A Simulation Study Comparing Weighted Estimation Equations with Multiple Imputation Based Estimating Equations for Longitudinal Binary Data

Naresh Khatiwada

promotor :
Drs. Cristina SOTTO, Drs. Caroline BEUNCKENS

Eindverhandeling voorgedragen tot het bekomen van de graad
Master of Science in Biostatistics

universiteit
hasselt

University Hasselt
Center for Statistics

A Simulation Study Comparing Weighted Estimating Equations with Multiple Imputation Based Estimating Equations for Longitudinal Binary Data

By
Naresh Khatiwada

Supervisors: Carolien Beunckens
Cristina Sotto

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics.

2006-2007

University Hasselt
Center for Statistics

A Simulation Study Comparing Weighted Estimating Equations with Multiple Imputation Based Estimating Equations for Longitudinal Binary Data

By

Naresh Khatiwada

Accepted by:

Supervisor: Carolien Beunckens

Signature ………………………..

Date ………………………………

Supervisor: Cristina Sotto

Signature … ……………………..

Date ………………………………

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics

2006-2007

Acknowledgement

I would like to acknowledge my supervisor Cristina Sotto, Hasselt University, Belgium for her extensive and continuous guidance and support during the preparation of this report.

I would also like to thank my supervisor, Carolien Beunckens, Hasselt University, Belgium for her support and help.

I would be remiss in my duty if I did not thank Dr. Peter De Valk, Professor, Institute of Social Studies, The Hague; The Netherlands for his support and encouragement, Mrs. Martine Michaels for her administrative and secretariat support, VLIR for scholarship and Mrs. Tapasya Aryal for her forever support, encouragement and good wishes….

'JAY JAY JAY NEPAL'

Tables of Contents

Lists of Tables

Table in Appendix

List of Figures

Figure in Appendix

Abstract

Introduction: Missingness has become an inevitable scenario in longitudinal data and this often complicates the proposed analyses. Different methods have been proposed, among which are likelihood-based methods where analysts rest assured that the missingness is taken care of, when, for instance, Linear, Generalized Linear, or Non-Linear Models are considered due to their validity under the missing at random assumption. Statistical methods which take no notice of the mechanism for dropout will show the way to biased inference. Likelihood-based methods have computational complexity when taking into consideration longitudinal binary data. Weighted Generalized Estimating Equations (WGEE) is one of the common methods for handling dropouts that is MAR and is more usually used in marginal models for discrete longitudinal data. Alternatively, multiple imputations can be used to pre-process incomplete data, after which standard GEE is applied (MI-GEE).

Objective: The objective of this thesis was to compare weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data.

Method: In this study, both approaches WGEE and MI-GEE were compared for incomplete binary data, through so-called asymptotic simulation study as well as small-sample simulation. Bias, variances and mean square error (MSE) were the bases for the comparison between those two approaches.

Results and conclusion: - The results provide evidence for the fact that MI-GEE is less biased and more accurate in small and moderate samples sizes, while WGEE is asymptotically unbiased and has only shown better performance for data having more percentage of dropout.

Key words: Multiple-Imputation, Inverse Probability Weighting, asymptotic simulation,

1. Introduction

Missing or incomplete data are a common scenario occurring in many studies. It is very common for sets of quantitative data to be incomplete, in the sense that not all planned observations are actually made. This is especially true when studies are conducted on human subjects. An observation is considered an incomplete case if the value of any of the variables is missing. Even with the best design and monitoring, the observations can be incomplete usually due to the following possible reasons: missing by design, censoring and drop-out, or non-response, etc.

Mostly in longitudinal studies, there is a large amount of missingness either due to patient dropout or intermittent missed visits. The first one is monotone missingness, in which a study subject completely drops out of the study once they become missing for the first time. A second type is an intermittent missing pattern, in which a study subject can drop out of the study one time and then resume in the study at a later visit. This study will focus on the first type of missingness.

Most statistical packages exclude incomplete cases from analysis by default. This approach is easy to implement but has serious problems. Firstly, the loss of any information on incomplete cases may lower the desired efficiency in the study. Secondly, such exclusions may lead to substantial biases in analyses. Thus, missing data are important to consider in the analyses.

Rubin (1976) provided a recognized frame for the field of incomplete data by introducing the important categorization of missing data mechanisms, consisting of *missing completely at random* (MCAR), *missing at random* (MAR), and *missing not at random* (MNAR). A non-response process is said to be *missing completely at random* (MCAR) if the missingness is independent of both unobserved and observed data. A non-response process is said to be *missing at random* (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements. Thus, MAR mechanism depends on the observed outcomes and perhaps also on the covariates, but not further on unobserved measurements. A process that is neither MCAR nor MAR is termed nonrandom (MNAR).When an MNAR mechanism is operating, missingness depends on the unobserved measurements, perhaps in addition to dependencies on covariates and/or on observed outcomes. Thus, there is no simplification of the joint distribution.

Missingness frequently complicates the analysis of longitudinal data. In many clinical trials and other settings, the standard methodology used to analyze incomplete longitudinal data is based on such methods as *complete case analysi*s (CC), *last observation carried forward method* (LOCF) or simple form of imputation. This is often done without questioning the possible influence of these assumptions on the final results (Molenberghs and Verbeke 2005).

Sophisticated and theoretically robust methods for handling missing data in statistical analysis have existed for many years. A popular solution for dealing with incomplete longitudinal data is the use of likelihood-based methods like linear, generalized linear, or non-linear mixed models due to their validity under the assumption of *missing at random*

(MAR). Similarly, non-likelihood methods like generalized estimating equations (GEE) require the assumption of *missing completely at random* (MCAR). Thus, Weighted GEE (WGEE) has been proposed as a way to make certain validity under MAR. This method takes into account the missingness by using inverse probability weighting for the analysis of the incomplete sequences under the MAR assumption. An alternative to the WGEE approach uses Multiple Imputation (MI). In multiple imputations, several augmented data sets are generated by random replacement of missing values with samples from appropriate distributions in order to obtain more stable estimates of the parameters of interest. The combination of MAR-based MI together with a final GEE is called MI-GEE.

Although, Inverse probability weighting and imputation based estimating equations are valid under MAR assumption, the ways of dealing missingness in the data for these two methods are pretty dissimilar. Thus, one of the concentrations would be the strength of these methods under various circumstances. In other words, the pros and cons of these methods would be the primary interest of the researcher. Several researchers have started work towards this end in various ways. A most recent work was done by Beunckens et al (2007) involving a simulation study comparing weighted estimating equations with multiple imputations based estimating equations for longitudinal binary data. They compared both approaches, inverse probability weighting and MI-based, GEE through the use of so-called asymptotic, as well as small-sample simulations, in a variety of correctly specified as well as incorrectly specified models. They carried out the study based on a sample of size 100. The results from their study provides striking evidence that MI-GEE is both less biased and more accurate in small to moderate sample, in spite of asymptotic unbiasedness property of WGEE.

The present study aims to provide extensions to the study carried out by Beunckens et al (2007). In this study, the sample size was allowed to vary from N=50 to 500 under the same settings of study of Beunckens et al (2007). Further an additional setting with more missing data was also explored.

1.1 Objectives of the study

The objective of this thesis is to compare weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. In this study, both approaches WGEE and MI-GEE will be compared for the incomplete binary data, through so-called asymptotic as well as small-sample simulations. Bias, variances and mean square error (MSE) will be the bases of the comparison of these two approaches. Misspecification of either dropout model, measurement model or imputation model will be investigated throughout the process, in line with and using the same settings of the study of Beunckens et al (2007).

1.2 Structure of the report

This report is set up in five Sections. Section one gives some background information about missing data in the longitudinal setting and the objectives of this study. In Section two, a brief descriptions of the methods used to analyze incomplete longitudinal binary data is presented. In Section three, brief description of asymptotic and small-sample simulation methods are given. Section four describes the result of the asymptotic and small-sample simulations, Section five describes the results of same study with some modifications, and finally, discussion and conclusion is placed in Section six.

2. Methods handling Non-Gaussian Longitudinal Data

2.1 Model Families

In non-Gaussian longitudinal studies, multiple assessments of the same subject at different time points are used and the within-subject responses are then correlated. This correlation must be accounted for by analysis methods appropriate to the data. Several models have been proposed for the analysis of such data. Most of them are extensions of the well-known logistic regression that is a particular case of generalized linear models with a logit link function. In general, in the non-Gaussian data setting, there are three model families to analyze the data and these are marginal, random effects and conditional models. In a marginal model, the entire response vector is modeled marginally on a set of covariates, the association structure is then typically captured via a set of association parameters, such as correlations, odds ratios, etc.

Marginal models, also called population-averaged models, represent a situation in which the parameters characterize the marginal probabilities of the entire set of outcomes, without conditioning on the other outcomes. In the full-likelihood marginal approach, the Bahadur model which has been proposed by Bahadur (1961), accounts for the association via marginal correlations. Molenberghs and Lesaffre (1994) and Lang and Agresti (1994) have proposed models that parameterize the association in terms of marginal odds ratios. Dale (1986) defined the bivariate global odds ratio model based on a bivariate Plackett distribution (Plackett 1965). Molenberghs and Lesaffre (1994, 1999) extended this model to multivariate ordinal outcomes. However, the main issue for full likelihood approaches is computational complexity, especially when a high-dimensional vector of correlated data arises.

As an alternative method, Liang and Zeger(1986) proposed generalized estimating equations (GEE),which require only the correct specification of the univariate marginal distributions provided one is willing to adopt "working" assumptions about the association structure. Second-order GEE, which extends the GEE approach by correct specification of the association structure in addition to correct specification of the univariate marginal distribution, have been proposed also. An alternative to GEE is given by alternating logistic regressions (Carey, Zeger, and Diggle 1993). Le Cessie and van Houseline (1994) suggested approximating the true likelihood by means of a pseudo-likelihood (PL) function that is easier to evaluate and to maximize. Both GEE and PL yield consistent and asymptotically normal estimators, with an empirically-corrected variance estimator.

In conditionally specified models, any response within the sequence of repeated measurements is modeled conditional upon (subsets of) the other outcomes .This could be the set of all past measurements or a subset thereof, in so-called transition models .The third type of model is subject-specific models, in which the responses are assumed independent, given a collection of subject-specific parameters.

2.2 Some Marginal Models for Longitudinal Binary Data

I.Bahadur Model

As explained above, Bahadur (1961) has proposed this model for binary data and can be introduced using the simpler regression notation. Let us assume a sequence of binary measurements Y_{ij} designed to be measured at occasions j for individual i under investigation. Let π_{ij} be the marginal probability, i.e., $E(Y_{ij}) = P(Y_{ij} = 1) = \pi_{ij}$. To describe the association, the pairwise probabilities $E(Y_{ij_1} Y_{ij_2}) = P(Y_{ij_1} = 1, Y_{ij_2} = 2) = \pi_{ij_1 j_2}$ need to be characterized. The success probability of the measurements taken in the same subject can be modeled in terms of the two marginal probabilities π_{ij_1} and π_{ij_2}, as well as an association parameter, this being the marginal coefficient of Bahadur's model.

The marginal correlation coefficient assumes the form

$$Corr(Y_{ij_1}, Y_{ij_2}) = \rho_{ij_1 j_2} = \frac{\pi_{ij_1 j_2} - \pi_{ij_1}\pi_{ij_2}}{\left[\pi_{ij_1}(1-\pi_{ij_1})\pi_{ij_2}(1-\pi_{ij_2})\right]^{1/2}}$$

Based on this expression, the joint probability can be calculated from the marginal correlation coefficient and the univariate probabilities and can be defined as

$$\pi_{ij_1 j_2} = \pi_{ij_1}\pi_{ij_2} + \rho_{ij_1 j_2}\left[\pi_{ij_1}(1-\pi_{ij_1})\pi_{ij_2}(1-\pi_{ij_2})\right]^{1/2}$$

The likelihood-based approach requires complete representation of the joint probabilities of the vector of binary responses in each unit. Thus, Bahadur used, apart from conventional two-way correlation coefficient, third and higher-order correlation coefficients to completely specify the joint distribution. For this, let us define standardized deviations

$$\varepsilon_{ij} = \frac{Y_{ij} - \pi_{ij}}{\left[\pi_{ij}(1-\pi_{ij})\right]^{1/2}} \text{ and } e_{ij} = \frac{y_{ij} - \pi_{ij}}{\left[\pi_{ij}(1-\pi_{ij})\right]^{1/2}},$$

where y_{ij} is an actual value of the binary response variable Y_{ij}.

Let us define

$$\rho_{ij_1 j_2} = E(\varepsilon_{ij_1}\varepsilon_{ij_2}),\ \rho_{ij_1 j_2 j_3} = E(\varepsilon_{ij_1}\varepsilon_{ij_2}\varepsilon_{ij_3}) \ldots \rho_{ij_1 j_2 \ldots\ldots j_{ni}} = E(\varepsilon_{ij_1}\varepsilon_{ij_2}\varepsilon_{ij_3}\ldots\ldots\ldots.\varepsilon_{ij_{ni}}),$$

where the parameters ρ_{ijk} are classical Pearson type correlation coefficients.

In Bahadur's model, the probability mass function is split into two components: the independence model and the correction factor. Thus, the general Bahadur model can be represented by the expression $f(\mathbf{y}_i) = f_I(\mathbf{y}_i)c(\mathbf{y}_i)$, (1)

where

$$f_I(y_i) = \prod_{j=1}^{n_i} \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}} \text{ and}$$

$$c(y_i) = 1 + \sum_{j_1<j_2}\rho_{ij_1 j_2} e_{ij_1} e_{ij_2} + \sum_{j_1<j_2<j_3}\rho_{ij_1 j_2 j_3} e_{ij_1} e_{ij_2} e_{ij_3} + \ldots\ldots\ldots\ldots + \rho_{i12\ldots\ldots n_i} e_{ij_1} e_{ij_2} e_{ij_3} \ldots\ldots\ldots\ldots e_{ij_{ni}}.$$

It can be seen that the Bahadur model has closed form but its use for non-Gaussian outcomes can be problematic due to prohibitive computational requirements.

II.Generalized Estimating Equations (GEEs)

As discussed in Section 2.1, the use of full likelihood methods for marginal models can be problematic due to prohibitive computational requirements. Therefore, GEE is a viable alternative within this family. For clustered and repeated data, Liang and Zeger (1986) proposed the Generalized Estimating Equations (GEEs). GEE is a marginal or population-averaged model for clustered and repeated data. It is a non-likelihood method, correcting for the clustering effect, using correlation to capture the association within the clusters. It focuses on modeling the mean structure:

$\log it(\pi_{ij}) = X'_{ij}\beta$ where β is the vector of model parameters.

For the classical form of GEE, the score equations to be solved when computing maximum likelihood estimates under a marginal non-Gaussian outcome are

$$S(\beta) = \sum \frac{\partial \boldsymbol{\mu}_i}{\partial \beta'} (A_i^{1/2} C_i A_i^{1/2})^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \qquad (2)$$

where A_i is the matrix with the marginal variances on the main diagonal and zeros elsewhere, and C_i is equal to the marginal correlation matrix. Typically, the correlation matrix C_i contains a vector α of unknown parameters that is replaced for practical purposes by a consistent estimate.

Assuming that the marginal mean $\boldsymbol{\mu}_i$ has been correctly specified as $h(\boldsymbol{\mu}_i) = X'_i\beta$, it can be shown that, under mild regularity conditions, the estimator $\hat{\beta}$, obtained by solving the estimating equations, is asymptotically normally distributed with mean β and with covariance matrix (Molenberghs and Verbeke, 2005):

$$\text{Var}(\hat{\beta}) = \text{I}_0^{-1}\text{I}_1\text{I}_0^{-1},$$

where $\text{I}_0 = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}'_i}{\partial \beta} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \beta'}$ and $\text{I}_1 = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}'_i}{\partial \beta} V_i^{-1} \text{var}(\mathbf{y}_i) V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \beta'}$, with $V_i = A_i^{1/2} C_i A_i^{1/2}$.

To investigate the effect of clustering, different working assumptions can be specified in the GEE model. GEE is interested in getting good estimates of the main effect. Based on Liang and Zeger (1986)'s results, GEE yields consistent main effect estimators even when the working correlation structure is misspecified. In this case, correlation parameters are treated as nuisance parameters.

However, severe misspecification may affect the efficiency of the estimators. Hence, distances between empirically-corrected and model-based standard errors are compared to choose a working assumption that is close to the true correlation structure. Since GEE allows for the misspecification of the correlation structure, the association obtained in terms of correlation cannot be trusted and is considered a nuisance, though the main effect estimates are consistent. As such, if the association is of scientific interest, GEE is less adequate.

2.3 Conditional Models for Longitudinal Binary Data

As explained in Section 2.1, in conditionally specified models, any response within the sequence of repeated measurements is modeled conditional upon the other outcomes on the same unit. A transition model is a well-known example of a conditionally specified model for longitudinal data.

In a transition model, a measurement Y_{ij} in the longitudinal series can be explained as the function of the previous outcome or history. In other words, a regression model for the outcome Y_{ij} can be written in terms of the previous outcome or history. Alternately, error terms can also be written as functions of previous error terms.

A form of a transition model for binary longitudinal outcomes is a stationary first-order autoregressive model and can be written as

$$\operatorname{logit}[P(Y_{ij}=1 \mid x_{ij}, Y_{i,j-1}=y_{i,j-1}, \boldsymbol{\beta}, \alpha)] = x'_{ij}\boldsymbol{\beta} + \alpha y_{i,j-1}. \qquad (3)$$

Second or higher-order extensions can be possible.

2.4 Missing Data Frameworks

In a longitudinal data setting, there are usually two types of missing patterns involved. The first one is monotone missingness, in which a study subject completely drops out of the study once they become missing for the first time. A second type is an intermittent missing pattern, in which a study subject can drop out of the study one time and then resume in the study at a later visit. This study only focuses on the first type of missingness pattern.

Let us assume a sequence of measurements Y_{ij} is designed to be measured on subject where $i = 1,2,\ldots\ldots\ldots N$ at occasions $j = 1,2,\ldots\ldots\ldots n_i$, which gives the outcome vector $\mathbf{Y}_i = (Y_{i1}, Y_{i2} \ldots\ldots Y_{in_i})'$.

Let us define R_{ij}, the missing data indicator, as:

$$R_{ij} = \begin{cases} 1, \text{if } Y_{ij} \text{ is observed} \\ 0, \text{otherwise} \end{cases} \quad \text{and} \quad \mathbf{Y}_i = \begin{cases} \mathbf{Y}_i^o, \text{ vector containing } Y_{ij} \text{ where } R_{ij}=1 \\ \mathbf{Y}_i^m, \text{ vector containing } Y_{ij} \text{ where } R_{ij}=0 \end{cases}$$

$\mathbf{Y}_i$ is referred to as the complete data which is made up of the observed $(\mathbf{Y}_i^o)$ and the missing data $(\mathbf{Y}_i^m)$ components. The full data $(\mathbf{Y}_i, \mathbf{R}_i)$ consists of the complete data, together with the missing data indicators, in which the complete data refers to the vector $\mathbf{Y}_i$ of planned measurements. For the dropout pattern of the missingness, the vector $\mathbf{R}_i$ vector can be replaced by a scalar variable D_i, called the dropout indicator. In this case each vector $\mathbf{R}_i$ is of the form (1,…0,…0) and we can write this scalar dropout indicator by the expression $D_i = 1 + \sum_{j=1}^{n_i} R_{ij}$, and for a complete sequence $D_i = n_i + 1$. In both the situations, D_i

is equal to 1 plus the length of the observed measurement sequences. For the dropout pattern, a balanced design which indicates the common set of measurement occasions is more relevant to provide meaningful definitions.

The joint model of the full data fall into three categories: the selection model, pattern-mixture model and the shared-parameter model. The selection model (Rubin, 1976; Little and Rubin, 1987) factorizes the joint distribution into two factors. The first factor is the marginal density of the measurement process and second factor is the density of the dropout process, conditional on the measurement process. The term originates from the econometric literature (Heckman 1976) and it can be thought of in terms of a subject's missing values being 'selected' through the probability model, given the measurements, whether observed or not(Molenberghs and Kenward,2007).The reverse factorization yields a pattern-mixture model(Little,1993,1994). The pattern-mixture model allows for a different response model for each pattern of missing values, the observed data being a mixture of these weighted by the probability of each missing value or dropout pattern. The third family is called shared-parameter models (Wu and Carrol, 1988; Wu and Bailey, 1989), where the measurement and dropout processes are assumed to be independent, given a certain set of shared parameters. The natural parameters of these three models have different interpretations, and transforming one statistical model from one of the frameworks to another is generally not straightforward (Molenberghs and Kenward, 2007). In this report, the common approach, the selection model, is taken into account.

2.5 Issues of Missingness in Non-Gaussian Data Setting

Missing data occur regularly in longitudinal studies. Subjects may drop out before the study terminates, or be lost to follow-up in such a way that no further measurements are provided after the time of dropout. Statistical methods which ignore the mechanism for dropout will lead to biased inference. Generally, which method is to be considered for handling incomplete data depends on which type of dropout mechanism it is. The focus in this report is on dropouts that are *missing at random*, i.e., the probability of dropout is related to the observed responses. As discussed in the previous Sections, the full-likelihood approaches are attractive due to their ignorability properties. Such approaches are valid under *missing at random* (MAR). But, the use of such approaches in the non-Gaussian setting leads to computational complexity, especially when a high-dimensional vector of correlated data arises. Thus, GEE has been proposed to overcome such complexity. However, as Liang and Zeger (1986) pointed out, inferences obtained using GEE are valid only under the strong assumption that the data are *missing completely at random* (MCAR) (Molenberghs and Verbeke, 2005). To allow the data to be MAR, Robins et al (1995) developed a class of weighted estimating equations which can seen as the extension of the GEE. Another alternative approach can be to combine multiple imputations, where the MAR assumption is deemed plausible, with GEE, resulting in MI-GEE. In the next Section, brief explanations of these two approaches are provided.

2.6 Methods for Incomplete Non-Gaussian Longitudinal Data

I.Weighted Generalized Estimating Equations (WGEE)

The generalized estimating equations (GEE) approach is commonly used to model incomplete longitudinal binary data. When dropouts are *missing at random* (MAR) through dependence on observed responses, GEE may give biased parameter estimates in the model for the marginal means. A weighted estimating equations approach (Robins, Rotnitzky, and Zhao, 1994), gives consistent estimation under MAR when the drop out mechanism is correctly specified. This can be viewed as an extension of GEE. The idea of weighted GEE is to weight each subject's measurements in the GEE by the inverse probability that subject drops out at that particular measurement occasion (Molenberghs and Verbeke, 2005). Such assigned weight can be calculated as

$$\upsilon_{ij} = P(D_i = j) = \prod_{k=2}^{j-1}[1 - P(R_{ik} = 0 \mid R_{i2} = \ldots = R_{i,k-1} = 1)] \times P(R_{ij} = 0 \mid R_{i2} = \ldots R_{i,j-1} = 1)^{I\{j \le n_i\}}.$$

The underlying idea behind this inverse probability weights (IPW) methodology is that each available observation at a particular occasion is given a weight that is inversely proportional to the cumulative probability of being observed at that time. In other words, the application of the inverse probability weights in a marginal model is to correct the bias that is caused by dropouts that are MAR.

Based on the above probability, the estimating equations are adjusted by adding the inverse of the weighted term to the score equations of standard GEE, which can now be written as:

$$\mathbf{S}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{d=2}^{n_i+1} \frac{I(D_i = d)}{\nu_{id}} \frac{\partial \boldsymbol{\mu}_i}{\partial \beta'}(d)(A_i^{1/2} C_i A_i^{1/2})^{-1}(d)(\mathbf{y}_i(d) - \boldsymbol{\mu}_i(d)) = \mathbf{0},$$

where $\mathbf{y}_\mathbf{i}(d)$ and $\boldsymbol{\mu}_i(d)$ are the first d-1 elements of $\mathbf{y}_\mathbf{i}$ and $\boldsymbol{\mu}_i$ respectively. The $\frac{\partial \boldsymbol{\mu}_i}{\partial \beta'}(d)$ and $(A_i^{1/2} C_i A_i^{1/2})^{-1}(d)$ terms are defined analogously, in line with the definitions in Robins et al (1995).

II.Multiple Imputation Techniques

Among the series of criticism on simple imputation techniques such as LOCF, at best, is that they ignore random variation by imputing fixed values. Multiple imputation methods proposed by Rubin (1987) is a technique to replace missing values with a set of M plausible values, that is, values generated from the distribution of one's data. This, combined with GEE on the computed data, is an alternative technique to direct likelihood and weighted GEE and, at least in its basic form, requires the missingness mechanism to be MAR. In multiple imputations, the analyst creates several different versions of a data set, replacing missing values with plausible random values, or imputations. The imputed data sets are analyzed separately, and the results are combined in a way that accounts for variation in the imputed values. More precisely, the multiple imputation technique has three basic phases:

- the missing values are filled in M times to obtained M complete data sets;
- the M complete data sets are analyzed by using standard procedures; and,
- the results from M analyses are combined into a single inference.

The process of multiple imputations is as follows. Suppose the interest is on the inferences about the $k \times 1$ parameter vector $\boldsymbol{\beta}$ from the substantial model and that one is able to make appropriate Bayesian posterior draws from the imputation model. Replacing the missing data by their corresponding imputation samples, M imputed data sets are constructed. Let $\hat{\boldsymbol{\beta}}^m$ and $\hat{V}^m$ be the estimate of $\boldsymbol{\beta}$ and its covariance matrix from m^{th} completed dataset $(\mathrm{m} = 1,2,\ldots\ldots\ldots \mathrm{M})$, respectively. Then, the MI estimate of $\boldsymbol{\beta}$ is the mean of the estimates

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

The estimates of the covariance matrix of $\hat{\boldsymbol{\beta}}*$ are calculated from the within and between imputation variability, i.e. ,

$$V = W + (\frac{M+1}{M})B,$$

where $W = \frac{1}{M}\sum_{m=1}^{M} V^m$ is the within-imputation variability, and

$B = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\beta}}^m - \hat{\boldsymbol{\beta}}^*)(\hat{\boldsymbol{\beta}}^m - \hat{\boldsymbol{\beta}}^*)'$ is the between-imputation variability.

Researchers often wish to know how many imputations are needed for each missing value. The established advice, however, is that 2 to 10 imputations suffice under the most realistic circumstances (Rubin, 1987). In many applications, as few as 3-5 imputations are sufficient to obtain excellent results (Molenberghs and Kenwards, 2007).

3. Design of Simulation Study

As explained in Section 2, many standard analyses, including maximum likelihood estimation and the GEE approach, can result in biased estimation when there is missingness in the dataset. In such situations, WGEE is a useful tool to analyze incomplete longitudinal non-Gaussian data under MAR.WGEE is unbiased for correctly specified dropout and mean structure of the measurement model. An alternative approach to WGEE is to combine MAR-based multiple imputation techniques together with final GEE analysis for the model and such combination is termed as MI-GEE. In this case, a correctly specified imputation model and estimation model is required. The main focus of this report is to compare these approaches: WGEE and MI-GEE. For such comparison, the interest can be the magnitude of bias incurred under these mechanisms and various types of misspecification can play a vital role. The comparison is done through the use of an asymptotic simulation study and small-sample simulations on various underlying data generating models. This process is split into two stages, namely the data generation stage and the analysis stage.

3.1 Data Generation

In this stage, data generating models are introduced. As described in Section 2.4, the selection model framework is considered, thus the generating model has two components a measurement model and a dropout model given the measurement model. Since the two methods WGEE and MI-GEE are based on a GEE analysis, it is required to generate data from a fully-specified model for the measurement and dropout mechanisms. In this simulation study, a binary outcome with three time points was considered. Also, a grouping variable with two levels, say treatment versus placebo, represented by a binary indicator, was also considered. For the outcomes, a binary outcome at three time points was generated from a Bahadur measurement model and also from a second-order autoregressive AR (2), transition model. In addition, continuous outcomes were also generated from the trivariate Gaussian distribution and then later dichotomized. In this situation, three sets of the data were generated using different measurement models. For the dropout, MAR mechanisms were considered. Thus, the results are three data-generating models, which will afterward be called GMI (Bahadur Measurement and MAR dropout model), GMII (AR (2) measurement model and MAR dropout model) and GMIII (Gaussian Measurement Model and MAR dropout model). The procedures used are explained in the following Sections.

I. GMI (Bahadur Measurement and MAR Dropout Model)

Let us define Y_{ij} to be the measurement of individual i at time point t_j. Let x_i be the treatment indicator having two factor levels, say 0 and 1. Since GMI is based on the Bahadur measurement model, the measurement model for GMI can be written as:

$$\operatorname{logit}(\pi_{ij}) = \operatorname{logit}[P(Y_{ij}=1 \mid x_i, t_j)] = \beta_0+\beta_x x_i+\beta_t t_j+\beta_{xt} x_i t_j, \qquad (4)$$

For this model the following parameters were chosen:
$\beta_0=-0.25, \beta_x=0.5, \beta_t=0.2,$ and $\beta_{xt}=-0.8,$ with two-way and three-way correlation coefficients $\rho_{ij_1j_2}=0.2$ and $\rho_{ij_1j_2j_3}=0,$ respectively. In this case, the missingness process is

assumed to be MAR, thus the probability of dropout at time point t_j given x_i and the measurement at the previous time point is modeled by a logistic regression and written as:

$$\operatorname{logit}[P(D_i = 1 | x_i, y_{i,j-1})] = \psi_0 + \psi_x x_i + \psi_{prev} y_{i,j-1}, \qquad (5)$$

where j=2, 3, 4 with parameters $\psi_0 = -0.5, \psi_x = -0.6$ and $\psi_{prev} = -3.5$.

Combining these two models, Bahadur measurement and MAR dropout model, yields GMI. For GMI, the missingness proportions are 68% for completers, 18 % with only first outcome observed (10% for x=0 and 8% for x=1), and 15 % of the cases the last observation missing (7% for x=0 and 8% for x=1).

II. GMII (AR (2) Transition Model and MAR Dropout Model)

In GMII, a second-order auto-regressive transition model was considered. In line with the terminology used above, the AR (2) transition model is written as

$$\begin{aligned} &P(x_i) = \mu_x \\ &\operatorname{logit}[P(Y_{i1} = 1 | x_i)] = \alpha_0 + \alpha_x x_i \\ &\operatorname{logit}[P(Y_{i2} = 1 | x_i, y_{i1})] = \phi_0 + \phi_x x_i + \phi_1 y_{i1} \\ &\operatorname{logit}[P(Y_{i3} = 1 | x_i, y_{i1}, y_{i2})] = \gamma_0 + \gamma_x x_i + \gamma_1 y_{i1} + + \gamma_2 y_{i2}, \end{aligned} \qquad (6)$$

For this model the following parameters were chosen $\mu_x = 0.5, \alpha_0 = -0.2, \alpha_x = 0.3, \phi_0 = -0.1, \phi_x = 0.5, \phi_1 = 0.7, \gamma_0 = -0.25, \gamma_x = 0.35, \gamma_1 = 0.35, \gamma_1 = 0.4$ and $\gamma_2 = 0.6$. The same MAR dropout model from equation (5), which was applied to GMI, was considered in this case also. Combining these two models, AR (2) transition and MAR dropout model, yields GMII. For GMII, the missingness proportions are 73% for completers, 17 % with only first outcome observed (11% for x=0 and 6% for x=1), and 11 % of the cases the last observation missing (7% for x=0 and 4% for x=1).

Since, AR (2) model is conditional rather than marginal, and the area of interest in this report is comparison of WGEE and MI-GEE, both marginal, the conditional model in (6) requires marginalization. This marginalization assumes that the corresponding underlying marginal model is of the form (4). Inasmuch as the underlying measurement model is in fact conditional, rather than marginal, there is no way to verify whether this assumed underlying marginal model is "true". For this, the marginalization has been done based on the AR (2) model from equation (6) to get marginal probabilities as follows

$$P(y_{i1}, y_{i2}, y_{i3}, x_i) = P(y_{i3} = 1 | x_i, y_{i1}, y_{i2}) P(y_{i2} = 1 | x_i, y_{i1}) P(y_{i1} = 1 | x_i) P(x_i). \qquad (7)$$

On a hypothetical data set consisting of all 16 possible combinations of the form $(y_{i1}, y_{i2}, y_{i3}, x_i)$, with corresponding probability weights $P(y_{i1}, y_{i2}, y_{i3}, x_i)$, a GEE model was fitted of form (4), the resulting marginalized "true" parameters for GMII were found to be $\beta_0 = -0.3658, \beta_x = 0.2673, \beta_x = 0.2265$ and $\beta_{xt} = 0.0790$.

III.GMIII (Gaussian Measurement Model and MAR Dropout Model)

For GMIII, the Gaussian measurement model is considered. A longitudinal Gaussian outcome with three time points and a two level covariate was generated from

$$\mu_{ij} = E(W_{ij}|x_i, t_j) = \eta_0 + \eta_x x_i + \eta_t t_j + \eta_{xt} x_i t_j, \qquad (8)$$

where i=0, 1 and j=1, 2, 3 with the parameters $\eta_0 = 3.5, \eta_x = 0, \eta_t = 1.75$, and $\eta_{xt} = 0.5$.

In other words, the mean vectors for each level of the treatment indicator x_i, say placebo and treatment, are, respectively, $\boldsymbol{\mu}_0 = (5.25, 7.00, 8.75)'$ and $\boldsymbol{\mu}_1 = (5.75, 8.00, 10.75)'$, with assumption of an unstructured covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.8 & 0.35 \\ 0.8 & 1 & 0.5 \\ 0.35 & 0.5 & 1 \end{pmatrix}.$$

In this case the missingness process is also assumed to be MAR, thus the probability of dropout is modeled by the following logistic regression

$$\operatorname{logit}[P(D_i = 1 | x_i, w_{i,j-1})] = \delta_0 + \delta_x x_i + \delta_{prev} w_{i,j-1}. \qquad (9)$$

For this model the following parameters were chosen as:
$\delta_0 = -0.15, \delta_x = 0.8, \gamma_1 = -0.35$.

Combining these two models, Gaussian measurement model and MAR dropout model, yields GMIII. For GMIII, on average, over all the 500 samples, the missingness proportions are 76% for completers, 17 % with only first outcome observed (7% for x=0 and 10% for x=1), and 7 % of the cases the last observation missing (3% for x=0 and 4% for x=1).

3.2 Simulation Study

A simulation is an imitation of some real thing, or a process. The act of simulating something generally entails representing certain key characteristics or behaviors of a selected physical or abstract system. In this study, asymptotic and small-sample simulations were considered.

I. Asymptotic Simulation

In an asymptotic simulation, first of all, a hypothetical dataset was created, which consists of all possible outcome sequences at each level of the covariates. As explained above, the focus here is on a binary outcome $\mathbf{y_i} = (y_{i1}, y_{i2}, y_{i3})'$ at three time points and a two-level covariate x_i. The hypothetical data set for one level of the covariate $(x_i = 0)$ is presented in Table 1.

Table 1: Hypothetical Dataset.

Obs	y_1	y_2	y_3	x_i
1	0	0	0	0
2	0	0	1	0
3	0	1	0	0
4	1	0	0	0
5	0	1	1	0
6	1	0	1	0
7	1	1	0	0
8	1	1	1	0

From Table 1, each level of the covariate has 8 possible sequences of the outcomes, which provide total of 16 possibilities in the setting. The probability mass $P(\mathbf{y}_i, x_i)$ for each of the sequences can be computed from the assumed measurement model. It is assumed that the dropout occurs after the first time point only. This indicates that the dropout can be in the second time point or in the third time point or no dropout (all complete). Thus, for each subject there are three possibilities to drop out which provides $16 \times 3 = 48$ possibilities. In other words, the probabilities $P(\mathbf{y}_i, x_i)$ are thus further split among the three missingness patterns according to the dropout probabilities. In particular, denoting by $P(D_i = 2 | D_i \geq 2)$, $P(D_i = 3 | D_i \geq 3)$ and $P(D_i = 4 | D_i \geq 4)$ the dropout probabilities at time points 2, 3 and 4 respectively, the probabilities for each of the 48 possible combinations are given as follows:

$$P(\mathbf{y}_i, x_i, D_i = 4 | D_i \geq 4) = P(\mathbf{y}_i, x_i) \prod_{j=2}^{4} [1 - P(D_i = j | D_i \geq j),$$

$$P(\mathbf{y}_i, x_i, D_i = 3 | D_i \geq 3) = P(\mathbf{y}_i, x_i) \prod_{j=2}^{3} [1 - P(D_i = j | D_i \geq j) P(D_i = 4 | D_i \geq 4), \text{ and}$$

$$P(\mathbf{y}_i, x_i, D_i = 2 | D_i \geq 2) = P(\mathbf{y}_i, x_i) [1 - P(D_i = 2 | D_i \geq 2) \prod_{j=3}^{4} P(D_i = j | D_i \geq j).$$

Asymptotic simulation was conducted on GMI and GMII based on these 48 possibilities, each weighted by the corresponding probabilities obtained from either from GMI or GMII. In this regard, the model is fitted to this hypothetical dataset with the application of probability weighting and the resulting inferences obtained are the asymptotic solutions.

II. Small-Sample Simulations

The second approach, which is called small-sample simulations, was considered. In this regard, S=500 samples of different sizes, N=50,100,200 and 500 subjects, were considered. A balanced design was chosen, which implies that equal numbers of subjects were taken for the two treatment groups. Three sets of data were simulated based on GMI, GMII, and GMIII, respectively. Samples generated from GMI and GMII are based on the underlying probabilities. In addition, for GMIII, S=500 samples were generated from trivariate Gaussian $N_3(\mu_0, \Sigma)$ and $N_3(\mu_1, \Sigma)$ for the two treatment groups, with equal size, and later dichotomized for analysis purposes. In this case, the underlying distribution for the outcome is continuous, which implies that the asymptotic simulation study is not possible. Thus, asymptotic simulations were conducted only for GMI and GMII, small-sample simulations were done for all three generation models.

3.3 Data Analysis and Model Misspecification

In the data analysis stage, MI-GEE and WGEE were applied on the data from the data generation stage. WGEE requires the specification of the marginal measurement model and a dropout model. For MI-Transition, specification of a conditional mean measurement model and a dropout model is required. In MI-GEE and MI-Transition case, the predictors of dropout are included in the imputation model. Based on the asymptotic simulation and small-sample simulations, in comparing MI-GEE and WGEE, different properties can be taken into account. One of the properties is bias, which is the difference between an estimator's expected value and the true value of the parameter being estimated. i.e., $Bias(\hat{\beta}) = E(\hat{\beta}) - \beta$. An estimator having nonzero bias is said to be biased.

In the asymptotic approach, as described above, the probability weights, which are computed from the respective generating model, are used to solve the estimating equations. The resulting inferences are asymptotic and the asymptotic bias ($Bias_{\infty}$) and the asymptotic variances (Var_{∞}) for the parameter estimators can be computed based on that.

In the small-sample simulation approach, where all GMI GMII and GMIII can be considered, the parameter estimates, their estimated and true variances for a sample of size N, and their MSE are calculated as:

$$\overline{\text{EST}} \equiv \overline{\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{S} \frac{\hat{\boldsymbol{\beta}}_i}{S}$$

$$\text{True Variance} \equiv \text{Var}_N = \text{Var}_N(\overline{\text{EST}}) = \frac{\text{Var}_{\infty}}{N},$$

$$\text{Estimated Variance} \equiv \hat{Var}_N = \hat{\text{V}}\text{ar}_N(\overline{\text{EST}}) = \sum_{i=1}^{S} \frac{(\hat{\boldsymbol{\beta}}_i - \overline{\hat{\boldsymbol{\beta}}})^2}{S-1}, \text{ and}$$

$$\text{Mean Square Error(MSE)} \equiv \text{MSE}(\overline{\text{EST}}) = \text{Bias}_N^2(\overline{EST}) + \hat{\text{V}}\text{ar}_N(\overline{\text{EST}}).$$

Based on these properties, MI-GEE and WGEE can be compared under the different data generating models GMI, GMII and GMIII. The behaviors of both methods in terms of bias and MSE can be studied under correctly specified and misspecified models. Due to this process, robustness of both models under misspecification of either the imputation model or dropout model or measurement model can be investigated.

In WGEE, all subjects are given weights, which are calculated based on the hypothesized dropout model. Thus any misspecification of this dropout model will affect all subjects and also the final result. In MI-GEE or MI-Transition, any misspecification in the imputation model affects only on the unobserved part of the data, not the observed part of the data. In addition, misspecification of the measurement model can also be assumed.

GMI is derived from the Bahadur measurement model associated with MAR dropout model, which is based on a logistic regression model where dropout depends on the previous outcome, as well as the treatment indicator. In addition, GEE methods are moment-based version of Bahadur model, thus, a GEE-based version with the same structure as that of the underlying measurement model would be suitable. To deal with the MAR nature of missingness, the GEE-based approach is supplemented with a weighting format, achieved from a model of the same form as that of the underlying dropout model, resulting now in WGEE. WGEE, using weights from a logistic dropout model with previous outcome and treatment indicator as covariates, would therefore yield a correctly-specified analysis for GMI. The important note is, under WGEE, the imputation model is not necessary, since the missingness is tackled, not by imputation, but, by means of a dropout model.

As explained in the previous Section, GMII is based on the AR (2) transition model for the mean structure and a conditional logistic model from the dropout. In this situation, fitting an AR (2) analysis model after multiple imputations, called MI-Transition, is consistent with the underlying data generating model. In other words, both measurement and the imputation model are correctly specified. In this situation, a dropout model is not relevant since the missingness is tackled, not by dropout weights, but, by means of imputation. In another scenario, a comparison of the two methods is done under the marginalized version of GMII. It can be recalled that marginalization was done based on (7), and afterward, MI-GEE and WGEE can be fitted and compared. In this case, the conditional model is forced to marginalize, thus, the measurement model is incorrectly specified.

In addition, GMIII based on the Gaussian measurement model and a logistic dropout model i.e., based on the treatment indicator and previous Gaussian outcome. In this situation, the generated outcomes are continuous in nature. Thus, the correct analysis model based on this generating model is MI-GEE, which needs a measurement model and imputation model, not a dropout model. For GMIII, imputing the missing observation using a Gaussian imputation model and afterward fitting standard GEE to the dichotomized outcome of the completed sets of data, results in MI-GEE with everything correctly specified. In this case, multiple imputations are based on the continuous outcome and the dichotomization is done afterward, which indicates the imputation model as well as measurement models are correctly specified. To compare with WGEE using an incorrectly specified dropout model, the weights are taken from a logistic dropout model with the treatment indicator and the binary version of the previous outcome as covariates. This is an apparent misspecification, since the underlying dropout model utilizes the continuous form of the previous outcome as covariate.

4. Results of Simulation Study

In this Section, both approaches, inverse probability weighting GEE and MI-based GEE, are compared through the use of so-called asymptotic simulation and small-sample simulations. The behaviors of both methods in terms of bias and mean square error, under various correctly and misspecified models, has been studied and the strength of each method under various types of misspecification has been examined.

4.1 Everything Correctly Specified

In this Section, all models which are correctly specified are considered. All three data generating models are taken into account. Since GMI consists of a Bahadur measurement model and logistic dropout model, as explained in Section 3, the correct analysis model for GMI would be WGEE, where GEE is modified by application of weighing scheme which is valid under MAR. In this regard, WGEE was applied to GMI with weights from a logistic dropout model with the treatment indicator and the previous outcome as covariates. Both asymptotic and small-sample simulations were considered and results are placed in Table 2.

Table 2: Asymptotic and small- sample simulation results for WGEE, with everything correctly specified, under GM1.

Asymtotic			Small-sample			
Parameter	**$Bias_\infty$**	**Var_∞**	**$Bias_N$**	**$Est(Var_N)$**	**Var_N**	**MSE**
			N=50			
β_0	0.0000	0.4409	-0.4824	1.2750	0.0088	1.5077
β_x	0.0000	1.1097	0.0234	2.4501	0.0222	2.4507
β_t	0.0000	0.1194	0.0952	0.2634	0.0024	0.2725
β_{xt}	0.0000	0.2782	-0.1124	0.5590	0.0056	0.5716
			N=100			
β_0	-	-	-0.3957	1.0779	0.0044	1.2345
β_x	-	-	0.1225	2.1108	0.0111	2.1258
β_t	-	-	0.1018	0.2388	0.0012	0.2492
β_{xt}	-	-	-0.1355	0.4441	0.0028	0.4625
			N=200			
β_0	-	-	-0.1313	0.7634	0.0022	0.7807
β_x	-	-	0.0035	1.7076	0.0055	1.7076
β_t	-	-	0.0398	0.1673	0.0006	0.1689
β_{xt}	-	-	-0.1019	0.3107	0.0014	0.3211
			N=500			
β_0	-	-	-0.0225	0.3031	0.0009	0.3036
β_x	-	-	0.0407	0.8947	0.0022	0.8963
β_t	-	-	0.0002	0.0662	0.0002	0.0662
β_{xt}	-	-	-0.0557	0.1785	0.0006	0.1816

In Table 2, asymptotic results include the asymptotic bias ($Bias_\infty$) and the asymptotic variance (Var_∞), while small-sample results include bias ($Bias_N$), estimated variance(EST (Var_N)), true variance (Var_N) and mean square error (MSE) of the parameter estimators. It can be seen that there is asymptotic unbiasedness of the WGEE parameters under correctly specified mean structure. In the small-sample simulation case, considerable quantity of bias was seen. For samples of size 50 and 100, large amounts of bias, as well as estimated variances of the parameter estimates, was observed. The amount of bias decreased when the sample of size N is increased and bias was very small when a sample of size 500 is taken. The estimated variances and MSE are also in line with the bias, in the sense that the estimated variance and MSE of the corresponding parameter estimates declined considerably as the sample size was increased. In addition, it was noted that the estimated variances are much larger than the true variances, which can be indicative of the inefficiency of WGEE. However, estimated variances seem to improve with larger sample sizes.

GMII is based on a second-order autoregressive transition model for the mean structure and a logistic model for dropout. In this situation, missing values are first imputed using an imputation model consistent with the underlying dropout model. That is, the predictors of dropout are taken into account into the imputation model. For analysis, an AR (2) transition model is fitted to the complete data resulting in MI-Transition. This ensures an analysis model that is consistent with the underlying GM. In other words, both measurement and the imputation model are correctly specified. For this GM, for the small-sample simulations M=5 imputations were considered and for asymptotic simulations, 500 imputations were considered. The results are displayed in Tables 3 and 4.

Table 3: Asymptotic simulation result for MI-Transition, with everything correctly specified, under GMII.

Parameter	**$Bias_\infty$**	**Var_∞**
α_0	0.0000	8.0803
α_x	0.0000	16.1003
ϕ_0	-0.0096	12.0925
ϕ_x	-0.0666	18.0150
ϕ_1	0.0343	18.1481
γ_0	0.0274	17.3872
γ_x	-0.0534	18.5088
γ_1	0.0130	18.8832
γ_2	-0.0749	19.7294

In Table 3, asymptotic results include the asymptotic bias ($Bias_\infty$) and asymptotic variance (Var_∞) of the parameter estimators. It can be seen that in the first segment, the parameter estimates are asymptotically unbiased. In the second and the third segments, some amount of bias was observed.

For the small-sample simulation case in Table 4, the results include bias ($Bias_N$), estimated variance(EST (Var_N)), true variance (Var_N) and MSE, of the parameter estimators. It can be seen that some amount of bias was observed. For samples of size N=50, some amount of bias was observed in the entire three segments but considerable amount of estimated variances, as well as MSE, were observed, and estimated variances were much larger in the third segment and quite far from the true variances. As the size N increases, it was found that the bias as well as estimated variances reduced sharply. For a sample of size N=500, it was found that the estimated variances are closer to the true variances and also the bias in all the three segments were very small, which indicates the effect of the sample size N on the bias and precision of the estimates.

Table 4: Small-sample simulation results for MI-Transition with everything correctly specified under GMII.

Parameter	$Bias_N$	Est(Var_N)	Var_N	MSE	$Bias_N$	Est(Var_N)	Var_N	MSE
N=50					**N=100**			
α_0	0.0184	0.1841	0.1616	0.1845	-0.0313	0.0925	0.0808	0.0935
α_x	-0.0137	0.3456	0.3220	0.3458	0.0369	0.1791	0.1610	0.1805
ϕ_0	0.0129	0.5168	0.2386	0.5170	0.0293	0.2140	0.1193	0.2148
ϕ_t	0.0703	1.9120	0.3581	1.9169	0.0061	0.2724	0.1791	0.2724
ϕ_1	0.0707	0.6210	0.3607	0.6260	0.0271	0.2768	0.1804	0.2775
γ_0	0.1746	13.3809	0.3453	13.4114	0.0768	0.3362	0.1727	0.3421
γ_x	-0.0315	2.4890	0.3693	2.4900	0.0183	0.3014	0.1846	0.3018
γ_1	-0.0134	2.5032	0.3759	2.5034	0.0496	0.3427	0.1880	0.3452
γ_2	-0.0470	9.3754	0.3918	9.3776	-0.1024	0.2495	0.1959	0.2600
N=200					**N=500**			
α_0	-0.0004	0.0449	0.0404	0.0449	-0.0009	0.0179	0.0162	0.0179
α_x	-0.0028	0.0893	0.0805	0.0893	-0.0017	0.0304	0.0322	0.0304
ϕ_0	-0.0024	0.1029	0.0597	0.1029	-0.0018	0.0401	0.0239	0.0401
ϕ_t	-0.0040	0.1362	0.0895	0.1362	-0.0030	0.0458	0.0358	0.0458
ϕ_1	0.0261	0.1283	0.0902	0.1290	0.0121	0.0507	0.0361	0.0509
γ_0	0.0641	0.1690	0.0863	0.1731	0.0499	0.0676	0.0345	0.0701
γ_x	0.0161	0.1545	0.0923	0.1547	0.0055	0.0603	0.0369	0.0603
γ_1	0.0415	0.1364	0.0940	0.1381	0.0132	0.0531	0.0376	0.0533
γ_2	-0.1243	0.1175	0.0979	0.1329	-0.1011	0.0459	0.0392	0.0561

As explained in Section 3.3, GMIII is based on a Gaussian measurement model and a logistic dropout model. In this situation, the generated outcomes are of a continuous nature but the main focus is on the binary outcome Y_{ij}. Thus, the binary outcome was taken based on a cut-off value of 6.5 for the continuous outcome W_{ij} (if W_{ij}>=6.5 then Y_{ij}=1, and Y_{ij}=0 otherwise). The true parameters corresponding to this binary outcome was calculated by applying a GEE model to the complete data. True parameters are placed in the Appendix

(Table 1). The analysis model for this GM with everything correctly specified consists of an imputation model that uses the predictors of dropout and a measurement model of the same form fitted to the dichotomized outcome. Thus, multiple imputations based on the continuous outcome were considered and later dichotomized based on the aforementioned cut-off value. Afterward, standard GEE method was applied and the results are placed in Table 5.

Table 5: Small-sample simulation result for MI-GEE, with everything correctly specified, under GMIII.

Parameter	$Bias_N$	$Est(Var_N)$	MSE	$Bias_N$	$Est(Var_N)$	MSE
N=50				N=100		
β_0	-0.0012	0.3664	0.3664	0.0017	0.1973	0.1973
β_x	0.0125	0.7577	0.7579	0.0043	0.3966	0.3966
β_t	0.001	0.1044	0.1044	-0.0003	0.0599	0.0599
β_{xt}	-0.013	0.2407	0.2408	-0.0047	0.1472	0.1472
N=200				N=500		
β_0	0.0059	0.094	0.094	0.0033	0.0361	0.0361
β_x	-0.0091	0.1946	0.1947	0.0007	0.0742	0.0742
β_t	-0.0051	0.0258	0.0258	-0.0024	0.0101	0.0101
β_{xt}	0.0089	0.074	0.0741	-0.0015	0.028	0.028

In this case, only small-sample simulation is possible and results include bias ($Bias_N$), estimated variance(EST (Var_N)) and mean square error (MSE) of the parameter estimators. In applying MI-GEE to the different samples, it was discovered that for N=50, in total 177 samples didn't converge. Similarly for N=100 and 200, in total 51 and 2 samples respectively, were not convergent. It was noticed that for the small sample sizes, more samples weren't convergent but for N=500, all samples were convergent. The main reason noticed was the inestimability of the treatment-by-time interaction in the model. Thus, only convergent samples were taken for the final analysis.

From Table 5, it can be seen that in all cases, bias of the estimates are quite small, which can be viewed as the measure of the strength of multiple imputation. But for N=50 and 100, since the number of non-convergent samples is quite large, parameter estimates, and also the estimated variances, may not be very reliable. From Table 5, it can also be noticed that the estimated variances are decreasing considerably when the sample size increases. Similarly, the bias of the estimates were found to get smaller as sample size increases, as expected.

4.2 Measurement Model and Dropout Models Correct but Incorrect Imputation Model

In this Section, correct measurement and dropout models are considered, but an incorrect imputation model in order to compare WGEE and MI-GEE. For WGEE, correct measurement model and dropout model is considered, while for MI-GEE, a correct measurement model but incorrect imputation model is considered. For this purpose, as discussed in Section 3.3, GMI can be taken. Using GMI, the measurement model and dropout model for WGEE are correct, but the imputation model of MI-GEE is incorrect.

The results of the both approaches are placed in Table 6. Results include bias ($Bias_N$), estimated variance(EST (Var_N)), true variance (Var_N) and MSE of the parameter estimators.

From Table 6, it can be seen that for each sample sizes, the amount of bias of the parameter estimates obtained from WGEE are larger than that of MI-GEE. Similarly, the estimates have large estimated variances in WGEE indicating less precise estimates than MI-GEE. As the sample size increases, the estimated variances decrease sharply for the WGEE approach, but even for a sample of a size 500, true variances and the estimated variances are not close to each other. For MI-GEE, for a sample of size N=500, the estimated variance is quite small.

Table 6: Small-sample simulation results for WGEE with correctly specified dropout model and measurement model, and MI-GEE with incorrectly specified imputation model, under GMI.

	WGEE				**MI-GEE**			
Parameter	**$Bias_N$**	**Est(Var_N)**	**Var_N**	**MSE**	**$Bias_N$**	**Est(Var_N)**	**Var_N**	**MSE**
N=50								
β_0	-0.4824	1.2750	0.0088	1.5077	0.0465	0.4916	0.0143	0.4937
β_x	0.0234	2.4501	0.0222	2.4507	-0.0142	0.9206	0.0368	0.9208
β_t	0.0952	0.2634	0.0024	0.2725	-0.0123	0.1221	0.0025	0.1223
β_{xt}	-0.1124	0.5590	0.0056	0.5716	-0.0182	0.2465	0.0072	0.2468
N=100								
β_0	-0.3957	1.0779	0.0044	1.2345	-0.0143	0.2302	0.0072	0.2304
β_x	0.1225	2.1108	0.0111	2.1258	0.0137	0.4729	0.0184	0.4731
β_t	0.1018	0.2388	0.0012	0.2492	0.0059	0.0538	0.0013	0.0538
β_{xt}	-0.1355	0.4441	0.0028	0.4625	-0.00197	0.1141	0.0026	0.11405
N=200								
β_0	-0.1313	0.7634	0.0022	0.7807	0.0105	0.1039	0.0036	0.1040
β_x	0.0035	1.7076	0.0055	1.7076	-0.0174	0.2200	0.0092	0.2203
β_t	0.0398	0.1673	0.0006	0.1689	-0.0046	0.0248	0.0006	0.0248
β_{xt}	-0.1019	0.3107	0.0014	0.3211	0.0054	0.0575	0.0018	0.0575
N=500								
β_0	-0.0225	0.3031	0.0009	0.3036	0.0127	0.0428	0.0014	0.0430
β_x	0.0407	0.8947	0.0022	0.8963	-0.0342	0.0747	0.0037	0.0759
β_t	0.0002	0.0662	0.0002	0.0662	-0.0096	0.0107	0.0003	0.0108
β_{xt}	-0.0557	0.1785	0.0006	0.1816	0.0227	0.0190	0.0007	0.0196

Based on the results obtained in Table 6, MI-GEE seemed more robust than WGEE, even though the former consists of a misspecification in the imputation model.

4.3 Measurement Model and Imputation Models Correct But Incorrect Dropout Model

In this Section, correct imputation and measurement models but a misspecified dropout model are considered. For this purpose GMIII is chosen. Since, GMIII is based on a Gaussian measurement model and a logistic dropout model, the correct analysis model for this generating model is MI-GEE which needs a measurement model and an imputation model, but not a dropout model.

In the MI-GEE, multiple imputations were carried on the continuous outcome first, which was later dichotomized, and then standard GEE was fitted on this imputed dataset. For the WGEE, weights were chosen based on the logistic dropout model with treatment indicator and the binary version of the previous measurement as covariates. In this case, dropout model is misspecified in the sense that the underlying dropout model uses the continuous version of the previous measurement as covariates. The results are placed in Table 7.

Table 7: Small-sample simulation results for WGEE with incorrectly specified dropout model, and for MI-GEE with correctly specified imputation model, under GMIII.

	WGEE			MI-GEE		
Parameter	**$Bias_N$**	**Est(Var_N)**	**MSE**	**$Bias_N$**	**Est(Var_N)**	**MSE**
N=50						
β_0	-0.1807	0.4825	0.5151	-0.0012	0.3664	0.3664
β_x	-0.0755	1.0028	1.0085	0.0125	0.7577	0.7579
β_t	0.2793	0.2095	0.2875	0.0010	0.1044	0.1044
β_{xt}	-0.0023	0.4001	0.4001	-0.0130	0.2407	0.2408
N=100						
β_0	-0.1855	0.3113	0.3457	0.0017	0.1973	0.1973
β_x	-0.1380	0.5644	0.5834	0.0043	0.3966	0.3966
β_t	0.3099	0.1376	0.2336	-0.0003	0.0599	0.0599
β_{xt}	0.0367	0.2312	0.2325	-0.0047	0.1472	0.1472
N=200						
β_0	-0.2450	0.4816	0.5416	0.0059	0.0940	0.0940
β_x	0.0959	0.7105	0.7197	-0.0091	0.1946	0.1947
β_t	0.1645	0.1379	0.1650	-0.0051	0.0258	0.0258
β_{xt}	-0.0339	0.2501	0.2513	0.0089	0.0740	0.0741
N=500						
β_0	-0.1323	0.1754	0.1929	0.0033	0.0361	0.0361
β_x	0.0189	0.2453	0.2457	0.0007	0.0742	0.0742
β_t	0.1029	0.0510	0.0616	-0.0024	0.0101	0.0101
β_{xt}	-0.0035	0.0867	0.0867	-0.0015	0.0280	0.0280

The results include the bias ($Bias_N$), estimated variance(EST (Var_N)) and mean square error (MSE) of the parameter estimators. It can be seen that the bias and estimated variance of the parameter estimates under MI-GEE are smaller compared to those under the WGEE approach. Bias in MI-GEE is very small even for a sample of size N=50. In WGEE, not only is bias larger, the estimated variance is also larger than that of MI-GEE, and the ratio of increase of the variances between two approaches increases as sample size N increases. In other words, for sample of size 100, the estimated variance for WGEE is almost 1.5 times that of MI-GEE, but for a sample of size N=500, estimated variance for WGEE is almost 5 times higher than that of MI-GEE. In both approaches, the amount of bias as well as estimated variances decreased as sample size increases. But in all the sample sizes considered, MI-GEE is superior to WGEE. This can be also being viewed as the effect of a correctly specified imputation model against the effect of a misspecification in the WGEE approach.

4.4 Dropout Model and Imputation Models Correct but Incorrect Measurement Model

In this Section, correct imputation and dropout model but misspecified measurement model are considered. For this purpose GMII is chosen. Since GMII is based on the AR(2) transition model for the mean structure and a logistic dropout model, in this case, in comparing WGEE and MI-GEE, the marginalized version of MI-transition was considered. Hence, the conditional model is forced to marginalize, and in this situation, the measurement model is incorrectly specified.

Table 8: Asymptotic and small-sample simulation results for marginalized MI-Transition, under GMII.

	Asymptotic		Small-sample			
	Bias$_\infty$	**Var$_\infty$**	**Bias$_N$**	**Est(Var$_N$)**	**Var$_N$**	**MSE**
N=50						
β_0	-0.0035	1.1160	-0.0592	1.1703	0.0223	1.1738
β_x	0.0469	2.3885	0.0159	2.5898	0.0478	2.5901
β_t	0.0041	0.2039	0.0648	0.2124	0.0041	0.2166
β_{xt}	-0.0493	0.4363	-0.0040	0.4746	0.0087	0.4746
N=100						
β_0	-	-	-0.0591	1.1200	0.0112	1.1235
β_x	-	-	0.0404	2.4671	0.0239	2.4687
β_t	-	-	0.0373	0.2048	0.0020	0.2061
β_{xt}	-	-	-0.0094	0.4505	0.0044	0.4506
N=200						
β_0	-	-	-0.0086	1.1162	0.0056	1.1163
β_x	-	-	-0.0082	2.4398	0.0119	2.4399
β_t	-	-	0.0083	0.2027	0.0010	0.2027
β_{xt}	-	-	0.0038	0.4436	0.0022	0.4436
N=500						
β_0	-	-	-0.0014	1.1168	0.0022	1.1168
β_x	-	-	-0.0001	2.4373	0.0048	2.4373
β_t	-	-	0.0013	0.2046	0.0004	0.2046
β_{xt}	-	-	-0.0015	0.4472	0.0009	0.4472

Before comparing these two approaches, firstly asymptotic and small-sample simulation results for marginalized MI-transition were fitted. Based on the correctly specified MI-transition model from Table 3 and 4, three distinct conditional probabilities were generated. In addition, based on those probabilities, marginal probabilities were derived in line with (7). These probabilities were used as weight for standard GEE on the hypothetical data having all possible outcome sequences including the treatment indicator; the results provide the marginalized MI-Transition. The results are placed in Table 8. These results are compared to a fitted GEE model of the form (4) on a hypothetical dataset which consists of all 16 possible combinations with corresponding probability weights obtained from (7) based on the true parameters defined in(6). The resulting marginalized true parameters for GMII were found to be $\beta_0 = -0.3658, \beta_x = 0.2673, \beta_x = 0.2265$ and $\beta_{xt} = 0.0790$.

From Table 8, it can be seen that asymptotic bias for the parameter estimates is generally small. For small-sample case, for samples of size 50 and 100, the bias for the parameter estimates is large, while for sample sizes 200 and 500, the bias for the parameter estimates is pretty small, close to zero. The difference between the true and estimated variance is quite noticeable, indicating some sort of inefficiency under the sample sizes considered.

Table 9: Small-sample simulation results for WGEE, with correctly specified dropout model and misspecified measurement model, and for MI-GEE, with correctly specified imputation model and misspecified measurement model under GMII.

	WGEE				**MI-GEE**			
Parameters	**Bias$_N$**	**Est(Var$_N$)**	**Var$_N$**	**MSE**	**Bias$_N$**	**Est(Var$_N$)**	**Var$_N$**	**MSE**
N=50								
β_0	-0.5802	1.2779	0.0094	1.6145	0.0058	0.4934	0.0152	0.4934
β_x	-0.1681	2.9703	0.0208	2.9986	-0.0087	0.8994	0.0345	0.8995
β_t	0.1374	0.2513	0.0027	0.2701	0.0122	0.1279	0.003	0.128
β_{xt}	0.0931	0.5878	0.0059	0.5964	-0.0042	0.225	0.0063	0.225
N=100								
β_0	-0.4229	1.131	0.0047	1.3098	-0.0552	0.2381	0.0076	0.2412
β_x	-0.1451	2.9804	0.0104	3.0014	0.048	0.4779	0.0173	0.4802
β_t	0.1241	0.2149	0.0014	0.2303	0.0334	0.057	0.0015	0.0581
β_{xt}	0.0792	0.5877	0.003	0.594	-0.019	0.1149	0.0031	0.1152
N=200								
β_0	-0.1887	0.7674	0.0024	0.803	-0.008	0.1185	0.0038	0.1186
β_x	-0.1904	2.3738	0.0052	2.41	-0.0045	0.2349	0.0086	0.2349
β_t	0.0684	0.1571	0.0007	0.1618	0.0075	0.0279	0.0007	0.028
β_{xt}	0.0842	0.4333	0.0015	0.4404	-0.0011	0.0581	0.0016	0.0581
N=500								
β_0	-0.0600	0.3064	0.0009	0.3099	-0.0013	0.048	0.0015	0.048
β_x	-0.1795	0.9594	0.0021	0.9916	0.0023	0.0832	0.0035	0.0832
β_t	0.0253	0.0636	0.0003	0.0642	0.0011	0.0119	0.0003	0.0119
β_{xt}	0.0913	0.1813	0.0006	0.1897	-0.0039	0.0203	0.0006	0.0203

As explained earlier, using the marginalized parameters to define the marginal model for GMII, both model WGEE and MI-GEE were fitted and the results are placed in Table 9.

As discussed in Section 3.3, for GMII, the marginalized parameters are used to define the measurement model, indicating that the outcomes are modeled marginally. From Table 9, it can be seen that bias and estimated variance of the parameter estimates in MI-GEE are quite smaller than that of WGEE. As sample size increases, the estimated variances decrease sharply in both approaches, but with high variability in WGEE, which indicates that MI-GEE performs better than WGEE. However, larger samples provide less bias and more precise results, as expected.

5. Effects of More Dropout on MI-GEE and WGEE

The results that were presented in Section 4 had about 68-76 % completer, with about 24-32% incomplete data and fairly equally distributed across the treatment arms. In this Section, exploration of other parameter values for the dropout models to generate more missingness (40-50% completers) using the same data setting, has been assessed. The main focus in this Section is to distinguish the effect of the amount of missingness on the performance of the methods. For this rationale, the same measurement models that were used in the previous Sections were chosen but the dropout models were altered and the analyses were conducted for N=100 only.

For GMI and GMII, the following parameters were selected for dropout model (5) with the same measurement models as those used in Section 4: $\psi_0 = -0.1, \psi_x = -1.2$ and $\psi_{prev} = -0.29$. Combining this dropout model with the measurement model in equation (4) yields, for GMI, 45 % completers, 33 % with only the first outcome observed (18% for x=0 and 15% for x=1) and 22 % with last observation missing (11% for x=0 and 11% for x=1). Likewise, combining this dropout model with the measurement model in equation (6) yields, for GMII, 49 % completers, 31% with only the first outcome observed (22% for x=0 and 9% for x=1) and 20 % with last observation missing (12% for x=0 and 8% for x=1).

In GMIII, the following parameters were chosen for the dropout model (9) with same Gaussian measurement model (8) used in Section 4: $\delta_0 = -0.15, \delta_x = 0.8,$ and $\gamma_1 = -0.15$. After combining this dropout model to the measurement model (8), on an average over 500 samples yields, for GMIII, 45 % completers, 37 % with only the first outcome observed (15% for x=0 and 22% for x=1) and 18 % with last observation is missing (8% for x=0 and 10%for x=1).

In this Section, both approaches, inverse probability weighting and MI-based GEE were considered and compared through the use of asymptotic and small-sample simulations. The behaviors of both methods in terms of bias and mean square error, under various correctly and misspecified models, has been studied and the strengths of each method under various types of misspecification has been examined. Results for this Section are also compared with the results in Section 4 for the samples of size N=100.

5.1 Everything Correctly Specified

In this Section, all models which are correctly specified were considered. All three data generating models are taken into account.

In line with the Section 4, the correct analysis model for GMI is WGEE. In this regard, WGEE was applied to GMI with weights from a logistic dropout model with the treatment indicator and the previous outcome as predictors. Both asymptotic and small-sample process were considered and results are placed in left panel of the Table 10. Right panel of Table 10 represent the results from Section 4 for the samples of size N=100.

Table 10: Asymptotic and small-sample simulation results for WGEE, with correctly specified, under GMI for N=100.

Case 1: Percentage of completers=45						Case 1: Percentage of completers=68					
Asymptotic			Small-sample			Asymptotic		Small-sample			
Parameter	**Bias$_\infty$**	**Var$_\infty$**	**Bias$_N$**	**Est(Var$_N$)**	**MSE**	**Bias$_\infty$**	**Var$_\infty$**	**Bias$_N$**	**Est(Var$_N$)**	**MSE**	
β_0	0.000	0.4410	-0.0273	0.2956	0.2964	0.000	0.4410	-0.3957	1.0779	1.2345	
β_x	0.000	1.1097	0.0674	0.6306	0.6351	0.000	1.1097	0.1225	2.1108	2.1258	
β_t	0.000	0.1194	0.0143	0.0846	0.0848	0.000	0.1194	0.3018	0.2388	0.2492	
β_{xt}	0.000	0.2782	-0.046	0.1914	0.1935	0.000	0.2782	-0.1355	0.4441	0.4625	

In the left panel of the Table 10, it can be seen that there is asymptotic unbiasedness property of the WGEE estimates under correctly specified mean structure. In the small-sample counterpart, some quantities of bias were seen. Comparing results of both panels, WGEE performed better for the data with more dropouts. In the right panel, the amounts of bias (Bias$_N$) and estimated variances (Est(Var$_N$)) were very large. Interesting results might be expected for various other samples, if considered.

As discussed earlier, the analysis model for GMII that is consistent with the underlying measurement model after multiple imputations and is MI-Transition. The results are displayed in the left panel of Table 11. The right panel of Table 11 represent the results from Section 4 for the samples of size N=100.

Table 11: Asymptotic and Small-sample simulation results for MI-Transition, with everything correctly specified under GMII for N=100.

Case 1: Percentage of completers=49						Case 2:Percentage of completers=73				
Asymptotic			Small-sample			Asymptotic		Small-sample		
Parameter	**Bias$_\infty$**	**Var$_\infty$**	**Bias$_N$**	**Est(Var$_N$)**	**MSE**	**Bias$_\infty$**	**Var$_\infty$**	**Bias$_N$**	**Est(Var$_N$)**	**MSE**
α_0	0.0000	8.0803	-0.0313	0.0925	0.0935	0.0000	8.0803	-0.0313	0.0925	0.0935
α_x	0.0000	16.1003	0.0369	0.1791	0.1805	0.0000	16.1003	0.0369	0.1791	0.1805
ϕ_0	-0.0104	12.3648	0.0300	0.2792	0.2801	-0.0096	12.0925	0.0293	0.2140	0.2148
ϕ_t	-0.0436	17.9467	-0.0019	0.3697	0.3697	-0.0666	18.0150	0.0061	0.2724	0.2724
ϕ_1	-0.1761	18.0521	0.0567	0.3365	0.3397	0.0343	18.1481	0.0271	0.2768	0.2775
γ_0	0.0542	18.3915	0.1602	2.0728	2.0985	0.0274	17.3872	0.0768	0.3362	0.3421
γ_x	0.0159	19.2350	-0.0025	2.2136	2.2136	-0.0534	18.5088	0.0183	0.3014	0.3018
γ_1	-0.1731	19.3857	0.0160	0.6057	0.6060	0.0130	18.8832	0.0496	0.3427	0.3452
γ_2	-0.2602	20.2183	-0.1715	0.2860	0.3154	-0.0749	19.7294	-0.1024	0.2495	0.2600

From the left panel of Table 11, in the asymptotic approach, it can be seen in the first segment, there is asymptotic unbiasedness of the parameter estimates. In the second and third segments, some amount of bias was observed, as expected. In the small-sample counterpart, some amount of bias was observed. Comparing these results with right panel, in the asymptotic case, more bias was noticed in segments 2 and 3, for this setting with more missingness, especially in the estimates of γ_1 and γ_2. More amount of bias was observed but that might be the effect of the more missingness in second and third time points in the data. For the small sample simulations, slightly more bias was observed for data setting with more missingness, in addition, estimated variances in second and third segments are also comparatively large.

Recalled that GMIII is based on the Gaussian measurement model and a logistic dropout model. The true parameters corresponding to the binary outcome (dichotomized version of continuous outcome) were calculated by applying a GEE model and the resulting parameters are β_0= -3.0388, β_x= 0.0353, β_x= 1.7835 and β_{xt}= 0.4532. The correct analysis model for this generating model is MI-GEE. Thus, multiple imputations based on the continuous outcome were considered and later dichotomized based on the predefined cut-off value. Afterward, standard GEE was applied and the results are placed in the left panel of Table 12. The right panel of Table12 represent the results from Section 4 for the samples of size N=100.

Table 12: Small-sample simulation result for MI-GEE under GMIII, with everything correctly specified, for N=100.

Case 1: Percentage of completer=45				**Case 2: Percentage of completer=76**		
Parameter	**Bias$_N$**	**Est(Var$_N$)**	**MSE**	**Bias$_N$**	**Est(Var$_N$)**	**MSE**
β_0	0.0169	0.1957	0.1959	0.0017	0.1973	0.1973
β_x	-0.0049	0.3950	0.3950	0.0043	0.3966	0.3966
β_t	-0.0099	0.0569	0.0570	-0.0003	0.0599	0.0599
β_{xt}	-0.0018	0.1407	0.1407	-0.0047	0.1472	0.1472

In the application of MI-GEE to the data having more dropouts (left panel of Table 12), it was discovered that total of 108 samples did not converge. As explained earlier, the main reason was the inestimability of the treatment-by-time interaction. Thus, only 392 convergent samples were taken for the final analysis. From the left panel of Table 12, it can be seen that in all cases, bias of the estimates are quite small which can be viewed as the effectiveness of multiple imputation. Comparing this result with right panel of Table 12, it was found that the amount of bias in both cases was small. Similarly in both the cases, the estimated variances are quite close to each other. This can be viewed as the consistency of MI-GEE over various percentage of missing data.

5.2 Measurement Model and Dropout Models Correct but Incorrect Imputation Model

In this Section, correct measurement and dropout models were considered, but, with an incorrect imputation model. For this purpose, as discussed in Sections 3.3 and 4.2, GMI can be taken. The results of both WGEE and MI-GEE are placed in left panel of Table 13. The right panel of Table 13 represent the results from Section 4 for samples of size N=100.

Table 13: Small-sample simulation results for WGEE, with correctly specified dropout and measurement model, and MI-GEE, with incorrectly specified imputation model, under GM1 for N=100.

Case 1: Percentage of completers=45				**Case 2: Percentage of completers=68**			
Parameter	**$Bias_N$**	**Est(Var_N)**	**MSE**	**Parameter**	**$Bias_N$**	**Est(Var_N)**	**MSE**
WGEE				**WGEE**			
β_0	-0.0273	0.2956	0.2964	β_0	-0.3957	1.0779	1.2345
β_x	0.0674	0.6306	0.6351	β_x	0.1225	2.1108	2.1258
β_t	0.0143	0.0846	0.0848	β_t	0.1018	0.2388	0.2492
β_{xt}	-0.0462	0.1914	0.1935	β_{xt}	-0.1355	0.4441	0.4625
MI-GEE				**MI-GEE**			
β_0	-0.0040	0.2781	0.2781	β_0	-0.0143	0.2302	0.2304
β_x	-0.0128	0.5557	0.5559	β_x	0.0137	0.4729	0.4731
β_t	-0.0010	0.0797	0.0797	β_t	0.0059	0.0538	0.0538
β_{xt}	0.0216	0.1594	0.1599	β_{xt}	-0.00197	0.1141	0.1141

In the left panel of Table 13, it can be seen that, the amount of bias and estimated variance from the WGEE approach are slightly larger than those of MI-GEE, but are very close to each other. Based on the result, MI-GEE seemed more robust than WGEE even though misspecifying the imputation model. Comparing these results with the results displayed in the right panel of Table 13, in the WGEE approach, the amount of bias and estimated variances are quite large in the right panel. This indicates better performance of the WGEE in data having more dropouts. However, comparing the MI-GEE results, in both panels, they are quite consistent, and bias and estimated variances of the parameter estimates are quite similar.

5.3 Measurement Model and Imputation Models Correct but Incorrect Dropout Model.

In this Section, correct imputation and measurement models but misspecified dropout model is considered. For this purpose GMIII is chosen. The results for both approaches MI-GEE and WGEE, are placed in the left panel of Table 14. The right panel of Table 14 represent the results from Section 4 for samples of size N=100.

In the left panel of Table 14, it can be seen that the bias and estimated variance of the parameter estimates under MI-GEE are smaller compared to WGEE, since for this case, MI-GEE is correctly specified and WGEE uses an incorrectly specified dropout model. Even with a misspecification in the dropout model, WGEE is not performing badly. Comparing these results with the results in the right panel of Table 14, the results for MI-GEE are similar, but results in WGEE are quite different. In the WGEE results, the amount of bias in the right panel is quite big as compared to left panel. This can be viewed as the better performance of WGEE in terms of bias in data with more missingness even though the dropout model is misspecified.

Table 14: Small-sample simulation results of WGEE, with incorrectly specified dropout model and for MI-GEE, with correctly specified imputation model under GMIII for N=100.

Case 1: Percentage of completer=45				Case 2: Percentage of completer=76			
Parameter	$Bias_N$	$Est(Var_N)$	MSE	Parameter	$Bias_N$	$Est(Var_N)$	MSE
WGEE				**WGEE**			
β_0	-0.0738	0.3553	0.3607	β_0	-0.1855	0.3113	0.3457
β_x	0.0605	0.6073	0.6110	β_x	-0.1380	0.5644	0.5834
β_t	0.0595	0.1227	0.1262	β_t	0.3099	0.1376	0.2336
β_{xt}	-0.0517	0.2468	0.2495	β_{xt}	0.0367	0.2312	0.2325
MI-GEE				**MI-GEE**			
β_0	0.0169	0.1957	0.1959	β_0	0.0017	0.1973	0.1973
β_x	-0.0049	0.3950	0.3950	β_x	0.0043	0.3966	0.3966
β_t	-0.0099	0.0569	0.0570	β_t	-0.0003	0.0599	0.0599
β_{xt}	-0.0018	0.1407	0.1407	β_{xt}	-0.0047	0.1472	0.1472

5.4 Dropout Model and Imputation Models Correct but Incorrect Measurement Model

In this Section, correct imputation and dropout models, but a misspecified measurement model, are considered. For this purpose GMII is chosen. Before comparing the MI-GEE and WGEE approaches, firstly asymptotic and small-sample simulation results for marginalized MI-transition were fitted in line with Section 4.3. The results are placed in the left panel of Table 15. These results are compared to fitted GEE model of the form (4) on a hypothetical dataset which consists of all 16 possible combinations with corresponding probability weights obtained from (7) based on the true parameters defined in (6). The resulting marginalized true parameters for GMII were found to be $\beta_0 = -0.3658, \beta_x = 0.2673, \beta_x = 0.2265$ and $\beta_{xt} = 0.0790$. The right panel of Table 15 represent the results from Section 4 for samples of size N=100.

In the left panel of Table 15, both asymptotic and small-sample case, the bias of the parameter estimates is small. But comparing these results with results of the right panel in Table 15, the results are quite similar in terms of bias and efficiency. This can be viewed as the consistency of multiple imputations under various amount of missingness.

Table 15: Asymptotic and small-sample simulation results for marginalized MI-Transition under GM II for N=100.

Case 1: Percentage of completer=49						Case 2: Percentage of completer=73				
Asymptotic			Small-sample			Asymptotic		Small-sample		
Parameter	$Bias_\infty$	Var_∞	$Bias_N$	$Est(Var_N)$	MSE	$Bias_\infty$	Var_∞	$Bias_N$	$Est(Var_N)$	MSE
β_0	0.0834	1.1399	-0.0767	1.1552	1.1611	-0.0035	1.1160	-0.0591	1.1200	1.1235
β_x	0.0072	2.4234	0.0583	2.5256	2.5290	0.0469	2.3885	0.0404	2.4671	2.4687
β_t	-0.0838	0.2216	0.0536	0.2110	0.2139	0.0041	0.2039	0.0373	0.2048	0.2061
β_{xt}	-0.0160	0.4712	-0.0246	0.4612	0.4618	-0.0493	0.4363	-0.0094	0.4505	0.4506

As discussed in Section 3.3, under GMII, the marginalized parameters are used to define the measurement model, indicating that the outcomes are modeled marginally. The results for MI-GEE and WGEE are placed in the left panel of Table 16. The right panel of Table16 represent the results from Section 4 for samples of size N=100.

Table 16: Small-sample simulation results of WGEE, with correctly specified dropout model with misspecified measurement model, and MI-GEE, with correctly specified imputation model with misspecified measurement model, under GMII for N=100.

Case 1: Percentage of completer=49				**Case 2: Percentage of completer=73**			
Parameter	**$Bias_N$**	**$Est(Var_N)$**	**MSE**	**Parameter**	**$Bias_N$**	**$Est(Var_N)$**	**MSE**
<u>WGEE</u>				**<u>WGEE</u>**			
β_0	-0.1152	0.2860	0.2993	β_0	-0.4229	1.131	1.3098
β_x	-0.0290	0.6367	0.6375	β_x	-0.1451	2.9804	3.0014
β_t	0.0723	0.0917	0.0969	β_t	0.1241	0.2149	0.2303
β_{xt}	0.0309	0.1886	0.1895	β_{xt}	0.0792	0.5877	0.594
<u>MI-GEE</u>				**<u>MI-GEE</u>**			
β_0	-0.0382	0.3239	0.3254	β_0	-0.0552	0.2381	0.2412
β_x	0.0442	0.6073	0.6093	β_x	0.048	0.4779	0.4802
β_t	0.0215	0.1043	0.1048	β_t	0.0334	0.057	0.0581
β_{xt}	-0.0171	0.1797	0.1800	β_{xt}	-0.019	0.1149	0.1152

From the left panel in Table 16, it can be seen that bias and estimated variance of the parameter estimates in MI-GEE are not so far from those of WGEE. But if these results are compared with results in the right panel in Table 16, WGEE is performing outstanding in the situation with more dropouts. The amounts of bias found in MI-GEE are similar in both cases, but slight differences were seen in the estimated variance. This result gives some interesting remarks for WGEE that seems to perform best for datasets having more missingness.

6. Discussion and Conclusion

Missing data happen habitually in longitudinal studies. Subjects may drop out before the study terminates, or be nowhere to be found to follow-up in such a way that no further measurements are made available after the time of dropout. Statistical methods which take no notice of the mechanism for dropout will show the way to biased inference. In the analysis of incomplete longitudinal data, there has been a shift, away from simple ad hoc methods that are valid only if the data are *missing completely at random* (MCAR), to more principled likelihood-based ignorable analyses, which are valid under the less restrictive *missing at random* (MAR) assumption, where the missing data simply don't contribute to estimation of parameters. Thus, which method is to be considered for handling incomplete data depends on which type of dropout mechanism it is. Inverse probability weights method is one of the common methods for handling dropout that is MAR and is more usually used in marginal models for discrete longitudinal data. Alternatively, multiple imputation can be used to preprocess incomplete data, after which standard GEE is applied (MI-GEE). When multiple imputation is considered, a number of analysis methods can be taken into account, e.g., MI-Transition.

This report is concerned with the comparison of the imputation techniques with inverse probability weighing techniques that are applied to incomplete binary longitudinal data with MAR dropout mechanism. This study explores further the work carried out by Beunckens et al (2007). In this study, the number of samples were diverse from N=50 to 500 in the same setting of the study of Beunckens et al (2007). In addition more percentage of missingness was also considered as an additional run. For this, asymptotic and small-sample simulations were considered, and inverse probability weighting and multiple imputations techniques were applied in different sets of circumstances. The applications of these approaches were classified into two parts. In the first part, 68-76% completers were considered for the different sample of sizes N=50,100,200 and 500. In the second part, just reversed, more missingness, with 45-49% completers, was considered for a sample of size N=100. The main focus in second part was to distinguish the effect of the amount of missingness on the performance of the methods. For this rationale, same measurement models that were used in the first part were chosen, but the dropout models were altered, and the analyses were conducted. In both parts, various types of data generating models were used in order to examine the effectiveness of approaches that are used to analyze incomplete longitudinal binary data.

First of all, let us discuss about the situation of the first part (Section 4), where the 68-76% subjects are completers. Although, WGEE has asymptotic unbiasedness property, even for correctly specified measurement and dropout models, in the small-sample simulations, WGEE has biased and imprecise estimates not only for small samples of size, 50 or 100, but a similar trend was observed in large samples of size 500. A clearer picture can be seen from Figure 1. Figure1 presents the plot of bias ($bias_N$) and MSE of the parameter estimates under GMI, where the left panel of the plot represents results of WGEE and the right panel is for results of MI-GEE. The only encouraging observation from the figure is that the amount of bias tends towards zero and MSE are reduced as the sample size increases. The amount of bias and MSE are larger in WGEE in compared to MI-GEE . The results shows steadiness starting from sample of size N=100.

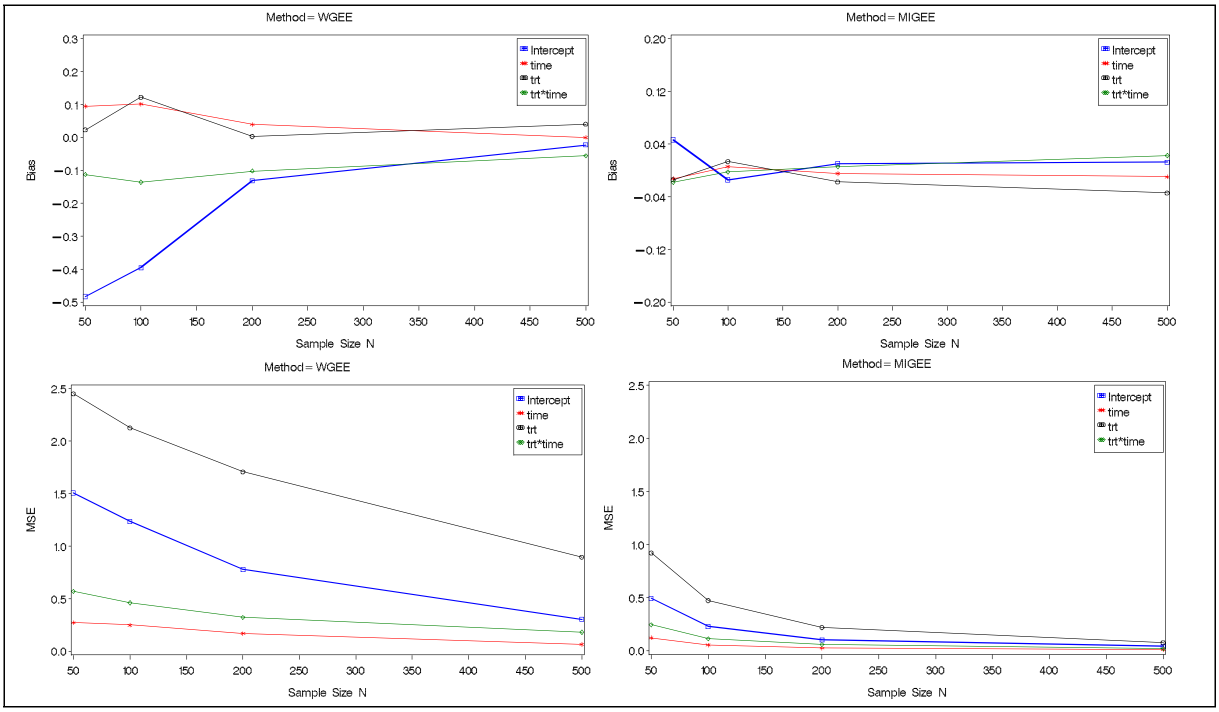

Figure 1: Plot of Bias and MSE of the parameters estimates under GMI with respect to the sample size, for GMI.

In addition, misspecification in dropout and measurement model also gave inefficiency of WGEE methods. Figures 2 and 3 (Appendix) shows the clear picture about this fact.

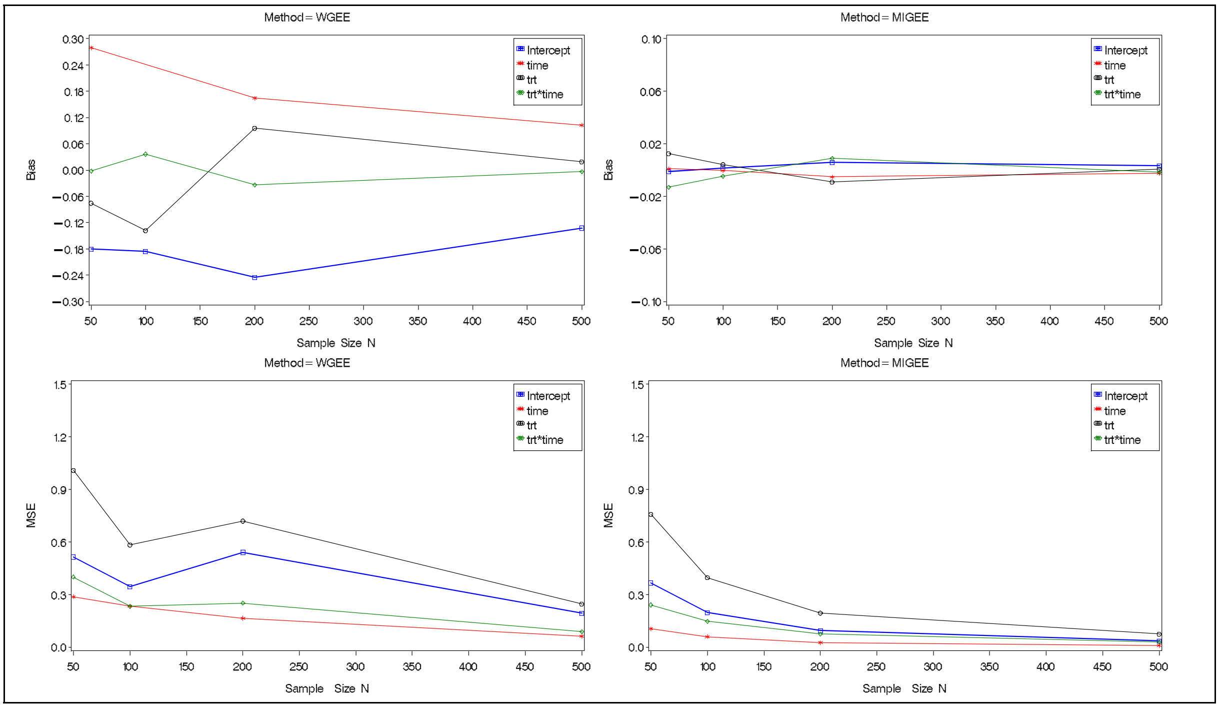

Figure 2 :Plot of Bias and MSE of the parameters estimates under GMI with respect to the sample size, for GMIII.

Figure 2 presents the plot of bias ($bias_N$) and MSE of the parameter estimates under GMIII, where the amount of bias and MSE are larger in WGEE in compared to MI-GEE in all the cases. On the other hand, MI-GEE provided positive extent of strength to misspecification in either measurement or imputation model. In both scenarios, the sample size plays role in the sense that an increase of sample size reduces the quantity of bias, as well as the estimated variances of the parameter estimates for MI-GEE. In conclusion, MI-GEE showed less biased and more precise results.

In the second part (Section 5), there were 45-49% of subjects are completers. In this case, WGEE has asymptotic unbiasedness property. It was found that even under incorrectly specified measurement and dropout model, in small-sample simulations, WGEE has performed better for the given sample size N=100. This is an interesting result compared to previous case. The results for the bias and the estimated variances from WGEE are not far from the results from MI-GEE. Though, MI-GEE has shown always less biased and more precise results in both scenarios of the missingness in the first and second parts, WGEE has shown equally better performance in the second part with larger amount of missingness. Since in the second part, only samples of size N=100 were considered, more interesting results might be expected for other sample sizes.

A simulation is an imitation of some real thing or process. The act of simulating something generally entails representing certain key characteristics or behaviors of a selected physical or abstract system. Thus, a simulation study cannot cover all possible alternatives.

The following aspects are recommended for further research. Since in the second part, only samples of size 100 were considered, it is recommend exploring more sample sizes and varying percentages of missingness. Extension to more time points would be preferred so as to conclude the result in a more general setting.

An important remark, asymptotic simulations were done to obtain the asymptotic bias and asymptotic variances, which have theoretical use only, and may provide the guidance as to what happens in large to very large samples. Supplementing them with small-sample simulations is therefore an attractive route.

In conclusion, the use of direct likelihood methods is attractive to analyze incomplete data and might be the user's ideal choice, but such methods have computational complexity, particularly arising when taking into consideration longitudinal binary data. Thus, weighted GEE and MI-GEE have been proposed for marginal models under the assumption of MAR. The results provide evidence for the fact that MI-GEE is less biased and more accurate in the small to moderate sample sizes, while WGEE is asymptotically unbiased and has shown only better performance for data having more percentage of dropout.

7. References

Agresti, A. (2002) *Categorical Data Analysis* (2nd edn). New York: John Wiley & Sons, Inc.

Bahadur, R.R., 1961. A representation of the joint distribution of responses to n dichotomous items. In*: Studies in Item Analysis and Prediction, Solomon,* H. (Ed.), Stanford Mathematical Studies in the Social Sciences, vol. VI. Stanford, CA: University Press, Stanford.

Beunckens, C., Sotto, C. and Molenberghs, G. (2007) A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics and Data Analysis*. In press.

Carey, V.C., Zeger, S. L., and Diggle, P.J. (1993) Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80,** 517-526.

Dale, J.R., 1986. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **42**, 909–917.

Diggle, P.J., Liang, K.-Y., and Zeger, S.L. (1994) *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Diggle, P.J., Heagerty, P.J., Liang, K-Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Heckman, J.J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement,* **5**, 475-492

Jansen, I., Beunckens, C., Molenberghs, G., Verbeke, G., Mallinckrodt, C. (2006a) Analyzing incomplete binary longitudinal clinical trial data. *Statistical Science*, **21**, 52–69.

Lang, J.B., Agresti, A., 1994. Simultaneously modelling joint and marginal distributions of multivariate categorical responses*. J. Amer. Statist. Assoc.* **89**, 625–632.

Le Cessie, S. and Van Houwelingen, J.C. (1994) Logistic regression for correlated binary data. Applied Statistics, **43,** 95-108.

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

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. (1994 a) A class of pattern-mixture models for normal incomplete data. *Biometrika, **81**, 471-483.*

Molenberghs, G., Lesaffre, E., 1994. Marginal modelling of correlated ordinal data using a multivariate Plackett distribution*. J. Amer. Statist. Assoc.* **89**, 633–644.

Molenberghs, G., Lesaffre, E., 1999. Marginal modelling of multivariate categorical data. *Statist. Med.* **18**, 2237–2255.

Molenberghs, G. and Kenward, M.G. (2007) *Missing Data in Clinical Studies*, John Wiley & Sons.

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

Plackett, R.L., 1965. A class of bivariate distributions. *J. Amer. Statist. Assoc.* **60**, 516–522.

Robins, J. M., Rotnitzky A., Zhao, L.P, 1995.Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J.Amer.Statist.Assoc*.**90**, 106-121.

Rotnitzky A., Wypij, D., 1994. A note on the bias of estimators with missing data. *Biometrics* **50**,106-121.

Rubin, D.B., 1978. Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. In: Imputation and Editing of Faulty or Missing Survey Data. U.S. Department of Commerce, Washington, DC, pp. 1–23.

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

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

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New-York: Springer.

Wu, M.C., Bailey, K.R., 1989. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45***,* 939-955.

Wu, M.C., Carrol, R.J., 1988. Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics* **44***,* 175-188.

8. Appendix

Table : True Parameter under GMIII for Section 4.

Parameter	N=50	N=100	N=200	N=500	S*
β_0	-3.0596	-3.0373	-3.0388	-3.0231	323
β_x	0.0605	0.0095	0.0112	0.0123	449
β_t	1.7987	1.7812	1.7774	1.7643	498
β_{xt}	0.4065	0.4828	0.4922	0.4947	500

S*= Number of convergent samples.

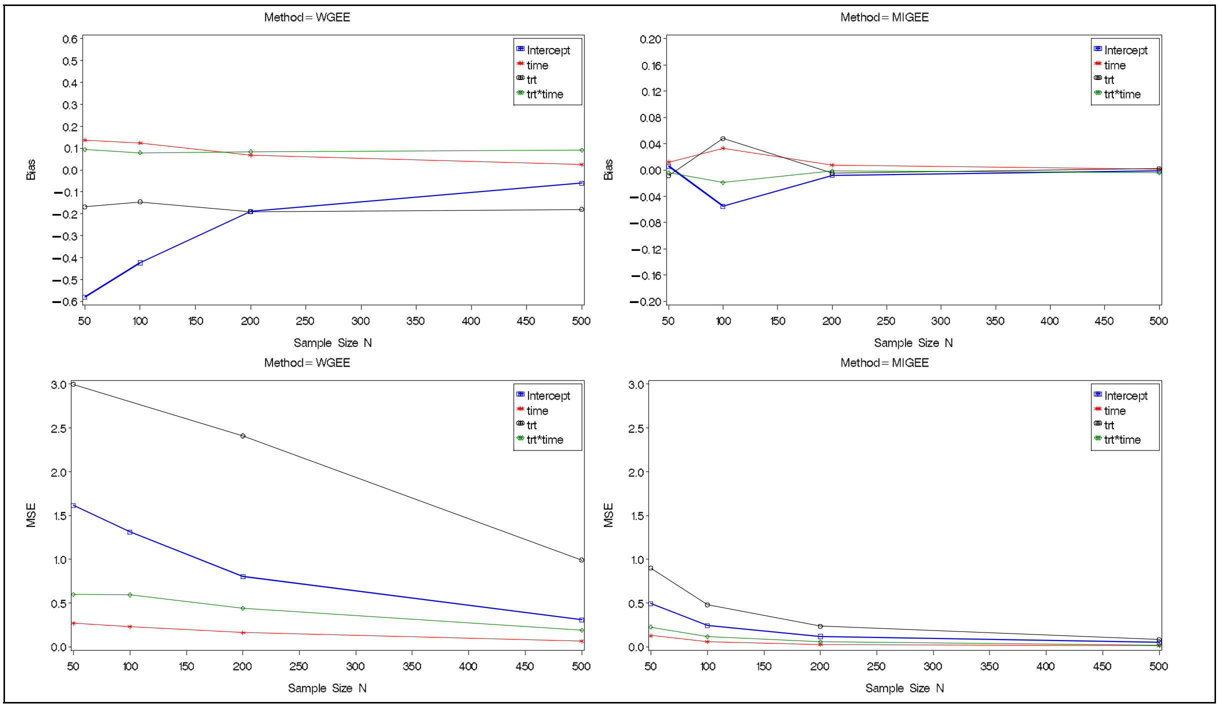

Figure 1: Plot of Bias and MSE of the parameters estimates under GMI with respect to the sample size, for GMIII.

Auteursrechterlijke overeenkomst

Opdat de Universiteit Hasselt uw eindverhandeling wereldwijd kan reproduceren, vertalen en distribueren is uw akkoord voor deze overeenkomst noodzakelijk. Gelieve de tijd te nemen om deze overeenkomst door te nemen, de gevraagde informatie in te vullen (en de overeenkomst te ondertekenen en af te geven).

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
A Simulation Study Comparing Weighted Estimation Equations with Multiple Imputation Based Estimating Equations for Longitudinal Binary Data
Richting: **Master of Science in Biostatistics** Jaar: **2007**
in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Ik ga akkoord,

Naresh Khatiwada

Datum: **27.08.2007**