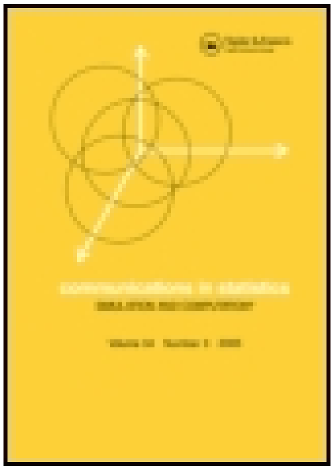

Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

A SIMULATION STUDY COMPARING MULTIPLE IMPUTATION METHODS FOR INCOMPLETE LONGITUDINAL ORDINAL DATA

A. F. Donneau [a] , M. Mauer [b] , G. Molenberghs [c] & A. Albert [a]

[a] Medical Informatics and Biostatistics , University of Liège , Sart Tilman B23, 4000, Liège , Belgium

[b] EORTC Headquarters, Departments of statistics and quality of life , Brussels , Belgium

[c] I-BioStat, University of Hasselt, Diepenbeek, Belgium - I-BioStat , Katholieke University of Leuven , Leuven , Belgium

Accepted author version posted online: 07 Apr 2014.

A SIMULATION STUDY COMPARING MULTIPLE IMPUTATION METHODS FOR INCOMPLETE LONGITUDINAL ORDINAL DATA

A.F. Donneau[1], M. Mauer[2], G. Molenberghs[3] and A. Albert[1]

[1] Medical Informatics and Biostatistics, University of Liège, Sart Tilman B23, 4000 Liège, Belgium,

[2]EORTC Headquarters, Departments of statistics and quality of life, Brussels, Belgium,

[3] I-BioStat, University of Hasselt, Diepenbeek, Belgium - I-BioStat, Katholieke University of Leuven, Leuven, Belgium

afdonneau@ulg.ac.be

Key Words: ordinal variables; longitudinal analysis; missing at random; multiple imputation.

ABSTRACT

Multiple imputation (MI) is now a reference solution for handling missing data. The default method for MI is the Multivariate Normal Imputation (MNI) algorithm which is based on the multivariate normal distribution. In the presence of longitudinal ordinal missing data, where the Gaussian assumption is no longer valid, application of the MNI method is questionable. This simulation study compares the performance of the MNI and ordinal imputation regression model for incomplete longitudinal ordinal data for situations covering various numbers of categories of the ordinal outcome, time occasions, sample sizes, rates of missingness, well-balanced and skewed data.

ACCEPTED MANUSCRIPT

1 Introduction

Longitudinal ordinal data arise naturally in many clinical settings. For example, in randomized treatment trials, the regular assessment of the patient's quality of life (QoL) by means of a Likert-type scale has become popular. In such longitudinal studies, however, subjects may drop out prematurely while others may miss one or more assessments. Rather than deleting missing values, it has been recommended to 'impute' them. The question of how to obtain valid inferences from imputed data was formally addressed by Rubin (1978) who introduced the multiple imputation (MI) method that replaces each missing value not only once but by a set of M ($M > 1$) plausible values whence reflecting the uncertainty about the prediction of the unknown missing values.

It is not uncommon in MI to rely on the assumption that the outcome variable follows a Normal distribution and hence ignore the categorical responses in the ordinal outcome. The present simulation study was designed to evaluate two MI methods for incomplete longitudinal ordinal data, one considering the outcome as continuous and the other as ordinal. The MI method for continuous outcome is based on the Markov Chain Monte Carlo (MCMC) method of data augmentation, while the MI method for ordinal outcome uses the proportional odds property of the ordinal logistic regression model. The paper will compare the performance of the two MI methods by focusing on the estimation of the parameters of the longitudinal ordinal logistic model. Both imputation methods were evaluated through Monte Carlo simulated artificial data sets. The simulations not only cover well-balanced data but also skewed distribution, as often observed in QoL studies.

The proportional odds model to analyze longitudinal ordinal data is briefly reviewed in Section 2, while a general overview of the problem of missing data is given in Section 3. Section 4 outlines the theoretical background of multiple imputation including those for continuous and ordinal variables. The simulation experimental design is described in Section 5 and results are presented in Section 6. Concluding remarks are given in Section 7.

2 Models for longitudinal ordinal data

2.1 The proportional odds model

Consider a sample of N subjects and let Y be an ordered variable with K categories assessed on T occasions on each subject. Then, let Y_{ij} denote the assessment of the ordinal variable Y for the ith subject ($i = 1, ..., N$) at the jth occasion ($j = 1, ..., T$). Hence, $\mathbf{Y}_i = (Y_{i1}, ..., Y_{iT})'$ is the vector of the repeated assessments of the ith subject. Associated with each subject, there is a $p \times 1$ vector of covariates, say $\mathbf{x}_{ij}$, measured at time j. Let $\mathbf{X}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$ denote the $T \times p$ design matrix of the ith subject. Covariates typically include time of measurement, age, gender, treatment group, interaction terms, and so on.

The ordinal nature of the outcome variable may be accounted for by considering the cumulative probabilities $Pr(Y_{ij} \leq k), k = 1, \cdots, K$. The cumulative proportional odds model is a popular choice to relate the marginal probabilities of Y to the covariate vector $\mathbf{x}$ (McCullagh, 1980). Specifically,

$$\text{logit}[\Pr(Y_{ij} \leq k|\mathbf{x}_{ij})] = \beta_{0k} + \mathbf{x}_{ij}'\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{\beta_0} = (\beta_{01}, ..., \beta_{0,K-1})'$ is the vector of the intercept parameters and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ the vector of coefficients ($i = 1, ..., N$; $j = 1, ..., T$; $k = 1, ..., K-1$). Under the proportional odds assumption, $\boldsymbol{\beta}$ does not depend on k.

2.2 Generalized estimating equations

Estimation of the regression parameters of marginal models can be approached by likelihood-based methods. One difficulty present with likelihood models resides in the complexity of the relationship

ACCEPTED MANUSCRIPT

between the parameters of the model and the joint probabilities that define the likelihood. Therefore, alternative solutions have been explored, in particular the generalized estimating equations (GEE), quite popular for the analysis of non-Gaussian correlated data. This approach circumvents the specification of the joint distribution of the repeated responses by means of a 'working' correlation matrix and only the marginal distributions are specified. Since the proportional odds model is not part of the regular generalized linear model family, some transformations are required before applying the GEE method. Following Lipsitz et al. (1994), a $(K-1)$-dimensional expanded vector of binary responses has to be created for each subject at each occasion, $\mathbf{Y}^*_{ij} = (Y^*_{i1j}, ..., Y^*_{i,(K-1),j})'$ where $Y^*_{ikj} = 1$ if $Y_{ij} \leq k$ and 0 otherwise. Now,

$$\text{logit}[\Pr(Y_{ij} \leq k|\mathbf{x}_{ij})] = \text{logit}[\Pr(Y^*_{ikj} = 1|\mathbf{x}_{ij})], \; k = 1, ..., K-1. \tag{2}$$

Since the logistic regression model is a member of the generalized linear model family, the GEE method applies and consistent estimates of the regression parameters can be obtained by solving the estimating equations

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\pi}'_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1}(\mathbf{Y}^*_i - \boldsymbol{\pi}_i) = \mathbf{0}, \tag{3}$$

where $\mathbf{Y}^*_i = (\mathbf{Y}^*_{i1}, ..., \mathbf{Y}^*_{iT})'$, $\boldsymbol{\pi}_i = E(\mathbf{Y}^*_i)$, $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}$ with $\mathbf{A}_i$ the diagonal matrix of the variance of the elements of $\mathbf{Y}^*_i$, and $\boldsymbol{\beta}$ the expanded vector of intercepts and regression coefficients. The matrix $\mathbf{R}_i$ is the 'working' correlation matrix that expresses the dependence among repeated observations over the subjects ranging from independence to exchangeable, banded, or unstructured.

3 Missingness

The profile of incomplete observations in a longitudinal data set may exhibit a variety of patterns. When an individual withdraws from the study before its completion time, we have a case of dropout. The missingness pattern may be monotone or non-monotone. In a monotone pattern, if

ACCEPTED MANUSCRIPT

Y_{ij} is missing for some j, then Y_{ik} is missing for all $k > j$. As a consequence, if Y_{ij} is known, so are all Y_{ik} $(k < j)$. By contrast, in a non-monotone pattern, there will be missing data before last available assessment. In line with the notation introduced previously, consider the missing data indicators, R_{ij}, defined as follows:

$$R_{ij} = \begin{cases} 1 \text{ if } Y_{ij} \text{ is observed,} \\ 0 \text{ otherwise,} \end{cases}$$

and let $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{iT})'$ the indicator vector corresponding to $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT})'$. Now $\mathbf{Y}_i$ can be split into two subvectors $(\mathbf{Y}_i^o, \mathbf{Y}_i^m)$ where $\mathbf{Y}_i^o$ refers to the observed component of $\mathbf{Y}_i$ and $\mathbf{Y}_i^m$ refers to the missing component part. For monotone dropout, $\boldsymbol{R}_i$ is of the form $(1, \ldots, 1, 0, \ldots, 0)$ and can be used to define the dropout indicator $D_i = 1 + \sum_{j=1}^{T} R_{ij}$ which represents the time at which subject i dropped out.

When missing data occur, we are concerned with the distribution of the measurement process together with the missing-data process. Little and Rubin (1987) and Little (1993, 1995) identified two broad classes of joint models: the selection model and the pattern-mixture model. In the selection model, the joint distribution $(\mathbf{Y}_i, \boldsymbol{R}_i)$ is split into the marginal distribution of the measurement and the distribution of the missingness process conditional on the measurement $\mathbf{Y}_i$. By contrast, the pattern-mixture model specifies the marginal distribution of $\boldsymbol{R}_i$ and the conditional distribution of $\mathbf{Y}_i$ given $\boldsymbol{R}_i$. Here we shall focus on the selection model approach in which Rubin (1987) and Little and Rubin (1987) made essential distinctions between the processes responsible for the missingness: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The determination of the mechanism responsible for missing data has a decisive implication on the choice of the statistical method used to analyze the data. Under the MCAR mechanism, the probability of an observation being missing is independent of both $\mathbf{Y}^o$ and $\mathbf{Y}^m$. Under the MAR mechanism, the probability of an observation being missing is independent of $\mathbf{Y}^m$ given $\mathbf{Y}^o$. When neither MCAR nor MAR holds, the missingness mechanism is said to be MNAR,

ACCEPTED MANUSCRIPT

so that the probability of an observation being missing depends on $\mathbf{Y}^m$.

Liang and Zeger (1986) pointed out that GEE are only valid under the restrictive assumption that the data are missing completely at random (MCAR). Alternative methods were investigated to allow the analysis of data under less strict missingness assumptions. Robins et al. (1995a, 1995b) developed an extension of the GEE, known as the weighted generalized estimating equations (WGEE), that provide consistent estimates of the regression parameters even under the MAR assumption. With their method, each subject's measurements is weighted in the GEE by the inverse probability of dropping out at that time point. Another alternative to analyze the data under the MAR assumption is multiple imputation based on GEE (MI-GEE). In this approach, missing values are imputed several times (Rubin, 1976, 1978) and the resulting completed datasets are analyzed using standard GEE methods. Using Rubin's rules, the final results obtained from the completed datasets are combined into a single inference. In the context of longitudinal binary data, Beunckens et al. (2008) showed by simulations that, in spite of the asymptotic unbiasedness of WGEE, the combination of GEE and multiple imputation is both less biased and more accurate in small to moderate sample sizes which typically arise in clinical trials. In this paper, focus will be on MI-GEE methods.

4 Multiple imputation

4.1 Theoretical framework

The idea behind multiple imputation is that instead of filling in a single value for each missing data, the technique is to replace each missing value with a set of $M > 1$ plausible values drawn from the conditional distribution of the missing data given the observed data. This conditional distribution represents the uncertainty about the right value to impute in the sense that the set of M imputed values properly represents the information about the missing value that is contained in the observed data. These M complete data sets are then analyzed by the method that would have been appropriate if the data had been complete. The model used in this last step is called the substantive model, while the model used in the imputation task is called the imputation model. Results derived from the substantive model are then combined using simple rules provided by Rubin (1987), resulting in a single inference about the parameters of interest that accounts for uncertainty due to missing data.

Using the notation introduced in previous sections, let $\boldsymbol{\theta}$ represent the parameter vector of the distribution of the response $\mathbf{Y}_i = (\mathbf{Y}_i^o, \mathbf{Y}_i^m)$. Note that $\boldsymbol{\theta}$ may differ from the parameters $\boldsymbol{\beta}$ of the substantive model. The observed data $\mathbf{Y}^o$ will be used to estimate the conditional distribution of $\mathbf{Y}^m$ given $\mathbf{Y}^o$, $f(\mathbf{Y}^{\mathbf{m}}|\mathbf{Y}^o, \boldsymbol{\theta})$. If $\boldsymbol{\theta}$ is known, the values for $\mathbf{Y}^m$ can be drawn from $f(\mathbf{Y}^m|\mathbf{Y}^o, \boldsymbol{\theta})$. For $\boldsymbol{\theta}$ unknown, an estimate is obtained from the data, say $\hat{\boldsymbol{\theta}}$; then missing values will be imputed using $f(\mathbf{Y}^m|\mathbf{Y}^o, \hat{\boldsymbol{\theta}})$. Frequentists incorporate uncertainty in $\boldsymbol{\theta}$ by using bootstrap or other methods. A Bayesian prior distribution for $\boldsymbol{\theta}$ can also be chosen. Given this distribution, a draw $\boldsymbol{\theta}^*$ is generated and now values for $\mathbf{Y}^m$ can be drawn from $f(\mathbf{Y}^m|\mathbf{Y}^o, \boldsymbol{\theta}^*)$. These two steps for the construction of the imputed data are the first phase of MI. Then the substantive model is applied to each of the M completed data $(\mathbf{Y}_i^o, \mathbf{Y}_i^{m*})$. Let $\hat{\boldsymbol{\beta}}_{\boldsymbol{m}}$ and $\hat{\boldsymbol{U}}_m$ be the vector of estimates and the corresponding variance-covariance matrix for the m^{th} imputed data set ($m = 1, \ldots, M$), respectively. The last step of MI is the combination of the M results. The MI point estimate for $\boldsymbol{\beta}$ is simply the average of the M complete-data point estimates (Rubin, 1987; Schafer, 1997),

$$\hat{\boldsymbol{\beta}}^* = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_m.$$

A measure of the precision of $\hat{\boldsymbol{\beta}}^*$ is obtained by Rubin's variance formula (Rubin, 1987) which combines the within- and the between-imputation variability. Define $\mathbf{W}$, the within-imputation variance, as the average of the M within imputation variance estimates $\hat{\boldsymbol{U}}_m$,

$$\mathbf{W} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{U}}_m,$$

and **B**, the between-imputation variance, measuring the variability across the imputed values,

$$\mathbf{B} = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}^*)(\hat{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}^*)'.$$

Then, the variance estimate associated with $\hat{\boldsymbol{\beta}}^*$ is the total variance

$$\mathbf{T} = \mathbf{W} + \left(1 + \frac{1}{M}\right)\mathbf{B},$$

where $\left(1 + \frac{1}{M}\right)$ is a correction factor for the finite number of imputations.

4.2 Multivariate Normal Imputation Method

MCMC methods have been considered to explore and simulate the entire joint posterior distribution of the unknown quantities through the use of Markov chains, and thereby obtain simulation-based estimates of virtually any feature of the posterior that are of interest. For this reason, MCMC methods are widely applied in the imputation phase of multiple imputation methods.

Assuming that data arise from a multivariate normal distribution, Schafer (1997) developed a method based on MNI for generating proper imputations that accounts for between-imputation variability. This approach, based on the algorithm of data augmentation (Tanner and Wong (1987)), is a procedure that iterates between an imputation step (I-step) and a posterior step (P-step). Let the T assessments of the ordinal outcome variable be viewed as a random vector, $(\mathbf{Y}_1, \cdots, \mathbf{Y}_T)'$ assumed to follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In the I-step, given starting values for $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, values for missing data $\mathbf{Y}^m$ are simulated by randomly drawing a value from the conditional multivariate normal distribution of $\mathbf{Y}^m$ given $\mathbf{Y}^o$, $f(\mathbf{Y}^m|\mathbf{Y}^o, \boldsymbol{\theta})$. The conditional mean, $\boldsymbol{\mu}_{m|o}$, and the conditional covariance matrix, $\boldsymbol{\Sigma}_{m|o}$, have to be derived. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_o, \boldsymbol{\mu}_m)$ be the mean vector of the variable calculated in the observed and in the missing part of the dataset. In the same way, suppose that the covariance matrix is partitioned as follows,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_o & \boldsymbol{\Sigma}_{o,m} \\ \boldsymbol{\Sigma}'_{o,m} & \boldsymbol{\Sigma}_m \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{o,m}$ denotes the covariance matrix between $\mathbf{Y}^o$ and $\mathbf{Y}^m$, $\boldsymbol{\Sigma}_o$ and $\boldsymbol{\Sigma}_m$ represent the variance matrix for $\mathbf{Y}^o$ and $\mathbf{Y}^m$, respectively. It has been shown (Goodnight (1979), Schafer (1997)) that the conditional covariance matrix, $\boldsymbol{\Sigma}_{m|o}$, can be expressed as:

$$\boldsymbol{\Sigma}_{m|o} = \boldsymbol{\Sigma}_m - \boldsymbol{\Sigma}'_{o,m}\boldsymbol{\Sigma}_o^{-1}\boldsymbol{\Sigma}_{o,m}. \tag{4}$$

Thus,

$$f(\mathbf{Y}^m|\mathbf{Y}^o, \boldsymbol{\theta}) \sim N(\boldsymbol{\mu}_{m|o}, \boldsymbol{\Sigma}_{m|o}),$$

with $\boldsymbol{\mu}_{m|o} = \boldsymbol{\mu}_m + \boldsymbol{\Sigma}'_{o,m}\boldsymbol{\Sigma}_o^{-1}(\mathbf{Y}^o - \boldsymbol{\mu}_o)$ and $\boldsymbol{\Sigma}_{m|o}$ given by (4).

After the first iteration, new values for $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are drawn from its posterior distribution. Assuming a noninformative prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, their posterior distribution at the tth iteration are given by a Normal and an inverted Wishart distribution (Schafer (1997)),

$$\boldsymbol{\mu}^{(t)}_{|\boldsymbol{\Sigma}} \sim N\left(\overline{\mathbf{Y}}, \frac{1}{n}\boldsymbol{\Sigma}^{(t)}\right), \tag{5}$$

$$\boldsymbol{\Sigma}^{(t)} \sim W^{-1}[n-1, (n-1)\mathbf{S}], \tag{6}$$

where $(\overline{\mathbf{Y}}, \mathbf{S})$ are both determined by the observed data and the missing data imputed in the last I-step, as follows, $\overline{\mathbf{Y}} = 1/n \sum_{i=1}^N \mathbf{y}_i = 1/n(\sum_{i=1}^N y_{i1}, \cdots, \sum_{i=1}^N y_{iT})'$ and $(n-1)\mathbf{S} = \sum_{i=1}^N (\mathbf{y}_i - \overline{\mathbf{Y}})(\mathbf{y}_i - \overline{\mathbf{Y}})'$. Both steps are iterated, which creates a Markov chain $(\mathbf{Y}^m_{(1)}, \boldsymbol{\theta}_{(1)}), (\mathbf{Y}^m_{(2)}, \boldsymbol{\theta}_{(2)}), \cdots$ where each step depends on the previous one, introducing dependency across the steps. The two steps are then iterated long enough until the distribution becomes stationary. Imputations from the last iteration are used to impute the missing values of the dataset. The Expectation-Maximization (EM) algorithm was used to derive initial mean and covariance estimates for the MNI method. More detail about this procedure can be found in Schafer (1997).

When proceeding this way for an ordinal outcome, the imputed values obtained are no longer integer values and need then to be rounded off to the nearest integer (category) or to the nearest plausible value. However, in the binary case, it was demonstrated that rounding is not recommended because the rounded imputed values may provide biased parameter estimates (Horton et al., 2003; Ake, 2005; Allison, 2005). In situations like ours, where one is concerned with presence of missing values for the outcome variable, unrounded values are physically not plausible. So, the rounding phase is unavoidable before application of the substantive model (e.g. GEE with proportional odds).

4.3 Ordinal imputation method

The ordinal imputation method (OIM) appears as an alternative to the MNI approach. To impute missing data for an ordinal outcome, one has to impose a probability model on the complete data. Multiple imputation of a longitudinal dataset with monotone missingness patterns consists in a sequential application of methods designed for univariate data by considering the previous fully observed assessment times as covariates.

In the presence of an ordinal outcome variable, a proportional odds model will be considered in the first step of the imputation phase to link the ordinal outcome to a set of q covariates. In a longitudinal setting, the covariates typically include those of the substantive model $\boldsymbol{X}_{ij}$, possible auxiliary covariates $\boldsymbol{A}_{ij}$, and the previous outcomes $\tilde{\boldsymbol{Y}}_{ij} = (Y_{i1}, ..., Y_{i,j-1})$. Specifically $\mathbf{X}_i^* = (\boldsymbol{X}_{ij}, \boldsymbol{A}_{ij}, \tilde{\boldsymbol{Y}}_{ij})'$ and the model is written as :

$$logit[Pr(Y_{ij} \leq k)|\mathbf{x}_{ij}^*] = \gamma_{0k} + \mathbf{x'}_{ij}^{*}\boldsymbol{\gamma}. \tag{7}$$

Regression coefficient estimates $\hat{\boldsymbol{\Gamma}} = (\boldsymbol{\gamma}_0', \boldsymbol{\gamma}')'$, where $\boldsymbol{\gamma}_0 = (\gamma_{01}, \cdots, \gamma_{0,(K-1)})$, and corresponding covariance matrix $\boldsymbol{V} = V(\hat{\boldsymbol{\Gamma}})$ are obtained by fitting the proportional odds model to the observed data. Based on these estimates, the algorithm to impute missing values at the jth assessment, Y_{ij}^m, operates as follows:

1. Draw new values for parameters $\mathbf{\Gamma}$, assuming large-sample normal approximation $N\left(\hat{\mathbf{\Gamma}}, V(\hat{\mathbf{\Gamma}})\right)$ of its posterior distribution assuming the noninformative prior $Pr(\mathbf{\Gamma}) \propto const$. In other words, compute

$$\mathbf{\Gamma}^* = \hat{\mathbf{\Gamma}} + \boldsymbol{C}'\mathbf{Z},$$

where $\boldsymbol{C}'$ is the upper triangular matrix of the Cholesky decomposition, $\boldsymbol{V} = \boldsymbol{C}'\boldsymbol{C}$ and $\mathbf{Z}$ is a $(K-1)+q$ vector of independent random normal variates.

2. For an observation with missing Y_{ij}^m and corresponding covariates $\mathbf{X}_{ij}^*$, from (Eq. 7) compute the expected probabilities, $\pi_k = P[Y_{ij} = k|\mathbf{x}_{ij}^*]$ $(k = 1, ..., K)$.

3. For each observation with missing Y_{ij}^m, draw a random variate from a multinomial distribution with the vector of probabilities $(\pi_1, \cdots, \pi_K)$ derived in the previous step.

4. Repeat steps 1 to 3 to obtain M sets of imputed values $(Y_{ij}^{(1)}, Y_{ij}^{(2)}, \cdots, Y_{ij}^{(M)})$, $(i = 1, \cdots, N;$ $j = 1, \cdots, T)$.

5 Simulation study

To assess the performance of both imputation methods (MNI and OIM), we conducted a large simulation study as described hereafter.

5.1 Longitudinal ordinal data-generating model

Correlated ordinal responses were generated with the SAS macro developed by Ibrahim and Suliadi (2011) and based on Lee's algorithm (Lee, 1997). The basic measurement model utilized in this study includes as covariates a binary group effect ($X = 0$ or 1), an assessment time (T) and an

ACCEPTED MANUSCRIPT

interaction term between group and time, so that the proportional odds model (Eq. 1) is written as $(i = 1, \cdots, N;\ j = 1, \cdots, T;\ k = 1, \cdots, K-1)$:

$$\text{logit}[\Pr(Y_{ij} \leq k | x_i, t_j)] = \beta_{0k} + \beta_x x_i + \beta_t t_j + \beta_{tx} x_i t_j. \tag{8}$$

The required arguments in Ibrahim's macro are: the marginal probabilities at each time point, the correlation structure and the sample size. To generate the longitudinal data in the two groups defined by the binary variable (i.e., $X = 0, 1$), the macro was applied twice. The corresponding marginal probabilities at each time point were derived using Eq. (8). As an example, for the group defined by $X = 1$, the value of the group parameter, x_i, was fixed to 1, the value of the time parameter, t_j, was fixed to $1, \cdots, T$ and the interaction term, $x_i t_j$, was given by the product of the two previously fixed parameters. Based on these values and using the theoretical values of the model parameters displayed in Table 1, the probabilities to be in each modalities of the ordinal outcome Y at each time point was determined for the group $X = 1$. For the correlation structure, the following exchangeable correlation structure was assumed:

$$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0.2 & j \neq k \end{cases}$$

$(i = 1, \cdots, N;\ j, k = 1, \cdots, T)$.

Within the GEE framework, Liang and Zeger (1986) demonstrated that consistent estimates are obtained whatever the choice of the working correlation matrix. As a consequence, the correlation structure chosen in the simulations will have no impact on the derived results.

5.2 Missing data generating mechanisms

The mechanism used to generate MAR missingness data is based on the following binary logistic regression model $(i = 1, \cdots, N;\ j = 1, \cdots, T;\ k = 1, \cdots, K-1)$:

$$\text{logit}[\Pr(D_i = j|x_i, y_{i,(j-1)})] = \psi_0 + \psi_x x_i + \psi_{prev} y_{i,(j-1)}. \quad (9)$$

Thus, the probability of drop out at a certain time point j depends on the binary covariate X_i and the outcome value at the previous time point $Y_{i,(j-1)}$. Verbiage about how to choose the population parameters to generate missing data was added in Appendix 8.2.

5.3 Simulation patterns

Theoretical values of the model parameters (see (Eq. 8)) considered in our simulations are given in Table 1 for well-balanced and skewed distributions. As an illustration, Figure 1 displays the distribution of the theoretical probabilities derived from Table 1 in each group and at each time point for $K = 4$ for a short study ($T = 3$) under well-balanced and skewed settings.

Three distinct sample sizes N were considered for the simulation: 100, 300 and 500, equally distributed between both groups. This covers small (50 subjects/arm) to large studies (250 subjects/arm). For the assessment time points T, two possibilities were envisaged corresponding to short ($T = 3$) or long ($T = 5$) longitudinal study. Note that for skewed data, only $T = 3$ was considered. The ordinal outcome variable Y covered various numbers of categories $K = 2, 3, 4, 5$ and 7, respectively. Finally, the population parameters of (Eq. 9) ($\psi_0, \psi_x, \psi_{prev}$) were chosen to yield a rate of missingness approximatively equal to 10%, 30% and 50%, respectively. The complete data case (0% missingness) was also considered. Thus, both imputation methods (MNI and OIM) were assessed on 90 different combination patterns. For each pattern, S = 500 random samples were generated. The two MI methods (MNI and OIM) were applied to impute missing data on the same incomplete dataset allowing a paired comparaison of the two approaches. A GEE model was then fitted to the resulting multiply imputed datasets. For each MI method, the number of multiple imputation was fixed to $M = 20$ (Rubin, 1987; Graham et al., 2007). As the generation of the MAR missingness was based on the binary covariate X, the latter had to be included in the imputation model. In the GEE model, the same working correlation as the one used in the generation data process was considered, that is an exchangeable correlation matrix. The MI based on MNI and on OIM were carried out using the SAS MI procedure. The GEE SAS macro based on the extension of Lipsitz et al. method (1994) and implemented by Williamson et al. (1999) was used to analyze the imputed datasets. Ultimately, the SAS MIANALYZE procedure was used to pool the results obtained.

5.4 Evaluation criteria

For each simulation pattern, the relative bias $RB = \hat{\beta}/\beta$ expressed in percent was averaged over the $S = 500$ replicated datasets. Likewise, the mean square error was calculated as

$$MSE = Bias^2 + Var(\hat{\beta}),$$

with $Var(\hat{\beta}) = \sum_{s=1}^{S} \frac{(\hat{\beta}_s - \bar{\hat{\beta}})^2}{(S-1)}$, $\bar{\hat{\beta}} = \sum_{s=1}^{S} \frac{\hat{\beta}_s}{S}$ and $Bias = \bar{\hat{\beta}} - \beta$.

The effect of the modeling parameters on RB was assessed by multiple regression analysis and so was the difference between RB obtained by MCMC and OIM, respectively. This statistical scheme was applied to both kinds of generated ordinal data, well-balanced and skewed distribution.

6 Results

The values of the relative bias (%) and the MSE calculated over the 500 replicate samples are detailed in Appendices for both imputation methods. For clarity, results for intercepts were omitted.

6.1 Well-balanced distributions

Relative bias. Table 2 presents the mean (±SD) of RB of each regression parameter derived from both imputation methods as well as their difference. Globally, the MNI method yielded highly underestimated values of the model parameters, whereas for the OIM method estimates were almost unbiased. Therefore, the RB difference between the two imputation methods was highly significant ($p < 0.0001$) for all parameters, ranging from 9% to 16%. A closer look at the results revealed that for the binary effect parameter, β_x, the relative bias using MNI was unchanged for K, N and rate of missingness, and varied only slightly with the number of time points. Specifically, RB was lower in long term than in short term studies (92.3 ± 12.0 % vs 86.5 ± 13.5 %; $p = 0.034$). The RB for the time effect parameter, β_t, decreased significantly with the number of categories K ($p < 0.0001$) and with the percentage of missingness ($p < 0.0001$) but was unaffected by N and T. It decreased from 96.4 ± 5.31 % for K=2 to 76.6 ± 9.07 % for K=7 and from 90.9 ± 4.08 % for 10% of missingness to 80.2 ± 14.0 % for 50% of missingness. The same observation was made for the interaction term, β_{tx}, except that a significant effect was noted for T (91.7 ± 5.82 % vs 89.4 ± 5.47 %; $p = 0.007$). By contrast, when focusing on the OIM appraoch, the relative bias behaved similarly for each regression parameter. RB decreased significantly with the number of categories K ($p < 0.0001$), as well as with the number of time occasions T ($p < 0.05$) but increased with the sample size N ($p < 0.05$). Contrary to the MNI method, no effect of the percentage of missingness was observed. Looking at the RB differences between the two approaches, results for model parameters were comparable except for the time parameter β_t where the bias was substantiably larger for $T = 3$ as compared to $T = 5$ ($p = 0.001$).

Mean square error. The mean square error (mean ± SD) of each regression parameters under both imputation methods and their difference are given in Table 3. Globally, although results were highly significant ($p < 0.0001$), difference between MNI and OIM were minute and not practically relevant. From this perspective, MNI and OIM were similar. As expected, under both imputation methods and for each model parameter, the MSE decreased significantly (p < 0.0001) with the sample size N. A decrease was also observed with T ($p < 0.0001$). MSE values also got lower as the number of categories K increased but the relationship did not always reach statistical significance. The rate of missingness did not really affect MSE except for the time parameter in

ACCEPTED MANUSCRIPT

both imputation methods (MNI: $p = 0.015$ and OIM: $p = 0.0005$).

6.2 Skewed distributions

As already mentioned, the case of skewed ordinal data was investigated in the context of a short term study only, that is $T = 3$. Simulation results are summarized in the Appendices.

Multiple linear regression of the MNI relative bias on all modeling parameters (K, N and missingness) showed that, except for the time effect, RB increased significantly with the number of categories K (β_x: $p < 0.0001$, β_t: $p = 0.068$, β_{tx}: $p = 0.0002$) and with the percentage of missingness (β_x: $p < 0.0001$, β_t: $p = 0.57$, β_{tx}: $p = 0.0005$). No relationship was observed between the OIM relative bias and the modeling parameters. Contrary to the well-balanced case, the MNI method overestimated the binary and the interaction term parameters of the model, while at the same time underestimated the time parameter β_t. As before, the OIM method yielded less biased estimates (see Figures 2 and 3). The RB of the time parameter, β_t, was more affected by the skewness of the ordinal outcome than the other model parameters and this effect was even more marked with the MNI method. In fact, the lowest RB value of β_t was equal to 42.1% and the highest RB value was equal to 265.5%; both extremes were obtained under the MNI method. The corresponding OIM relative biases were equal to 103.9% and 169.7%, respectively.

The MSE of each regression parameter under both imputation methods and their differences are displayed in Table 4. Comparison of the MSE calculated in presence of skewed ordinal outcomes with those derived in well-balanced setting showed that MSE values were larger in presence of skewness. The conclusions made previously on MSE values in case of well-balanced distributions can be transported here. Specifically, MNI and OIM mean square errors were similar and differences of MSE under both methods were not meaningful, even if statistically significant. As expected, the MSE decreased significantly ($p < 0.0001$) with the sample size. The MSE decreased

ACCEPTED MANUSCRIPT

with the number of categories of the ordinal outcome for the binary effect under OIM ($p = 0.009$) and less markedly for both time effect and interaction terms of the model. The effect of the rate of missingness on MSE was significant for the time effect parameter, under both imputation methods (MNI: $p = 0.001$ and OIM: $p < 0.0001$) and the MSE of the interaction term of the model derived under OIM (p = 0.017). Although not relevant, the difference in the MSE of both imputation methods decreased with the sample size for the time and the interaction term of the model. The MSE difference for the latter further deteriorated with higher rates of missingness ($p < 0.0001$). The number of categories of the ordinal outcome affected differently the MSE difference of the binary and the interaction terms of the model. For the binary effect of the model, the difference in MSE increased with the number of categories of the ordinal outcome ($p < 0.0001$), while for the interaction term the MSE difference decreased ($p = 0.0009$).

7 Discussion

This paper compared the performance of two imputation methods available in statistical packages, namely the MNI algorithm and the ordinal imputation regression model, in the context of longitudinal ordinal datasets with missing values. The comparison was based on a comprehensive simulation plan covering a wide range of real life situations. Specifically, the parameters of the experimental design included the number of categories of the ordinal outcome (K), the number of time points (T), the sample size (N) and the rate of missingness (%) but also the form of the distribution (well-balanced or skewed) of the ordinal outcome data. The performance of the two methods (MNI and OIM) was appraised by the relative bias and the mean square error of the regression parameters of the model. The latter included a group effect and a time effect, as well as their interaction.

In the well-balanced setting, the estimates of the model parameters obtained with the MNI approach were markedly underestimated (RB << 100%). By contrast, estimates derived with the

ACCEPTED MANUSCRIPT

OIM method were almost unbiased. These general observations however have to be tempered according to the study pattern. For example, RB differences between MNI and OIM for the binary and the interaction model parameters vanished with increasing K, the number of categories. By contrast, for the time effect parameter, the RB difference increased with K but decreased with T, the number of time points. For all regression parameters, the MSE of both imputation methods were almost equal but departed slightly for larger sample sizes and higher missingness rates. For skewed data, estimates under MNI method were positively biased, except for the time effect, and MSE were comparable.

In conclusion, based on the results of this large simulation study, the MNI method is not really recommended to analyze longitudinal ordinal data with missing values. Preferably, it is advisable to impute missing ordinal data using an appropriate regression model. The OIM method however requires the construction of an imputation model. Meng (1994) showed that, as long as the imputation model is not grossly misspecified, MI approach will perform well. From a practical point of view, the imputation model should at least include any variable structure (e.g. interaction) present in the substantive model (Fay, 1992). The inclusion of other available covariates, which are not necessarily of interest in the substantive model, is unlikely to produced biased results. Therefore, Rubin's rule which consists in including as many variables as possible when performing multiple imputation (Rubin,1996) is recommended. Furthermore, in the binary setting with MAR missingness, Beunckens et al. (2008) demonstrated the robustness of MI-GEE when misspecifying either the imputation or measurement model. Those findings were extended in the MNAR case (Birhanu et al., 2011).

As a final remark, we should noted that, contrary to the MNI method, the use of the OIM method is limited to the situation of monotone missingness. In the presence of non-monotone missingness, a solution to minimize the imputation's bias could be to iterate between application

ACCEPTED MANUSCRIPT

of MNI and OIM methods to first monotonize the dataset before application of the OIM method. This proposal, however, requires additional researches.

BIBLIOGRAPHY

Ake, C. (2005). Rounding after multiple imputation with non-binary categorical covariates. *Paper presented at SAS Users Group international*, Thirty annual conference, Philadelphia.

Allison, P. (2005). Imputation of categorical variables with PROC MI. *Paper presented at SAS Users Group international*, Thirty annual conference, Philadelphia.

Beunckens, C., Sotto, C. and Molenberghs, G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics and Data Analysis,* **52**, 1533–1548.

Birhanu, T., Molenberghs, G., Sotto, C. and Kenward, M. G. (2011). Doubly robust and multiple-imputation-based generalized estimating equations. *Journal of Biopharmaceutical Statistics*, **21**, 202–225.

Fay, R. E. (1992). When are inferences from multiple imputation valid?. *U.S. Bureau of the Census*, Washington, DC 20233-4001.

Goodnight, J. H (1979). A Tutorial on the SWEEP Operator. *The American Statistician* **33**, 149–158.

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

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

Ibrahim, N. and Suliadi, S. (2011). Generating correlated discrete ordinal data using R and SAS

ACCEPTED MANUSCRIPT

IML. *Computer Methods and Programs in Biomedicine*, **104**, 122–132.

Lee, A. J. (1997). Some simple methods for generating correlated categorical variates. *Computational Statistics and Data Analysis*, **26**, 133–148.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lipsitz, SR., Kim, K. and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley: New York.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**, 125–134.

Little, R. J. A.. (1995). Modelling the drop-out mechanism in repeated measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, **42**, 109–142.

Meng, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input (with discussion). *Statistical Science*, **9**, 538–573.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, **90**, 122–129.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106–121.

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

ACCEPTED MANUSCRIPT

Rubin, D. (1978). Multiple imputation in sample surveys - a phenomenological bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 20–34.

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

Rubin, D. B. (1996). Multiple imputation after 18+years. *Journal of the American Statistical Association*, **91**, 473–489.

Schafer, J. L. (1997) *Analysis of Incomplete Multivariate Data*,Chapman & Hall

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of American Statistical Association*, **82**, 528–550.

Williamson, J., Lipsitz, S. and Kim, K. (1999). GEECAT and GEEGOR: computer programs for the analysis of correlated categorical response data. *Computer Methods and Programs in Biomedicine*, **58**, 25–34.

ACCEPTED MANUSCRIPT

Table 1: Values of the model parameters used for generating longitudinal ordinal dataset (well-balanced and skewed distributions)

Distribution	K	β_{01}	β_{02}	β_{03}	β_{04}	β_{05}	β_{06}	β_x	β_t	β_{tx}
Well-balanced										
	2	-0.25	-	-	-	-	-	0.10	0.10	-0.15
	3	-0.71	0.66	-	-	-	-	0.10	0.10	-0.15
	4	-1.10	0.00	1.10	-	-	-	0.10	0.10	-0.15
	5	-1.39	-0.41	0.41	1.39	-	-	0.10	0.10	-0.15
	7	-1.79	-0.92	-0.29	0.29	0.92	1.79	0.10	0.10	-0.15
Skewed										
	2	1.00	-	-	-	-	-	0.80	0.10	-0.25
	3	-2.20	-0.85	-	-	-	-	0.80	0.10	-0.25
	4	-0.41	0.00	0.41	-	-	-	0.80	0.10	-0.25
	5	-0.85	-0.20	0.20	0.85	-	-	0.80	0.10	-0.25
	7	-1.39	-0.66	-0.16	0.16	0.66	1.39	0.80	0.10	-0.25

Table 2: Relative bias (mean ± SD) of the parameters of the substantive model after imputation of the ordinal outcome using MCMC and OIM methods. Globally and according to the modeling parameters

		β_x			β_t			β_{tx}		
		MNI	OIM	MNI-OIM	MNI	OIM	MNI-OIM	MNI	OIM	MNI-OIM
Global		89.4 ± 13.1	99.5 ± 15.5	-10.1 ± 8.91	84.6 ± 10.4	100.9 ± 8.95	-16.4 ± 9.58	90.6 ± 5.73	99.7 ± 5.37	-9.10 ± 4.60
				< 0.0001			< 0.0001			< 0.0001
K	2	91.7 ± 14.3	106.7 ± 11.8	-15.0 ± 7.39	96.4 ± 5.31	104.3 ± 7.74	-7.91 ± 4.26	92.9 ± 5.18	101.2 ± 2.93	-8.35 ± 4.29
	3	95.5 ± 10.9	106.9 ± 13.1	-11.4 ± 5.78	89.4 ± 5.69	103.8 ± 5.34	-14.4 ± 6.13	94.1 ± 2.98	103.4 ± 4.23	-9.35 ± 4.34
	4	81.3 ± 17.8	94.9 ± 19.6	-13.6 ± 5.28	80.0 ± 8.52	102.1 ± 6.79	-22.1 ± 9.92	88.0 ± 6.71	99.1 ± 6.05	-11.1 ± 4.66
	5	86.4 ± 7.95	96.4 ± 11.1	-9.94 ± 7.09	80.5 ± 8.36	102.6 ± 8.36	-22.1 ± 11.2	89.1 ± 5.36	99.5 ± 3.09	-10.4 ± 4.70
	7	92.1 ± 7.51	92.6 ± 15.7	-0.52 ± 10.6	76.6 ± 9.07	92.0 ± 10.3	-15.4 ± 7.14	88.7 ± 5.56	95.0 ± 6.12	-6.34 ± 3.87
		0.63	0.0005	< 0.0001	< 0.0001	< 0.0001	0.0014	< 0.0001	< 0.0001	0.034
T	3	92.3 ± 12.0	103.0 ± 12.8	-10.7 ± 8.06	85.1 ± 11.4	103.5 ± 9.08	-18.4 ± 11.2	91.7 ± 5.82	100.9 ± 5.34	-9.26 ± 4.73
	5	86.5 ± 13.5	96.0 ± 17.2	-9.46 ± 9.73	84.0 ± 9.31	98.4 ± 8.12	-14.3 ± 7.23	89.4 ± 5.47	98.4 ± 5.14	-8.94 ± 4.51
		0.034	0.018	0.39	0.46	0.001	0.009	0.007	0.009	0.61
N	100	87.4 ± 17.1	93.2 ± 20.7	-5.82 ± 10.8	84.1 ± 11.5	97.4 ± 10.4	-13.3 ± 8.16	90.5 ± 6.60	97.7 ± 6.73	-7.22 ± 4.18
	300	91.0 ± 12.2	102.8 ± 13.0	-11.8 ± 7.26	84.6 ± 9.88	102.1 ± 8.02	-17.5 ± 9.58	90.9 ± 5.37	100.8 ± 4.77	-9.88 ± 4.48
	500	89.8 ± 8.67	102.5 ± 8.96	-12.6 ± 6.82	85.0 ± 9.98	103.4 ± 7.24	-18.4 ± 10.4	90.2 ± 5.29	100.4 ± 3.85	-10.2 ± 4.67
		0.47	0.012	0.0003	0.61	0.002	0.008	0.74	0.027	0.0002
Missingness	10	92.6 ± 11.3	99.5 ± 11.5	-6.89 ± 1.68	90.9 ± 4.08	99.8 ± 3.24	-8.91 ± 3.54	95.4 ± 2.65	100.1 ± 2.47	-4.64 ± 0.94
	30	87.9 ± 11.9	99.9 ± 14.0	-12.0 ± 6.08	82.6 ± 7.26	101.2 ± 5.59	-18.6 ± 7.36	89.9 ± 3.23	99.9 ± 3.57	-9.94 ± 2.21
	50	87.7 ± 15.4	99.1 ± 20.2	-11.4 ± 13.7	80.2 ± 14.0	101.8 ± 14.2	-21.6 ± 11.1	86.3 ± 6.29	99.0 ± 8.31	-12.7 ± 4.92
		0.14	0.90	0.014	< 0.0001	0.29	< 0.0001	< 0.0001	0.37	< 0.0001

ACCEPTED MANUSCRIPT

Table 3: Mean square error (mean ± SD) of the parameters of the substantive model after imputation of the ordinal outcome using MCMC and OIM methods. Globally and according to the modeling parameters

		β_x			β_t			β_{tx}		
		MNI	OIM	MNI-OIM	MNI	OIM	MNI-OIM	MNI	OIM	MNI-OIM
Global		0.123 ± 0.098	0.119 ± 0.095	0.004 ± 0.009	0.011 ± 0.012	0.013 ± 0.014	-0.001 ± 0.003	0.020 ± 0.022	0.022 ± 0.024	-0.001 ± 0.003
				< 0.0001			< 0.0001			< 0.0001
K	2	0.136 ± 0.107	0.141 ± 0.112	-0.005 ± 0.007	0.013 ± 0.015	0.015 ± 0.017	-0.002 ± 0.003	0.024 ± 0.026	0.027 ± 0.029	-0.002 ± 0.004
	3	0.131 ± 0.106	0.128 ± 0.104	0.003 ± 0.003	0.012 ± 0.013	0.013 ± 0.015	-0.002 ± 0.003	0.022 ± 0.024	0.024 ± 0.027	-0.002 ± 0.003
	4	0.124 ± 0.111	0.117 ± 0.104	0.007 ± 0.008	0.011 ± 0.013	0.012 ± 0.015	-0.001 ± 0.002	0.020 ± 0.023	0.021 ± 0.024	-0.001 ± 0.002
	5	0.111 ± 0.086	0.105 ± 0.081	0.007 ± 0.007	0.010 ± 0.011	0.012 ± 0.014	-0.002 ± 0.004	0.018 ± 0.018	0.019 ± 0.020	-0.001 ± 0.002
	7	0.114 ± 0.085	0.104 ± 0.078	0.009 ± 0.011	0.009 ± 0.009	0.011 ± 0.011	-0.001 ± 0.002	0.017 ± 0.017	0.017 ± 0.018	-0.000 ± 0.001
		0.063	0.005	< 0.0001	0.068	0.095	0.51	0.034	0.013	0.0003
T	3	0.159 ± 0.113	0.154 ± 0.111	0.004 ± 0.010	0.018 ± 0.014	0.021 ± 0.016	-0.003 ± 0.003	0.033 ± 0.024	0.036 ± 0.026	-0.002 ± 0.003
	5	0.087 ± 0.063	0.084 ± 0.060	0.004 ± 0.008	0.004 ± 0.003	0.004 ± 0.004	-0.000 ± 0.001	0.007 ± 0.005	0.008 ± 0.006	-0.000 ± 0.001
		< 0.0001	< 0.0001	0.78	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
N	100	0.242 ± 0.077	0.234 ± 0.078	0.009 ± 0.014	0.022 ± 0.015	0.025 ± 0.018	-0.003 ± 0.004	0.040 ± 0.027	0.042 ± 0.030	-0.002 ± 0.004
	300	0.080 ± 0.026	0.078 ± 0.026	0.002 ± 0.004	0.007 ± 0.005	0.008 ± 0.006	-0.001 ± 0.001	0.013 ± 0.009	0.014 ± 0.010	-0.001 ± 0.001
	500	0.047 ± 0.016	0.046 ± 0.016	0.001 ± 0.002	0.004 ± 0.003	0.005 ± 0.004	-0.000 ± 0.001	0.008 ± 0.005	0.008 ± 0.006	-0.000 ± 0.001
		< 0.0001	< 0.0001	0.0002	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Missingness	10	0.115 ± 0.094	0.113 ± 0.093	0.002 ± 0.003	0.009 ± 0.010	0.010 ± 0.010	-0.000 ± 0.000	0.018 ± 0.020	0.018 ± 0.020	-0.000 ± 0.001
	30	0.123 ± 0.099	0.119 ± 0.097	0.004 ± 0.007	0.011 ± 0.011	0.012 ± 0.013	-0.001 ± 0.001	0.020 ± 0.022	0.021 ± 0.023	-0.001 ± 0.002
	50	0.131 ± 0.103	0.125 ± 0.099	0.006 ± 0.013	0.013 ± 0.014	0.017 ± 0.018	-0.003 ± 0.004	0.023 ± 0.024	0.025 ± 0.027	-0.003 ± 0.004
		0.15	0.29	0.024	0.015	0.0005	< 0.0001	0.099	0.028	< 0.0001

Table 4: Mean square error (mean ± SD) of the parameters of the substantive model after imputation of the ordinal outcome using MNI and OIM methods, globally and according to the modeling parameters (skewed distribution)

		β_x			β_t			β_{tx}		
		MNI	OIM	MNI-OIM	MNI	OIM	MNI-OIM	MNI	OIM	MNI-OIM
Global		0.192 ± 0.144	0.184 ± 0.144	0.008 ± 0.018	0.022 ± 0.016	0.023 ± 0.017	-0.000 ± 0.006	0.038 ± 0.029	0.042 ± 0.034	-0.004 ± 0.006
				< 0.0001			0.61			< 0.0001
K	2	0.241 ± 0.197	0.254 ± 0.210	-0.013 ± 0.017	0.026 ± 0.018	0.027 ± 0.020	-0.000 ± 0.003	0.050 ± 0.041	0.058 ± 0.049	-0.008 ± 0.010
	3	0.173 ± 0.126	0.171 ± 0.125	0.002 ± 0.003	0.033 ± 0.020	0.026 ± 0.019	0.007 ± 0.007	0.036 ± 0.026	0.040 ± 0.029	-0.004 ± 0.005
	4	0.186 ± 0.138	0.170 ± 0.129	0.016 ± 0.012	0.017 ± 0.012	0.020 ± 0.015	-0.003 ± 0.004	0.035 ± 0.027	0.040 ± 0.030	-0.005 ± 0.005
	5	0.194 ± 0.138	0.178 ± 0.130	0.016 ± 0.010	0.017 ± 0.013	0.020 ± 0.017	-0.003 ± 0.006	0.038 ± 0.027	0.041 ± 0.031	-0.003 ± 0.005
	7	0.169 ± 0.131	0.148 ± 0.117	0.021 ± 0.019	0.017 ± 0.013	0.019 ± 0.016	-0.002 ± 0.004	0.032 ± 0.025	0.033 ± 0.026	-0.000 ± 0.002
		0.088	0.010	< 0.0001	0.0005	0.012	0.04	0.026	0.007	0.0009
N	100	0.384 ± 0.070	0.371 ± 0.089	0.013 ± 0.028	0.040 ± 0.012	0.044 ± 0.012	-0.004 ± 0.007	0.076 ± 0.017	0.085 ± 0.024	-0.009 ± 0.009
	300	0.119 ± 0.015	0.112 ± 0.016	0.007 ± 0.010	0.016 ± 0.009	0.015 ± 0.004	0.001 ± 0.005	0.024 ± 0.004	0.026 ± 0.005	-0.002 ± 0.002
	500	0.075 ± 0.014	0.070 ± 0.016	0.005 ± 0.008	0.010 ± 0.007	0.009 ± 0.003	0.001 ± 0.005	0.015 ± 0.003	0.016 ± 0.005	-0.001 ± 0.002
		< 0.0001	<0.0001	0.11	<0.0001	<0.0001	0.020	<0.0001	<0.0001	<0.0001
Missingness	10	0.177 ± 0.138	0.174 ± 0.38	0.003 ± 0.006	0.017 ± 0.012	0.017 ± 0.013	-0.000 ± 0.001	0.035 ± 0.027	0.036 ± 0.028	-0.001 ± 0.001
	30	0.192 ± 0.147	0.183 ± 0.145	0.010 ± 0.015	0.021 ± 0.014	0.021 ± 0.015	0.000 ± 0.003	0.038 ± 0.029	0.041 ± 0.032	-0.003 ± 0.003
	50	0.209 ± 0.155	0.197 ± 0.158	0.012 ± 0.027	0.028 ± 0.020	0.030 ± 0.021	-0.002 ± 0.010	0.043 ± 0.032	0.051 ± 0.040	-0.008 ± 0.009
		0.19	0.36	0.092	0.001	<0.0001	0.45	0.11	0.017	<0.0001

8 Appendices

8.1 Simulation results for the MI-GEE based MNI and OIM methods

Table 5: Simulation results for the MI-GEE based MNI and OIM methods (K = 2 - Well-balanced distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	117.4	0.392	115.9	0.337	122.8	0.340	119.8	0.342	132.0	0.358	110.7	0.341	130.4	0.369
		β_t	98.6	0.038	92.7	0.036	95.4	0.037	95.8	0.040	102.4	0.044	103.5	0.052	114.0	0.064
		β_{tx}	99.5	0.081	99.9	0.072	103.5	0.073	100.1	0.078	107.6	0.085	93.4	0.085	106.0	0.101
3	300	β_x	117.4	0.125	94.7	0.122	100.5	0.123	76.5	0.120	89.1	0.123	78.8	0.116	103.3	0.121
		β_t	98.6	0.013	97.8	0.014	101.4	0.014	93.3	0.013	101.3	0.015	106.9	0.017	122.9	0.021
		β_{tx}	99.5	0.024	98.9	0.026	102.5	0.027	88.5	0.027	96.6	0.029	89.0	0.027	104.2	0.032
3	500	β_x	98.4	0.072	97.6	0.066	87.7	0.067	102.4	0.065	79.5	0.067	101.9	0.070	104.4	0.075
		β_t	98.0	0.007	97.4	0.007	95.1	0.007	100.2	0.008	108.1	0.009	103.6	0.010	123.8	0.013
		β_{tx}	97.1	0.014	98.2	0.014	93.0	0.014	101.2	0.015	87.3	0.016	101.6	0.017	101.9	0.021
5	100	β_x	115.6	0.190	98.2	0.181	106.0	0.183	92.5	0.186	108.6	0.193	77.1	0.197	96.8	0.205
		β_t	108.0	0.008	97.4	0.009	100.6	0.009	97.5	0.010	105.8	0.011	92.7	0.011	104.0	0.013
		β_{tx}	106.0	0.016	97.5	0.016	101.0	0.016	93.9	0.017	102.1	0.018	87.3	0.020	98.4	0.022
5	300	β_x	109.6	0.058	104.4	0.061	111.5	0.061	91.2	0.064	110.0	0.067	84.8	0.066	111.7	0.072
		β_t	104.7	0.003	96.3	0.003	99.4	0.003	92.2	0.003	100.3	0.003	90.9	0.003	102.8	0.004
		β_{tx}	102.6	0.005	97.4	0.005	100.7	0.005	91.3	0.006	86.5	0.006	99.4	0.006	99.9	0.007
5	500	β_x	110.3	0.035	92.2	0.034	99.2	0.035	81.2	0.037	97.4	0.037	68.0	0.035	92.6	0.037
		β_t	101.6	0.002	96.7	0.001	99.8	0.002	92.6	0.002	100.3	0.002	87.7	0.002	99.9	0.002
		β_{tx}	102.0	0.003	95.7	0.003	99.1	0.003	90.4	0.004	98.6	0.004	83.3	0.004	96.8	0.004

ACCEPTED MANUSCRIPT

Table 6: Simulation results for the MI-GEE based on MNI and OIM methods (K = 3 - Well-balanced distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	103.3	0.294	84.0	0.322	90.1	0.321	80.9	0.339	91.3	0.330	97.8	0.348	99.8	0.339
		β_t	113.5	0.029	91.3	0.033	100.5	0.034	85.0	0.037	100.9	0.040	100.8	0.045	107.4	0.057
		β_{tx}	103.8	0.059	95.6	0.066	99.8	0.068	91.1	0.072	99.8	0.075	100.9	0.079	109.5	0.091
3	300	β_x	116.7	0.094	90.6	0.108	97.4	0.107	87.9	0.114	98.4	0.111	106.9	0.121	117.5	0.119
		β_t	111.5	0.010	87.9	0.010	97.6	0.010	83.3	0.012	99.6	0.013	97.9	0.014	111.1	0.019
		β_{tx}	107.0	0.020	95.3	0.021	100.2	0.022	92.1	0.024	101.6	0.025	98.6	0.026	110.6	0.030
3	500	β_x	109.7	0.053	84.9	0.066	92.1	0.065	86.7	0.069	99.2	0.066	99.8	0.070	112.7	0.066
		β_t	106.7	0.006	87.7	0.006	96.9	0.006	82.9	0.007	100.3	0.007	96.4	0.008	113.7	0.010
		β_{tx}	104.4	0.011	93.7	0.013	98.5	0.014	91.0	0.014	101.4	0.015	95.3	0.015	108.4	0.017
5	100	β_x	117.9	0.155	107.7	0.164	113.8	0.161	105.6	0.180	119.4	0.174	122.6	0.180	134.7	0.176
		β_t	105.5	0.007	96.6	0.007	103.6	0.007	90.0	0.008	105.2	0.008	90.4	0.010	110.7	0.012
		β_{tx}	103.0	0.013	97.5	0.014	101.5	0.014	93.1	0.015	102.6	0.016	94.0	0.017	107.3	0.019
5	300	β_x	89.7	0.047	91.1	0.053	98.0	0.053	81.6	0.061	98.7	0.059	101.0	0.063	123.7	0.062
		β_t	100.5	0.002	91.6	0.002	98.6	0.002	82.2	0.003	100.0	0.003	84.4	0.003	110.4	0.004
		β_{tx}	99.5	0.003	95.2	0.005	99.3	0.005	89.0	0.006	100.1	0.006	91.6	0.006	108.8	0.007
5	500	β_x	88.3	0.027	94.7	0.030	102.2	0.030	94.9	0.032	111.3	0.031	100.9	0.036	124.6	0.036
		β_t	99.6	0.001	92.1	0.001	99.3	0.001	85.2	0.002	102.9	0.002	83.6	0.002	109.2	0.002
		β_{tx}	99.0	0.002	95.6	0.002	99.8	0.002	92.1	0.003	103.2	0.003	91.9	0.003	109.2	0.004

Table 7: Simulation results for the MI-GEE based on MNI and OIM methods (K = 4 - Well-balanced distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	122.3	0.264	78.3	0.313	88.3	0.303	73.1	0.346	91.0	0.322	60.7	0.373	80.3	0.345
		β_t	106.7	0.027	83.0	0.031	95.4	0.030	69.2	0.037	95.7	0.039	62.3	0.045	95.6	0.053
		β_{tx}	107.8	0.052	90.7	0.061	97.1	0.061	85.4	0.068	98.1	0.069	74.7	0.077	89.7	0.082
3	300	β_x	107.3	0.098	109.8	0.084	117.8	0.082	102.7	0.097	120.9	0.091	110.5	0.105	132.2	0.099
		β_t	103.7	0.008	91.1	0.009	104.2	0.009	81.0	0.010	109.3	0.011	76.5	0.013	114.8	0.016
		β_{tx}	100.4	0.018	99.8	0.016	105.4	0.017	93.8	0.020	106.8	0.020	94.6	0.023	112.6	0.027
3	500	β_x	105.2	0.058	101.4	0.055	110.1	0.054	97.7	0.060	116.0	0.057	93.9	0.070	114.8	0.066
		β_t	101.7	0.005	93.7	0.005	106.6	0.005	80.7	0.006	110.8	0.006	72.6	0.008	112.7	0.010
		β_{tx}	100.1	0.011	97.5	0.011	103.2	0.011	91.7	0.012	105.0	0.013	87.4	0.016	105.0	0.018
5	100	β_x	96.5	0.146	62.1	0.148	98.9	0.143	67.1	0.156	72.1	0.149	59.2	0.183	72.9	0.169
		β_t	95.2	0.005	86.6	0.006	95.4	0.006	79.0	0.007	94.9	0.008	70.0	0.009	91.4	0.009
		β_{tx}	99.0	0.010	87.9	0.011	95.5	0.011	84.3	0.013	92.5	0.013	78.9	0.016	92.7	0.016
5	300	β_x	82.9	0.046	71.5	0.049	79.6	0.046	60.6	0.052	76.2	0.048	64.6	0.057	77.9	0.052
		β_t	96.5	0.002	89.3	0.002	99.1	0.002	79.1	0.003	100.4	0.002	76.6	0.003	103.2	0.003
		β_{tx}	96.9	0.003	91.1	0.004	96.4	0.004	83.4	0.005	95.1	0.004	80.6	0.0005	96.3	0.005
5	500	β_x	82.1	0.027	88.5	0.028	96.7	0.027	80.2	0.029	95.8	0.027	80.8	0.033	95.9	0.030
		β_t	96.8	0.001	92.3	0.001	102.1	0.001	80.2	0.002	101.9	0.002	76.0	0.002	103.8	0.002
		β_{tx}	97.2	0.002	93.9	0.002	99.0	0.002	86.8	0.003	98.9	0.003	82.2	0.004	97.9	0.003

ACCEPTED MANUSCRIPT

Table 8: Simulation results for the MI-GEE based on MNI and OIM methods (K = 5 - Well-balanced distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	87.8	0.251	92.5	0.253	97.2	0.247	87.1	0.273	92.0	0.262	80.8	0.276	82.5	0.251
		β_t	104.1	0.024	92.8	0.027	103.0	0.028	80.8	0.031	103.2	0.036	75.2	0.036	107.9	0.055
		β_{tx}	99.4	0.050	98.1	0.048	102.0	0.048	90.0	0.055	98.3	0.059	84.9	0.060	95.0	0.069
3	300	β_x	90.7	0.085	96.2	0.079	105.1	0.076	90.1	0.087	105.3	0.083	84.1	0.097	104.1	0.088
		β_t	105.5	0.008	89.6	0.009	104.0	0.009	80.1	0.010	111.4	0.011	73.5	0.014	115.6	0.018
		β_{tx}	97.4	0.017	96.8	0.015	103.0	0.015	90.0	0.018	102.7	0.019	84.5	0.023	102.6	0.025
3	500	β_x	85.5	0.051	92.0	0.048	100.4	0.046	84.1	0.053	102.4	0.050	71.8	0.059	95.7	0.054
		β_t	101.4	0.005	90.0	0.006	104.4	0.006	77.4	0.006	109.8	0.007	70.1	0.009	117.0	0.012
		β_{tx}	96.3	0.010	94.4	0.010	100.3	0.010	87.8	0.011	102.1	0.012	80.1	0.015	99.6	0.015
5	100	β_x	91.3	0.148	77.5	0.153	84.0	0.148	74.8	0.169	80.5	0.156	74.3	0.195	68.2	0.174
		β_t	102.8	0.005	88.3	0.006	97.0	0.006	80.3	0.007	97.8	0.007	65.8	0.010	81.4	0.013
		β_{tx}	101.0	0.010	93.7	0.012	98.5	0.012	88.6	0.013	77.4	0.013	81.5	0.016	90.7	0.018
5	300	β_x	103.4	0.047	86.7	0.048	94.0	0.047	89.3	0.053	101.7	0.050	88.5	0.061	99.8	0.056
		β_t	100.2	0.002	89.1	0.002	98.6	0.002	80.7	0.003	101.1	0.003	70.1	0.003	98.1	0.003
		β_{tx}	100.2	0.003	92.7	0.004	98.0	0.004	89.1	0.004	100.5	0.004	82.5	0.006	97.2	0.006
5	500	β_x	103.5	0.028	98.5	0.030	107.1	0.029	91.7	0.032	106.8	0.030	96.1	0.037	108.2	0.034
		β_t	99.6	0.001	92.2	0.001	102.4	0.001	81.1	0.002	102.4	0.001	71.9	0.002	96.7	0.002
		β_{tx}	99.8	0.002	95.6	0.002	101.2	0.002	89.4	0.003	101.4	0.002	84.6	0.003	100.1	0.003

ACCEPTED MANUSCRIPT

Table 9: Simulation results for the MI-GEE based on MNI and OIM methods (K = 7 - Well-balanced distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	104.3	0.274	92.3	0.250	96.2	0.242	92.5	0.268	96.5	0.254	96.6	0.288	76.1	0.258
		β_t	102.7	0.024	86.4	0.024	97.3	0.024	79.7	0.027	99.3	0.030	68.1	0.034	75.0	0.043
		β_{tx}	109.7	0.051	93.8	0.047	97.7	0.046	89.9	0.050	96.4	0.052	84.8	0.060	83.8	0.062
3	300	β_x	126.0	0.090	101.4	0.089	107.2	0.086	101.0	0.101	110.3	0.097	94.9	0.107	99.2	0.097
		β_t	102.4	0.008	83.2	0.009	95.1	0.009	72.4	0.010	96.6	0.010	62.0	0.012	84.6	0.015
		β_{tx}	107.7	0.017	94.5	0.017	99.0	0.017	89.4	0.020	98.7	0.021	83.1	0.022	92.8	0.023
3	500	β_x	124.9	0.050	102.3	0.053	108.7	0.052	95.1	0.059	104.3	0.057	90.9	0.063	96.7	0.058
		β_t	100.9	0.005	88.8	0.006	101.2	0.006	75.2	0.007	99.7	0.007	70.6	0.008	95.6	0.010
		β_{tx}	106.2	0.010	95.1	0.011	99.7	0.011	88.2	0.012	97.4	0.012	82.5	0.014	92.8	0.014
5	100	β_x	122.5	0.149	80.6	0.149	83.2	0.142	78.0	0.168	68.9	0.153	80.4	0.184	52.3	0.140
		β_t	100.8	0.005	66.5	0.006	93.4	0.006	71.5	0.007	81.5	0.008	64.8	0.008	65.1	0.012
		β_{tx}	100.2	0.011	79.0	0.010	99.3	0.010	87.3	0.012	91.7	0.013	81.5	0.015	79.0	0.014
5	300	β_x	126.2	0.049	97.7	0.048	105.2	0.046	97.8	0.055	102.2	0.050	93.6	0.063	86.0	0.055
		β_t	101.8	0.002	91.0	0.002	102.2	0.002	79.3	0.003	99.4	0.002	69.4	0.003	85.6	0.004
		β_{tx}	102.8	0.003	96.4	0.003	102.1	0.003	90.4	0.004	99.6	0.004	82.2	0.006	91.6	0.005
5	500	β_x	119.8	0.028	92.9	0.029	100.7	0.028	88.2	0.033	96.6	0.030	80.6	0.037	77.0	0.033
		β_t	101.1	0.001	88.8	0.001	99.9	0.001	76.4	0.002	98.5	0.001	66.5	0.003	86.0	0.002
		β_{tx}	102.2	0.002	94.7	0.002	100.3	0.002	87.6	0.003	98.4	0.002	79.0	0.004	89.7	0.003

Table 10: Simulation results for the MI-GEE based on MNI and OIM methods (K = 2 - Skewed distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	101.3	0.584	105.3	0.627	105.8	0.640	106.3	0.652	107.7	0.676	99.8	0.674	100.9	0.751
		β_t	94.0	0.053	97.9	0.052	109.1	0.054	75.8	0.060	107.6	0.064	47.0	0.082	92.9	0.094
		β_{tx}	99.8	0.116	105.2	0.123	105.9	0.129	105.8	0.133	108.1	0.142	96.1	0.148	96.4	0.196
3	300	β_x	99.4	0.175	98.7	0.174	99.3	0.175	97.1	0.178	98.4	0.188	96.8	0.172	99.0	0.187
		β_t	95.4	0.017	87.0	0.018	101.9	0.018	60.9	0.022	97.8	0.021	42.1	0.028	103.9	0.027
		β_{tx}	99.0	0.035	98.6	0.036	100.2	0.037	95.2	0.038	97.6	0.043	95.0	0.038	98.1	0.047
3	500	β_x	96.7	0.102	99.8	0.115	100.4	0.116	99.2	0.115	100.6	0.121	98.1	0.122	100.5	0.134
		β_t	89.1	0.010	89.2	0.011	101.9	0.011	66.3	0.014	103.8	0.014	42.4	0.019	105.6	0.017
		β_{tx}	94.9	0.021	99.4	0.023	100.5	0.024	98.5	0.025	101.1	0.028	97.1	0.028	100.6	0.035

Table 11: Simulation results for the MI-GEE based on MNI and OIM methods (K = 3 - Skewed distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	106.4	0.381	106.8	0.323	105.2	0.322	110.5	0.336	107.0	0.327	110.6	0.358	106.5	0.356
		β_t	111.4	0.041	143.6	0.038	119.6	0.039	209.6	0.069	154.5	0.076	253.8	0.075	178.4	0.090
		β_{tx}	110.6	0.079	113.1	0.065	110.1	0.067	120.7	0.069	114.9	0.076	119.1	0.075	115.0	0.090
3	300	β_x	102.7	0.115	103.1	0.113	101.8	0.114	105.8	0.115	102.9	0.112	108.2	0.115	103.1	0.111
		β_t	103.0	0.014	140.2	0.014	114.3	0.013	199.2	0.023	135.2	0.017	265.5	0.043	169.7	0.025
		β_{tx}	105.0	0.025	106.5	0.022	103.7	0.023	111.2	0.023	106.2	0.025	115.5	0.025	106.9	0.030
3	500	β_x	101.8	0.065	101.1	0.066	99.5	0.065	103.3	0.066	99.5	0.065	105.1	0.069	100.2	0.068
		β_t	102.3	0.008	132.7	0.008	105.2	0.008	193.4	0.017	126.3	0.010	254.7	0.033	156.3	0.015
		β_{tx}	102.8	0.014	104.0	0.013	100.8	0.014	108.4	0.014	101.2	0.016	112.4	0.015	104.1	0.018

ACCEPTED MANUSCRIPT

Table 12: Simulation results for the MI-GEE based on MNI and OIM methods (K = 4 - Skewed distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	101.5	0.301	107.1	0.352	103.9	0.338	111.7	0.383	102.9	0.357	114.1	0.367	99.9	0.324
		β_t	100.3	0.028	81.5	0.032	97.4	0.033	79.5	0.033	108.1	0.039	95.8	0.034	124.0	0.048
		β_{tx}	100.6	0.064	106.3	0.069	103.4	0.071	109.9	0.073	101.0	0.082	111.0	0.069	94.6	0.084
3	300	β_x	101.7	0.104	104.9	0.107	102.2	0.102	109.1	0.114	100.7	0.102	111.7	0.129	97.0	0.110
		β_t	101.5	0.010	83.3	0.011	102.0	0.011	76.5	0.011	108.4	0.012	84.2	0.012	112.4	0.017
		β_{tx}	101.2	0.022	104.7	0.020	102.7	0.021	108.1	0.021	99.9	0.023	108.9	0.024	90.8	0.030
3	500	β_x	101.3	0.065	104.6	0.064	101.9	0.061	108.6	0.075	100.7	0.067	110.5	0.079	96.2	0.069
		β_t	101.3	0.006	87.0	0.006	105.8	0.006	76.2	0.007	108.0	0.008	90.4	0.007	122.2	0.011
		β_{tx}	101.3	0.014	105.0	0.012	103.2	0.013	108.0	0.013	100.4	0.015	106.9	0.014	89.3	0.019

Table 13: Simulation results for the MI-GEE based on MNI and OIM methods (K = 5 - Skewed distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	99.7	0.279	106.6	0.324	103.4	0.313	111.8	0.357	102.7	0.325	113.8	0.433	96.1	0.401
		β_t	100.5	0.028	91.9	0.029	105.5	0.029	81.9	0.031	100.8	0.034	72.6	0.040	50.8	0.058
		$\beta_t x$	100.4	0.057	104.7	0.065	100.8	0.067	112.0	0.068	100.2	0.072	115.2	0.086	86.9	0.102
3	300	β_x	100.5	0.094	105.9	0.109	102.9	0.104	111.2	0.128	102.9	0.116	115.0	0.149	101.1	0.130
		β_t	99.0	0.009	86.0	0.009	101.5	0.009	73.0	0.011	94.5	0.012	69.5	0.014	67.0	0.018
		$\beta_t x$	100.5	0.019	104.9	0.022	101.9	0.023	110.9	0.025	101.0	0.027	116.2	0.030	96.0	0.034
3	500	β_x	99.5	0.056	105.0	0.068	102.2	0.064	111.0	0.080	103.1	0.069	114.7	0.097	101.2	0.081
		β_t	97.6	0.006	81.5	0.005	96.7	0.005	72.8	0.006	94.5	0.006	73.6	0.008	75.8	0.011
		$\beta_t x$	100.0	0.011	102.8	0.013	100.2	0.013	111.2	0.015	101.8	0.015	116.9	0.019	97.5	0.021

ACCEPTED MANUSCRIPT

Table 14: Simulation results for the MI-GEE based on MNI and OIM methods (K = 7 - Skewed distribution)

			0%		10%				30%				50%			
					MNI		OIM		MNI		OIM		MNI		OIM	
T	N	Parm	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE	RB(%)	MSE
3	100	β_x	101.0	0.291	102.4	0.290	99.8	0.277	109.7	0.342	101.8	0.303	115.2	0.384	103.5	0.323
		β_t	100.7	0.028	75.1	0.027	91.8	0.027	63.1	0.031	89.2	0.033	64.1	0.041	77.9	0.052
		β_{tx}	101.8	0.056	100.3	0.050	97.7	0.050	110.1	0.062	100.5	0.064	121.9	0.076	108.1	0.080
3	300	β_x	99.2	0.098	103.4	0.085	101.2	0.081	110.1	0.102	103.7	0.089	117.8	0.129	106.7	0.103
		β_t	102.5	0.009	80.3	0.010	97.5	0.009	74.7	0.011	106.7	0.011	81.5	0.013	108.6	0.017
		β_{tx}	100.0	0.018	103.5	0.016	101.6	0.017	114.3	0.020	107.1	0.020	127.3	0.027	113.2	0.027
3	500	β_x	100.1	0.059	103.1	0.048	100.9	0.046	110.1	0.059	103.2	0.050	117.1	0.081	106.2	0.058
		β_t	102.7	0.005	82.8	0.005	100.2	0.005	73.4	0.007	104.7	0.007	80.1	0.007	107.8	0.009
		β_{tx}	100.4	0.011	103.0	0.009	101.1	0.009	113.4	0.011	105.4	0.011	126.5	0.017	112.8	0.014

8.2 Selection of population parameters to generate missing data

The population parameters in (Eq. 9) were chosen using the following pragmatic way. First, rewrite the dropout probability model as follows,

$$\Pr(D_i = j | x_i, y_{i,(j-1)}) = \frac{e^{\psi_0 + \psi_x x_i + \psi_{prev} y_{i,(j-1)}}}{1 + e^{\psi_0 + \psi_x x_i + \psi_{prev} y_{i,(j-1)}}}.$$

Let us assume that the ordinal outcome Y has K categories and that their probabilities of occurring, $p_y(y)$, are known. Let us also assume that X has two categories and that we know their probabilities of occurrence $p_x(x)$. In line with our simulation plan, assume that these two occurrences are independent, so that $p(x, y) = p_x(x)p_y(y)$. Then, we chose parameter values for ψ_x and ψ_{prev}, leaving only ψ_0 unspecified. Let the proportion of missingness aimed for be π (e.g. 10%, 30%, or 50%), we then found the values for ψ_0 (by trial and error) that satisfied

$$\pi = \sum_x \sum_y p(x, y) \frac{e^{\psi_0 + \psi_x x_i + \psi_{prev} y_{i,(j-1)}}}{1 + e^{\psi_0 + \psi_x x_i + \psi_{prev} y_{i,(j-1)}}}.$$

ACCEPTED MANUSCRIPT

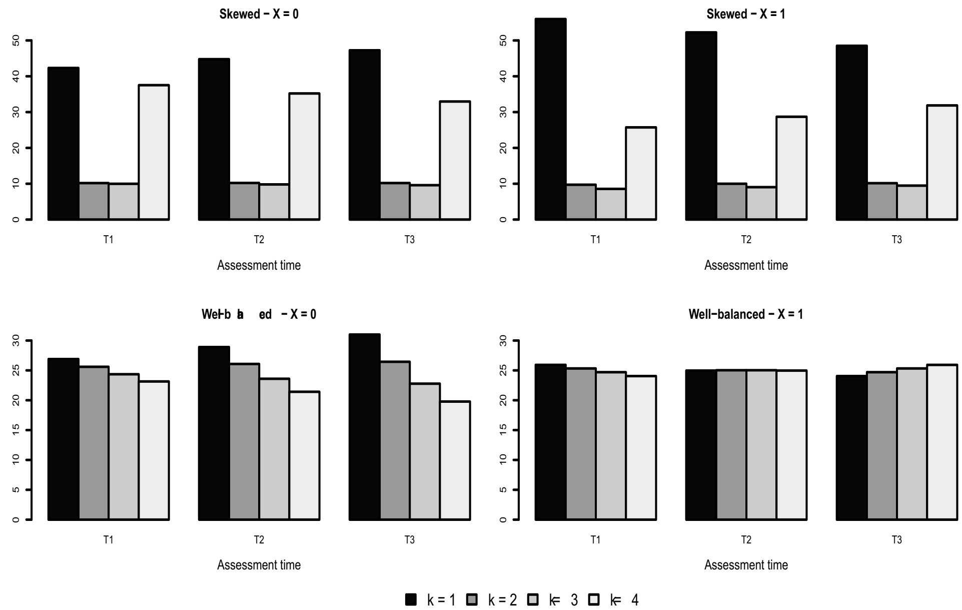

Figure 1: Distribution of the theoretical probabilities under well-balanced and skewed setting - K = 4 - T = 3

ACCEPTED MANUSCRIPT

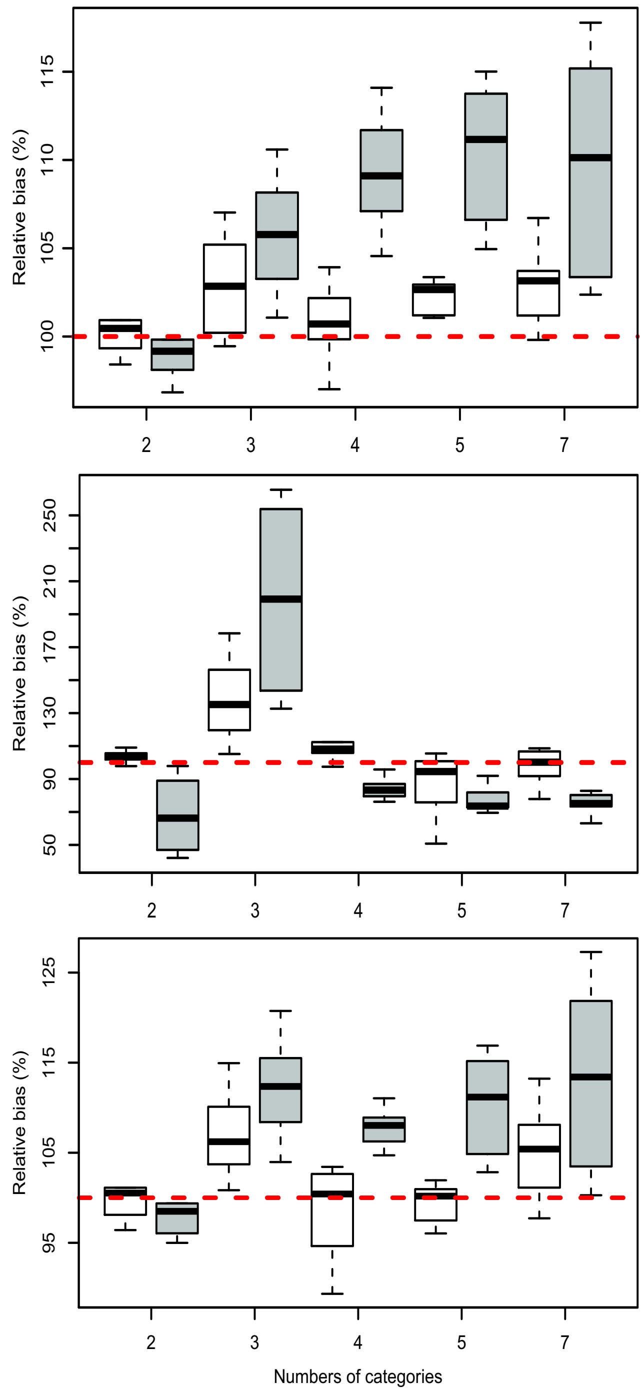

Figure 2: Relative bias (%) of the model parameters (top to bottom: $\beta_x, \beta_t, \beta_{tx}$) according to K the number of categories of the ordinal outcome (MNI= shaded boxplot - OIM=empty boxplot)

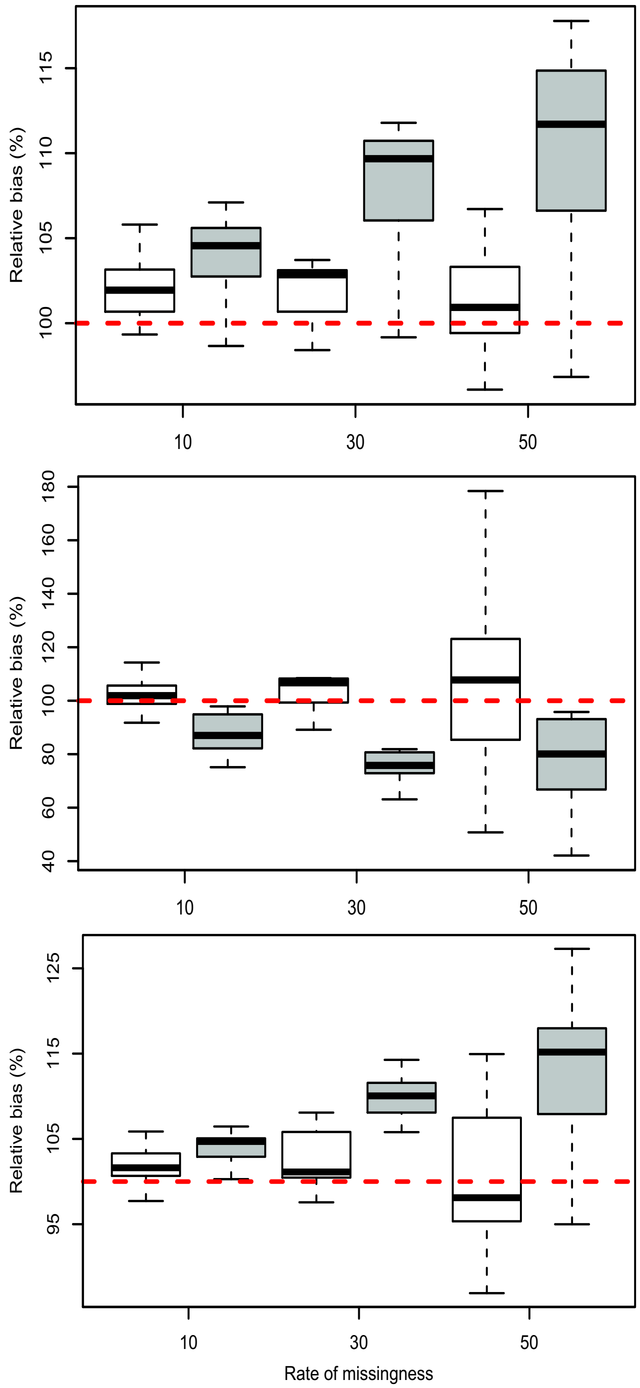

Figure 3: Relative bias (%) of the model parameters (top to bottom: β_x, β_t, β_{tx}) according to the rate of missingness (MNI= shaded boxplot - OIM=empty boxplot)