

ESTIMATING THE EFFECT OF MULTIPLE IMPUTATION ON INCOMPLETE
LONGITUDINAL DATA WITH APPLICATION TO A RANDOMIZED CLINICAL
STUDY

Daniel Y. T. Fong¹, Shesh N. Rai², and Karen S. L. Lam³

¹*School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong,
Pokfulam Road, Hong Kong*

²*Department of Bioinformatics & Biostatistics, University of Louisville, Louisville, KY
40202*

³*Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong,
Pokfulam Road, Hong Kong*

Correspondence to:

Daniel YT Fong

School of Nursing, Li Ka Shing Faculty of Medicine,

The University of Hong Kong, Pokfulam Road, Hong Kong SAR, PR China

Email: dytfong@hku.hk

Tel: +852 2819 2645

Fax: +852 2872 6079

Short title:

Multiple Imputation on Incomplete Longitudinal Data

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the CRCG Small Project Funding of The University of Hong Kong and the American Syrian Lebanese Associated Charities (ALSAC).

ESTIMATING THE EFFECT OF MULTIPLE IMPUTATION ON INCOMPLETE LONGITUDINAL DATA WITH APPLICATION TO A RANDOMIZED CLINICAL STUDY

Abstract

For analyzing incomplete longitudinal data, there has been a recent interest in comparing estimates with and without the use of multiple imputation along with mixed effects model and generalized estimating equations. Empirically, the additional use of multiple imputation generally led to overestimated variances and may yield more heavily biased estimates than the use of last observation carried forward. Under ignorable or non-ignorable missing, mixed effects model or generalized estimating equations alone yielded more unbiased estimates. The different methods were also assessed in a randomized controlled clinical trial.

Key Words: Generalized estimating equations; Missing values; Mixed effects model

1. INTRODUCTION

Missing values are a fact of life in experimental studies. They are frequently encountered in clinical trials with repeated measurements for many reasons, including subject dropouts, noncompliance, and adverse events. Missing values can be classified as missing completely at random (MCAR) when the probability of having a missing value does not depend on the observed values or on the missing values had they been observed, missing at random (MAR) when the probability depends only on the observed values, or missing not at random (MNAR) when the probability depends only on the missing values had they been observed (Rubin, 1976).

In superiority trials, the clear consensus has been to require use of the intention-to-treat (ITT) principle to analyze all randomized subjects, regardless of whether data for the subjects were measured at all follow-up visits. This practice reduces the risk of overestimating the efficacy of a treatment, especially when data are MAR or MNAR (ICH, 1998). Common methods of longitudinal data analysis consistent with the ITT principle are the mixed effects model (MEM) and generalized estimating equations (GEE) (Chan et al., 2005; Dahmen & Ziegler, 2004; Edwards, 2000; Goldstein, Browne, & Rasbash, 2002; Wagner et al., 2005). The MEM is a conditional or subject-specific model, whereas the GEE is a marginal or population-averaged model. The models differ in the way the effects are interpreted, and there has been controversy about their use in analyzing longitudinal data (Carriere & Bouyer, 2002; Heagerty, 2002; Lee & Nelder, 2004; Lindsey & Lambert, 1998). Nevertheless, the two methods often produce similar estimates and standard errors when missing values are MCAR or MAR, despite their unpredictable results for

dichotomous outcomes (Twisk, 2004). Moreover, it was recently argued that the difference in interpreting conditional and marginal estimates is meaningless, and marginal predictions can often be made from conditional models (Heagerty, 2002; Lee & Nelder, 2004; Lindsey, 2000). Therefore, the MEM and GEE appear comparable for the analysis of incomplete longitudinal data.

Alternatively, missing values could be imputed so that methods such as the repeated measures analysis of variance (ANOVA), which accommodate only subjects with complete observations, can be used. Imputation methods can generally be classified as single imputation (e.g., the last observation carried forward [LOCF]), when imputation is performed once, or multiple imputation (MI) when imputation is performed more than once (Schafer, 1999). Single imputation methods do not take into account the uncertainty about the values imputed. In general, they may also vary widely in their assumptions and may be used without due consideration of their appropriateness (Wood, White, Hillsdon, & Carpenter, 2005). Moreover, the use of single imputation may often lead to larger bias than MI in longitudinal data analysis (Shieh, 2003).

Although imputation methods were developed to facilitate methods that analyze only subjects with complete observations, there has been a recent interest in examining the use of imputation methods together with MEM or GEE (Kang, Kraft, Gauderman, & Thomas, 2003; Shieh, 2003; Twisk & de Vente, 2002). Kang et al. (2003) showed that MI with MEM may yield biased variance component estimates in real and simulated longitudinal datasets. Shieh (2003) conducted a simulation study examining various imputation methods with MEM in analyzing cohort studies with MCAR and MAR values

and concluded that the use of MEM alone on all available subjects was an effective and flexible way to deal with missing values. The same conclusion was also derived in a dataset on orthodontic growth (Beunckens, Molenberghs, & Kenward, 2005). However, they had not examined the situation when values are MNAR. Twisk & de Vente (2002) evaluated the use of imputation methods with GEE. They considered a dataset from the Amsterdam Growth and Health study, an observational longitudinal study, with and without generating MCAR, MAR, and MNAR values to compare seven imputation methods with GEE. They found that using MI with GEE produced comparable estimates but generally larger standard errors than using GEE without imputation. Therefore, they concluded that the use of GEE alone was adequate. However, they focused on the analysis on a real dataset and results may not be generalized to other situations.

In summary, two issues have not been sufficiently examined in the literature. First, the use of MI with MEM or GEE has not been adequately tested in a clinical trial setting in which the objective was to determine the treatment effect in specific epochs, which is common in clinical trials. Examining this would provide a more complete picture of the performance of analysis strategies in clinical trials to facilitate the preparation of a statistical analysis plan. Second, to our knowledge, the performance of MI with MEM or GEE when values were MNAR and with GEE under all missing value scenarios had not been assessed by simulation. Therefore, we aimed to assess the performance of the MEM and GEE methods with and without MI in estimating treatment effects on a continuous outcome during different epochs by using simulation and a real dataset from a randomized clinical study.

2. THE SETTING

We consider the setting of a typical clinical trial in which n study subjects are randomly allocated to receive either a test or a control treatment and are followed up for T visits after the baseline visit. For subject i ($i = 1, 2, \dots, n$), let y_{it} be the observation at visit t ($t = 0, 1, \dots, T$) and g_i be an indicator for the allocated group (0 for the control group or 1 for the treatment group). Without a loss of generality, we assume y_{it} may be missing in any follow-up visits except for the baseline visit. The pattern of missing values is not restricted and may occur intermittently. That is, a subject with y_{it} missing in one visit may or may not have values missing in subsequent visits.

In this setting, we consider the objective of estimating the treatment effect at a follow-up visit, which is often pursued after adjusting for the baseline value. Before we discuss the methods of analysis, we further define the dummy variables $v_{it}^k = 1$ if $t = k$ and $v_{it}^k = 0$ if otherwise, for $k = 1, 2, \dots, T - 1$. That is, v_{it}^k indicates the t th measurement from subject i taken at visit k .

2.1. Analysis by MEM or GEE

The following linear MEM with adjustment for baseline value is considered:

$$y_{it} = \mu + \gamma_0 y_{i0} + \gamma_1^\top v_{it} + \gamma_2 g_i + \gamma_3^\top v_{it} g_i + a_i + e_{ij} \quad (1)$$

for $i = 1, 2, \dots, n$; $t = 1, 2, \dots, T$, where $v_{it} = (v_{it}^1, v_{it}^2, \dots, v_{it}^{T-1}, 0)^\top$,

$\gamma_1 = (\gamma_{11}, \gamma_{12}, \dots, \gamma_{1T})^\top$, $\gamma_3 = (\gamma_{31}, \gamma_{32}, \dots, \gamma_{3T})^\top$, and $a_i \sim N(0, \sigma_a^2)$ and $e_{ij} \sim N(0, \sigma^2)$ are

statistically independent. The incorporation of v_{it} in (1) allows the estimation of treatment effect at each clinical visit.. The a_i is the random effect used to account for subject-to-subject heterogeneity, which induces an exchangeable correlation structure for the responses from the follow-up visits. Note that (1) may also be written as

$$Y_i = X_i^\top \beta + b_i + e_i; \quad b_i \sim N(0, \sigma_a^2 \mathbf{1}_T \mathbf{1}_T^\top), \quad e_i \sim N(0, \sigma^2 I_T)$$

where $Y_i = (y_{i1}, y_{i2}, \dots, y_{iT})^\top$, $X_i = (x_{i1}, x_{i2}, \dots, x_{iT})$, $x_{it} = (\mathbf{1}, y_{i0}, v_{it}^\top, g_i, v_{it}^\top g_i)^\top$,

$\beta = (\mu, \gamma_0, \gamma_1^\top, \gamma_{2i}, \gamma_3^\top)^\top$, $\mathbf{1}_T$ is the $T \times 1$ vector of 1, and I_T is the $T \times T$ identity matrix.

Estimation of the unknown parameters is often pursuit by restricted maximum likelihood or maximum likelihood.

For the analysis by GEE, we need to specify the first two moments of y_{it} . Specifically, we consider

$$E(y_{it} | x_{it}) = x_{it}^\top \beta, \quad \text{Var}(y_{it} | x_{it}) = \sigma^2 \quad (2)$$

for $t = 1, 2, \dots, T$, and an exchangeable working covariance matrix W . Estimate of β is

then obtained by solving the GEE $\frac{1}{n\sigma^2} \sum_{i=1}^n X_i^\top W^{-1} (Y_i - X_i^\top \beta) = 0$.

In either (1) or (2), the treatment effect at visit t ($t = 1, 2, \dots, T-1$) is given by $\gamma_2 + \gamma_{3t}$ and the effect at visit T is γ_2 . They are denoted by θ_t , for $t = 1, 2, \dots, T$. The code used in Statistical Analysis System (SAS) Version 9.2 for the analysis is provided in Appendix A.

2.2. Analysis by Using LOCF with MEM or GEE

The analysis is performed by replacing the missing values using the LOCF approach (i.e., the last observed value before a missing value is used to impute the missing value) before the MEM in (1) or the GEE in (2).

2.3. Analysis by using MI with MEM or GEE

The MI proceeds by first imputing the missing values for m (> 1) times, and thereby generates m complete datasets. This is opposed to single imputation which does not account the uncertainty due to imputation. There are various imputation methods depending on the missing value pattern (Rubin, 1987; Yang, Li, & Shoptaw, 2008). Because any intermittent y_{it} for a subject may be missing, the Markov Chain Monte Carlo (MCMC) method is used for imputation.(Schafer, 1997) In each imputation, the MCMC method generates a chain of sequentially associated values until they stabilized and the converged values are used to impute the missing values (Schafer, 1997). Then, each of the m complete datasets is analyzed by the linear MEM in (1) or the GEE in (2); thereby, m sets of parameter estimates are obtained. The m sets of estimates are combined for inference about θ_t . Specifically, if $\hat{\theta}_t^b$ and $\hat{\omega}_t^b$ ($b = 1, 2, \dots, m$) are respectively the estimate and estimated variance of θ_t from the b th dataset, the combined estimate and variance estimate of θ_t are $\bar{\theta}_t = \frac{1}{m} \sum_{b=1}^m \hat{\theta}_t^b$ and $\omega_t = \frac{1}{m} \sum_{b=1}^m \hat{\omega}_t^b + \frac{m+1}{m} \left[\frac{1}{m-1} \sum_{b=1}^m (\theta_t^b - \bar{\theta}_t)^2 \right]$, respectively.(Schafer, 1997) Note however the MCMC method is adequate only when

missing values are at most MAR. This facilitates fair comparisons with the use of MEM and GEE alone, which are valid only up to MAR and MCAR respectively.

In our applications, m is taken as 5 and 500 to assess the effects of using small and moderate sizes of imputations. The SAS code used for the analysis is provided in Appendix A.

3. DATA GENERATION AND ASSESSMENTS

The simulation study was programmed in SAS. Generation of the data was based on the setting of a clinical trial described in Section 2. The number of follow-up visits T was assumed to be 3. Then, $\{y_{it}\}_{t=0,1,2,3}$ was generated from

$$y_{it} = a_i + y_{i0} + \theta_t g_i + e_{it}$$

for $t = 1, 2, 3$, where y_{i0} , a_i and e_{it} were identically and independently distributed as $N(0, 1)$. That is, conditioned on y_{i0} , (y_{i1}, y_{i2}, y_{i3}) followed a multivariate normal distribution with mean $(\theta_1, \theta_2, \theta_3)g_i$ and covariance matrix $V = I_3 + 1_3 1_3^\top$.

The $(\theta_1, \theta_2, \theta_3)$ represented the treatment effect at the three follow-up visits, which was taken as $(1, 0.5, 0)$ and $(1, 0, 0)$. The first corresponds to a steady decrease of treatment effect, which does not favor the use of LOCF to handle missing values at all visits. The second corresponds to a sharply diminished treatment effect at the second follow-up visit but no change between visits 2 and 3, which favors the use of LOCF to handle missing values at the third follow-up visit. The generated dataset had no missing values and was referred as the complete dataset.

Then, incomplete datasets under the three missing value scenarios were generated from the complete dataset. Without a loss of generality, only values at visits 1, 2, and 3 of the complete dataset could be made missing. By specifying $p_m \times 100\%$ of missing values to be generated, a dataset with MCAR values (MCAR dataset) was obtained by randomly deleting values from the complete dataset with probability p_m . To generate a dataset with MAR values (MAR dataset), values in a visit were deleted if they corresponded to the upper $p_m \times 100\%$ observed values in the previous visit. For the dataset with MNAR values (MNAR dataset), the upper $p_m \times 100\%$ of values in each visit were deleted. The p_m was chosen as 0.30.

The number of subjects, n , was assumed to be 200 with 100 subjects in each group. With this sample size, the powers of detecting treatment effects of 1 at visit 1 and 0.5 at visit 2 were 0.9988 and 0.7054, respectively, when either the GEE or the linear MEM was used. Thus, the sample size would enable us to examine the performance of various methods when the power was high or of moderate size. When 30% of the values were MCAR, the powers became 0.9869 and 0.5524, respectively, and thus exaggerated the difference in power. Details of the calculation are shown in Appendix B. Note: We did not calculate the power when values were MAR and MNAR, because those situations may result in biased estimates and the power would not really reflect the chance of detecting a discernible effect.

The data generation procedure was repeated to generate 1000 sets of complete and missing datasets under various missing value scenarios. By denoting them as $(\hat{\theta}_{1s}, \hat{\theta}_{2s}, \hat{\theta}_{3s})$

($s = 1, 2, \dots, 1000$), the following summary statistics were computed for assessing each method of analysis.

1. Mean estimate, i.e. $\sum_{s=1}^{1000} \hat{\theta}_{1s} / 1000$, $\sum_{s=1}^{1000} \hat{\theta}_{2s} / 1000$ and $\sum_{s=1}^{1000} \hat{\theta}_{3s} / 1000$.
2. SE, the square root of the mean estimated variances of the estimates.
3. Bias, i.e. $\sum_{s=1}^{1000} \hat{\theta}_{1s} / 1000 - \theta_1$, $\sum_{s=1}^{1000} \hat{\theta}_{2s} / 1000 - \theta_2$ and $\sum_{s=1}^{1000} \hat{\theta}_{3s} / 1000 - \theta_3$.
4. Mean squared error (MSE), i.e. $\sum_{s=1}^{1000} (\hat{\theta}_{1s} - \theta_1)^2 / 1000$, $\sum_{s=1}^{1000} (\hat{\theta}_{2s} - \theta_2)^2 / 1000$ and $\sum_{s=1}^{1000} (\hat{\theta}_{3s} - \theta_3)^2 / 1000$. The MSE measures the sampling variance of each estimator.
5. Ratio, i.e., a mean estimated variance divided by the corresponding sampling variance.
6. 95% coverage, i.e., the proportion of estimated 95% confidence intervals covering the corresponding true values.

4. SIMULATION RESULTS

Table 1 compares the bias of various methods of estimating treatment effects across visits under the three missing value scenarios when $(\theta_1, \theta_2, \theta_3) = (1, 0.5, 0)$ and the number of imputations was 5.

Under MCAR, the MEM and GEE methods yielded unbiased estimates, but any combination of the LOCF or MI with either the MEM or the GEE method resulted in substantial biases in all visits. The bias was the largest when LOCF was used. At visit 1, when LOCF was used, a missing value from a subject was replaced by the observed value of the subject at visit 0. Hence, the imputed value was 0, on average, for both the control and treatment groups. Therefore, treatment effect at visit 1 was underestimated. By the

same token at visit 2, the use of LOCF led to overestimation of the treatment effect because missing values in the treatment group were imputed mostly by 1, which is larger than the true value of 0.5, whereas missing values in the control group were still mostly imputed by 0. Similarly, using LOCF also overestimated the treatment effect at visit 3. In contrast, MI for a missing value of a subject made use of all observed values from visits other than when the value was missing. Therefore, naïve expected values used to impute missing values at visits 1, 2, and 3 would be $(0+0.5+0)/3 = 0.17$, $(0+1+0)/3 = 0.33$, and $(0+1+0.5)/3 = 0.50$, respectively, for the treatment group, and all 0s for the control group. With the true treatment effects of $(\theta_1, \theta_2, \theta_3) = (1, 0.5, 0)$ and 30% missing values, the biases at the three follow-up visits should be approximately $(0.17-1)(0.3) = -0.25$, $(0.33-0.5)(0.3) = -0.05$ and $(0.50-0)(0.3) = 0.15$, respectively. These values were consistent with those reported in Table 1. Subsequently, we envisaged that MI may yield a larger bias than LOCF when the last observation was closer to the true value of the missing value than the average of observations from all other visits was. Indeed, in the simulation study with $(\theta_1, \theta_2, \theta_3) = (1, 0, 0)$, treatment effect at visit 3 was slightly more overly estimated by MI (bias = 0.0668) than by LOCF (bias = 0.0619). Regarding the precision of the estimates, SEs from all methods were only moderately higher than the corresponding values in the analysis of complete datasets.

Under MAR, there was a very slightly higher general bias by the MEM and GEE methods, but the estimates remained essentially unbiased. The LOCF and MI again yielded substantially biased estimates. However, MI under MAR resulted in smaller bias as well as a smaller SE than under MCAR because the accuracy of imputation by other observed

values was expectedly higher when the probability of data missing is in fact related to the observed values.

Under MNAR, there was a general reduction in SEs due to reduced variance resulted from deleting “large” values. However, there was considerable bias in all methods and a general underestimation of treatment effects when the treatment effect was positive (i.e., at visits 1 and 2). This happened because values in the treatment group were likely to be higher than those in the control group and thus more likely to be missing. Bias from the use of MEM or GEE alone remained the smallest when treatment effect was positive. At visit 3 when there was no hypothesized treatment effect, the bias was smaller because values in the two groups had equal probabilities of being missing.

In general, the MEM and GEE methods behaved very similarly, regardless of whether LOCF or MI was used and regardless of missing value scenarios. SEs when LOCF was used were generally smaller than those resulting from the other methods due to the use of initial values for imputation that increased the estimation precision when the initial value was also adjusted in the analysis.

Table 2 shows the performance of the methods of analysis on MSE, ratio and 95% coverage under the various missing value scenarios when $(\theta_1, \theta_2, \theta_3) = (1, 0.5, 0)$. Under MCAR or MAR, the use of LOCF resulted in the poorest MSE and 95% coverage. Although MI may yield a smaller MSE than the other methods, it had a more overly estimated variance. MEM and GEE used alone gave small MSEs, did not severely underestimate the variance compared with the sampling variance, and had good coverage.

Under MNAR when there were positive treatment effects at visits 1 and 2, the MSE was generally inflated due to the increase in bias. Particularly, both LOCF and MI resulted in inaccurate estimated variance compared with the sampling variance and thus yielded poor 95% coverage. In contrast, MEM and GEE used alone gave unbiased variance estimates, and the corresponding 95% coverage was the best compared with other methods, regardless of whether there was a positive treatment effect.

When the number of imputations used in MI was 500, similar results were obtained. When $(\theta_1, \theta_2, \theta_3) = (1, 0, 0)$, similar phenomena were also observed except where indicated.

5. AN EXAMPLE

A randomized, controlled clinical trial was conducted to examine the effects of the Chinese exercise qigong vs. a control exercise in patients with type 2 diabetes mellitus (Lee et al., 2003; Lee et al., 2002). The study was conducted in the Queen Mary Hospital of Hong Kong with a study protocol and informed consent forms approved by the hospital's Research Ethics Committee.

There were several outcomes of interest, but we focused only on measurements taken by the 36-item Medical Outcome Study Short Form (Hong Kong version), a generic health-related quality of life questionnaire making up eight distinct constructs: physical function (PF), social function (SF), role physical (RP), role emotional (RE), vitality (VT), mental health (MH), general health (GH), and bodily pain (BP). The questionnaire was administered to 100 study patients (37 men and 63 women) before randomization, as well

as 6, 10, and 18 weeks afterward. Patients allocated to the qigong group learned qigong during the first four weeks and were advised to practice qigong daily. Effects of qigong on the eight scales have been examined elsewhere (Lee et al., 2002). In particular, there was no evidence that qigong improved PF. However, it is of interest to examine whether men or women, after practicing qigong, had more improvement in PF. Therefore, we focused on the analysis of how the patient's sex affected PF over time in the 50 patients (27 men and 23 women) who practiced qigong.

Of the 50 patients, one dropped out after randomization and another withdrew after week 10. Therefore, there were only a few missing values in the raw dataset, and their effects were deemed negligible. About 15% of missing values were generated using the MCAR, MAR, and MNAR scenarios, and the MEM and GEE with and without LOCF and MI were applied.

Table 3 summarizes the estimated effect of the patient's sex on PF at each visit. Based on the analysis of the raw dataset, female patients had higher PF values than male patients at 10 weeks but not at other visits.

Under the missing value scenarios used in our estimates, the use of MEM and GEE alone on the incomplete datasets generally provided the same conclusions as when they were applied on the raw dataset. In contrast, the additional use of LOCF or MI generally yielded substantially different p-values or estimates. In particular, LOCF gave a few false-positive errors because it generally led to biased estimates. The MI generally led to insignificant results along with larger standard errors. MI yielded, at visit 3 in the MCAR

dataset, a biased estimate higher than that of LOCF. Therefore, the use of MEM or GEE alone appeared to be sufficient.

6. DISCUSSIONS

In response to the recent interest in examining the use of MI together with MEM or GEE, we performed the first simulation study to assess the use of MI with GEE for longitudinal analysis when there are missing values. Moreover, we also assessed for the first time, the performance of using MI with MEM when values are MNAR. When MEM or GEE are used for the analysis of incomplete longitudinal trials, the use of LOCF or MI appears to be unnecessary under all missing value scenarios.

In general, using LOCF with either MEM or GEE may lead to substantial bias compared with other methods. The bias resulting from using LOCF in longitudinal studies has been well documented in the literature (Cook, Zeng, & Yi, 2004; Liu & Gould, 2002; Siddiqui & Ali, 1998). Despite this, LOCF has been very commonly used in the analysis of clinical trials. MI emerged as an attractive alternative imputation method, taking into account the uncertainty about the imputed values. However, in our simulation study, the use of MI with MEM or GEE also resulted in biased effect estimates. The bias may even be worse than that of LOCF in certain scenarios under the MCAR, a missing value scenario often considered negligible. Moreover, MI often yields overestimated variances, possibly due to imputation uncertainty. Nielsen (2003) showed that multiple imputation methods can sometimes be improper and that even a proper multiple imputation method can be inefficient. Moreover, the use of different methods of handling missing values may

influence sample size requirements (Auleley et al., 2004). Therefore, the use of MEM or GEE without MI or LOCF is generally sufficient.

When values are MNAR, the use of MEM or GEE alone may result in considerable bias. Because MNAR and MAR are, unfortunately, indistinguishable unless external information about the missing value scenario is available, sensitivity analysis is often advisable to guard against having overly optimistic treatment effect estimates (Molenberghs et al., 2004). The sensitivity analysis can be performed on different sets of subjects according to their compliance, or imputation methods that have a predictable direction of bias (e.g., the worst-case method, in which missing values in the treatment and control groups are replaced by the respective worst- and best-case values) can be used. Nevertheless, the use of MI by MCMC would generally result in larger bias and poorer 95% coverage. Although the contrary was observed when treatment effect is nil, the bias was small and the coverage was reasonable when only MEM or GEE was used. Therefore, the use of MI by MCMC does not remarkably improve the estimation of treatment effects. Note however we have not assessed the use of MI by algorithms based on MNAR as this does not enable a fair comparison with the use of MEM or GEE alone which is only valid up to MAR or MCAR respectively (Yang et al., 2008).

There were no notable differences in the results between the use of 5 and 500 imputations. Generally, the number of imputations should be guided by the relative efficiency. However, there is often no practical benefit to using more than 5 to 10 imputations unless missing values are unusually many (Schafer, 1999). Indeed, with 50%

missing values, the relative efficiency for using 5 imputations is 1.049 which is not remarkable.

Although it was not our intention to compare MEM and GEE in our simulation study, we did find that the two methods performed quite similarly in all missing value scenarios. However, we examined only situations in which correlations among the repeated measurements were exchangeable. GEE has the well-known flexibility and robustness to be the choice of the covariance matrix for repeated measurements, but it assumes the MCAR scenario. On the other hand, MEM may also accommodate a wide range of covariance structures by using random effects (Fong, Lam, Lawless, & Lee, 2001; Lawless & Fong, 1999). Moreover, MEM uses likelihood-based inference and thus produces consistent estimates under both MCAR and MAR. In general, MEM appears to have more advantages for the analysis of explicative studies.

APPENDIX A: SAS CODE FOR THE ESTIMATION METHODS USED

ID is the subject identification number, VISIT is the number of the visit (0, 1, 2, 3), GROUP is the number of the treatment group (0 = control; 1 = treatment), SCORE0 is the baseline measurement, and SCORE n is the measurement at visit number n .

1. The SAS code for the analysis by the linear MEM given in (1) is:

```
proc mixed;

class ID VISIT GROUP;

model SCORE = SCORE0 VISIT GROUP VISIT*GROUP ;

estimate 'Group at Visit 1'  GROUP -1 1 VISIT*GROUP -1 1 0 0 0 0;

estimate 'Group at Visit 2'  GROUP -1 1 VISIT*GROUP 0 0 -1 1 0 0;

estimate 'Group at Visit 3'  GROUP -1 1 VISIT*GROUP 0 0 0 0 -1 1;

random intercept /subject=ID;

run;
```

2. The SAS code for the analysis by the GEE given in (2) is:

```
proc genmod;

class ID VISIT GROUP;

model SCORE = SCORE0 VISIT GROUP VISIT*GROUP;

estimate 'Group at Visit 1'  GROUP -1 1 VISIT*GROUP -1 1 0 0 0 0;

estimate 'Group at Visit 2'  GROUP -1 1 VISIT*GROUP 0 0 -1 1 0 0;

estimate 'Group at Visit 3'  GROUP -1 1 VISIT*GROUP 0 0 0 0 -1 1;

repeated subject=ID /type=exch;
```

```
run;
```

3. The SAS code for the analysis by MI and MEM is:

```
proc mi out=sim.mi nimpute=5 seed=18039;
```

```
var SCORE0–SCORE3; run;
```

```
data sim.ana; set sim.mi;
```

```
array dum [3] SCORE1 SCORE2 SCORE3;
```

```
do i = 2 to 4; score = dum[i-1]; VISIT=i-1; output; end;
```

```
drop SCORE1–SCORE3 i; run;
```

```
proc sort data=sim.ana; by _imputation_ ID descending VISIT ; run;
```

```
proc mixed data=sim.ana covtest noitprint noprofile noinfo noclprint;
```

```
by _imputation_;
```

```
class id VISIT GROUP;
```

```
model score = SCORE0 VISIT GROUP VISIT*GROUP /s covb;
```

```
random intercept /subject=ID;
```

```
ods output solutionF=sim.sol covb=sim.covb;
```

```
run;
```

```
data sim.sol(type=EST); set sim.sol;
```

```
if effect='visit' and visit=1 then effect='visit1';
```

```
if effect='visit' and visit=2 then effect='visit2';
```

```
if effect='visit' and visit=3 then delete;
```

```
if effect='group' and group=0 then effect='group0';
```

```

if effect='group' and group=1 then delete;

if effect='visit*group' then do;
    if visit=1 and group=0 then effect='v1_g0'; else
    if visit=2 and group=0 then effect='v2_g0'; else delete;
end; run;

data sim.covb(type=covb); set sim.covb;

if effect='visit' and visit=1 then effect='visit1';
if effect='visit' and visit=2 then effect='visit2';
if effect='visit' and visit=3 then delete;
if effect='group' and group=0 then effect='group0';
if effect='group' and group=1 then delete;
if effect='visit*group' then do;
    if visit=1 and group=0 then effect='v1_g0'; else
    if visit=2 and group=0 then effect='v2_g0'; else delete;
end;

drop col5 col7 col9 col11-col13;

rename col6=Col5 col8=Col6 col10=Col7;

if row=6 then row=5; if row=8 then row=6; if row=10 then row=7;

run;

proc mianalyze parms=sim.sol covb(effectvar=rowcol)=sim.covb;
    modeleffects intercept score0 visit1 visit2 group0 v1_g0 v2_g0;

```

```
Group: test -group0-v1_g0, -group0-v2_g0, -group0;  
run;
```

4. The SAS code for the analysis by MI and GEE is:

```
proc genmod data=sim.ana ; by _imputation_ ;  
  
class ID VISIT GROUP;  
  
model score = SCORE0 VISIT GROUP VISIT*GROUP /covb;  
  
repeated subject=ID /type=exch;  
  
ods output ParameterEstimates=sim.sol Covb=sim.covb ParmInfo=sim.info;  
  
run;  
  
data sim.sol; set sim.sol;  
  
if parameter='visit' and level1=1 then parameter='visit1';  
if parameter='visit' and level1=2 then parameter='visit2';  
if parameter='visit' and level1=3 then delete;  
if parameter='group' and level1=0 then parameter='group0';  
if parameter='group' and level1=1 then delete;  
if parameter='visit*group' then do;  
if level1=1 and level2=0 then parameter='v1_g0'; else  
if level1=2 and level2=0 then parameter='v2_g0'; else delete;  
end; run;  
  
data sim.covb(type=covb); set sim.covb;  
  
rename Prm6=Prm5 Prm8=Prm6 Prm10=Prm7;
```

```

if RowName='Prm6' then RowName='Prm5';

if RowName='Prm8' then RowName='Prm6';

if RowName='Prm10' then RowName='Prm7'; run;

data sim.info; set sim.info;

if effect='visit' and visit=1 then effect='visit1';

if effect='visit' and visit=2 then effect='visit2';

if effect='visit' and visit=3 then delete;

if effect='group' and group=0 then effect='group0';

if effect='group' and group=1 then delete;

if effect='visit*group' then do;

    if visit=1 and group=0 then effect='v1_g0'; else

    if visit=2 and group=0 then effect='v2_g0'; else delete;

end;

if Parameter='Prm6' then Parameter='Prm5';

if Parameter='Prm8' then Parameter='Prm6';

if Parameter='Prm10' then Parameter='Prm7'; run;

proc mianalyze parms=sim.sol covb=sim.covb parminfo=sim.info;

    modeleffects intercept SCORE0 VISIT1 VISIT2 GROUP0 v1_g0 v2_g0;

    Group: test -group0-v1_g0, -group0-v2_g0, -group0;

run;

```


APPENDIX B: POWER ANALYSES

We aim to determine the power of detecting a treatment effect of size θ_t^0 at visit t with a sample size of n and a maximum false positive error rate of α under Model (3). In the sequel, we concern the testing of

$$H_0 : K_t \beta = 0 \text{ against } H_A : K_t \beta = \theta_t^0$$

where K_t is the t^{th} row of the matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Estimate of β may be obtained by using GEE, denoted by $\hat{\beta}_{GEE}$, and linear MEM, denoted by $\hat{\beta}_{MEM}$.

When there are no missing values, $\hat{\beta}_{GEE}$ asymptotically follows $N(\beta, n^{-1}\Sigma_\beta)$

where

$$\Sigma_\beta = E^{-1}(X_1 V^{-1} X_1^\top)$$

(Tu et al., 2007). $E(X_1 V^{-1} X_1^\top)$ can be easily evaluated with V known, $y_{10} \sim N(0, 1)$ and $g_1 \sim \text{Bin}(1, 1/2)$. Then, the power can be calculated as

$$1 - F_{\chi^2(a, c)}(1 - p_{1-\alpha}) \quad (4)$$

where $F_{\chi^2(a, c)}(p_{1-\alpha}) = 1 - \alpha$, $F_{\chi^2(a, c)}$ represents the cumulative distribution function of a χ^2 distribution with a degrees of freedom and noncentrality parameter c , and $c = n(\theta_t^0)^2 (K_t \Sigma_\beta K_t^\top)^{-1}$. With $n = 200$ and $\theta_t^0 = 1$, we have $K_t \Sigma_\beta K_t^\top = 8$, $c = 25$ and

power = 0.9988. With $n = 200$ and $\theta_2^0 = 0.5$, we have $K_t \Sigma_\beta K_t^\top = 8$, $c = 6.25$ and power = 0.7054. On the other hand, $\hat{\beta}_{MEM}$ asymptotically follows $N(\beta, n^{-1} \Sigma_\beta)$ again. Therefore, the power of using linear MEM is identical to that of using GEE.

When missing values are MCAR, let r_{it} be 1 if y_{it} is observed and 0 if otherwise. Then, with $R_i = \text{diag}(r_{it})$ and $A_i = 2I_3$, $\hat{\beta}_{GEE}$ asymptotically follows $N(\beta, n^{-1} \Sigma_\beta)$ but now

$$\Sigma_\beta = B^{-1} \Sigma_U B^{-1}$$

where $B = E(X_1 R_1 A_1^{-1} X_1^\top)$ and $\Sigma_U = E(X_1 A_1^{-1} R_1 V R_1 A_1^{-1} X_1^\top)$ (Tu et al., 2007). Again, with V known, $y_{10} \sim N(0, 1)$, $g_1 \sim \text{Bin}(1, 1/2)$ and $r_{11} \sim \text{Bin}(1, 1 - p_m)$, both B and Σ_U can be computed. Then, the power can be calculated as in (4). With $n = 200$, $p_m = 0.3$ and $\theta_1^0 = 1$, we have $K_t \Sigma_\beta K_t^\top = 11.43$, $c = 17.5$ and power = 0.9869. With $n = 200$, and $\theta_2^0 = 0.5$, we have $K_t \Sigma_\beta K_t^\top = 11.43$, $c = 4.375$ and power = 0.5524. On the other hand, as in the case with no missing values, $\hat{\beta}_{MEM}$ asymptotically follows $N(\beta, n^{-1} \Sigma_\beta)$ again and thus use of linear MEM results in the same power as in the use of GEE.

REFERENCES

- Auleley, G. R., Giraudeau, B., Baron, G., Maillefert, J. F., Dougados, M., & Ravaud, P. (2004). The methods for handling missing data in clinical trials influence sample size requirements. *Journal of Clinical Epidemiology*, *57*(5):447-453.
- Beunckens, C., Molenberghs, G., & Kenward, M. G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials*, *2*:379-386.
- Carriere, I., & Bouyer, J. (2002). Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons. *BMC Medical Research Methodology*, *2*(1):15.
- Chan, Y. M., Lee, P. W., Fong, D. Y., Fung, A. S., Wu, L. Y., Choi, A. Y., Ng, T. Y., Ngan, H. Y., & Wong, L. C. (2005). Effect of individual psychological intervention in chinese women with gynecologic malignancy: a randomized controlled trial. *Journal of Clinical Oncology*, *23*(22):4913-4924.
- Cook, R. J., Zeng, L., & Yi, G. Y. (2004). Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics*, *60*(3):820-828.
- Dahmen, G., & Ziegler, A. (2004). Generalized estimating equations in controlled clinical trials: Hypotheses testing. *Biometrical Journal*, *46*(2):214-232.
- Edwards, L. J. (2000). Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Pediatric Pulmonology*, *30*(4):330-344.

- Fong, D. Y., Lam, K. F., Lawless, J. F., & Lee, Y. W. (2001). Dynamic random effects models for times between repeated events. *Lifetime Data Analysis*, 7(4):345-362.
- Goldstein, H., Browne, W., & Rasbash, J. (2002). Multilevel modelling of medical data. *Statistics in Medicine*, 21(21):3291-3315.
- Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics*, 58(2):342-351.
- ICH. (1998). E9: Statistical Principles for Clinical Trials Retrieved 16 Nov 2007, from <http://www.ich.org/>
- Kang, T., Kraft, P., Gauderman, W. J., & Thomas, D. (2003). Multiple imputation methods for longitudinal blood pressure measurements from the Framingham Heart Study. *BMC Genetics*, 4 Suppl 1:S43.
- Lawless, J. F., & Fong, D. Y. T. (1999). State duration models in clinical and observational studies. *Statistics in Medicine*, 18(17-18):2365-2376.
- Lee, K. K., Leung, E., Fong, D., Wong, S., Tam, S. C. F., Wong, V., & Lam, K. S. L. (2003). The beneficial effects of Qigong and conventional exercise on blood pressure and anthropometric indices in type 2 Chinese diabetic patients. *Hong Kong Medical Journal*, 9 Suppl(1):34.
- Lee, K. K., Leung, E., Wong, S., Tam, S. C. F., Wong, V., & Lam, K. S. L. (2002). The effects of Qigong and conventional exercise on type 2 diabetic patients. *Diabetes*, 51 Suppl(2):A245.

- Lee, Y., & Nelder, J. (2004). Conditional and marginal models: Another view. *Statistical Science*, *19*(2):219-228.
- Lindsey, J. K. (2000). Obtaining marginal estimates from conditional categorical repeated measurements models with missing data. *Statistics in Medicine*, *19*(6):801-809.
- Lindsey, J. K., & Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, *17*(4):447-469.
- Liu, G., & Gould, A. L. (2002). Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *Journal of Biopharmaceutical Statistics*, *12*(2):207-226.
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M., Mallinckrodt, C., & Carroll, R. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, *5*(3):445-464.
- Nielsen, S. (2003). Proper and improper multiple imputation. *International Statistical Review*, *71*(3):593-607.
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, *63*(3):581-590.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (1st ed.). London: Chapman & Hall.
- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*(1):3-15.

- Shieh, Y.-Y. (2003). Imputation methods on general linear mixed models of longitudinal studies, *Federal Committee on Statistical Methodology Research Conference*. Arlington Virginia.
- Siddiqui, O., & Ali, M. W. (1998). A comparison of the random-effects pattern mixture model with last-observation-carried-forward (LOCF) analysis in longitudinal clinical trials with dropouts. *Journal of Biopharmaceutical Statistics*, 8(4):545-563.
- Tu, X. M., Zhang, J., Kowalski, J., Shults, J., Feng, C., Sun, W., & Tang, W. (2007). Power analyses for longitudinal study designs with missing data. *Statistics in Medicine*, 26(15):2958-2981.
- Twisk, J., & de Vente, W. (2002). Attrition in longitudinal studies. How to deal with missing data. *Journal of Clinical Epidemiology*, 55(4):329-337.
- Twisk, J. W. (2004). Longitudinal data analysis. A comparison between generalized estimating equations and random coefficient analysis. *European Journal of Epidemiology*, 19(8):769-776.
- Wagner, A. K., Fabio, A., Puccio, A. M., Hirschberg, R., Li, W., Zafonte, R. D., & Marion, D. W. (2005). Gender associations with cerebrospinal fluid glutamate and lactate/pyruvate levels after severe traumatic brain injury. *Critical Care Medicine*, 33(2):407-413.
- Wood, A. M., White, I. R., Hillsdon, M., & Carpenter, J. (2005). Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes. *International Journal of Epidemiology*, 34(1):89-99.

Yang, X., Li, J., & Shoptaw, S. (2008). Imputation-based strategies for clinical trial longitudinal data with nonignorable missing values. *Stat Med*, 27(15):2826-2849.

Table 1. Comparisons of the bias of using last observation carried forward (LOCF) and multiple imputation (MI) on mixed effects model (MEM) and generalized estimating equations (GEE) under various missing value scenarios

	Treatment effect at Visit 1 ($\theta_1 = 1$)			Treatment effect at Visit 2 ($\theta_2 = 0.5$)			Treatment effect at Visit 3 ($\theta_3 = 0$)		
	Mean estimate	SE**	Bias	Mean estimate	SE**	Bias	Mean estimate	SE**	Bias
Complete Dataset									
MEM	1.0044	0.2003	0.0044	0.5033	0.2003	0.0033	-0.0011	0.2003	-0.0011
GEE	1.0044	0.1992	0.0044	0.5033	0.1989	0.0033	-0.0011	0.1989	-0.0011
MCAR Dataset*									
MEM	1.0008	0.2297	0.0008	0.4979	0.2297	-0.0021	-0.0007	0.2297	-0.0007
GEE	1.0007	0.2279	0.0007	0.4978	0.2277	-0.0022	-0.0008	0.2275	-0.0008
LOCF + MEM	0.7032	0.1970	-0.2968	0.5762	0.1970	0.0762	0.1705	0.1970	0.1705
LOCF + GEE	0.7032	0.1730	-0.2968	0.5762	0.2088	0.0762	0.1705	0.2035	0.1705
MI + MEM	0.7482	0.2310	-0.2518	0.4317	0.2292	-0.0683	0.1054	0.2295	0.1054
MI + GEE	0.7482	0.2297	-0.2518	0.4317	0.2278	-0.0683	0.1054	0.2282	0.1054
MAR Dataset*									
MEM	1.0013	0.2289	0.0013	0.5063	0.2111	0.0063	0.0012	0.2219	0.0012
GEE	1.0019	0.2277	0.0019	0.5021	0.2090	0.0021	0.0031	0.2200	0.0031
LOCF + MEM	0.7005	0.2010	-0.2995	0.6674	0.2010	0.1674	0.1280	0.2010	0.1280
LOCF + GEE	0.7005	0.1728	-0.2995	0.6674	0.2142	0.1674	0.1280	0.2094	0.1280
MI + MEM	0.7714	0.2296	-0.2286	0.4871	0.2165	-0.0129	0.0723	0.2246	0.0723
MI + GEE	0.7714	0.2282	-0.2286	0.4871	0.2151	-0.0129	0.0723	0.2232	0.0723
MNAR Dataset*									
MEM	0.6431	0.1839	-0.3569	0.3024	0.1827	-0.1976	-0.0346	0.1822	-0.0346
GEE	0.6464	0.1814	-0.3536	0.3023	0.1796	-0.1977	-0.0380	0.1793	-0.0380
LOCF + MEM	0.4864	0.1563	-0.5136	0.1744	0.1564	-0.3256	-0.0382	0.1564	-0.0382
LOCF + GEE	0.4864	0.1391	-0.5136	0.1744	0.1655	-0.3256	-0.0382	0.1604	-0.0382
MI + MEM	0.4528	0.1831	-0.5472	0.2325	0.1811	-0.2675	0.0072	0.1797	0.0072
MI + GEE	0.4528	0.1822	-0.5472	0.2325	0.1801	-0.2675	0.0072	0.1788	0.0072

*MCAR, missing completely at random; MAR, missing at random; MNAR, missing not at random.

**SE, square root of the mean estimated variance over all generations of data.

Table 2. Comparisons of the mean squared error (MSE) and the 95% coverage of using last observation carried forward (LOCF) and multiple imputation (MI) on mixed effects model (MEM) and generalized estimating equations (GEE) under various missing value scenarios

	Treatment effect at Visit 1			Treatment effect at Visit 2			Treatment effect at Visit 3		
	$(\theta_1 = 1)$			$(\theta_2 = 0.5)$			$(\theta_3 = 0)$		
	MSE (*100)	Ratio**	95% coverage	MSE (*100)	Ratio**	95% coverage	MSE (*100)	Ratio**	95% coverage
Complete Dataset									
MEM	3.6191	1.11	95.80%	3.8061	1.05	94.90%	3.9488	1.02	95.40%
GEE	3.6191	1.10	95.60%	3.8061	1.04	94.70%	3.9488	1.00	95.00%
MCAR Dataset*									
MEM	4.9359	1.07	95.30%	5.1385	1.03	94.50%	5.0962	1.03	94.90%
GEE	4.9453	1.05	95.00%	5.1340	1.01	94.50%	5.1000	1.01	94.70%
LOCF + MEM	11.6432	1.37	70.30%	5.2797	0.83	91.00%	7.1774	0.91	85.10%
LOCF + GEE	11.6432	1.06	58.60%	5.2797	0.93	92.60%	7.1774	0.97	86.30%
MI + MEM	9.4530	1.71	87.20%	3.8647	1.54	97.20%	4.4604	1.57	96.90%
MI + GEE	9.4530	1.69	86.60%	3.8647	1.53	97.20%	4.4604	1.55	96.80%
MAR Dataset*									
MEM	4.6790	1.12	96.00%	4.3223	1.03	94.90%	4.8731	1.01	95.10%
GEE	4.6944	1.10	95.60%	4.3023	1.01	94.60%	4.8848	0.99	94.60%
LOCF + MEM	11.6051	1.53	71.80%	7.2085	0.92	85.20%	6.0139	0.92	89.10%
LOCF + GEE	11.6051	1.13	58.60%	7.2085	1.04	87.80%	6.0139	1.00	90.70%
MI + MEM	8.3919	1.66	89.30%	3.9587	1.19	96.80%	3.9990	1.45	97.60%
MI + GEE	8.3919	1.64	89.10%	3.9587	1.17	96.70%	3.9990	1.43	97.60%
MNAR Dataset*									
MEM	15.8586	1.08	49.10%	7.1624	1.02	80.70%	3.3981	1.01	94.40%
GEE	15.6176	1.06	48.80%	7.1599	0.99	79.80%	3.4135	0.98	93.40%
LOCF + MEM	28.0895	1.43	6.50%	13.5904	0.82	46.50%	3.3133	0.77	91.10%
LOCF + GEE	28.0895	1.13	3.40%	13.5904	0.92	50.10%	3.3133	0.81	91.40%
MI + MEM	31.6730	1.93	9.40%	9.0015	1.77	72.10%	1.9354	1.67	98.50%
MI + GEE	31.6730	1.91	9.30%	9.0015	1.75	71.50%	1.9354	1.65	98.50%

*MCAR, missing completely at random; MAR, missing at random; MNAR, missing not at random.

**Ratio of the mean estimated variance by each method to the corresponding sampling variance.

Table 3. Estimating gender effects on physical function in patients with type 2 diabetes mellitus who practiced qigong

	Female effect at Visit 1			Female effect at Visit 2			Female effect at Visit 3		
	Estimate**	Standard error	p-value	Estimate**	Standard error	p-value	Estimate**	Standard error	p-value
Raw Dataset									
MEM	3.0628	1.7868	0.090	4.3170	1.7868	0.018	3.2048	1.7975	0.078
GEE	3.0626	1.6966	0.071	4.3168	1.7914	0.016	3.2024	1.8043	0.076
MCAR Dataset*									
MEM	3.4870	1.9215	0.074	4.3178	1.8785	0.025	3.6058	1.9063	0.063
GEE	3.4411	1.7276	0.046	4.3347	1.8447	0.019	3.5857	1.8418	0.052
LOCF + MEM	4.1251	1.6658	0.015	4.0526	1.6658	0.017	4.0204	1.6658	0.018
LOCF + GEE	4.1251	1.4786	0.005	4.0526	1.7760	0.022	4.0204	1.7189	0.019
MI + MEM	2.7812	1.9229	0.150	3.6419	1.8846	0.055	2.8869	2.0863	0.173
MI + GEE	2.7812	1.8737	0.140	3.6419	1.8344	0.048	2.8869	2.0411	0.164
MAR Dataset*									
MEM	2.3189	2.0224	0.256	5.0998	2.1203	0.019	2.6264	2.2248	0.243
GEE	2.3155	1.9786	0.242	5.1311	2.0550	0.013	2.4232	2.1134	0.252
LOCF + MEM	1.4898	2.0222	0.463	1.9871	2.0451	0.334	1.4905	2.0451	0.468
LOCF + GEE	1.4899	1.8861	0.430	1.9890	1.9037	0.296	1.4924	2.0128	0.458
MI + MEM	0.9268	2.0234	0.648	3.0100	2.1645	0.170	2.4564	2.1740	0.264
MI + GEE	0.9268	1.9731	0.639	3.0100	2.1175	0.161	2.4564	2.1272	0.254
MNAR Dataset*									
MEM	2.3996	1.4315	0.098	2.8895	1.4515	0.051	0.8918	1.4567	0.543
GEE	2.4161	1.3889	0.082	2.8953	1.3860	0.037	0.8893	1.4136	0.529
LOCF + MEM	3.1927	1.3997	0.025	1.5261	1.3997	0.278	0.4391	1.3997	0.754
LOCF + GEE	3.1927	1.4243	0.025	1.5261	1.3386	0.254	0.4391	1.3544	0.746
MI + MEM	2.3577	1.5284	0.129	2.9356	1.3302	0.027	1.3167	1.5105	0.387
MI + GEE	2.3577	1.4949	0.121	2.9356	1.2914	0.023	1.3167	1.4766	0.376

*MCAR, missing completely at random; MAR, missing at random; MNAR, missing not at random.

**Females over males.