Published in final edited form as:
Biom J. 2009 August ; 51(4): 627–643. doi:10.1002/bimj.200800186.

On the Impact of Parametric Assumptions and Robust Alternatives for Longitudinal Data Analysis

Naiji Lu[1,2,3], **Wan Tang**[1,2], **Hua He**[1,2,3], **Qin Yu**[1], **Paul Crits-Christoph**[4], **Hui Zhang**[1], and **Xin Tu**[1,2,3]

[1]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14623

[2]Department of Psychiatry, University of Rochester, Rochester, NY 14623

[3]Center of Excellence, Canandaigua, NY 14424

[4]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104

Summary

Models for longitudinal data are employed in a wide range of behavioral, biomedical, psychosocial, and health-care related research. One popular model for continuous response is the linear mixed-effects model (LMM). Although simulations by recent studies show that LMM provides reliable estimates under departures from the normality assumption for complete data, the invariable occurrence of missing data in practical studies renders such robustness results less useful when applied to real study data. In this paper, we show by simulated studies that in the presence of missing data estimates of the fixed-effect of LMM are biased under departures from normality. We discuss two robust alternatives, the weighted generalized estimating equations (WGEE) and the augmented WGEE (AWGEE), and compare their performances with LMM using real as well as simulated data. Our simulation results show that both WGEE and AWGEE provide valid inference for skewed non-normal data when missing data follows the missing at random (MAR), the most popular missing data mechanism for real study data.

Keywords

Augmented weighted generalized estimating equations; double robust estimate; missing at random; surrogacy assumption; weighted generalized estimating equations

1 Introduction

Models for longitudinal data are employed in a wide range of behavioral, biomedical, psychosocial, and health-care related research studies. One popular modeling paradigm is the latent variable, or random-effect, based approach for addressing correlated responses arising from such study data (Demidenko, 2004; Fitzmaurice et al. 2004; Raudenbush and Bryk, 2002). For example, for continuous response, the linear mixed-effects model (LMM) is one of the most widely used methods for modeling longitudinal data. By using random-effects to account for correlations across multiple within-subject outcomes, the LMM extends the classic multiple linear regression for modeling cross-sectional data to a general longitudinal data setting.

One major drawback of LMM as well as other random-effects based models is their dependence on distribution assumptions for inference. In recent years, many software packages have implemented robust methods to improve the validity of inference in the presence of departures from assumed parametric models. For example, simulation studies of

Maas and Hox (2004; 2005) have found that estimates of the fixed-effects, or population parameters, of LMM are quite robust when model errors depart from the assumed normal distribution. However, these simulation studies have all been carried out under complete data. In most real studies, missing data invariably occurs and the robustness of estimates from LMM requires reassessment in the presence of missing data.

In this paper, we show by simulated data that the robustness of fixed-effect estimates of LMM is compromised under non-random missing data when the error term of the model deviates from the normality assumption. To address this key limitation, we investigate the performance of two robust alternatives, the weighted generalized estimating equations (WGEE) and augmented WGEE (AWGEE), for inference for non-normal longitudinal data. In Section 2, we briefly review the parametric LMM and the distribution-free formulation of this model. In Section 3, we discuss inference by the two dueling paradigms under missing data. In Section 4, we use simulated data to compare the performance of the various approaches for fitting non-normal longitudinal data under non-random missing data. In Section 5, we give our concluding remarks.

2 Models for Longitudinal Data

Over the past three decades, studies in biomedical and behavioral sciences have evolved from simple cross-sectional study designs to modern day longitudinal trials. As longitudinal study designs use subjects as their own controls, they provide a unique opportunity to study changes of outcomes of interest over time, causal effects and disease progression, in addition to providing more power for assessing treatment differences. In this section, we briefly review the two most popular approaches for modeling longitudinal data with continuous response.

2.1 Parametric Linear Mixed-effects Model

First, consider a relatively simple longitudinal study design with only two assessments, i.e., the so-called pre-post design. Let n denote the number of subjects and y_{it} some continuous outcome of interest from the ith subject at time t (= 1, 2). We are interested in modeling the mean response $\mathrm{E}(y_{it})$ of y_{it} over time. One popular approach is the linear mixed-effects model (LMM), which for such a pre-post study design, has the simple form:

$$y_{it}=\beta_0+\beta_1 I_{\{t=2\}}+b_i+\varepsilon_{it},\ b_i \sim N(0,\sigma_b^2),\ \varepsilon_{it} \sim N(0,\sigma^2),\ 1 \le i \le n,\ 1 \le t \le 2, \tag{1}$$

where $I_{\{\cdot\}}$ denotes a set indicator, and $N(\mu, \sigma^2)$ denotes a normal with mean μ and variance σ^2. In the LMM above, since $\mathrm{E}(y_{it}) = \beta_0 + \beta_1 I_{\{t=2\}}$, β_t represents the (population) mean of y_{it} at pre-(t = 1) and post-treatment (t = 2) and is known as the fixed or population effect. The latent b_i in (1) accounts for the variation across the individual subjects around the fixed effect or population mean and is known as the random effect.

Let $\mathrm{y}_i = (y_{i1}, y_{i2})^\top$. It then follows from (1) that

$$V=\mathrm{Var}(\mathrm{y}_i)=\begin{pmatrix} \sigma^2+\sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma^2+\sigma_b^2 \end{pmatrix}=(\sigma^2+\sigma_b^2)C_2(\rho), \tag{2}$$

where $\rho=\sigma_b^2/(\sigma_b^2+\sigma^2)$ is known as the intraclass correlation (ICC) and $C_m(\rho)$ denotes a $m \times m$ compound symmetry correlation matrix with a correlation coefficient ρ. The LMM for the pre-post study design will be used to illustrate the performance of LMM in Section 4.

More generally, the LMM for a longitudinal data with n assessments has the following form:

$$y_{it}=\mathrm{x}_{it}^{\top}\beta+\mathrm{z}_{it}^{\top}\mathrm{b}_i+\varepsilon_i,\ \mathrm{b}_i \sim N(\mathbf{0}, \sum\nolimits_b),\ \varepsilon_i \sim N(\mathbf{0}, \sigma^2\mathrm{I}_m),\ 1 \le i \le n,\ 1 \le t \le m, \tag{3}$$

where $\mathrm{x}_{it} = (1, x_{i1t} \ldots, x_{ipt})^{\top}$ ($\mathbf{z_{it}} = (1, z_{i1t} \ldots, z_{ilt})^{\top}$) denotes a $p \times 1$ ($l \times 1$) vector of covariates, $\mathbf{b}_i = (b_{i1},\ldots, b_{il})^{\top}$ a $l \times 1$ vector of normal random-effect, $\varepsilon_i = (\varepsilon_{i1},\ldots, \varepsilon_{im})^{\top}$ the error term for the model, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a multivariate normal with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, I_m the $m \times m$ identity matrix. Under (3), we have:

$$V_i=\mathrm{Var}(\mathrm{y}_i|\mathrm{x}_i, \mathrm{z}_i)=\mathrm{Z}_i\sum\nolimits_b\mathrm{Z}_i^{\top}+\sigma^2\mathrm{I}_m. \tag{4}$$

By setting $\mathrm{x}_{it}^{\top}=(1, I_{\{t=2\}})^{\top}$ and $\mathbf{z}_{it} = 1$ in (4), we immediately obtain (2).

Maximum likelihood (ML) is the most popular inference procedure for LMM (Demidenko, 2004; Raudenbush and Bryk, 2002). One major drawback of ML estimate is the dependence on the parametric assumptions; if data does not follow the normality assumptions in (1), model estimates may become biased or inconsistent. In recent years, many packages have adopted the sandwich variance estimate to address this issue (Goldstein, 1995; Rasbash et al., 2000; Raudenbush and Bryk, 2002). In this case, these procedures essentially yield variance estimates equivalent to those from the generalized estimating equations (GEE), which we review next.

2.2 Distribution-free Linear Model

Since the seminal work of Liang and Zeger (1986), the generalized estimating equations (GEE) approach has been widely used as an alternative for modeling longitudinal data. By modeling the marginal mean of the response at each assessment time, GEE eliminates both layers of distribution assumptions for the random effect and error term, thereby providing consistent estimates regardless of data distributions and the complexities of structure of correlated responses.

Within our context, consider the following linear model:

$$\mathrm{E}(y_{it}|\mathrm{x}_{it})=\mu_{it}=\mathrm{x}_{it}^{\top}\beta,\ 1 \le i \le n,\ 1 \le t \le m. \tag{5}$$

In (5), only the marginal mean at each time t is specified, which models the fixed-effect of LMM in (1) at time t. Under GEE, inference is based on the following score-like vector equation:

$$\mathrm{W}_n(\beta)=\sum_{i=1}^{n}\mathrm{W}_{ni}(\beta)=\sum_{i=1}^{n}\mathrm{G}_i(\mathrm{x}_i)\mathrm{S}_i(\beta)=\sum_{i=1}^{n}\mathrm{G}_i(\mathrm{x}_i)(\mathrm{y}_i-\mu_i)=0, \tag{6}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im})^\top$, $\mathbf{G}_i(\mathrm{x}_i)$ is some matrix function of $\mathrm{x}_i = \left(\mathrm{x}_{i1}^\top, \ldots, \mathrm{x}_{im}^\top\right)$. If the model in (5) is correct, then the GEE in (6) is unbiased, i.e., $\mathrm{E}[\mathbf{W}_n(\beta)] = \mathbf{0}$, regardless of the choice of $\mathbf{G}_i(\mathrm{x}_i)$, ensuring the consistency of estimate of $\boldsymbol{\beta}$ obtained as the solution to (6) (Liang and Zeger, 1986;Diggle et al. 2002). In most applications, $\mathbf{G}_i(\mathrm{x}_i)$ has the form:

$$\mathrm{G}_i(\mathrm{x}_i) = \mathrm{D}_i \mathrm{V}^{-1}(\alpha),\ \mathrm{D}_i = \frac{\partial}{\partial \beta}\mu_i,\ 1 \le i \le n, \tag{7}$$

where $\mathbf{V}(\boldsymbol{\alpha})$ denotes a working variance matrix parameterized by α.

The phrase working variance is used to emphasize the fact that $\mathbf{V}(\boldsymbol{\alpha})$ is not necessarily the true variance matrix of $\mathbf{y}_i$. For example, the simplest choice is $\mathbf{V} = \mathrm{diag}_t(\mathrm{Var}(y_{it}))$, where $\mathrm{diag}_t(\alpha_t)$ denotes a diagonal matrix with α_t on the tth diagonal. In this case, the correlated responses y_{it} are treated as if they are independent. In addition, there is no parameter associated with this particular working independence model. Another popular choice is the exchangeable or uniform compound symmetry correlation matrix, $V(\alpha) = \mathrm{diag}_t\left(\sqrt{\mathrm{Var}(y_{it})}\right) C_m(\rho)\, \mathrm{diag}_t\left(\sqrt{\mathrm{Var}(y_{it})}\right)$, where $C_m(\rho)$ denotes a $m \times m$ matrix correlation matrix with a common correlation ρ for any pair of the component responses of $\mathbf{y}_i$.

In addition to β, (6) also depends on $\boldsymbol{\alpha}$, though we have suppressed this dependence to highlight the fact that (6) is the equation for estimating β. Thus, before proceeding with inference about β, $\boldsymbol{\alpha}$ must be estimated (except for the working independence model). Although the consistency of $\hat{\boldsymbol{\beta}}$ does not depend on how $\boldsymbol{\alpha}$ is estimated, judicious choices of the type of estimates of $\boldsymbol{\alpha}$ are required to ensure the asymptotic normality of $\hat{\beta}$. In particular, if $\hat{\boldsymbol{\alpha}}$ is $\sqrt{n}$-consistent, $\hat{\boldsymbol{\beta}}$ is asymptotically normal with its asymptotic variance $\boldsymbol{\Sigma}_\beta$ given by (e.g. Liang and Zeger, 1986;Kowalski and Tu, Chap. 4, 2007):

$$\sum\nolimits_\beta = \mathrm{B}^{-1}\mathrm{E}\left(\mathrm{G}_i \mathrm{S}_i \mathrm{S}_i^\top \mathrm{G}_i^\top\right)\mathrm{B}^{-\top},\ \mathrm{B} = \mathrm{E}\left(\frac{\partial^\top}{\partial \beta}\mathrm{W}_{ni}(\beta, \alpha)\right).$$

A consistent estimate of $\boldsymbol{\Sigma}_\beta$ is obtained by substituting consistent estimates in place of the respective quantities above. For example, we can estimate $\mathbf{B}$ by $\widehat{\mathrm{B}} = \sum_{i=1}^n \frac{\partial^\top}{\partial \beta}\mathrm{W}_{ni}\left(\widehat{\beta}, \widehat{\alpha}\right)/n$. Since moment estimates are $\sqrt{n}$-consistent, $\boldsymbol{\alpha}$ is readily estimated for the working independence and exchangeable correlation models (e.g. Liang and Zeger, 1986).

3 Inference under Missing Data

In longitudinal studies, missing data are inevitable; subjects may simply quit or they may not show up at follow-up visits. We characterize the impact of missing data on model estimates through assumptions or missing data mechanisms, which allow us to ignore the multitude of reasons for missing data and focus on addressing their impact on estimation of model parameters. The missing completely at random (MCAR) assumption models a class of missing data that does not affect model estimates when completely ignored. For example, in a treatment study, missing data resulting from patient's relocation or scheduling conflict falls into this category. However, MCAR is not a plausible model when missing data are associated with treatment interventions such as patients' deteriorated or improved health

conditions due to treatment. By modeling the occurrence of missing data as a function of observed responses prior to the assessment point, the missing at random (MAR) assumption addresses this class of treatment related or response-dependent missing data.

Within the longitudinal study setting in Section 2, define a missing (or rather, observed) data indicator as:

$$\mathrm{r}_{it}=\begin{cases} 1 & \text{if } y_{it} \text{ is observed} \\ 0 & \text{if } y_{it} \text{ is missing} \end{cases}, \ \mathrm{r}_i=(r_{i1},\ldots,r_{im})^\top, \ 1\le t\le m, \ 1\le i\le n.$$

We assume no missing data at baseline $t = 1$ such that $r_{i1} = 1$ for all $1 \le i \le n$. Below, we first briefly review inference for the parametric LMM in (3) and then turn our attention to the distribution-free version in (5).

3.1 Parametric Model

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})^\top$ and $\mathrm{x}_i=\left(\mathrm{x}_{i1}^\top,\ldots,\mathrm{x}_{im}^\top\right)^\top$. Let y_i^o and y_i^m denote the observed and missing responses, respectively. Under likelihood based parametric inference, we jointly model the response $\mathbf{y}_i$ and missing data indicator $\mathbf{r}_i$.

The joint density function, $f(\mathrm{y}_i, \mathbf{r}_i \mid \mathrm{x}_i)$, can be factored into the product of marginal and conditional distributions:

$$f(\mathrm{y}_i,\mathrm{r}_i|\mathrm{x}_i)=f(\mathrm{y}_i|\mathrm{x}_i)\ f(\mathrm{r}_i|\mathrm{y}_i,\mathrm{x}_i). \tag{8}$$

Under MAR, the distribution of r_i depends only on the observed response, y_i^o, and thus:

$$f(\mathrm{r}_i|\mathrm{y}_i,\mathrm{x}_i)=f(\mathrm{r}_i|\mathrm{y}_i^o,\mathrm{y}_i^m,\mathrm{x}_i)=f(\mathrm{r}_i|\mathrm{y}_i^o,\mathrm{x}_i). \tag{9}$$

It follows from (8) and (9) that:

$$\begin{aligned} f(\mathrm{y}_i^o,\mathrm{r}_i|\mathrm{x}_i) &= \int f(\mathrm{y}_i^o,\mathrm{y}_i^m|\mathrm{x}_i)\ f(\mathrm{r}_i|\mathrm{y}_i^o,\mathrm{x}_i)\,d\mathrm{y}_i^m \\ &= f(\mathrm{r}_i|\mathrm{y}_i^o,\mathrm{x}_i,\theta_{\mathrm{y}|\mathrm{r}})\int f(\mathrm{y}_i^o,\mathrm{y}_i^m|\mathrm{x}_i)\,d\mathrm{y}_i^m \\ &= f(\mathrm{y}_i^o|\mathrm{x}_i,\theta_{\mathrm{y}})\ f(r_i|\mathrm{y}_i^o,\mathrm{x}_i,\theta_{y|r}). \end{aligned} \tag{10}$$

If $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_{y|r}$ are assumed disjoint, then following (10) the log-likelihood based on the joint observations $(\mathrm{y}_i^o, \mathrm{r}_i)$ is given by:

$$l(\theta)=\sum_{i=1}^{n}\log\left(f(\mathrm{y}_i^o|\mathrm{x}_i,\theta_{\mathrm{y}})\right)+\sum_{i=1}^{n}\log\left(f(\mathrm{r}_i|\mathrm{y}_i^o|\mathrm{x}_i,\theta_{y|r})\right)=l_1(\theta_y)+l_2(\theta_{y|r}).$$

Thus, inference about $\boldsymbol{\theta}_y$ can simply be based on the log-likelihood $l_1(\boldsymbol{\theta}_y)$.

Most packages provide inference about the parameters of interest $\boldsymbol{\theta}_y$ based on maximizing the likelihood function $l_1(\boldsymbol{\theta}_y)$. Under the model assumptions of LMM, estimates are consistent under both MCAR and MAR. When study data fail to follow the parametric assumptions, maximum likelihood estimates are no longer guaranteed to be consistent. We examine bias from such estimates using simulated data in Section 4.

3.2 Distribution-free Model

3.2.1 Weighted Generalized Estimating Equations—In the presence of missing data, we may apply the GEE in (6) to the observed responses, i.e.,

$$\mathrm{W}_n(\beta)=\sum_{i=1}^{n}\mathrm{W}_{ni}(\beta)=\sum_{i=1}^{n}\mathrm{G}_i(\mathrm{x}_i)\mathrm{R}_i\mathrm{S}_i=\sum_{i=1}^{n}\mathrm{G}_i(\mathrm{x}_i)\boldsymbol{\Delta}_i\,(\mathrm{y}_i-\mu_i)=0, \tag{11}$$

where $\boldsymbol{\Delta_i} = \mathrm{diag}_t(r_{it})$. However, the vector estimating equation in (11) is generally biased, i.e., $\mathrm{E}(\mathbf{W}_n(\boldsymbol{\beta})) \neq \mathbf{0}$, unless missing data follow the MCAR model. To obtain consistent estimates of β under MAR, we must revise the GEE above.

To illustrate the basic idea for modification, consider the relatively simple, pre-post design, with a homogeneous sample. We are interested in estimating the mean response at pre- and post-assessment, $\boldsymbol{\mu} = \mathrm{E}(\mathrm{y}_i) = (\mathrm{E}(y_{i1}), \mathrm{E}(y_{i2}))^\top$. By selecting the $\mathbf{G}_i(\mathrm{x}_i)$ according to (7), it follows from (11) that

$$\mathrm{W}_n(\mu)=R(\alpha)^{-1}\begin{pmatrix}1 & 0\\ 0 & r_{i2}\end{pmatrix}\left[\sum_{i=1}^{n}\begin{pmatrix}y_{i1}-\mu_1\\ y_{i2}-\mu_2\end{pmatrix}\right]=\mathbf{0}. \tag{12}$$

Solving the equations above for $\boldsymbol{\mu}$ yields:

$$\widehat{\mu}=(\widehat{\mu}_1,\widehat{\mu}_2)^\top=\left(\frac{1}{n}\sum_{i=1}^{n}y_{i1},\frac{1}{n_2}\sum_{i=1}^{n}r_{i2}y_{i2}\right)^\top,\ n_2=\sum_{i=1}^{n}r_{i2}. \tag{13}$$

If the missingness of y_{i2} depends on y_{i1}, it is readily checked that $\mathrm{E}(\hat{\mu_2}) \neq \mu_2$, implying that $\hat{\mu_2}$ is not a consistent estimate. This is also clear on intuitive grounds. For example, if y_{i1} and y_{i2} are positively correlated with higher values of y_{i1} leading to missing y_{i2}, $\hat{\mu_2}$ in (13) will be downwardly biased, since it only averages over the observed y_{i2} corresponding to lower values of y_{i1}. In treatment studies, this type of response-dependent missingness often occurs if a patient feels that his/her health condition has improved (or deteriorated) during study and decides not to undergo any additional treatment.

Under MAR, the missingness of y_{i2} only depends on y_{i1}, i.e.,

$$\pi_{i2}=\mathrm{P}(r_{i2}=1|y_{i1},y_{i2})=\mathrm{P}(r_{i2}=1|y_{i1}).$$

This probability π_{i2} selects which y_{i2}'s are to be observed based on the values of y_{i1}. Thus, each ith subject observed at $t = 2$ represents a subgroup of $1/\pi_{i2}$ subjects with the same baseline value y_{i1}, but unobserved at post-treatment because of the selection process defined

by π_{i2}. By augmenting each observed response y_{i2} at $t = 2$ with the weight function $1/\pi_{i2}$, we can statistically include the missing responses in the estimation of μ_2 by using a weighted GEE (WGEE):

$$\mathrm{W}_n(\mu)=\mathrm{R}(\alpha)^{-1}\begin{pmatrix} 1 & 0 \\ 0 & \frac{r_{i2}}{\pi_{i2}} \end{pmatrix}\left[\sum_{i=1}^{n}\begin{pmatrix} y_{i1}-\mu_1 \\ y_{i2}-\mu_2 \end{pmatrix}\right]=\mathbf{0}. \tag{14}$$

It is readily checked that $\mathrm{E}(\mathbf{W}_n(\boldsymbol{\mu})) = \mathbf{0}$, enabling (14) to yield a consistent estimate of μ_2, $\widehat{\mu_2}=\left(\sum_{i=1}^{n}\frac{r_{i2}}{\pi_{i2}}y_{i2}\right)/n$ (e.g. Kowalski and Tu, Chap. 4, 2007). We can also directly verify this:

$$\widehat{\mu_2}=\frac{1}{n}\sum_{i=1}^{n}\frac{r_{i2}}{\pi_{i2}}y_{i2}\rightarrow_p \mathrm{E}\left(\frac{r_{i2}}{\pi_{i2}}y_{i2}\right)=\mathrm{E}\left[y_{i2}\frac{1}{\pi_{i2}}E(r_{i2}|y_{i1},y_{i2})\right]=\mathrm{E}\left[y_{i2}\frac{1}{\pi_{i2}}E(r_{i2}|y_{i1})\right]=\mu_2,$$

where $\rightarrow_p$ denotes convergence in probability.

By comparing (12) with (14), it is seen that the latter differs from the former only in the definition of $\boldsymbol{\Delta}_i$. By carrying this modification over to a general setting with m assessments, we obtain the WGEE for inference about $\boldsymbol{\beta}$ for the distribution-free LMM in (5), which is defined by the same vector equation in (11) except for substituting the following modified $\boldsymbol{\Delta}_i$:

$$\pi_{it}=\mathrm{P}(r_{it}=1|\mathrm{x}_i,\mathrm{y}_i),\ \boldsymbol{\Delta}_{it}=\frac{r_{it}}{\pi_{it}},\ \boldsymbol{\Delta}_i=\mathrm{diag}_t(\boldsymbol{\Delta}_{it}),\ 1\le t\le m,\ 1\le i\le n. \tag{15}$$

It is again readily checked that $\mathrm{E}\,[\mathbf{G}_i\,(\mathrm{x}_i)\,\boldsymbol{\Delta}_i\mathbf{S}_i] = \mathbf{0}$, ensuring that the WGEE yields consistent estimates of β (e.g. Robins et al. 1995; Kowalski and Tu, Chap. 4, 2007).

To use WGEE, we must know or have estimates of π_{it}. In some cases, subject dropout is created by study design and π_{it} are known. For example, in some multi-stage trials, patients can only enter the next stage of the study if they satisfy certain criteria such as response to treatment at the previous stage. However, as noted earlier, in most studies, missing data patterns are defined by a host of factors not directly related to study design. We discuss estimation of π_{it} for the general setting after introducing another robust approach for the distribution-free LMM.

3.2.2 Augmented Weighted Generalized Estimating Equations—The WGEE discussed in Section 3.2.1 depends on the model for missing data in (15). In most studies, π_{it} are unknown and must be modeled and estimated. If such a model is misspecified, the WGEE estimate may be inconsistent. In applications, reliable models may also exist for directly relating the missing response to the observed ones and other covariates. The augmented WGEE (AWGEE) is developed to take advantage of this additional source of modeling information to ensure valid inference when the model for π_{it} may be incorrect (Robins et al., 1995;Tsiatis, 2006).

To illustrate, consider again the pre-post study design. Suppose that we can predict y_{i2} directly based on y_{i1} using a linear regression:

$$\mathrm{E}(y_{i2}|y_{i1})=\eta_0+\eta_1 y_{i1},\ 1\le i\le n. \tag{16}$$

Then, we can estimate μ_2 without using WGEE by $\widetilde{\mu}_2=\sum_{i=1}^{n}(\widehat{\eta}_0+\widehat{\eta}_1 y_{i1})/n$. This new estimate is consistent if (16) is a correct model, since

$$\widetilde{\mu}_2=\widehat{\eta}_0+\widehat{\eta}_1\frac{1}{n}\sum_{i=1}^{n}y_{i1}\to_p\eta_0+\eta_1\mathrm{E}(y_{i1})=\mathrm{E}\left[\mathrm{E}(y_{i2}|y_{i1})\right]=E(y_{i2})=\mu_2.$$

By combining both the prediction model in (16) and the WGEE in (14), we obtain an augmented WGEE to estimate μ as follows:

$$\mathrm{W}_n(\mu)=\sum_{i=1}^{n}\mathrm{R}(\alpha)^{-1}\left(\mathbf{\Delta}_i\mathrm{S}_i-\mathbf{\Delta}_i^c\widetilde{S}_i\right),\ \mathbf{\Delta}_i^c=\mathbf{\Delta}_i-\mathrm{I}_2, \tag{17}$$

$$\widetilde{S}_i=(y_{i1}-\mu_1,\widehat{y}_{i2}-\mu_2)^{\top},\widehat{y}_{i2}=\eta_0+\eta_1 y_{i1}.$$

It is readily checked that $\mathrm{E}\left[\mathrm{G_i(x_i)}\left(\mathbf{\Delta}_i\mathrm{S}_i-\mathbf{\Delta}_i^c\widetilde{\mathrm{S}}_i\right)\right]=\mathbf{0}$ if either (15) or (18) or both are correct. Thus, the AWGEE above yields consistent estimates of $\boldsymbol{\mu}$ if at least one of these models is correct. Further, when both models are correct, the AWGEE estimate from (17) may also be more efficient than the WGEE estimate (Robins et al, 1995;Tsiatis, 2006).

The above is readily extended to a more general setting where the prediction model in (16) also involves other baseline covariates. Let u_i be a set of baseline variables including y_{i1} and the prediction model be defined by:

$$\mathrm{E}(y_{i2}|\mathrm{u}_i)=\eta_0+\mathrm{u}_i^{\top}\eta_1,\ 1\le t\le 2,\ 1\le i\le n. \tag{18}$$

The AWGEE is defined by

$$\mathrm{W}_n(\beta)=\sum_{i=1}^{n}\mathrm{W}_{ni}(\beta)=\sum_{i=1}^{n}\mathrm{G}_i(\mathrm{x}_i)\left(\mathbf{\Delta}_i\mathrm{S}_i-\mathbf{\Delta}_i^c\widetilde{S}_i\right)=\mathbf{0}, \tag{19}$$

where $\widehat{y}_{i2}=\eta_0+\mathrm{u}_i^{\top}\eta_1$. To ensure consistent estimation for the regression model, we assume a surrogacy-type assumption, $[y_{i2}\mid \mathrm{u}_i, \mathrm{x}_i]=[y_{i2}\mid \mathrm{u}_i]$, where $[y_{i2}\mid \mathrm{v}_i]$ denotes the conditional distribution of y_{i2} given v_i (e.g. Prentice, 1989; Kowalski and Tu, 2002). Of course, the condition holds if u_i includes x_i. Under the surrogacy condition, $\mathrm{E}\left[\mathrm{G}_i(x_i)(\mathbf{\Delta}_i\mathrm{S}_i-\mathbf{\Delta}_i^c\widetilde{\mathrm{S}}_i)\right]=\mathbf{0}$, if

either (15) or (18) or both are correct. Thus, the AWGEE in (15) yields consistent estimates of β if at least one of the missing data models is correct.

Although feasible in principle, it is more complex to implement AWGEE for a general longitudinal study with more than two assessments. For example, for $m = 3$, we need to consider two missing data patterns when predicting missing y_{i3}: one with observed y_{i1} and y_{i2}, and the other with observed y_{i1} only. The number of prediction models grows rapidly as the frequency of assessments increases. Further, it is more intricate to specify the prediction models than models for the missing response probabilities π_{it}.

3.2.3 Estimation of Weight Function and Augmented Term—Under MCAR, $\mathbf{r}_i$ are independent of x_i and y_i and $\pi_{it} = P[r_{it} = 1] = \pi_t$. In this case, π_t are readily estimated by the sample moment: $\widehat{\pi}_t = \left(\sum_{i=1}^{n} r_{it}\right)/n \ (1 \le t \le m)$. In many studies, however, π_{it} are dependent of either x_i or y_i or both. It is difficult to model π_{it} as a function of x_i and y_i without imposing some additional assumptions regarding the relationship between them. As in the literature, we focus on the MAR mechanism.

As noted earlier, missing data in longitudinal trials often occur as the result of subject dropout due to deteriorated/improved health conditions and other related conditions, exhibiting the so-called monotone missing data pattern (MMDP). The structured patterns under MMDP make it possible to model π_{it} in most studies.

Under MMDP, if y_{it} is observed at time t, then all y_{is} at all earlier times s ($< t$) are also observed. Let

$$H_{it} = \{\tilde{x}_{is}, \tilde{y}_{is}; 1 \le s \le t-1\},\ \tilde{x}_{it} = (x_{i1}^{\top}, \ldots, x_{i1(t-1)}^{\top})^{\top},$$
$$\tilde{y}_{it} = (y_{i1}, \ldots, y_{i(t-1)})^{\top},\ 2 \le t \le m.$$

The subset $\mathbf{H}_{it}$ contains all observed data prior to time t. Under MAR,

$$\pi_{it} = P(r_{it} = 1 | x_i, y_i) = P(r_{it} = 1 | H_{it}) = P(r_{it} = 1 | \tilde{x}_{it}, \tilde{y}_{it}). \tag{20}$$

Thus, under MMDP and MAR, π_{it} are a function of observed data only, making it possible to estimate these selection probabilities.

Let $p_{it} = E(r_{it} = 1 \mid r_{i(t-1)} = 1, \mathbf{H}_{it})$ denote the one-step transition probability of the occurrence of missing data. Then, by invoking MMDP, it is readily checked that

$$\pi_{it} = \prod_{s=2}^{t} p_{is},\ 2 \le t \le m,\ 1 \le i \le n. \tag{21}$$

Thus, we can estimate π_{it} by modeling the p_{it}. Since p_{it} is the probability of a binary response, we model that using logistic regression:

$$\text{logit}\,(p_{it}\,(\alpha_t))=\text{logit}\,(\mathrm{E}(r_{it}=1|r_{i(t-1)}=1,\mathrm{H}_{it}))=\boldsymbol{\xi}_t^{\top}\tilde{w}_{it},\ 2\le t\le m, \tag{22}$$

where $\widetilde{\mathrm{w}}_{it}=\left(1,\widetilde{\mathrm{x}}_{it}^{\top},\widetilde{\mathrm{y}}_{it}^{\top}\right)^{\top}$ and $\widetilde{\mathrm{w}}_i=\left(\widetilde{\mathrm{w}}_{i2}^{\top},\ldots,\widetilde{\mathrm{w}}_{im}^{\top}\right)^{\top}$. For each t, we can estimate $\boldsymbol{\xi}_t$ by maximum likelihood or GEE conditional on the observed $\tilde{\mathbf{w}}_{it}$ at $t-1$ ($2\le t\le m$).

For AWGEE, we again consider the pre-post study design. In this case, we can readily estimate $\boldsymbol{\eta}$ in (18) using GEE, where $\eta=(\eta_0,\eta_1^{\top})^{\top}$. With an estimate $\hat{\eta}$, we can predict y_{it} By using $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\xi}}=\widehat{\boldsymbol{\xi}}_2^{\top}$, we can construct the AWGEE in (19).

3.2.4 Inference for WGEE and AWGEE Estimates—For inference based on WGEE, let $\boldsymbol{\Delta}_i$, ($\boldsymbol{\xi}$) be modeled as in (21) and (22). If $\boldsymbol{\xi}_t$ is estimated by GEE or maximum likelihood, $\widehat{\xi}=(\widehat{\xi}_2^{\top},\ldots,\widehat{\xi}_m^{\top})^{\top}$ is the solution to the following vector estimating equation:

$$\sum_{i=1}^{n}\mathrm{q}_i(\boldsymbol{\xi})=\sum_{i=1}^{n}\left(\mathrm{q}_{i2}^{\top},\ldots,\mathrm{q}_{im}^{\top}\right)^{\top}=0,\ \mathrm{q}_{it}=\frac{\partial}{\partial\boldsymbol{\xi}_t}\left\{r_{i(t-1)}\ \left[r_{it}\log(p_{it})-(1-r_{it})\log(1-p_{it})\right]\right\}. \tag{23}$$

Now, let

$$\mathrm{B}=\mathrm{E}\left(\frac{\partial^{\top}}{\partial\beta}\mathrm{W}_{ni}\right),\ \mathrm{C}=\mathrm{E}\left(\frac{\partial^{\top}}{\partial\boldsymbol{\xi}}\mathrm{W}_{ni}\right),\ \mathrm{F}=\mathrm{E}\left(\frac{\partial^{\top}}{\partial\boldsymbol{\xi}}\mathrm{q}_i(\boldsymbol{\xi})\right),$$

where $\mathbf{W}_{ni}$ is defined in (11). Then, as shown in Appendix A, under $\sqrt{n}$-consistency of $\hat{\boldsymbol{\alpha}}$, the WGEE estimate $\hat{\beta}$ is asymptotically normal with the asymptotic variance given by

$$\sum\nolimits_{\beta}=\mathrm{B}^{-1}(\mathrm{Var}(\mathrm{W}_{ni})+\boldsymbol{\Phi})\mathrm{B}^{-\top},\ \boldsymbol{\Phi}=\mathbf{C}\mathbf{F}^{-1}\mathrm{Var}(\mathrm{q}_i)\mathrm{F}^{-\top}\mathrm{C}^{\top}-\mathrm{E}\left(\mathrm{W}_{ni}\mathrm{q}_i^{\top}\mathrm{F}^{-\top}\mathrm{C}^{\top}\right)-\left[\mathrm{E}\left(\mathrm{W}_{ni}\mathrm{q}_i^{\top}\mathrm{F}^{-\top}\mathrm{C}^{\top}\right)\right]^{\top}. \tag{24}$$

In (24), $\boldsymbol{\Phi}$ accounts for the variability of estimated $\hat{\boldsymbol{\xi}}$. We can estimate $\boldsymbol{\Sigma}_{\beta}$ by substituting consistent estimates in place of the respective quantities.

For AWGEE inference, we also need to estimate η for the prediction model in (18). Using GEE, the vector estimating equation is given by:

$$\sum_{i=1}^{n}\mathrm{g}_i(\eta)=\sum_{i=1}^{n}\frac{\partial}{\partial\eta}\left\{r_{i2}\left[y_{i2}-\left(\eta_0+\mathrm{u}_i^{\top}\eta_2\right)\right]\right\}=\mathbf{0}. \tag{25}$$

Let $\widehat{\varphi}=\left(\widehat{\varphi}_1^{\top},\widehat{\varphi}_2^{\top}\right)^{\top}$ with $\boldsymbol{\phi}_1=\boldsymbol{\xi}$ and $\boldsymbol{\phi}_2=\boldsymbol{\eta}$. Then, by combining (23) and (25), $\hat{\boldsymbol{\phi}}$ can be expressed as the solution to the following joint vector estimating equation:

$$\sum_{i=1}^{n}\mathrm{s}_i=\sum_{i=1}^{n}\left(\mathrm{q}_i^{\top}(\phi_1),\mathrm{g}_i^{\top}(\phi_2)\right)^{\top}=\mathbf{0},\ \mathrm{s}_i=\left(\mathrm{q}_i^{\top}(\phi_1),\mathrm{g}_i^{\top}(\phi_2)\right)^{\top}.$$

The AWGEE estimate $\hat{\boldsymbol{\beta}}$ is also asymptotically normal under $\sqrt{n}$-consistency of $\hat{\boldsymbol{\alpha}}$, with the asymptotic variance having the same form as in (24) except for substituting $\mathbf{s}_i$ for $\mathbf{q}_i$ and redefining $\mathrm{C} = \mathrm{E}\left(\frac{\partial^\top}{\partial\varphi}\mathrm{W}_{ni}\right)$ and $\mathrm{F}=\mathrm{E}\left(\frac{\partial^\top}{\partial\varphi}\mathrm{S}_i(\varphi)\right)$. Again, we can estimate the asymptotic variance by substituting consistent estimates in place of the respective parameters.

4 Application

We illustrate our considerations with both real and simulated data. We first present an application to data from a longitudinal study in depression research and then investigate the performance of the approach with small to moderate sample sizes by simulation. In all the examples, we set the statistical significance at $\alpha = 0.05$. All analyses are carried out using a code we have developed for implementing the proposed approach using the R software platform (Free Software Foundation, 1999). This code is available from the author upon request.

4.1 Real Study

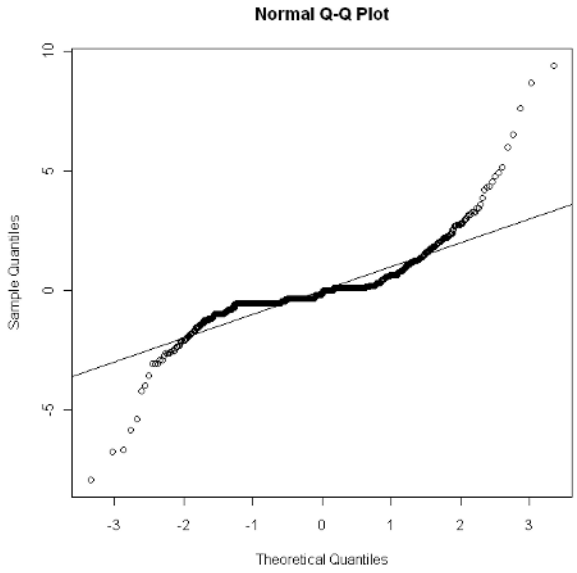
In a study on geriatric depression and associated medical comorbidities for old primary care patients, 744 subjects were enrolled from private practices and University-affiliated clinics in general internal medicine, geriatrics, and family medicine in Monroe County, New York (Lyness et al., 2007). All patients age 65 years and older who presented for care on selected days and were capable of giving informed consent. Enrolled subjects underwent semi-structured interviews, administered by trained raters in the subjects' homes or in research offices at the UR Medical Center. The raters' subject interviews included assessments of cognition, functional status, and psychopathology, the latter including the Structured Clinical Interview for DSM-IV (SCID) (Spitzer et al. 1986). Interviews and chart reviews were conducted at study intake, and again at one- and two-year follow-up time points.

In geriatric research, overall functional disability is of particular importance, as it reflects both the mental and physical health conditions of the individual. Primary measures of overall functional status include the Instrumental Activities of Daily Living (IADL), Physical Self-Maintenance Scales (PSMS), Global Assessment of Functioning (GAF), and the Karnofsky Performance Status Scale (KPSS) (Lawton MP and Brody, 1969, Karnofsky DA and Burchenal JH, 1949, Ware JE, Jr. and Sherbourne CD). For illustration purposes, we analyzed the change of IADL from baseline to one-year follow-up, as this measure assesses instrumental activities such as shopping or using the telephone and is particularly popular in geriatric research. Further, we only included the baseline value as a predictor when modeling the missingness of this outcome as well as the outcome itself at the follow-up using the respective logistic (20) and linear (18) models.

Of the 744 enrolled, 468 completed the IADL at the one-year follow-up. Shown in Table 1 are the estimates of the intercept and slope from the fitted logistic regression for modeling the missingness and the linear regression for modeling the outcome of IADL as a function of its baseline value at the follow-up. The baseline IADL was significant in both models, indicating that it did predict the occurrence of missing IADL as well as the outcome itself at the follow-up. Note that the negative sign of the estimate of the coefficient for baseline IADL in the logistic model indicates that the subjects with lower baseline IADL were more likely to come for assessment at the one-year follow-up. As lower IADL is associated with poorer functioning status, the observed sample at the follow-up visit seemed to be biased towards those with more severe overall functional disability at baseline.

We fit the LMM in (1) and the distribution-free alternative in (5) to examine the change of IADL from baseline to the one-year follow-up. Shown in Table 2 are the estimates of the intercept β_0 and slope β_1 for the respective models under the different inference procedures. As the estimates of β_1 were positive across the board, the mean IADL increased at the follow-up visit, indicating better functioning status for the old primary care patients in this observational study.

However, the magnitude of the estimate of β_1 did vary substantially — not only between the models, but also across the different procedures within the same distribution-free linear model. The WGEE and AWGEE yielded quite similar estimates, with the latter AWGEE also providing improved efficiency, as indicated by smaller asymptotic standard errors. The GEE performed poorly, with a whopping 50% downward bias, as compared to its counterparts WGEE and AWGEE estimates. For the between-model comparison, the ML estimate of β_1 from the fitted LMM also incurred a downward bias, albeit with a much smaller magnitude relative to the GEE estimate. The downward bias in both cases was consistent with the fact that those assessed at the follow-up visit represented a subgroup with more severe overall functional disability at baseline.



Shown in the Figure is the normal-based Q-Q plot of the conditional residuals obtained from the estimated fixed and random effects for the fitted LMM model (Nobre and Singer, 2007). The plot indicates clearly that the residuals did not follow a normal distribution, which may explain the difference in the estimates of the fixed effects between the parametric (LMM) and distribution-free models (WGEE and AWGEE).

4.2 Simulation Study

Given the discrepant estimates between LMM and WGEE (AWGEE) for the real study data in 4.2, we conducted a simulation study with a pre-post study design to investigate this issue further. We considered two non-normal distributions for the model error term of the LMM: a rescaled central chi-square with one degree of freedom and a uniform between −1 and 1. Since the results are quite similar, we only discuss and report the results from the chi-square-distributed error. To examine the performance of the models under small, moderate and large samples, we performed the simualation study with three sample sizes: n = 50, 100 and 2,000. All simulations were performed with a Monte Carlo sample of 1,000 using the R software (R Development Core Team, 2007).

We considered the pre- and post-treatment design and simulated the outcome according to the LMM in (1) by setting $\beta_0 = \beta_1 = 1$ and ϵ_{it} ($t = 1, 2$) to follow the rescaled chi-square distribution, $(\chi_1^2+1)\sqrt{\sigma^2/2}$, where χ_1^2 denotes a central chi-square with one degree of freedom. We varied σ_b^2 and σ^2 to control the within-subject correlation $\rho=\sigma_b^2/(\sigma_b^2+\sigma^2)$. We assumed no missing data at baseline $t = 1$ and simulated the missing response at post-treatment $t = 2$ under MAR according to the following logistic regression:

$$\text{logit}\,(\pi_{i2})=\text{logit}\,(\text{P}\,(r_{i2}=1|y_{i1}))=\xi_0+\xi_1 y_{i1},\ \ 1 \le i \le n. \tag{26}$$

We set $\xi_0 = 0.5$ and $\xi_1 = 1.2$ to create about 25% missing response y_{i2} at $t = 2$.

Under (1), it is readily checked that regardless of the distributions for b_i and ϵ_{it} (see Appendix)

$$\text{E}\,(y_{i2}|y_{i1})=\eta_0+\eta_1 y_{i1},\ \ \eta_0=\beta_0\left(1-\frac{\sigma_b^2}{\sigma_b^2+\sigma^2}\right)+\beta_1,\ \ \eta_1=\rho=\frac{\sigma_b^2}{\sigma_b^2+\sigma^2},\ \ 1 \le i \le n. \tag{27}$$

The above was used to predict missing y_{i2} from y_{i1} for AWGEE inference about β as discussed in Section 3.2.4. To study the effect of wrong weight function on WGEE and the robustness of AWGEE in such a scenario, we also estimated π_{i2} under an incorrect model by leaving out y_{i1} in (26).

We considered the null $H_0 : \beta_0 = \beta_1$, i.e., the mean at post-treatment is twice that at pretreatment. We tested H_0 using the Wald statistic, $\text{Q}_n^2=n\widehat{\beta}^\top \text{K}^\top(\text{K}\widehat{\sum}_\beta \text{K}^\top)^{-1}\text{K}\widehat{\beta}$, which has an asymptotic central χ_1^2 distribution. We estimated the type I error rate α based on the distribution of Q_n^2 from 1, 000 Monte Carlo (MC) replications, $\widehat{\alpha}=(\sum_{j=1}^{1000} I_{\{Q_{nj}\ge q^{0.95}\}})/1000$, where Q_{nj}^2 denotes the value of Q_n^2 from the jth MC replication and $q_{0.95}$ the 95th percentile of χ_1^2. The maximum likelihood (ML) inference about $\beta = (\beta_0,\beta_1)^\top$ for LMM was obtained from the LME procedure in R, while the WGEE and AWGEE estimates for the distribution-free alternative in (5) were computed based on the asymptotic results in Section 3.2.

Shown in Table 3 are the estimates of β and associated asymptotic standard errors averaged over 1,000 MC replications obtained from ML, GEE, WGEE and AWGEE for the respective models, with a within-subject correlation $\rho = 0.1$ (or $\sigma_b^2=2/9$ and $\sigma^2 = 2$). The results confirmed that the baseline mean β_0 is consistently estimated by all four procedures. For β_1, ML yielded consistent estimates for the normal distributed error, but biased estimates under the rescaled χ_1^2 error. As expected, GEE estimates were biased under both types of error distributions. Note that in the rescaled χ_1^2 error case, the standard errors of these estimates did not increase much, making the upwardly biased estimates yield false significant results in practice.

Under the correct weight function, WGEE performed well. When the incorrect constant weight function was used, WGEE yielded biased estimates, while the AWGEE estimates remained close to the true value of β_1 under the correct prediction model. However,

AWGEE did not show any significant gain in efficiency; in fact, the estimate of the slope β_1 had a larger standard error under AWGEE than under WGEE across small sample sizes n.

To further investigate the relative efficiency between WGEE and AWGEE, we replicated the above analysis by increasing the within-subject correlation. For example, shown in Tables 4 and 5 are the estimates from one such replicated analysis with $\rho = 0.3$ (or σ_b^2=1 and $\sigma^2 = 2$ and 0.6 (or σ_b^2=1.5 and $\sigma^2 = 1$), respectively. It is seen that all conclusions above remain the same, except for the relative efficiency between WGEE and AWGEE under the correct weight function and prediction model. As ρ increased to 0.3, AWGEE have smaller standard errors than its counterpart WGEE when $n = 2000$. At $\rho = 0.6$, not only did AWGEE show smaller standard errors than WGEE across all sample sizes, their differences also widened as compared to those with $\rho = 0.1$.

5 Discussion

We investigated the two primary modeling strategies for longitudinal continuous response, the linear mixed-effects model (LMM) and the distribution-free linear model, with respect to their performance under missing data, and illustrated our considerations using real as well as simulated study data. Our results show that LMM and the GEE procedure for the distribution-free alternative generally yield biased estimates under MAR when the normality assumption for LMM is violated. Further, as indicated by the simulation results, the standard errors of these estimates do not increase to reflect model misspecification, making inference prone to misleading findings. Thus, when modeling longitudinal data, it is important to test the MCAR assumption as discussed in Section 3.2 before applying any of the models and inference procedures considered. If the null of MCAR is rejected, WGEE and/or AWGEE should be considered, unless there is strong evidence to support the use of the alternative linear mixed-effects model.

Our simulation results also indicate that the gain in efficiency by AWGEE over WGEE depends on the magnitude of the within-subject correlation ρ. Within the context of the particular simulation model considered, AWGEE is less efficient than WGEE for small sample size under small ρ such as 0.1. But, AWGEE edged out WGEE to be a more efficient procedure as ρ increased to 0.6.

Acknowledgments

This research was supported in part by NIH grants R01-DA012249, R21-AG023956, UL1 RR024160 and R24-MH071604. We thank Ms. Bliss-Clark for her careful proofreading of the manuscript, and an anonymous, an Associate Editor and Editor Prof. Leonhard Held for their constructive comments that led to a substantial improvement in the presentation of this research.

Appendix

Under (1), we have

$$\rho_{yb}=\text{Corr}\,(y_{i1}, b_i)=\frac{\text{Cov}\,(y_{i1}, b_i)}{\sqrt{\text{Var}\,(y_{i1})\,\text{Var}\,(b_i)}}=\sqrt{\frac{\sigma_b^2}{\sigma_b^2+\sigma^2}}. \quad (28)$$

Further, the above holds regardless of the distributions for b_i and $\in_{it}$. Now, consider the linear regression

$$b_i=\phi_0+\phi_1 y_{i1}+\delta_i,\ \ \delta_i \sim (0,\sigma_\delta^2),\ \ 1 \le i \le n,$$

Where $(0,\sigma_\delta^2)$ denotes a distribution with mean 0 and variance σ_δ^2. It follows from (28) and the relationship between linear regression coefficients and correlation that

$$\phi_1=\rho_{yb}\sqrt{\frac{\mathrm{Var}\,(b_i)}{\mathrm{Var}\,(y_{i1})}}=\sqrt{\frac{\sigma_b^2}{\sigma_b^2+\sigma^2}}\sqrt{\frac{\sigma_b^2}{\sigma_b^2+\sigma^2}}=\frac{\sigma_b^2}{\sigma_b^2+\sigma^2}.$$

Also, since

$$0=\mathrm{E}\,(b_i)=\phi_0+\phi_1\mathrm{E}\,(y_{i1})=\phi_0+\phi_1\beta_0,$$

it follows that $\varphi_0=-\beta_0\sigma_b^2/(\sigma_b^2+\sigma^2)$.

Under (1) and regardless of the distributions for b_i and $\in_{it}$, we have

$$\mathrm{E}\,(y_{i2}|y_{i1})=\beta_0+\beta_1+\mathrm{E}\,(b_i|y_{i1})=\beta_0+\beta_1+\phi_0+\phi_1 y_{i1}.$$

By substituting the expressions of φ_0 and φ_0 into the above and combining the coefficients, we obtain (27).

References

1. Demidenko, E. Mixed Models: Theory and Applications. New York: Wiley; 2004.
2. Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. Analysis of Longitudinal Data. 2nd. Oxford University Press; 2002.
3. Fitzmaurice, GM.; Laird, NM.; Ware, JH. Applied Longitudinal Analysis. New York: Wiley; 2004.
4. Goldstein, H. Multilevel Statistical Models. Edward Arnold; London; Halsted, New York: 1995.
5. Karnofsky, DA.; Burchenal, JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod, CM., editor. Evaluation of Chemotherapeutic Agents. New York: Columbia; 1949.
6. Kowalski J, Tu XM. A GEE approach to modeling longitudinal data with incompatible data formats and measurement error: application to HIV immune markers. Journal of the Royal Statistical Society, Series C. 2002; 51:91–114.
7. Kowalski, J.; Tu, XM. Modern Applied U Statistics. New York: Wiley; 2007.
8. Lawton, MP.; Brody, EM., editors. Gerontologist. Vol. 9. 1969. Assessment of older people; self-maintaining and instrumental activities of daily living; p. 179-186.
9. Lyness JM, Niculescu A, Tu XM, Reynolds CF III, Caine ED. The relationship of medical comorbidity to depression in older primary care patients. Psychosomatics. 2007; 47:435–439. [PubMed: 16959933]
10. Maas CJM, Hox JJ. The influence of violations of assumptions on multilevel parameter estimates and their standard errors. Computational Statistics & Data Analysis. 2004; 46:427–440.
11. Maas CJM, Hox JJ. Sufficient Sample Sizes for Multilevel Modeling. Methodology. 2005; 1:86–92.
12. Nobre JS, Singer JM. Residual analysis for linear mixed models. Biometrical Journal. 2007; 49:863–875. [PubMed: 17638292]

13. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. Statistics in Medicine. 1989; 8:431–440. [PubMed: 2727467]
14. Rasbash, J.; Browne, W.; Goldstein, H.; Yang, M.; Plewis, I.; Healy, M.; Woodhouse, G.; Draper, D.; Langford, I.; Lewis, T. A User's Guide to MLwiN. Multilevel Models Project. University of London; London: 2000.
15. Raudenbush, SW.; Bryk, AS. Hierarchical Linear Models. 2nd. Sage; Thousand Oaks, CA: 2002.
16. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2007. URL http://www.R-project.org
17. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association. 1995; 90:106–121.
18. Spitzer, RL.; Williams, JB.; Gibbon, M. Structured Clinical Interview for DSM-III-R (SCID). New York: Biometrics Research; 1986.
19. Tsiatis, AA. Semiparametric Theory and Missing Data. New York: Springer; 2006.
20. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care. 1992; 30:473–483. [PubMed: 1593914]
21. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986; 42:121–130. [PubMed: 3719049]

Table 1

Estimates of parameters of (1) logistic regression for modeling missingness at one-year follow-up, and (2) linear model for predicting IADL at one-year follow-up for the study on geriatric depression and associated medical comorbidities.

Estimates of models for missingness and outcome at one-year follow-up			
Predictors	**Estimate**	**Standard error**	**p-value**
Logistic regression for missingness at one year follow-up			
Intercept	0.635	0.086	<0.0001
Baseline IADL	−0.0491	0.018	0.007
Linear regression for predicting missing IADA at one year follow-up			
Intercept	0.685	0.124	<0.0001
Baseline IADL	1.01	0.03	<0.0001

Table 2

Estimates of parameters of (1) linear mixed-effects model (ML), and (2) distribution-free linear model (GEE, WGEE and AWGEE) for change of IADL from baseline to one-year follow-up for the study on geriatric depression and medical comorbidities.

Estimates of models for change of IADL from baseline to one-year follow-up				
Methods	**β_0 (s.e.)**	**p-value for β_0**	**β_1 (s.e.)**	**p-value for β_1**
ML	2.11(0.16)	<0.0001	0.054(0.009)	<0.0001
GEE	2.11(0.15)	<0.0001	0.032(0.014)	0.02
WGEE	2.11(0.15)	<0.0001	0.061(0.016)	0.0002
AWGEE	2.11(0.15)	<0.0001	0.060(0.013)	0.0001

Table 3

Averaged estimates of β over 1,000 Monte Carlo replications along with asymptotic standard errors (s.e.) and type I error rates α for sample size 50, 100, 2000, with about 25% missing data at post-treatment based on ML for linear mixed-effects model, and GEE, WGEE and AWGEE for distribution-free linear model, with true $\beta_0 = 1$ and $\beta_1 = 1$ and within-subject correlation $\rho = 0.1$.

Methods	Weight Function	Prediction Model	Normal Distribution			Chi-square Distribution		
			β_0 (s.e.)	β_1 (s.e.)	α	β_0 (s.e.)	β_1 (s.e.)	α
Sample size = 50								
ML			1.00(0.21)	0.98(0.30)	0.06	1.00(0.22)	1.10(0.23)	0.07
GEE			1.01(0.21)	1.05(0.31)	0.05	1.00(0.20)	1.05(0.30)	0.05
WGEE	Right		0.99(0.21)	1.00(0.33)	0.08	1.01(0.21)	1.00(0.31)	0.06
	Wrong		1.01(0.21)	1.04(0.31)	0.06	1.00(0.20)	1.04(0.30)	0.06
AWGEE	Right	Right	1.00(0.21)	1.00(0.39)	0.07	1.00(0.21)	1.01(0.38)	0.07
	Wrong	Right	1.01(0.21)	0.99(0.32)	0.08	1.01(0.20)	0.99(0.32)	0.09
	Right	Wrong	0.99(0.21)	1.00(0.40)	0.06	0.99(0.21)	1.00(0.39)	0.06
	Wrong	Wrong	1.01(0.21)	1.05(0.33)	0.07	1.01(0.21)	1.05(0.34)	0.07
Sample size = 100								
ML			1.00(0.15)	1.00(0.21)	0.06	1.00(0.15)	1.09(0.17)	0.06
GEE			1.00(0.15)	1.04(0.22)	0.06	1.00(0.15)	1.05(0.22)	0.05
WGEE	Right		1.00(0.15)	1.00(0.23)	0.05	1.00(0.15)	1.00(0.22)	0.06
	Wrong		1.00(0.15)	1.04(0.22)	0.07	1.00(0.15)	1.05(0.22)	0.05
AWGEE	Right	Right	1.00(0.15)	1.00(0.29)	0.04	1.00(0.15)	1.00(0.30)	0.04
	Wrong	Right	1.00(0.15)	1.00(0.23)	0.05	1.00(0.15)	1.00(0.22)	0.06
	Right	Wrong	1.00(0.15)	1.01(0.30)	0.04	1.00(0.15)	1.00(0.30)	0.04
	Wrong	Wrong	1.00(0.15)	1.04(0.24)	0.07	1.00(0.15)	1.06(0.23)	0.07
Sample size = 2000								
ML			1.00(0.033)	1.00(0.049)	0.06	1.00(0.035)	1.10(0.038)	0.59
GEE			1.00(0.033)	1.04(0.049)	0.14	1.00(0.033)	1.05(0.049)	0.21
WGEE	Right		1.00(0.033)	1.00(0.054)	0.05	1.00(0.033)	1.00(0.050)	0.05
	Wrong		1.00(0.033)	1.04(0.049)	0.12	1.00(0.033)	1.05(0.049)	0.17
AWGEE	Right	Right	1.00(0.033)	1.00(0.052)	0.05	1.00(0.033)	1.00(0.051)	0.05

Methods	Weight Function	Prediction Model	Normal Distribution			Chi-square Distribution		
			β_0 (s.e.)	β_1 (s.e.)	α	β_0 (s.e.)	β_1 (s.e.)	α
	Wrong	Right	1.00(0.033)	1.00(0.048)	0.06	1.00(0.033)	1.00(0.047)	0.06
	Right	wrong	1.00(0.033)	1.00(0.053)	0.05	1.00(0.033)	1.00(0.052)	0.05
	Wrong	Wrong	1.00(0.033)	1.04(0.049)	0.12	1.00(0.033)	1.05(0.048)	0.15

Table 4

Averaged estimates of β over 1,000 Monte Carlo replications along with asymptotic standard errors (s.e.) and type I error rates α for sample size 50, 100, 2000, with about 25% missing data at post-treatment based on ML for linear mixed-effects model, and GEE, WGEE and AWGEE for distribution-free linear model, with true $\beta_0 = 1$ and $\beta_1 = 1$ and within-subject correlation $\rho = 0.3$.

Methods	Weight Function	Prediction Model	Normal Distribution			Chi-square Distribution		
			β_0 (s.e.)	β_1 (s.e.)	α	β_0 (s.e.)	β_1 (s.e.)	α
Sample size = 50								
ML			1.00(0.25)	1.00(0.31)	0.06	1.01(0.24)	1.08(0.32)	0.06
GEE			1.00(0.24)	1.18(0.32)	0.08	1.00(0.24)	1.22(0.33)	0.08
WGEE	Right		0.99(0.25)	1.02(0.36)	0.05	1.02(0.24)	1.00(0.35)	0.04
	Wrong		1.00(0.24)	1.18(0.33)	0.09	1.01(0.24)	1.24(0.32)	0.11
AWGEE	Right	Right	1.00(0.24)	1.01(0.40)	0.04	1.01(0.24)	0.99(0.41)	0.03
	Wrong	Right	1.01(0.24)	0.99(0.35)	0.05	1.02(0.24)	1.00(0.34)	0.05
	Right	Wrong	0.99(0.24)	1.00(0.41)	0.03	1.00(0.24)	1.01(0.41)	0.04
	Wrong	Wrong	1.00(0.24)	1.17(0.34)	0.08	1.01(0.24)	1.21(0.34)	0.10
Sample size = 100								
ML			0.99(0.17)	1.01(0.22)	0.07	1.00(0.17)	1.12(0.23)	0.07
GEE			1.00(0.17)	1.19(0.23)	0.15	1.00(0.17)	1.23(0.23)	0.15
WGEE	Right		1.00(0.17)	1.01(0.26)	0.06	1.00(0.17)	1.02(0.25)	0.03
	Wrong		1.00(0.17)	1.20(0.24)	0.13	1.00(0.17)	1.23(0.23)	0.18
AWGEE	Right	Right	1.00(0.17)	0.99(0.28)	0.05	1.00(0.17)	1.00(0.27)	0.04
	Wrong	Right	1.00(0.17)	1.01(0.24)	0.06	1.00(0.17)	0.99(0.25)	0.06
	Right	Wrong	1.00(0.17)	1.01(0.29)	0.05	1.00(0.17)	0.99(0.30)	0.04
	Wrong	Wrong	1.00(0.17)	1.18(0.25)	0.11	1.00(0.17)	1.22(0.25)	0.16
Sample size = 2000								
ML			1.00(0.039)	1.00(0.050)	0.06	1.00(0.038)	1.12(0.052)	0.64
GEE			1.00(0.039)	1.20(0.052)	0.97	1.00(0.038)	1.23(0.052)	0.99
WGEE	Right		1.00(0.039)	1.00(0.068)	0.06	1.00(0.039)	1.00(0.059)	0.05
	Wrong		1.00(0.039)	1.20(0.053)	0.95	1.00(0.039)	1.23(0.052)	1.00
AWGEE	Right	Right	1.00(0.039)	1.00(0.055)	0.05	1.00(0.039)	1.00(0.054)	0.06

Methods	Weight Function	Prediction Model	Normal Distribution			Chi-square Distribution		
			β_0 (s.e.)	β_1 (s.e.)	α	β_0 (s.e.)	β_1 (s.e.)	α
	Wrong	Right	1.00(0.039)	1.00(0.050)	0.06	1.00(0.039)	1.00(0.050)	0.06
	Right	wrong	1.00(0.039)	1.00(0.057)	0.05	1.00(0.039)	1.00(0.056)	0.05
	Wrong	Wrong	1.00(0.039)	1.19(0.051)	0.94	1.00(0.039)	1.23(0.052)	1.00

Table 5

Averaged estimates of β over 1,000 Monte Carlo replications along with asymptotic standard errors (s.e.) and type I error rates α for sample size 50, 100, 2000, with about 25% missing data at post-treatment based on ML for linear mixed-effects model, and GEE, WGEE and AWGEE for distribution-free linear model, with true $\beta_0 = 1$ and $\beta_1 = 1$ and within-subject correlation $\rho = 0.6$.

Methods	Weight Function	Prediction Model	Normal Distribution			Chi-square Distribution		
			β_0 (s.e.)	β_1 (s.e.)	α	β_0(s.e.)	β_1(s.e.)	α
Sample size = 50								
ML			1.00(0.22)	1.01(0.23)	0.06	1.00(0.22)	1.08(0.24)	0.07
GEE			1.01(0.22)	1.30(0.26)	0.19	1.00(0.22)	1.32(0.25)	0.25
WGEE	Right		1.00(0.22)	1.02(0.29)	0.05	0.99(0.22)	1.04(0.29)	0.04
	Wrong		1.01(0.22)	1.30(0.27)	0.21	1.01(0.22)	1.32(0.25)	0.24
AWGEE	Right	Right	1.00(0.22)	1.01(0.27)	0.06	1.01(0.22)	1.01(0.28)	0.05
	Wrong	Right	1.01(0.22)	0.99(0.26)	0.06	1.00(0.22)	1.02(0.27)	0.06
	Right	Wrong	0.99(0.22)	1.02(0.28)	0.07	1.00(0.22)	0.99(0.28)	0.05
	Wrong	Wrong	1.01(0.22)	1.28(0.27)	0.20	1.01(0.22)	1.30(0.26)	0.22
Sample size = 100								
ML			1.00(0.16)	1.00(0.16)	0.06	1.00(0.16)	1.10(0.17)	0.09
GEE			1.00(0.16)	1.30(0.18)	0.36	1.01(0.16)	1.32(0.18)	0.42
WGEE	Right		1.00(0.16)	1.02(0.21)	0.04	1.00(0.16)	1.01(0.21)	0.03
	Wrong		1.00(0.16)	1.29(0.18)	0.35	1.00(0.16)	1.33(0.18)	0.44
AWGEE	Right	Right	1.00(0.16)	1.00(0.18)	0.05	1.00(0.16)	1.00(0.19)	0.04
	Wrong	Right	1.00(0.16)	1.01(0.16)	0.06	1.00(0.16)	1.01(0.16)	0.06
	Right	Wrong	1.00(0.16)	0.99(0.19)	0.05	1.00(0.16)	1.00(0.19)	0.05
	Wrong	Wrong	1.00(0.16)	1.26(0.17)	0.34	1.00(0.16)	1.31(0.16)	0.44
Sample size = 2000								
ML			1.00(0.038)	1.00(0.046)	0.06	1.00(0.035)	1.10(0.038)	0.59
GEE			1.00(0.035)	1.30(0.041)	1.00	1.00(0.035)	1.32(0.041)	1.00
WGEE	Right		1.00(0.035)	1.00(0.057)	0.05	1.00(0.035)	1.00(0.054)	0.04
	Wrong		1.00(0.035)	1.30(0.041)	1.00	1.00(0.035)	1.32(0.041)	1.00
AWGEE	Right	Right	1.00(0.035)	1.00(0.043)	0.05	1.00(0.035)	1.00(0.044)	0.05

Methods	Weight Function	Prediction Model	Normal Distribution			Chi-square Distribution		
			β_0 (s.e.)	β_1 (s.e.)	α	β_0(s.e.)	β_1(s.e.)	α
	Wrong	Right	1.00(0.035)	1.00(0.040)	0.06	1.00(0.035)	1.00(0.040)	0.06
	Right	wrong	1.00(0.035)	1.00(0.043)	0.06	1.00(0.035)	1.00(0.045)	0.05
	Wrong	Wrong	1.00(0.035)	1.27(0.041)	1.00	1.00(0.035)	1.30(0.042)	1.00