Biometrics 65, 478–486
June 2009
DOI: 10.1111/j.1541-0420.2008.01102.x

Mixed-Effect Hybrid Models for Longitudinal Data with Nonignorable Dropout

Ying Yuan[1,*] and Roderick J. A. Little[2]

[1]Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.
[2]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
[*]*email:* yyuan@mdanderson.org

Summary. Selection models and pattern-mixture models are often used to deal with nonignorable dropout in longitudinal studies. These two classes of models are based on different factorizations of the joint distribution of the outcome process and the dropout process. We consider a new class of models, called mixed-effect hybrid models (MEHMs), where the joint distribution of the outcome process and dropout process is factorized into the marginal distribution of random effects, the dropout process conditional on random effects, and the outcome process conditional on dropout patterns and random effects. MEHMs combine features of selection models and pattern-mixture models: they directly model the missingness process as in selection models, and enjoy the computational simplicity of pattern-mixture models. The MEHM provides a generalization of shared-parameter models (SPMs) by relaxing the conditional independence assumption between the measurement process and the dropout process given random effects. Because SPMs are nested within MEHMs, likelihood ratio tests can be constructed to evaluate the conditional independence assumption of SPMs. We use data from a pediatric AIDS clinical trial to illustrate the models.

Key words: Longitudinal data; Missing data; Nonignorable dropout; Shared-parameter model.

1. Introduction

Missing data are a common problem in longitudinal studies, and are frequently caused by dropout, whereby subjects are lost to follow-up and thereafter their responses are not observed. In many cases, dropout depends on either unobserved values or an underlying response process, resulting in nonignorable missing data (Little and Rubin, 2002). Analysis of such data requires jointly modeling the outcome process and the missing data mechanism. Likelihood based and estimation equation methods have been proposed to handle this problem. For the likelihood-based approach, Little (1995) identified two broad classes of nonignorable models based on different factorizations of the likelihood of the outcome process and the dropout process: selection models (Wu and Carroll, 1988; Diggle and Kenward, 1994; Follman and Wu, 1995, among others) and pattern-mixture models (Wu and Bailey, 1989; Little, 1993, 1994; Hogan and Laird, 1997a; Fitzmaurice, Laird, and Schneyer, 2001, among others). Approaches based on estimation equations include inverse probability methods (Robins, Rotnitzky, and Zhao, 1995; Rotnitzky, Robins, and Scharfstein, 1998; Scharfstein, Robins, and Rotnitzky, 1999). Recent reviews of methods handling nonignorable dropout in longitudinal data can be found in Little (1995, 2008), Verbeke and Molenberghs (2000), and Hogan, Roy, and Korkontzelou (2004).

Little (2008) defined a new class of likelihood-based models, namely, mixed-effect hybrid models (MEHMs), based on a new factorization of the likelihood of the outcome process and the dropout process. Unlike selection models and pattern-mixture models, we factorize the likelihood of the outcome process and the dropout process into the marginal distribution of random effects, the conditional distribution of the dropout pattern given random effects, and the conditional distribution of the outcome given both random effects and the dropout pattern. The resulting MEHMs have features of selection models in that they directly model the dropout process, and also have features of pattern-mixture models in that the sample is stratified by the missing data patterns and the outcome process is modeled over these patterns. As a result, the MEHM directly models the missing data mechanism, an attractive feature of selection models, and it shares with pattern-mixture models the desirable feature of computational simplicity. MEHMs also provide a generalization of the shared-parameter model (SPM; Wu and Carroll, 1988; De Gruttola and Tu, 1994; Follman and Wu, 1995). Because SPMs are nested within MEHMs, MEHMs provide a convenient way to test the conditional independence assumption and goodness of fit of SPMs via likelihood ratio tests.

We apply our methods to a double-blind randomized pediatric AIDS trial (Brady et al., 1996), previously analyzed by Hogan, Lin, and Herman (2004). The primary objective of the trial was to investigate whether a lower dosage (90 mg/m^2/dose) of zidovudine was equally effective as a higher dosage (180 mg/m^2/dose) of zidovudine to treat HIV-infected children (3 months to 12 years of age) with mild to moderate symptoms. A total number of 424 subjects were

enrolled and randomized into a low-dose group ($n = 216$) and a high-dose group ($n = 208$). Participants were followed up to 5 years to collect a number of endpoints, such as CD4 cells count, neuropsychologic evaluation, and serum p24 antigen level. In this article, we are interested in comparing the longitudinal trajectory of CD4 cell counts between the two dose groups. Subjects were scheduled for measurements of CD4 count before entry and every 12 weeks up to week 200; however, the actual measurement times may have varied. This trial experienced a significant amount of dropout: only 52% and 45% of the subjects completed 3 years of follow-up for the low- and high-dose groups, respectively. There were also some intermittent missing values. As the intermittent missing visits were often due to reasons unrelated to the response variable (i.e., CD4 count), we assume that the intermittent missing data are missing at random. Because methods proposed here are likelihood based, we can ignore the intermittent missing data and only focus on nonignorable missing data due to dropout.

In AIDS studies, dropout is often associated with the level or underlying rate of change of the CD4 counts, thus leading to nonignorable missing data (Wu and Carroll, 1988; De Gruttola and Tu, 1994; Hogan and Laird, 1997b). Figure 1 shows the estimated individual least-square slopes versus dropout time for the children in the HIV trial. It is clear that the lower slopes are associated with early dropout, suggesting that the dropouts may be nonignorable. In this case, the standard random-effects model (REM) leads to biased estimates (Little and Rubin, 2002), and we need to model the dropout process to obtain consistent estimates.

The remainder of the article is organized as follows. In Section 2, we define MEHMs, and in Section 3, we provide estimation procedures. In Section 4, we present a simulation study comparing these methods with other methods. In

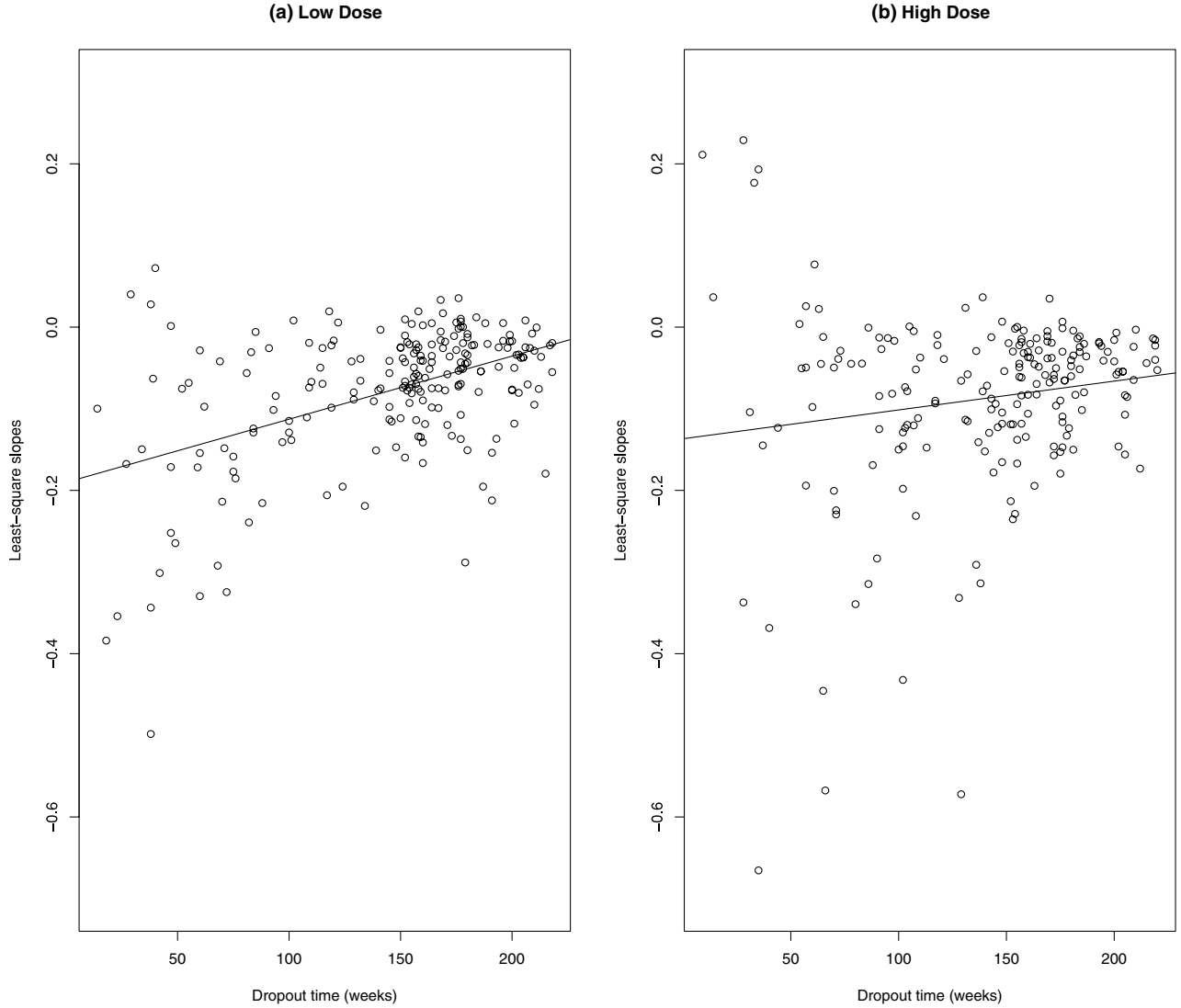

Figure 1. Subject-specific least-square slopes for square root of CD4 count as a function of dropout time for (a) low dose and (b) high dose with least-square regression line.

Section 5, we illustrate our methods by analyzing the pediatric AIDS trial data. We provide conclusions in Section 6.

2. Models

Suppose a longitudinal study is designed to collect K repeated measurements of a response Y and covariates X for each of n subjects. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iK})$ denote a vector of repeated measurements planned for subject i, and $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iK})$ with $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})'$ denote the associated $K \times p$ covariates matrix. We assume that $\boldsymbol{X}_i$ is fully observed (e.g., external or fixed by study design), but Y_{ij} is subject to missingness due to dropout. Let D_i index dropout patterns, such that $D_i = K$ for complete cases, and $D_i = k$ if the subject i drops out between the kth and $(k+1)$th measurement time, i.e., $y_{i1}, \ldots, y_{i,k}$ are observed and $y_{i,k+1}, \ldots, y_{iK}$ are missing, for $k = 1, \ldots, K$. We assume at least one observation for each subject, as subjects without any observations have no information and are often excluded from the analysis. We use $\boldsymbol{Y}_i^o$ to denote the observed part of $\boldsymbol{Y}_i$, and $\boldsymbol{Y}_i^m$ to denote the missing part of $\boldsymbol{Y}_i$.

Longitudinal data are often modeled by mixed-effects models (Laird and Ware, 1982) where random effects, say $\boldsymbol{b}_i$, are used to model the correlation of repeated measurements on the same subject. The likelihood of the observed data is obtained by integrating the missing data $\boldsymbol{Y}_i^m$ and random effects $\boldsymbol{b}_i$ from the joint distribution of $(\boldsymbol{Y}_i,\ D_i,\ \boldsymbol{b}_i)$:

$$L(\boldsymbol{\theta} \mid \boldsymbol{D}, \boldsymbol{Y}^o, \boldsymbol{X}) = \prod_{i=1}^{N} \int\int f(D_i, \boldsymbol{Y}_i, \boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\theta})\, d\boldsymbol{Y}_i^m\, d\boldsymbol{b}_i, \quad (1)$$

where $\boldsymbol{\theta}$ denotes a vector of unknown parameters.

2.1 *Selection and Pattern-Mixture Models*

A rich class of models can be formed by different factorization of $f(D_i, \boldsymbol{Y}_i, \boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\theta})$ in equation (1). In particular, Little (1995) calls models based on the factorization

$$\begin{aligned} f(D_i, \boldsymbol{Y}_i, \boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\theta}) = {} & f_B(\boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\gamma}_1) f_{Y\mid B}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, \boldsymbol{\gamma}_2) \\ & \times f_{D\mid Y,B}(D_i \mid \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{b}_i, \boldsymbol{\psi}) \end{aligned} \quad (2)$$

mixed-effect selection models, where $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$, and $\boldsymbol{\psi}$ are parameter vectors indexing the corresponding densities. Models based on the factorization

$$\begin{aligned} f(D_i, \boldsymbol{Y}_i, \boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\theta}) = {} & f_D(D_i \mid \boldsymbol{X}_i, \boldsymbol{\delta}) f_{B\mid D}(\boldsymbol{b}_i \mid \boldsymbol{X}_i, D_i, \boldsymbol{\nu}_1) \\ & \times f_{Y\mid B,D}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, D_i, \boldsymbol{\nu}_2) \end{aligned} \quad (3)$$

are called mixed-effect pattern-mixture models, with $\boldsymbol{\delta}, \boldsymbol{\nu}_1$, and $\boldsymbol{\nu}_2$ parameter vectors indexing the corresponding densities.

Mixed-effect selection models provide a natural way of factoring the model, with $f_B f_{Y\mid B}$ the mixed-effects model for the data in the absence of missing values, and $f_{D\mid Y,B}$ the model for the missing-data mechanism that specifies how the dropout depends on the outcome variable Y and latent random effects $\boldsymbol{b}_i$. As a special case of mixed-effect selection models, SPMs (Wu and Carroll, 1988; De Gruttola and Tu, 1994; Follman and Wu, 1995; Ten Have et al., 1998) assume that dropout is independent of outcome $\boldsymbol{Y}_i$ when conditioned on the underlying random effects $\boldsymbol{b}_i$, i.e., $f_{D\mid Y,B}(D_i \mid \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{b}_i, \boldsymbol{\psi}) = f_{D\mid B}(D_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, \boldsymbol{\psi})$. SPMs implicitly correlate the dropout process with the longitudinal outcome process by modeling both processes with shared random effects.

Mixed-effect pattern-mixture models are based on factorization (3). This approach stratifies the sample by the patterns of missing data (e.g., by the times of dropout) and then models the difference in the distribution of $\boldsymbol{Y}_i$ over these patterns. Pattern-mixture models are particularly suitable for situations in which it is not substantively meaningful to consider nonresponse as missing data, and it may make better sense to restrict the inference to the subpopulation of cases with observed values. Other desirable features of pattern-mixture models are that in some cases they do not require full specification of the missing data mechanism, and they are often computationally simpler than selection models. For comparisons of selection and pattern-mixture models, see for example Little (1995), Michiels, Molenberghs, and Lipsitz (1999), and Verbeke and Molenberghs (2000).

2.2 *Mixed-Effects Hybrid Model*

Little (2008) proposed a new class of models based on another factorization of the joint distribution of $\boldsymbol{Y}_i,\ \boldsymbol{b}_i$, and D_i:

$$\begin{aligned} f(D_i, \boldsymbol{Y}_i, \boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\theta}) = {} & f_B(\boldsymbol{b}_i \mid \boldsymbol{X}_i, \boldsymbol{\gamma}_1) f_{D\mid B}(D_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, \boldsymbol{\phi}) \\ & \times f_{Y\mid B,D}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, D_i, \boldsymbol{\nu}_2). \end{aligned} \quad (4)$$

This factorization has features of both equations (2) and (3). In particular, the first two factors explicitly model the dropout process, a feature of mixed-effects selection models, and the third factor models the longitudinal outcome process conditional on the pattern of missing data, a feature of pattern-mixture models. Therefore, we call models based on the factorization (4) MEHMs. Little (2008) did not consider these models in any detail or provide examples.

SPMs are special cases of MEHMs obtained by assuming the same outcome processes across different missing patterns given $\boldsymbol{b}_i$, $f_{Y\mid B,D}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, D_i, \boldsymbol{\nu}_2) = f_{Y\mid B}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{b}_i, \boldsymbol{\nu}_2)$, that is, conditional independence between Y_i and D_i given the random effects $\boldsymbol{b}_i$. This crucial assumption underlying the SPM is often difficult to test. Pulkstenis, Ten Have, and Landis (1998) proposed testing the conditional independence by including observed data $\boldsymbol{Y}_i^o$ as covariates in the model of the dropout process. Because the SPM is nested within the MEHM, a likelihood ratio test can be used to test the validity of the conditional independence assumption of SPMs.

The MEHM combines some desirable features of selection models and pattern-mixture models. In common with selection models, the MEHM directly models the dropout process. This is natural and easily understood by researchers, thereby facilitating the elicitation of expert opinions to formulate reasonable dropout models. Incorporating outside information about the missing data mechanism is especially important for nonignorable models because the missing data mechanism is untestable based on observed data. On the other hand, as with pattern-mixture models, the computation for MEHMs is relatively simple, because the missing data $\boldsymbol{Y}_i^m$ can be integrated out from the joint distribution of $(D_i, \boldsymbol{Y}_i^o, \boldsymbol{Y}_i^m, \boldsymbol{b}_i)$ under the factorization (4). As we describe in Section 3, some

MEHMs can be fitted conveniently by standard software such as `SAS Proc NLMIXED`.

The MEHM is also advantageous when it is meaningful to model the outcomes in strata defined by the patterns of missing data. An important example is when dropout is caused by death, and it is not meaningful to define values of the outcome for individuals who do not survive. This is often the case with clinical trials of AIDS or other severe diseases, where many of the dropouts are attributable to death. However, in the pediatric AIDS clinical trial we consider here, patients enrolled were a relatively healthy group of HIV-infected children, and only 18% of dropouts were due to death. The majority of dropouts were a result of parental or physician request because of concern of disease progression or interest in newer antiretroviral agents (Brady et al., 1996). In our analysis we make inferences about marginal effects of the drug by averaging over missing data patterns, while noting that if we had information to distinguish dropout by death from other causes, the analysis would be improved by conditioning on the subset of cases who survived. The MEHM model provides flexibility to make conditional inference or marginal inference (with respect to the missing data strata), depending on which is substantively meaningful.

We now describe the particular form of MEHM used to analyze these data. We model the longitudinal outcome process using a linear mixed-effects model:

$$\begin{aligned} \boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{b}_i, D_i = k &\sim N_K\left(\boldsymbol{X}_i\boldsymbol{\beta}^{(k)} + \boldsymbol{Z}_i\boldsymbol{b}_i, \boldsymbol{\Sigma}^{(k)}\right) \quad (5) \\ \boldsymbol{b}_i \,|\, \boldsymbol{\Gamma} &\sim N_q(\boldsymbol{0}, \boldsymbol{\Gamma}) \end{aligned}$$

where $N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the K-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\boldsymbol{\beta}^{(k)}$ is a vector of unknown parameters characterizing the conditional effects of covariates $\boldsymbol{X}_i$ given the missing pattern k; $\boldsymbol{\Sigma}^{(k)}$ and $\boldsymbol{\Gamma}$ are unknown covariance matrices; and $\boldsymbol{Z}_i$ is a known $K \times q$ design matrix associated with random effects $\boldsymbol{b}_i$.

To model the missing data mechanism, we define the discrete hazard rate of dropout $\lambda_{ik} = \Pr(D_i = k \,|\, D_i \geq k; \boldsymbol{w}_{ik}, \boldsymbol{b}_i)$, for $k = 1, \ldots, K$, where $\boldsymbol{w}_{ik}$ are fixed covariates and $\lambda_{i,K} = 1$. We model the discrete time dropout process by the continuation-ratio logit model (Agresti, 2002):

$$\operatorname{logit}(\lambda_{ik}) = \boldsymbol{w}'_{ik}\boldsymbol{\gamma} + \boldsymbol{\phi}'\boldsymbol{b}_i, \quad (6)$$

where $\boldsymbol{\phi}$ is a vector of parameters that govern the relationship between the random effects and the dropout process. The conditional probability of the ith subject who drops out right after the kth measurement given $\boldsymbol{w}_{ik}$ and the random effects $\boldsymbol{b}_i$ is

$$\Pr(D_i = k \,|\, \boldsymbol{b}_i, \boldsymbol{w}_{ik}) = \lambda_{ik}\prod_{j=1}^{k-1}(1 - \lambda_{ij}). \quad (7)$$

Note that if $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{(k)} = \boldsymbol{\Sigma}$ in model (5), the above MEHM becomes an SPM.

When the effect of covariates $\boldsymbol{X}_i$ on the marginal mean of Y is of interest, it can be obtained by averaging the conditional mean over the random effects $\boldsymbol{b}_i$ and the missing pattern D_i. For the model (5) this yields

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{Z}_i) &= E_{\boldsymbol{b}_i} E_{D_i|\boldsymbol{b}_i}\left(\boldsymbol{X}_i\boldsymbol{\beta}^{(k)} + \boldsymbol{Z}_i\boldsymbol{b}_i\right) \\ &= \boldsymbol{X}_i \sum_{k=1}^{K} E_{\boldsymbol{b}_i}(\pi_{ik})\boldsymbol{\beta}^{(k)}, \quad (8) \end{aligned}$$

where $\pi_{ik} = \Pr(D_i = k \,|\, \boldsymbol{b}_i, \boldsymbol{w}_{ik})$. The marginal effect of the covariates $\boldsymbol{X}_i$ is a weighted average of $\boldsymbol{\beta}^{(k)}$. The weight $E_{\boldsymbol{b}_i}(\pi_{ik})$ can be obtained by numerically integrating out random effects $\boldsymbol{b}_i$ from the conditional dropout probability (7). More simply, if this is assumed independent of the covariates, it can be estimated by n_k/n, where n_k is the number of subjects who dropped out between the kth and $(k+1)$th measurement time.

Nonignorable models typically suffer from identification problems. The MEHM (5) specifies distinct fixed parameters $\boldsymbol{\beta}^{(k)}$ and $\boldsymbol{\Sigma}^{(k)}$ and is clearly not identified, because the parameters of the complete-data distribution of $\boldsymbol{Y}_i$ are not estimable for incomplete patterns. For instance, for $D_i = k$ with $k \neq K$, some components of the covariance structure $\boldsymbol{\Sigma}^{(k)}$ are not estimable. Particular structures might be specified for $\boldsymbol{\Sigma}^{(k)}$ or/and $\boldsymbol{\beta}^{(k)}$ to identify the model. For $\boldsymbol{\Sigma}^{(k)}$, one common assumption in mixed-effects models is that $\{Y_{i1}, \ldots, Y_{iK}\}$ are independent given the random effects $\boldsymbol{b}_i$, yielding $\boldsymbol{\Sigma}^{(k)} = \sigma_k^2\boldsymbol{I}$, where $\boldsymbol{I}$ is a $K \times K$ identity matrix. More often, we assume that $\{Y_{i1}, \ldots, Y_{iK}\}$ are identically independent distributed conditional on $\boldsymbol{b}_i$, i.e., $\boldsymbol{\Sigma}^{(k)} = \sigma^2\boldsymbol{I}$. If the conditional independence of $\{Y_{i1}, \ldots, Y_{iK}\}$ does not hold, we may assume that the covariance matrix is the same across missing patterns, i.e., $\boldsymbol{\Sigma}^{(k)} = \boldsymbol{\Sigma}$, to identify the parameters, but this approach requires that some subjects complete the study. The identifiability of $\boldsymbol{\beta}^{(k)}$ depends on the nature of the associated covariates $\boldsymbol{X}_i$. If $\boldsymbol{X}_i$ are time independent, $\boldsymbol{\beta}^{(k)}$ is individually estimable as long as the number of subjects in each missing pattern exceeds the number of parameters. If the covariate is time or time dependent, $\boldsymbol{\beta}^{(1)}$ is not identified, because the estimation of straight lines requires at least two observed time points, and this pattern contains only one observation. These cases might simply be omitted from the analysis. Alternatively, we may impose some structure on $\boldsymbol{\beta}^{(k)}$, for example, assuming that the intercept is independent of the pattern, and the slope is linearly related with the dropout time (Wu and Bailey, 1989). We adopt this approach to obtain a parsimonious MEHM for the pediatric AIDS trial, for which the number of missing patterns is large.

3. Estimation

Under the MEHM, the estimates of unknown parameters $\boldsymbol{\beta}^{(k)}$, σ_k, $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\phi}$ are obtained by maximizing the marginal likelihood (1). We obtained this marginal likelihood by numerical integration over the random-effects distribution using the adaptive Gaussian quadrature approximation (Pinheiro and Bates, 1995). We then maximized the marginal likelihood with respect to the unknown parameters using a quasi-Newton approach. To choose appropriate initial values of parameters in the quasi-Newton maximization algorithm, we first fit the hybrid model by a two-stage estimation procedure: at the first stage, fit the model (5) without considering the dropout process; and at the second stage, fit the dropout model (6) by replacing $\boldsymbol{b}_i$ with the estimates obtained from the first stage. Estimates of parameters resulting from

the two-stage estimation can then be used as initial values in the quasi-Newton algorithm. These estimates are biased, but they serve well as initial values for maximization. The adaptive Gaussian quadrature and quasi-Newton algorithms have been implemented in `Proc NLMIXED` in `SAS 9.1`. We used this procedure to fit the MEHM.

4. Simulation Study

We carried out a simulation study to compare the performance of MEHM with several alternative models. We assumed $n = 200$ subjects with four repeated measures, and two covariates X and Z, where X was generated from $N(1, 1)$, and $Z = \{1, 2, 3, 4\}$ denotes the vector of measurement times. The outcome variable Y was generated according to the following model

$$Y_{ij} \mid X_i, Z_{ij}, b_{0i}, b_{1i}, D_i = k \sim N\left(b_{0i} + b_{1i}Z_{ij} + \beta^{(k)}X_i, \sigma_k^2\right), \quad k = 1, \ldots, 4.$$

$$b_{0i} \mid \alpha_0, \tau_0 \sim N\left(\alpha_0, \tau_0^2\right), \qquad b_{1i} \mid \alpha_1, \tau_1 \sim N\left(\alpha_1, \tau_1^2\right).$$

We assumed that the dropout probability depends only on the random slope b_{1i} and covariate X via $\text{logit}(\lambda_{ik} \mid b_{1i}) = \phi_0 + \phi_1 b_{1i} + \gamma X_i$. We set $\phi_0 = -4.1, \phi_1 = 1, \gamma = -0.6, \alpha_1 = 3$, and $\tau_1^2 = 2$, so that on average 45% of subjects drop out prematurely. In terms of how the outcome process depends on dropout patterns, we simulated four scenarios: (i) both the regression parameter and variance vary across dropout patterns, i.e., $\boldsymbol{\beta} = \{\beta^{(1)}, \beta^{(2)}, \beta^{(3)}\beta^{(4)}\} = \{1, 2, 4, 5\}$ and $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\} = \{1, 2, 4, 6\}$; (ii) the regression parameter varies across the dropout pattern but the variance does not, i.e., $\boldsymbol{\beta} = \{1, 2, 4, 5\}$ and $\boldsymbol{\sigma}^2 = \{2, 2, 2, 2\}$; (iii) the variance varies across the dropout pattern but the regression parameter does not, i.e., $\boldsymbol{\beta} = \{3.62, 3.62, 3.62, 3.62\}$ and $\boldsymbol{\sigma}^2 = \{1, 2, 4, 6\}$; and (iv) both the regression parameter and variance are constant across dropout patterns, i.e., $\boldsymbol{\beta} = \{3.62, 3.62, 3.62, 3.62\}$, and $\boldsymbol{\sigma}^2 = \{2, 2, 2, 2\}$. We set other parameters as $\alpha_0 = 2$ and $\tau_0^2 = 1.5$. Dropouts are nonignorable in the above four scenarios. For comparison, we also simulated another scenario in which dropouts are missing at random by setting $\phi_1 = 0$, $\boldsymbol{\beta} = \{3.62, 3.62, 3.62, 3.62\}$, and $\boldsymbol{\sigma}^2 = \{2, 2, 2, 2\}$. Under each of these scenarios, 1000 data sets were generated. We analyzed each simulated data set using a standard REM (Laird and Ware, 1982), an SPM, and an MEHM. We were interested in making inference about the marginal effect of Z and X, i.e., α_0, α_1, and β, where β is the marginal effect of X obtained by averaging $\beta^{(k)}$ over the dropout patterns. In our simulation, the true values of α_0, α_1, and β are 2, 3, and 3.62, respectively.

The results are reported in Table 1. Across all scenarios, the MEHM yielded unbiased estimates of α_0, α_1, and β and sound coverage probabilities close to the nominal value (0.950). In contrast, the SPM led to biased estimates of α_0, α_1, and β when $\beta^{(k)}$ varied across missing patterns (scenarios 1 and 2), i.e., the conditional independence assumption was violated. Although α_0 and α_1 were constant across missing patterns, misspecification of $\beta^{(k)}$ also caused bias in the estimates of these parameters. In these situations, coverage probabilities of 95% confidence intervals of these parameters were generally poor (i.e., less than 70%). When regression parameters were constant and only the residual variance σ_k^2 varied across

Table 1

Standardized bias (SB) (i.e., bias/standard error), average standard error (SE), and coverage probability (CP) of the 95% confidence interval for estimates of regression parameters based on the standard REM, the SPM, and the MEHM. The biased estimates are highlighted by boldface.

	α_0(=2)			α_1(=3)			β (=3.62)		
Model	SB	SE	CP	SB	SE	CP	SB	SE	CP
Scenario 1: $\beta^{(k)} = \{1, 2, 4, 5\}$, $\sigma_k^2 = \{1, 2, 4, 6\}$									
REM	**1.64**	0.34	0.608	**−3.86**	0.14	0.030	**2.19**	0.21	0.449
SPM	**1.31**	0.36	0.700	**−2.50**	0.18	0.262	**2.19**	0.21	0.464
MEHM	0.03	0.30	0.948	0.00	0.16	0.947	0.05	0.22	0.946
Scenario 2: $\beta^{(k)} = \{1, 2, 4, 5\}$, $\sigma_k^2 = \{2, 2, 2, 2\}$									
REM	**2.14**	0.29	0.401	**−4.58**	0.12	0.004	**2.11**	0.19	0.445
SPM	**1.73**	0.30	0.577	**3.29**	0.14	0.111	**2.11**	0.19	0.448
MEHM	0.00	0.14	0.951	0.07	0.14	0.951	−0.10	0.20	0.938
Scenario 3: $\beta^{(k)} = \{3.62, 3.62, 3.62, 3.62\}$, $\sigma_k^2 = \{1, 2, 4, 6\}$									
REM	**1.61**	0.31	0.623	**−3.86**	0.14	0.042	0.11	0.19	0.940
SPM	0.06	0.32	0.942	0.00	0.17	0.942	0.00	0.19	0.947
MEHM	0.00	0.30	0.943	0.00	0.16	0.951	0.00	0.18	0.934
Scenario 4: $\beta^{(k)} = \{3.62, 3.62, 3.62, 3.62\}$, $\sigma_k^2 = \{2, 2, 2, 2\}$									
REM	**1.79**	0.24	0.609	**−3.17**	0.12	0.115	−0.27	0.15	0.945
SPM	0.04	0.24	0.946	0.00	0.14	0.947	0.00	0.15	0.942
MEHM	0.04	0.24	0.944	0.00	0.14	0.941	0.00	0.16	0.956
Scenario 5: Missing at random									
REM	0.04	0.24	0.952	0.07	0.13	0.947	0.00	0.15	0.943
SPM	0.04	0.24	0.957	0.07	0.13	0.951	0.00	0.15	0.939
MEHM	−0.04	0.24	0.956	0.07	0.14	0.951	0.00	0.16	0.943

missing patterns (scenarios 3 and 4), SPM yielded estimates of regression parameters with little empirical bias and close to nominal coverage probabilities, but estimates of the variance components σ_k, τ_0, and τ_1 were biased (results are not shown). As expected, the estimates from REM were generally biased except when missing data were missing at random. In the case in which regression parameters were constant and only the residual variance varied across missing patterns (scenarios 3 and 4), the REM estimates of α_0 and α_1 were biased but the estimate of β was largely unbiased because β was not associated with the dropout process. In summary, the simulation demonstrates that estimates based on SPM can be seriously biased if the conditional independent assumption is violated, and that MEHM is a more general model that does not require the conditional independence assumption to yield consistent estimates.

The estimation procedure proposed in Section 3 is rather efficient. Under our simulation settings, it took about 20 seconds to fit the MEHM in a PC with a Pentium 3.00 GHZ CPU, and about 98.5% iterations successfully converged. The algorithm failed to converge in a small percentage of simulations, because the number of observations with some missing data patterns was occasionally small (e.g., less than 5).

To assess robustness, we assessed the behavior of our models as the distribution of the response Y_{ij} departures from normal. We simulated data by assuming that the error of Y_{ij} was from a t distribution with a degree of freedom of 2, or from a gamma distribution with the scale parameter of 2 and the shape parameter of 2. The simulation results (Web Table 1) suggest that the proposed model is not particularly robust to the severe violation of the normality assumption. Where possible, an appropriate transformation should be considered to improve the normality of data.

5. Application

In this section, we describe our analysis of the pediatric AIDS data using three different models: REM, SPM, and MEHM. The square root transformation is applied to the CD4 count to improve the normality, and we fitted each model separately by dose.

5.1 *Models*

A simple but reasonable approach to analyze the pediatric AIDS data is to use an REM:

$$Y_{ij} \mid b_{0i}, b_{1i} \sim N\left(\beta_0 + \beta_1 t^*_{ij} + b_{0i} + b_{1i} t^*_{ij}, \sigma^2\right) \tag{9}$$
$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{11,} & \tau_{12} \\ \tau_{12,} & \tau_{22} \end{pmatrix}\right),$$

where Y_{ij} is the square root of the CD4 count at the jth measurement time t_{ij}; β_0 and β_1 are fixed-effects parameters; and b_{0i} and b_{1i} are subject-specific random intercepts and slopes. Following Hogan, Lin, and Herman (2004), we rescaled the time using $t^*_{ij} = t_{ij}/\text{range}(t_{ij})$ so that the new time scale has a range of 1. The rescaling makes the estimates more stable by increasing the variance of individual slopes away from zero, and also improves the interpretation of parameters. In particular, β_1 represents the average total change in the square root of the CD4 count from baseline to the longest follow-up time for the study. The linear relationship between the square root of the CD4 count Y_{ij} and the measurement time t_j for both treatment arms is supported by the plots (Web Figure 1).

The association between the dropout time and the individual least-square estimate of the slopes (Figure 1) motivated us to fit an SPM to reflect that relationship. SPMs have been previously used to account for nonignorable (or informative) dropouts in AIDS clinical trials by Wu and Carroll (1988) and De Gruttola and Tu (1994). SPMs consist of two components: a model for the outcome process and a model for the dropout process. We model the measurement process using the REM (9), and model the dropout process using the continuation-ratio logit model (6) with random intercepts and slopes as regressors.

We also fit the proposed MEHM to the pediatric AIDS data. Given that there are 17 possible missing patterns and some of them have sparse data, we adopt the approach of Wu and Carroll (1989) and assume that

$$\beta_0^{(k)} = \gamma_1 + \gamma_2 t^{(k)^*}, \quad \beta_1^{(k)} = \gamma_3 + \gamma_4 t^{(k)^*}, \tag{10}$$

where $t^{(k)^*} = (t^{(k)} - \frac{1}{K}\sum_k t^{(k)})/\text{range}(t_{ij})$ with $t^{(k)}$ denotes the dropout time of the kth missing pattern. We centered the dropout time $t^{(k)}$ so that γ_1 and γ_3 have desirable marginal interpretations (Fitzmaurice et al., 2001). Note that if $\gamma_2 = \gamma_4 = 0$, the MEHM becomes an SPM.

5.2 *Results*

Table 2 shows estimates and associated standard errors for the marginal regression coefficients. In the REM and SPM, β_0 and

Table 2
Estimates of intercept and slope characterizing the marginal mean of the CD4 trajectory under the standard REM, the SPM, and the MEHM. Standard errors of the estimates appear in parentheses.

		Zidovudine dose			
Model	Parameter	Low (90 mg)	High (180 mg)	Difference	p-value
REM	β_0	28.6 (0.8)	30.1 (0.9)		
	β_1	−12.7 (0.8)	−18.2 (1.4)	5.5 (1.6)	0.0007
SPM	β_0	28.8 (0.8)	30.1 (0.9)		
	β_1	−13.6 (0.9)	−19.0 (1.4)	5.4 (1.6)	0.0007
MEHM	γ_1	29.0 (0.8)	30.1 (0.9)		
	γ_3	−16.7 (1.0)	−20.7 (1.5)	4.0 (1.8)	0.03

Table 3
Estimates of parameters characterizing the dropout process and the overall likelihood under the SPM and the MEHM. Standard errors of the estimates appear in parentheses.

		Zidovudine dose	
Dosage	Parameter	SPM	MEHM
Low	ϕ_0	−2.39 (0.07)	−2.40 (0.07)
(90 mg)	ϕ_1	−0.017 (0.007)	0.001 (0.009)
	ϕ_2	−0.013 (0.005)	0.001 (0.009)
	τ_{11}	143.1 (14.3)	136.9 (14.7)
	τ_{22}	122.4 (15.9)	117.9 (14.9)
	τ_{12}	−68.7 (12.4)	−81.2 (12.5)
	Likelihood	−8705	−8667
High	ϕ_0	−2.35 (0.07)	−2.37 (0.07)
(180 mg)	ϕ_1	−0.017 (0.007)	0.003 (0.009)
	ϕ_2	−0.010 (0.005)	−0.004 (0.006)
	τ_{11}	152.9 (15.8)	152.5 (17.3)
	τ_{22}	338.3 (42.3)	338.3 (50.4)
	τ_{12}	−115.4 (20.3)	−116.0 (21.6)
	Likelihood	−8033	−8011

β_1 are marginal intercept and slope regression parameters. In the MEHM, as described in equation (8), we need to average over the dropout patterns to obtain the marginal regression parameters. Because we centered the dropout time in equation (10), γ_1 and γ_3 directly represent the marginal intercept and slope parameters. As expected, the slope estimates for SPM and MEHM are smaller than for REM, because SPM and MEHM take into account the fact that early dropouts are associated with lower slopes. There is no appreciable difference in the estimates of the intercept in the three models. When comparing the slopes between two treatment arms, REM and SPM yield very similar results, i.e., the estimated mean difference between the low- and high-dose arms in total change in the square root of the CD4 count is about 5.5, with a highly significant p-value of 0.0007, suggesting that the low dose of zidovudine is superior to the high dose because the decline in the CD4 count is less steep. Under MEHM, the estimated mean difference between the low- and high-dose arms in total change in the square root of the CD4 count is decreased to 4.0, with a p-value of 0.03. Although we draw the same conclusion with MEHM, the superiority of the low dose is substantially

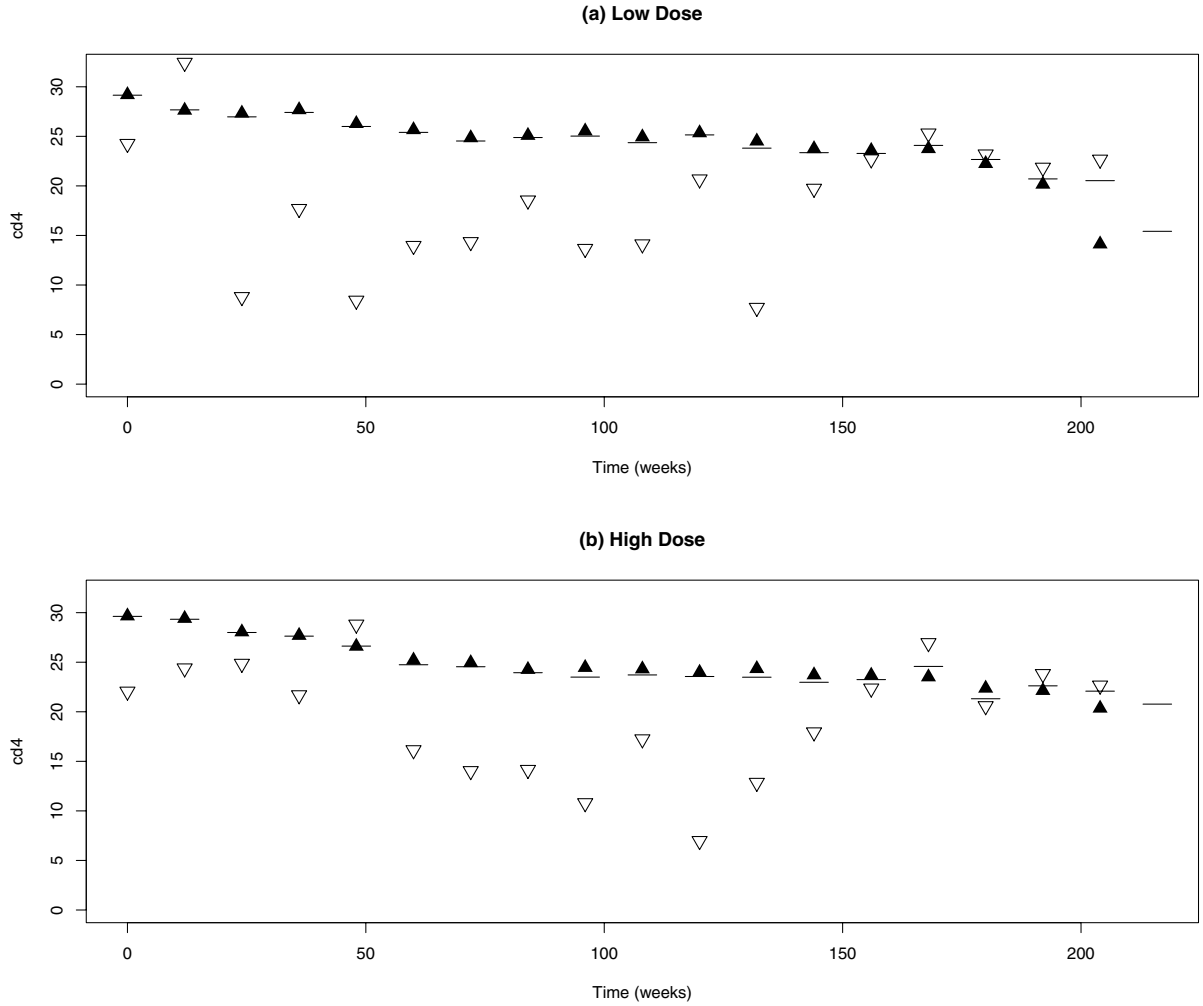

Figure 2. Sample means of the square root of the CD4 count at each measurement time. The horizontal line segments are the overall means; the solid triangles pointing upward are the means for subjects who do not drop out at the subsequent visit; and the open triangles pointing downward are the means for subjects who drop out at the subsequent visit.

weakened using this approach, i.e., the effect size is decreased from 3.38 to 2.22.

The estimates of slope based on the SPM and MEHM are substantially different in the AIDS trials (Table 2). To explore this difference, we display estimates of other parameters in Table 3. Noticeably, in the SPM, the regression parameters of the dropout process ϕ_1 and ϕ_2 are both significant in both the high- and low-dose arms, suggesting that the dropout process depends on the random intercept and slopes in the outcome process. However, in MEHM, the dependence between dropout and random effects is weakened, and both ϕ_1 and ϕ_2 are not significant when the measurement process is allowed to depend on dropout pattern after conditioning on random effects. The likelihood ratio test can be employed to assess the conditional independence assumption and goodness of fit of the SPM. For the low-dose arm, the maximized likelihood for SPM is -8705 and for MEHM is -8667, yielding a likelihood ratio test statistics of 76, which is highly significant; for the high-dose arm, the maximized likelihood for SPM is -8033 and for MEHM is -8011, yielding a highly significant likelihood ratio test statistic of 44. These results suggest that the conditional independence assumption of the SPM may not hold for the AIDS data set, i.e., the dropout process may still depend on the missing outcome conditional on random effects. As empirical evidence, we plotted the sampling mean of the CD4 count for subjects who had not dropped out at the subsequent visit, and subjects who had dropped out at the subsequent visit, across all visits (Figure 2). Clearly, subjects who had remained in the study tended to have larger CD4 counts than those who had dropped out by the subsequent visit, suggesting that the dropout process may directly depend on the value of the CD4 count at the time when a subject drops out of the study.

5.3 *Sensitivity Analysis*

The MEHM is more general and provides a better fit than the SPM for the pediatric AIDS trial data. However, as is common when modeling nonignorable missing data, the identification of MEHM is heavily driven by model assumptions. The observed data contain weak information about ϕ_1 and ϕ_2, and estimation of these parameters relies on the normality assumption of the random effects. Unfortunately, these distributional assumptions are untestable based on the observed data. In this case, a sensible strategy is to perform a sensitivity analysis. Following Rotnitzky et al. (1998), we set ϕ_1 and ϕ_2 at a series of fixed values (rather than estimating them based on the observed data), and then evaluated the sensitivity of the inference to the value of ϕ_1 and ϕ_2. If the inference showed no essential change, the interpretation of the results would be straightforward. Otherwise, there would be some residual ambiguity in the interpretation.

Figure 3 shows the results of the sensitivity analysis. We set $\phi_0 = \phi_1$ and varied their values from -0.4 to 0.4 for each dose arm. Compared with the estimates of these parameters

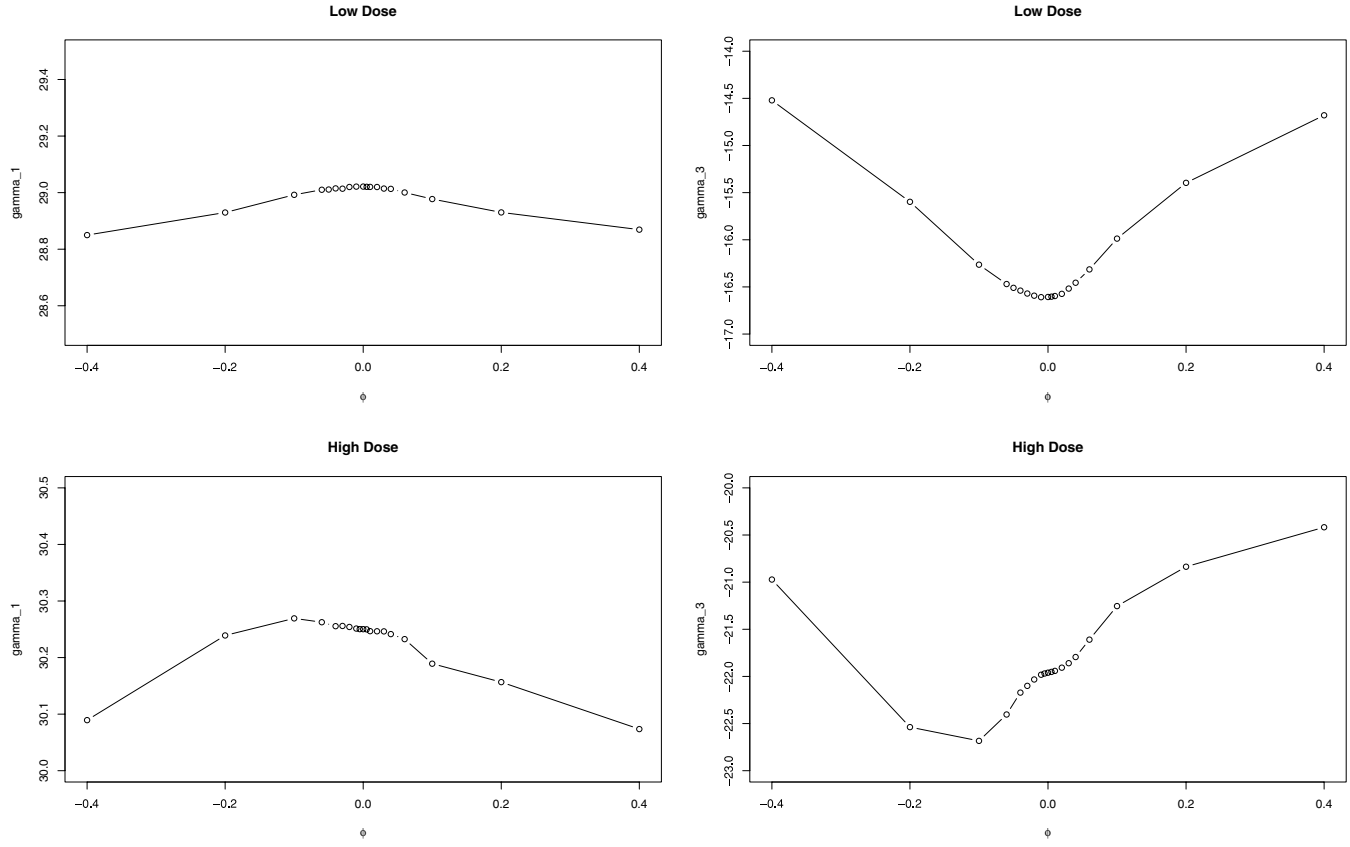

Figure 3. Sensitivity analysis of MEHM for the AIDS data. This figure depicts the change of regression parameters γ_1 and γ_3 by setting $\phi_1 = \phi_2 = \phi$ at various values.

based on the MEHM (Table 3), this range is rather wide, representing about a 24-fold change in each direction. As ϕ_0 and ϕ_1 vary from -0.4 to 0.4, for the low-dose arm, the value of γ_1 varies from 28.85 to 29.02, and the value of γ_3 varies from -16.61 to -14.52; for the high-dose arm, the value of γ_1 varies from 30.07 to 30.27, and the value of γ_3 varies from -22.68 to -20.42. Clearly, the estimates of γ_1 and γ_3 are quite stable, suggesting that our results based on the MEHM is not sensitive to the model assumptions for the AIDS data.

6. Conclusion

We have considered a new class of models, MEHMs, for analyzing longitudinal data with nonignorable dropout based on a new factorization of the joint distribution of random effects, the dropout process, and the measurement process. The MEHM has features of both selection models and pattern-mixture models, such as directly modeling the missing data mechanism, and computational simplicity. The MEHM can be viewed as a generalization of an SPM without making the conditional independence assumption of the dropout and measurement processes given random effects. Because SPMs are nested within MEHMs, by fitting the MEHM, the conventional likelihood ratio test can be used to test the conditional independence assumptions underlying the SPMs. Although we have noted some useful features of these models, it should be emphasized that these models do not resolve issues of lack of identifiability that are inherent with nonignorable nonresponse. Also, models based on other factorizations of the joint distribution of the measurement and missing-data process, such as random-effects selection models and random-effects pattern-mixture models, remain valid alternatives. Despite these caveats, MEHMs do provide another useful tool for modeling nonignorable missing data in longitudinal studies.

7. Supplementary Materials

Web Figures and Tables referenced in Section 4 and 5 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

The authors would like to thank Joseph Hogan for kindly providing the pediatric AIDS data, and an associate editor and two referees for useful comments on an earlier draft.

References

Agresti, A. (2002). *Analysis of Categorical Data*, 2nd edition. New York: Wiley.

Brady, M. T., McGrath, N., Brouwers, P., et al., for the Pediatric AIDS clinical trial. (1996). Randomized study of the tolerance and efficacy of high- versus low-dose zidovudine in human immunodeficiency virus-infected children with mild to moderate symptoms (ACTG 128). *Journal of Infectious Disease* **173,** 1097–1106.

De Gruttola, V. and Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50,** 1003–1014.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* **43,** 49–73.

Fitzmaurice, G. M., Laird, N. M., and Schneyer, L. (2001). An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. *Statistics in Medicine* **20,** 1009–1021.

Follman, D. and Wu, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51,** 151–168.

Hogan, J. W. and Laird, N. M. (1997a). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16,** 239–257.

Hogan, J. W. and Laird, N. M. (1997b). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16,** 259–272.

Hogan, J. W., Lin, X., and Herman, B. (2004). Mixture of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics* **60,** 854–864.

Hogan, J. W., Roy, J., and Korkontzelou, C. (2004). Biostatistics tutorial: Handling dropout in longitudinal data. *Statistics in Medicine* **23,** 1455–1497.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

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

Little, R. J. A. (1994). A class of pattern mixture models for normal missing data. *Biometrika* **81,** 471–483.

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

Little, R. J. A. (2008). Selection and pattern-mixture models. In *Advances in Longitudinal Data Analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (eds). London: CRC Press.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd edition. New York: Wiley.

Michiels, B., Molenberghs, G., and Lipsitz, S. R. (1999). Selection models and pattern-mixture models for incomplete data with covariates. *Biometrics* **55,** 978–983.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4,** 12–35.

Pulkstenis, E., Ten Have, T. R., and Landis, J. R. (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association* **93,** 438–450.

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

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association* **93,** 1321–1339.

Scharfstein, D., Robins, J., and Rotnitzky, A. (1999). Adjusting for nonignorable nonresponse using semiparametric nonresponse models with time dependent covariates (with discussion). *Journal of the American Statistical Association* **94,** 1096–1146.

Ten Have, T. R., Pulkstenis, E., Kunselman, A., and Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics* **54,** 367–383.

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

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

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44,** 175–188.

Received January 2008. *Revised April* 2008.
Accepted April 2008.