

Missing Covariates in Longitudinal Data with Informative Dropouts: Bias Analysis and Inference

Jason Roy

Department of Biostatistics and Computational Biology, University of Rochester, Rochester,
New York 14642, U.S.A.

email: jason_roy@urmc.rochester.edu

and

Xihong Lin

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

email: xlin@sph.umich.edu

SUMMARY. We consider estimation in generalized linear mixed models (GLMM) for longitudinal data with informative dropouts. At the time a unit drops out, time-varying covariates are often unobserved in addition to the missing outcome. However, existing informative dropout models typically require covariates to be completely observed. This assumption is not realistic in the presence of time-varying covariates. In this article, we first study the asymptotic bias that would result from applying existing methods, where missing time-varying covariates are handled using naive approaches, which include: (1) using only baseline values; (2) carrying forward the last observation; and (3) assuming the missing data are ignorable. Our asymptotic bias analysis shows that these naive approaches yield inconsistent estimators of model parameters. We next propose a selection/transition model that allows covariates to be missing in addition to the outcome variable at the time of dropout. The EM algorithm is used for inference in the proposed model. Data from a longitudinal study of human immunodeficiency virus (HIV)-infected women are used to illustrate the methodology.

KEY WORDS: Asymptotic bias; EM algorithm; Missing data; Random effects; Sensitivity analysis; Transition model.

1. Introduction

A substantial proportion of subjects often drop out from longitudinal studies prior to completion of the study. Dropouts may be nonignorable, in the sense that methods that ignore the mechanism that leads to dropout will often yield biased results. A variety of methods have been proposed for analyzing these types of data, including parametric (e.g., Diggle and Kenward, 1994; Little, 1995; Ibrahim, Chen, and Lipsitz, 2001) and semiparametric (e.g., Rotnitzky, Robins, and Scharfstein, 1998; Scharfstein, Rotnitzky, and Robins, 1999) approaches. It is commonly assumed in the longitudinal dropout literature that the outcomes are missing at the time of dropout, but all covariates are completely observed. For example, the approaches reviewed by Little (1995) require the covariates to be completely observed. However, time-varying covariates are common in longitudinal studies. These covariates, along with the outcome variable, are generally not observed at the time of dropout. Hence, the assumption of completely observed covariates is often not realistic in the presence of time-varying covariates. To illustrate, consider the following example.

Tashima et al. (2001) reported results from a longitudinal study of human immunodeficiency virus (HIV)-infected women. This was a substudy of the HIV Epidemiological Research Study (HERS; Smith et al., 1997), where interest was in determining whether the use of protease inhibitors (PIs) reduced the number of hospitalizations and emergency department visits. The HERS recruited 1310 participants from four cities, beginning in 1993, with follow-up visits every 6 months. Tashima et al. (2001) found a decreased risk of hospitalization for patients treated with PI, but only for those with CD4 count less than 200 cells/mL. However, a substantial number of women dropped out from the study, and a preliminary look at the data suggests they were more likely to have been hospitalized in the past 6 months. We are therefore interested in models that account for the possibility that missing data are missing not at random. That is, the probability of dropout may depend on whether or not a subject had a recent hospitalization. A further complication is certain PI use changes over time and therefore has the same pattern of missingness as the response. Models that account for both missing response and covariate data are therefore necessary.

Roy and Lin (2002) and Stubbendick and Ibrahim (2003) proposed maximum likelihood methods for dealing with both dropout-related missing response and missing covariates in a longitudinal setting. These methods require the specification of models for the dropout mechanism and covariates that have at least some missing values. Roy and Lin (2002) specifically focused on continuous time-varying covariates that are missing due to dropout. Both continuous and discrete covariates subject to missingness were modeled via a series of one-dimensional conditional distributions in Stubbendick and Ibrahim (2003). A limitation of their modeling approaches is they are only appropriate for normally distributed outcomes. However, it is very common to have a discrete response variable along with time-varying covariates of discrete or mixed type. For cross-sectional data, maximum likelihood methods have been developed for generalized linear models (GLMs) with nonignorable missing covariates (Ibrahim, Lipsitz, and Chen, 1999), but these methods have not yet been extended to the longitudinal setting. Further, although several naive methods have been proposed in the literature to handle missing time-varying covariates, such as the last observation carried forward (LOCF) method, little is known about how these naive methods perform, particularly in comparison to more complicated modeling strategies such as is proposed in this manuscript.

In this article, we consider the situation where dropouts are informative and both the outcome and time-varying covariates are missing at the time of dropout. We study the asymptotic bias that results from dealing with missing time-varying covariates using one of the following naive approaches: (1) using baseline measures throughout; (2) using the LOCF, i.e., for units that dropped out, assume the value of the time-varying covariates has not changed since the previous time point; and (3) assuming dropouts are ignorable. Our asymptotic bias analysis shows situations in which these naive approaches yield substantially biased estimators of model parameters.

In view of the asymptotic bias analysis results, we propose a model that allows for nonignorable dropouts as well as missing covariates. This is done by generalizing the models for dealing with dropout-related missing covariates proposed by Roy and Lin (2002) and Stubbendick and Ibrahim (2003) to allow a generalized linear mixed model (GLMM) for the response. The covariates may be continuous, discrete, or mixed type. To account for possibly nonignorable dropouts, we utilize a selection model. The time-varying covariates that are subject to missingness are assumed to follow a transition model. In view of multidimensional integration required for a full likelihood analysis, a Monte Carlo EM algorithm (Wei and Tanner, 1990) is developed for inference in the proposed model. Because the observed data cannot distinguish between missing at random (MAR) and informative missingness (Little, 1995), we proceed with a sensitivity analysis (Rotnitzky et al., 1998; Verbeke et al., 2001), which is commonly recommended.

The remainder of the article is organized as follows. Section 2 introduces the model. In Section 3, an asymptotic bias analysis is performed to study the impact of the use of naive approaches to handle missing covariates. Section 4 describes the EM algorithm for inference in the pro-

posed model. Two simulation studies are carried out in Section 5, comparing the proposed methods to the naive approaches in finite samples. This is done for continuous and binary data. We illustrate the proposed approach in Section 6 using the hospitalization data. A discussion is given in Section 7.

2. Model Specifications

Suppose that n units are sampled repeatedly over time. The intent is to sample each unit K times, but due to dropouts the i th unit is sampled at $K_i \leq K$ time points. Therefore, we observe the continuous or discrete response variable Y_{ik} , covariates X_{ik} ($p \times 1$) and Z_{ik} ($q \times 1$), associated with the fixed effects and the random effects, respectively, and a dropout indicator R_i for units $i = 1, \dots, n$ and time points $k = 1, \dots, K_i$. The variable R_i takes value k if unit i dropped out after time point $k - 1$ (i.e., the first missing value was at time point k). We distinguish between two types of covariates: those that are known even at the time of dropout S_{ik} ($p_1 \times 1$) (i.e., time-invariant covariates), and those whose values are unknown at the time of dropout T_{ik} ($p_2 \times 1$) (i.e., time-varying covariates). Write $X_{ik} = (S_{ik}^T, T_{ik}^T)^T$. We assume that the covariates associated with the random effects Z_{ik} are completely observed.

We assume a GLMM (Breslow and Clayton, 1993) for the outcome variable Y_{ik} . Specifically, given X_{ik} , Z_{ik} , and the random effects b_i , the Y_{ik} are independent with means μ_{ik} and variances $\phi_y a_{ik}^{-1} v(\mu_{ik})$, where ϕ_y is a scale parameter, a_{ik} is a weight, and $v(\cdot)$ is a variance function. The conditional mean is related to the covariates through the following GLM:

$$g(\mu_{ik}) = S_{ik}^T \beta_1 + T_{ik}^T \beta_2 + Z_{ik}^T b_i, \quad (1)$$

where $g(\cdot)$ is a link function, β_1 and β_2 are $p_1 \times 1$ and $p_2 \times 1$ vectors of unknown parameters, and the $q \times 1$ random effects b_i are normally distributed with mean 0 and variance $D(\theta)$, where θ is a vector of variance components.

The dropout indicator R_i is assumed to be associated with the complete repeated measures data through a logistic model:

$$\text{logit}(P_{ik}) = \alpha_{0k} + H_{i(k-1)}^T \alpha_1 + \alpha_2 Y_{ik}, \quad (2)$$

where $P_{ik} = \Pr(R_i = k | R_i \geq k, H_{i(k)})$, α_{0k} and α_1 are unknown parameters, and $H_{i(k)}$ denotes the history data and is some subset of the covariates and response up to time point k , i.e., some subset of $\{S_{i(k)}, T_{i(k)}, Y_{i(k)}\}$, where (k) denotes the history up to time point k . Because the data contain little information about the dropout parameter α_2 , estimation of α_2 would be heavily driven by modeling assumptions instead of by the data (Little, 1995). A sensitivity analysis by fixing α_2 at different values is often recommended (Rotnitzky et al., 1998; Verbeke et al., 2001). Model (2) can be easily extended to allow the dropout probability P_{ik} to depend on (S_{ik}, T_{ik}) . For more details, see discussion in Section 7.

The above standard selection model requires covariates to be completely observed. However, time-varying covariates are common in longitudinal studies. They are often missing at the time of dropout in addition to the outcome variable Y_{ik} . As a result, the conventional assumption of completely observed covariates is often violated. Likelihood-based inference therefore requires the specification of a model for these missing covariates.

For simplicity, we here assume the p_2 time-varying covariates T_{ik} are fully observed at the baseline. We propose a transition model to model missing time-varying covariates T_{ik} . We assume the l th time-varying covariate follows a first-order transitional model ($l = 1, \dots, p_2; k = 2, \dots, K$)

$$h(\nu_{ikl}) = \lambda_{0l} + \lambda_{1l}T_{i,k-1,l} + \lambda_{2l}^T S_{i(k)}, \quad (3)$$

where $\nu_{ikl} = E(T_{ikl})$, $h(\cdot)$ is a known link function, and $\lambda_l = (\lambda_{0l}, \lambda_{1l}, \lambda_{2l}^T)^T$ is an unknown parameter vector. As $[T_{ikl} | T_{i(k-1)l}, S_{i(k)}]$ follows a GLM, there may be an associated scale parameter ϕ_{Tl} . This model assumes that the l th time-varying covariate at the current time depends on its value at the previous time point and the history of the covariates $S_{i(k)}$. In addition, we assume the p_2 time-varying covariates are independent of each other conditional on their history and the history of the complete covariates, i.e., T_{ikl} is independent of $T_{ikl'}$ conditional on the history $T_{i(k-1)}$ and $S_{i(k)}$. However, we do not make any assumptions about the joint distribution of the p_2 covariates at baseline T_{i1} . It follows that even under the conditional independence assumption, the p_2 covariates T_i are still allowed to be correlated marginally at each time point.

The assumptions of model (3) could be relaxed by: (i) allowing the current value of each time-varying covariate T_{ikl} to depend on the other time-varying covariates $T_{ikl'}$ ($l' \neq l$); or (ii) by allowing the time-varying covariates to be both continuous and discrete. For more details, see discussion in Section 7.

Let $Y_{\text{obs},i} = (Y_{i1}, \dots, Y_{iK_i})^T$, $T_{\text{obs},i} = (T_{i1}, \dots, T_{iK_i})^T$ denote the observed values of Y_i and T_i , with $S_{\text{obs},i}$ and $Z_{\text{obs},i}$ defined similarly. Further, for units that dropped out from the study, let $Y_{\text{mis},i} = Y_{i,K_i+1}$ and $T_{\text{mis},i} = T_{i,K_i+1}$ denote the missing observations. Then, define $Y_i = Y_{\text{obs},i}$ if $K_i = K$ (i.e., unit i did not drop out) and $Y_i = (Y_{\text{obs},i}^T, Y_{\text{mis},i})^T$ otherwise (i.e., unit i drops out at time point $K_i + 1$). Define T_i similarly. Let $S_i = (S_{i1}, \dots, S_{i,K_i+1})^T$ and Z_i defined similarly. If $\alpha_2 = 0$, then equation (2) does not depend on missing data, and the dropout mechanism is ignorable. Inference could then be based on the integrated quasilielihood $L(Y_{\text{obs},i} | T_{\text{obs},i}, S_{\text{obs},i}, Z_{\text{obs},i}; \beta, \theta)$, as in Breslow and Clayton (1993).

When $\alpha_2 \neq 0$, the missing data mechanism is nonignorable, and inference needs to be based on the joint integrated quasilielihood

$$\begin{aligned} & L(Y_{\text{obs},i}, T_{\text{obs},i}, R_i | S_i, T_{i1}; \Omega) \\ &= \int L(Y_i, T_i, R_i | S_i, T_{i1}; \Omega) db_i dT_{\text{mis},i} dY_{\text{mis},i} \\ &= \int L(Y_i | S_i, T_i, b_i; \beta, \phi_y) L(T_i | T_{i1}, S_i; \lambda, \phi_T) L(b_i; \theta) \\ & \quad \times L(R_i | Y_i, T_{\text{obs},i}, S_{\text{obs},i}; \alpha) db_i dT_{\text{mis},i} dY_{\text{mis},i}, \quad (4) \end{aligned}$$

where $\lambda = (\lambda_1^T, \dots, \lambda_{p_2}^T)^T$, $\alpha = (\alpha_0^T, \alpha_1^T)^T$, and Ω is a vector containing all the parameters in the model. The dimension of integration in the likelihood (4) is $q + 1 + p_2$ for units that dropped out and q for those that did not. Denote the log likelihood $\ell(\cdot) = \ln L(\cdot)$. The log likelihoods in the integrand in (4) are

$$\begin{aligned} \ell(Y_i | S_i, T_i, b_i; \beta, \phi_y) &= - \sum_{k=1}^{(K_i+1) \wedge K} d(Y_{ik}, \mu_{ik}), \\ \ell(T_i | T_{i1}; \lambda, \phi_T) &= - \sum_{l=1}^{p_2} \sum_{k=2}^{(K_i+1) \wedge K} d(T_{ikl}, \nu_{ikl}), \\ \ell(b_i | \theta) &= - \frac{1}{2} \ln |D(\theta)| - \frac{1}{2} b_i^T D(\theta)^{-1} b_i, \\ \ell(R_i | Y_i, T_{\text{obs},i}, S_{\text{obs},i}; \alpha) &= \sum_{k=2}^{K_i} \ln(1 - P_{ik}) \\ & \quad + I[K_i < K] \ln(P_{i,K_i+1}), \end{aligned}$$

where $d(y, \mu) = -2 \int_y^\mu \{\phi a_{ik}^{-1} v(u)\}^{-1} (y - u) du$ is the deviance function, P_{ik} is defined in (2), and $I[K_i < K]$ is an indicator function.

3. Asymptotic Bias Analysis

In this section, we investigate the asymptotic bias of the model parameters if missing time-varying covariates are handled using naive approaches. Three naive approaches are considered: (1) using the baseline measures throughout by assuming the time-varying covariates have not changed since baseline; (2) carrying forward the last observation, i.e., assuming their values at the time of dropout are the same as the previous values; and (3) ignoring missing data completely by assuming MAR. These naive approaches enable one to fit models using the existing approaches that assume covariates are completely observed. For example, Diggle and Kenward's (1994) method could be applied to the data filled in using the baseline or LOCF approaches; a standard GLMM could be applied if we assume ignorable missingness (IM). The question of interest is how much asymptotic bias would arise by doing so.

3.1 The Specific Model Considered in the Bias Analysis

To demonstrate the fundamental impact of the naive approaches, we consider a simple case in the asymptotic bias analysis. Specifically, we assume there is a single time-varying covariate and the outcomes are Gaussian. The outcomes Y_{ik} are assumed to follow a linear random intercept model

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + b_i + \epsilon_{ik}, \quad k = 1, \dots, K, \quad (5)$$

where b_i and ϵ_{ik} are independently distributed as $N(0, \theta)$ and $N(0, \tau^2)$, respectively. The time-varying covariate is assumed to follow a transition model

$$X_{ik} = \lambda_0 + \lambda_1 X_{i,k-1} + e_{ik}, \quad (6)$$

where $X_{i1} \sim N(\mu_1, \delta)$ and $e_{ik} \sim N(0, \sigma^2)$ for $k = 2, \dots, K$. Finally, the dropout model is

$$\text{logit}(P_{ik}) = \alpha_{0k} + \alpha_1 Y_{i,k-1} + \alpha_2 Y_{ik}, \quad (7)$$

where P_{ik} is defined in Section 2.

The naive model assumes

$$Y_{ik} = \beta_{0,\text{naive}} + \beta_{1,\text{naive}} X_{ik}^* + b_i + \epsilon_{ik}, \quad (8)$$

where $b_i \sim N(0, \theta_{\text{naive}})$ and $\epsilon_{ik} \sim N(0, \tau_{\text{naive}}^2)$, and the covariate X_{ik} is made complete by setting it equal to X_{ik}^* using one of

the three naive approaches. We further assume that when using either the baseline or LOCF approach, the dropout mechanism (7) is correctly specified with α being known. Therefore, the naive model only differs from model (5) in the way missing covariates are handled. Note that the true model accounts for missing X_{ik} by assuming the transition model (6).

Let $\xi = (\theta, \tau^2)^T$. Denote by $\gamma = \{\beta^T, \xi^T\}^T$ the true values and $\gamma_{\text{naive}} = (\beta_{\text{naive}}^T, \xi_{\text{naive}}^T)^T$ their asymptotic limits as $n \rightarrow \infty$ using one of the naive approaches. Denote by $\ell_{\text{naive}}(Y_i, R_i)$ the log likelihood of γ for unit i under each of the three naive approaches. The naive score function is $U_{\text{naive}}(Y, R) = n^{-1} \sum_{i=1}^n \partial \ell_{\text{naive}}(Y_i, R_i) / \partial \gamma = n^{-1} \sum_{i=1}^n U_{\text{naive}}(Y_i, R_i)$. It follows that the asymptotic limits of the naive MLEs γ_{naive} solve

$$E\{U_{\text{naive}}(Y_i, R_i | \beta_{\text{naive}}, \xi_{\text{naive}})\} = 0, \quad (9)$$

where the expectation is taken under the true models (5)–(7). Hence, the solution $\{\beta_{\text{naive}}, \xi_{\text{naive}}\}$ is a function of the true parameters $\{\beta, \xi\}$. Equation (9) generally does not have a closed-form solution and needs to be solved numerically. Details about these calculations can be obtained from the authors upon request.

3.2 The Results

We now consider numerical calculations of bias for the simple case where $K = 2$. Equation (9) does not have a closed-form solution. The Gauss–Hermite quadrature was used to evaluate the integrals and the Newton–Raphson method was used to solve the equations.

We assumed in our numerical calculations the true values of the parameters were $\beta_0 = 1, \beta_1 = 1, \theta = 1, \tau^2 = 1, \mu_1 = 1, \delta = 1, \sigma^2 = 1, \alpha_0 = -10$, and $\alpha_1 = 0$. We then made different assumptions about the relationship between X at time points 1 and 2 by varying $(\lambda_0, \lambda_1) = (0.5, 0.5), (1.0, 0.0), (0.0, 1.0)$. In each case, we calculated the asymptotic biases by increasing the values of α_2 . This allowed us to investigate how the asymptotic biases changed as the marginal probability of dropout increased. Note that the probability of dropout is essentially 0 when $\alpha_2 \leq 0$. Therefore, in the plots that follow, all dropouts can be thought of as arising from a nonignorable mechanism.

Figure 1 shows the relative bias plots for the baseline, LOCF, and IM approaches for $(\lambda_0, \lambda_1) = (0.5, 0.5)$, which assumes the value of X at the first and second time points are positively correlated. When the marginal probability of dropout is 0, the LOCF and IM approaches yield unbiased parameters. This is because all of the data are used. However, the baseline approach still yields biased estimates even when there are no dropouts. As the marginal probability of dropout goes to 1, the amount of bias from the baseline and LOCF approaches converge to each other (curves not shown), since the methods are equivalent when all units drop out after time point 1. The LOCF approach yields parameters with a smaller bias than the baseline approach. The IM approach results in similar biases in the estimates of β_1 and τ to those using the LOCF approach, but a more biased estimate of the variance component θ . All three naive approaches give attenuated estimates of β_1 . Both the baseline and IM approaches underestimate the variance components θ while the bias of the LOCF estimate of θ is small. Both the baseline and LOCF approaches overestimate τ while the IM estimate of τ has little bias.

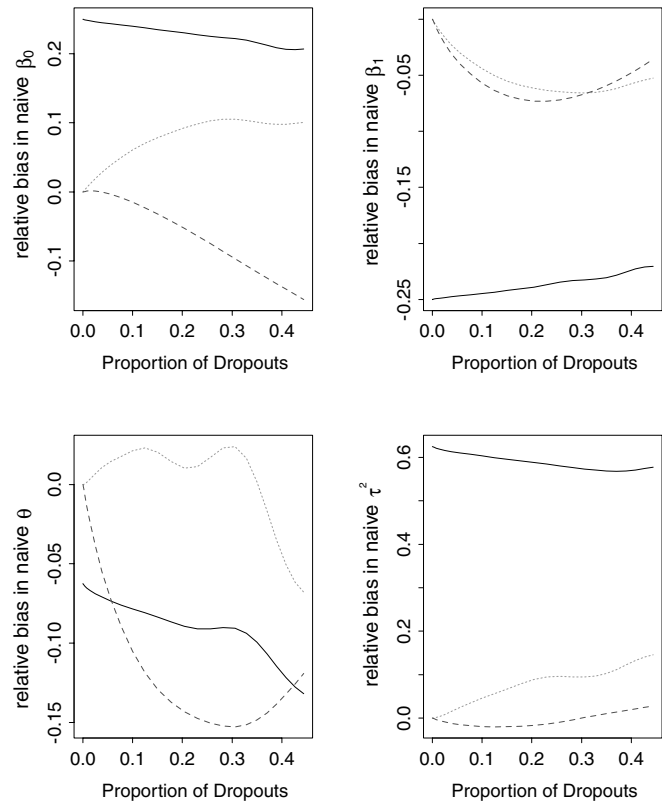


Figure 1. Asymptotic relative biases in the naive parameters of β, θ , and τ^2 when the transition model for the time-varying covariate holds. The true parameter values are $\beta_0 = 1, \beta_1 = 1, \theta = 1, \tau^2 = 1, \lambda_0 = 0.5, \lambda_1 = 0.5, \mu_1 = 1, \delta = 1, \sigma^2 = 1, \alpha_0 = -10$, and $\alpha_1 = 0$. The three curves in each plot are — baseline; \cdots LOCF; --- IM.

We also calculated bias curves under other scenarios. When we assumed that the values of X at time points 1 and 2 are uncorrelated (e.g., $\lambda_0 = 1, \lambda_1 = 0$), the shape of the curves looked similar to Figure 1, but the magnitude of the biases was larger. In the situation when X_2 is equal to X_1 plus random noise (i.e., $\lambda_0 = 0$ and $\lambda_1 = 1$), other than for τ^2 , the magnitude of the biases is smaller than in Figure 1. For τ^2 the bias plot is nearly identical to that of Figure 1.

Our asymptotic bias analysis shows that the three naive approaches to handle missing time-varying covariates often yield biased estimates in the model parameters. The amount of bias depends on the values of the true parameters of the model. The naive estimates in the regression parameter of main interest β_1 are attenuated and the bias in β_1 increases as the correlation between X at time points 1 and 2 decreases. Our results hence suggest that appropriate analysis needs to account for missing time-varying covariates in the presence of informative dropouts. Such an estimation procedure is proposed in the next section.

4. Estimation and Inference

In view of the asymptotic bias results in Section 3, we model missing time-varying covariates for longitudinal data with informative dropouts using the selection–transition

model (1)–(3). We develop in this section an estimation procedure using the EM algorithm. Specifically, statistical inference for model (1)–(3) is challenged by the possibly high-dimensional integrals in the joint quasilielihood (4). For a Gaussian outcome Y_i with an identity link function, i.e., Y_i follows a linear mixed model, the quasilielihood (4) can be simplified and only involves one-dimensional integral. However, for non-Gaussian data the dimension of integration of the joint quasilielihood (4) is $q + 1 + p_2$, which may be quite large. Estimation by directly maximizing (4) is computationally difficult. We hence develop an EM algorithm approach for estimation.

The complete data for subject i are $\{Y_i, R_i, T_i, S_i, b_i\}$ and the observed data are $\{Y_{\text{obs},i}, R_i, T_{\text{obs},i}, S_i\}$. The complete data log likelihood is $\ell_c(\Omega) = \sum_{i=1}^n \{\ell(Y_i | S_i, T_i, b_i; \beta, \phi) + \ell(T_i | T_{i1}, S_i; \lambda, \sigma) + \ell(b_i; \theta) + \ell(R_i | Y_i, T_{\text{obs},i}, S_{\text{obs},i}; \alpha)\}$, where each of the terms was defined in Section 2.

The M-step is as follows. Let $\Omega^{(k)}$ denote the estimate of Ω at the k th iteration. The updated estimates of β are obtained by solving

$$\sum_{i=1}^n \sum_{k=1}^{(K_i+1) \wedge K} \text{E}\{X_{ik} \mu'_{ik} V_{ik}^{-1} (Y_{ik} - \mu_{ik}) | Y_{\text{obs},i}, T_{\text{obs},i}, S_i, R_i; \Omega^{(k)}\} = 0,$$

$$\text{E}\{h(b_i, Y_{\text{mis},i}, T_{\text{mis},i}) | Y_{\text{obs},i}, T_{\text{obs},i}, R_i\} = \frac{\int h(b_i, Y_{\text{mis},i}, T_{\text{mis},i}) L(Y_i | T_i, S_i) L(R_i | Y_i, T_i, S_i) L(T_i) L(b_i) dY_{\text{mis},i} dT_{\text{mis},i} db_i}{\int L(Y_i | T_i, S_i) L(R_i | Y_i, T_i, S_i) L(T_i) L(b_i) dY_{\text{mis},i} dT_{\text{mis},i} db_i},$$

where $\mu'_{ik} = 1/g'(\mu_{ik})$ and $V_{ik} = \phi_y a_{ik}^{-1} v(\mu_{ik})$. In some cases this will have closed form (e.g., linear mixed models for normal data), but in other cases (e.g., GLMMs for nonnormal data) the equation will have to be solved iteratively using the Fisher scoring algorithm. The updated estimate of ϕ_y is obtained similarly (i.e., using standard GLM maximization techniques). The parameters λ and ϕ_T are updated in an analogous manner, since $[T_{ikl} | T_{i(k)l}]$ is assumed to follow a GLM. If $D(\theta)$ is unstructured, then it is updated using $D^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \text{E}(b_i b_i^T | Y_{\text{obs},i}, T_{\text{obs},i}, S_i, R_i; \Omega^{(k)})$. The parameters in the dropout model $\alpha = (\alpha_{02}, \dots, \alpha_{0K}, \alpha_1^T)^T$ are updated iteratively using the following Newton–Raphson algorithm

$$\alpha^{(k+1)} = \alpha^{(k)} - \left(\frac{\partial^2 \ell_R}{\partial \alpha \partial \alpha^T} \right)^{-1} \frac{\partial \ell_R}{\partial \alpha} \Big|_{\alpha^{(k)}},$$

where

$$\begin{aligned} \frac{\partial \ell_R}{\partial \alpha} = \sum_{i=1}^n \left[\left\{ \sum_{k=2}^{K_i} -\text{E}(P_{ik} \Delta_{ik} | Y_{\text{obs},i}, T_{\text{obs},i}, S_i, R_i) \right\} \right. \\ \left. + \text{E}\{(1 - P_{i,K_i+1}) \Delta_{i,K_i+1} | Y_{\text{obs},i}, T_{\text{obs},i}, S_i, R_i\} \right. \\ \left. \times I[K_i < K] \right], \end{aligned}$$

$$\frac{\partial^2 \ell_R}{\partial \alpha \partial \alpha^T} = \sum_{i=1}^n \sum_{k=2}^{(K_i+1) \wedge K} -\text{E}[P_{ik} (1 - P_{ik}) \Delta_{ik} \Delta_{ik}^T | Y_{\text{obs},i}, T_{\text{obs},i}, S_i, R_i],$$

and $\Delta_{ik} = (\delta_k, H_{i(k-1)}^T)^T$ and δ_k is a $(K - 1) \times 1$ vector with all elements equal to 0 except that the k th element equal to 1.

The E-step involves taking expectations of functions of the “missing data” conditional on the observed data and take the form $\text{E}\{h(b_i, Y_{\text{mis},i}, T_{\text{mis},i}) | Y_{\text{obs},i}, T_{\text{obs},i}, S_i, R_i\}$, where h is some function of the missing data (e.g., $h(b_i, Y_{\text{mis},i}, T_{\text{mis},i}) = b_i b_i^T$). For subjects that did not drop out, the expectation is as follows:

$$\begin{aligned} \text{E}\{h(b_i) | Y_i, T_i, R_i\} \\ = \frac{\int h(b_i) L(Y_i | T_i, S_i) L(R_i | Y_i, T_i, S_i) L(T_i) L(b_i) db_i}{\int L(Y_i | T_i, S_i) L(R_i | Y_i, T_i, S_i) L(T_i) L(b_i) db_i}, \end{aligned}$$

where the integrals are q -dimensional. Because we assume the random effects have a normal distribution, the Gaussian quadrature can be used to evaluate the integral. For subjects that did drop out of the study, the expectations take the form

where the integrals have higher dimensions and are $q + 1 + p_2$ -dimensional. In some special cases the integral can be evaluated directly. For example, if the response is binary integration over missing Y is just a two-part summation. If the response is normal, Gaussian quadrature can be used. Similarly, components of T that are binary or normal could be integrated over directly, if the dimension of T is relatively small. Otherwise a Monte Carlo approximation (Wei and Tanner, 1990; Booth and Hobert, 1999) is used for all or part of the integrals. Following Ibrahim et al. (2001), we first repeatedly draw values from the conditional distribution of $[b_i, Y_{\text{mis},i}, T_{\text{mis},i} | Y_{\text{obs},i}, T_{\text{obs},i}, R_i]$ using the Gibbs sampler along with rejection sampling. We then evaluate the integrands in the numerator and denominator for each set of draws, and average over the draws. At the t th iteration, we draw M_t values from the conditional distribution. It is often recommended to select a small value for M_t at early iterations, then increase it as the parameters get closer to their ML estimates (Booth and Hobert, 1999). For the final iterations a large M_t should be chosen to minimize Monte Carlo error. Convergence of EM is declared when the maximum relative change in the parameter estimates in consecutive iterations is smaller in magnitude than some predetermined value (e.g., $< 10^{-4}$). The method of Louis (1982) is used for obtaining estimates of the standard errors.

5. Simulations

We carry out two sets of simulations studies—one for continuous data and another for binary data. The purpose is twofold: (1) to evaluate the finite sample performance of the EM algorithm for the proposed model; and (2) to further assess the bias that results from implementing naive methods for filling in missing covariate data. In the previous section, we considered asymptotic bias for data with a continuous response, continuous time-varying covariate and only two time points. The simulation study will provide insight into the bias that arises in finite samples, with more than two time points for both continuous and discrete response and time-varying covariates.

For both sets of simulations, we generated complete data from $n = 100$ subjects with $K = 3$ observations each. The vector S_{ik} consisted of an intercept and time itself. One time-varying covariate T_{ik} was generated. Data were missing after time point k with probability 1 if data were missing at time point $k - 1$, and with probability P_{ik} otherwise, where

$$\text{logit}(P_{ik}) = \alpha_{0k} + \alpha_1 Y_{i,k-1} + \alpha_2 Y_{ik}$$

for $k = 2, 3$.

Each simulated data set was analyzed using four methods—MLE based on the correct model and three naive models. Two of the naive methods filled in missing T_{ik} (using baseline values or carrying forward the last observation), and then fitted a selection model using the filled-in covariate data. The other naive method assumed the missing data were ignorable, and a random intercept model for binary data was fitted to the observed data.

5.1 Continuous Response

For this simulation study, we assumed both a continuous response and continuous time-varying covariate. First, we generated T_{i1} from a standard normal distribution. We then generated $T_{ik} = \lambda_0 + \lambda_1 T_{i,k-1} + e_{i,k}$ for $k = 2, 3$, where $e_{i,k} \sim N(0, 1)$. The response was then $Y_{ik} = \beta_0 + \beta_s k + \beta_T T_{ik} +$

ε_{ik} , where b_i and ε_i were independently distributed as $N(0, 1)$.

We set $\beta_0 = 1, \beta_s = 0.5, \beta_T = 1, \alpha_{02} = \alpha_{03} = 0.5$, and $\alpha_1 = 0$. We simulated data under various values for α_2 and λ in order to change the amount of missing data, the dependence of missingness on Y , and the degree to which T_{ik} depends on the past. Percentage bias and mean square error (MSE) are given in Table 1. First, note that the proposed method (which is based on a correctly specified model) yielded nearly unbiased estimates of β_s and β_T (<2% bias in all cases). The baseline approach performed poorly in all cases except for the parameter β_T when $\lambda_0 = 0$ and $\lambda_1 = 1$. For those values of λ , future values of T were strongly predicted by the past, and therefore imputing using baseline values was not too costly, although the method still performed poorly for estimating β_s . The LOCF approach outperformed the baseline approach in every instance. The amount of bias was 5% or less when $\alpha_2 = 0.5$, and for $\alpha_2 = 1$ when $\lambda_1 = 1$ (serial observations of T highly correlated). Not surprisingly, the baseline and LOCF methods performed the worst when there was no correlation in the repeated measurements of T . The IM approach led to bias of less than 10% for β_T in each scenario. However, the method performed very poorly for estimating the slope β_s , with bias ranging from 13% to 51%. In terms of MSE, LOCF was competitive with the proposed approach when $\alpha_2 = 0.5$, but when dropout more heavily depended on missing values (i.e., $\alpha_2 = 1$), the proposed approach tended to have the smaller MSE.

5.2 Binary Response

We next simulated data with a binary response and binary time-varying covariate. We generated T_{i1} from a Bernoulli distribution with success probability equal to 0.5. We then generated T_{ik} from a Bernoulli distribution with success probability $\text{logit}^{-1}(\lambda_0 + \lambda_1 T_{i,k-1})$ for $k = 2, 3$. Finally, the response Y_{ik} was equal to 1 with probability $\text{logit}^{-1}(\beta_0 + \beta_s k + \beta_T T_{ik} + b_i)$ for $k = 1, 2, 3$, where $b_i \sim N(0, 1)$.

Table 1

Results from simulation study with continuous response. Percentage bias and MSE in estimation of β_s (coefficient of time) and β_T (coefficient of time-varying covariate) from the proposed method, along with the three naive approaches. Results are based on 1000 replications each.

α_2	(λ_0, λ_1)	Statistic	Method							
			Baseline		LOCF		IM		Proposed	
			β_s	β_T	β_s	β_T	β_s	β_T	β_s	β_T
0.5	(0.5,0.5)	Bias	72.8	-30.9	1.6	-2.9	-15.3	-1.7	-0.4	0.0
		MSE	0.145	0.117	0.006	0.006	0.018	0.006	0.008	0.007
	(0,1)	Bias	-12.2	4.2	-0.1	0.1	-12.8	-1.5	-1.0	0.1
		MSE	0.016	0.021	0.005	0.006	0.013	0.005	0.006	0.005
	(1,0)	Bias	104.2	-51.7	5.1	-5.0	-15.9	-2.4	0.2	-0.6
		MSE	0.287	0.284	0.008	0.007	0.017	0.005	0.009	0.006
1.0	(0.5,0.5)	Bias	60.4	-24.0	8.8	-10.9	-46.6	-5.5	0.1	-0.5
		MSE	0.114	0.079	0.015	0.021	0.066	0.010	0.013	0.010
	(0,1)	Bias	-42.0	3.0	-1.9	-0.1	-25.4	-0.2	-0.8	-0.4
		MSE	0.058	0.023	0.008	0.005	0.026	0.006	0.009	0.007
	(1,0)	Bias	108.3	-40.5	30.6	-20.9	-51.4	-8.6	-1.4	-0.5
		MSE	0.309	0.189	0.041	0.053	0.080	0.015	0.017	0.009

Table 2

Results from simulation study with binary response. Percentage bias and MSE in estimation of β_s (coefficient of time) and β_T (coefficient of time-varying covariate) from the proposed method, along with the three naive approaches. Results are based on 1000 replications each.

α_2	(λ_0, λ_1)	Statistic	Method							
			Baseline		LOCF		IM		Proposed	
			β_s	β_T	β_s	β_T	β_s	β_T	β_s	β_T
0.5	(-0.5,1)	Bias	2.4	-43.7	3.8	4.5	-9.7	-42.4	1.6	1.3
		MSE	0.037	0.358	0.059	0.130	0.061	0.352	0.046	0.168
	(-1.5,3)	Bias	4.7	-16.2	3.2	3.0	-8.0	-21.3	0.9	1.6
		MSE	0.053	0.166	0.050	0.128	0.044	0.206	0.056	0.170
	(0,0)	Bias	6.5	-40.1	6.5	4.5	-11.4	-51.8	1.1	1.2
		MSE	0.058	0.299	0.047	0.135	0.051	0.401	0.049	0.159
1.0	(-0.5,1)	Bias	8.2	-32.0	4.2	4.1	-24.7	-80.2	1.0	1.5
		MSE	0.061	0.240	0.051	0.169	0.067	0.784	0.063	0.179
	(-1.5,3)	Bias	10.3	-13.2	4.3	3.6	-21.2	-19.7	1.2	2.2
		MSE	0.039	0.218	0.046	0.194	0.058	0.210	0.058	0.204
	(0,0)	Bias	8.1	-37.9	7.2	8.9	-26.8	-46.7	0.6	1.6
		MSE	0.052	0.308	0.051	0.176	0.071	0.362	0.045	0.154

In each set of simulations, we set $\beta_0 = -1$, $\beta_S = -0.5$, $\beta_T = 1$, $\alpha_{02} = \alpha_{03} = -2$, and $\alpha_1 = 0$. We varied α_2 as 0.5 or 1 in order to affect the probability of dropout and the dependence of dropout on the current value of the response. We also considered several values of λ in order to assess how the methods performed as serial correlation in T varied. Table 2 displays the results from the simulations. The proposed method generally yielded parameters with little bias (<3%). The bias that is present is a finite-sample bias, as it does substantially decrease if we instead generated data with $n = 500$ (results not shown). Assuming the missing data are MAR (third column) results in a substantial amount of bias for both β_S and β_T . Parameters from the baseline approach are nearly as biased as those assuming MAR. As expected, the percentage bias with the baseline approach is smallest when serial correlation in T is large (i.e., when $\lambda = (-1.5, 3)$). Surprisingly, the LOCF approach performed fairly well in these simulations, with <10% bias for β_S and β_T in all cases. This is different than what we found in the continuous case, where LOCF performed poorly when there was no serial correlation in T . The MSEs in most cases were smallest for the LOCF approach, although the MSEs from the full model were not much larger. The MSEs corresponding to β_T were much larger for the baseline and IM approaches, than for LOCF or the proposed model.

6. Application

We next analyzed a subset of the HERS data, which were briefly described in Section 1. We were interested in whether or not the use of PIs affected the risk of hospitalization among women in the HERS data who had acquired immune deficiency syndrome (AIDS) at baseline (defined here as $CD4 < 200$ cells/mL). This subset of women is of particular interest, because Tashima et al. (2001) showed a beneficial effect of PIs for this group only. In addition, subjects with $CD4$ cell counts < 200 have the most advanced disease, and are prime candidates for the therapy.

At baseline, 126 women from the HERS had $CD4$ cell count < 200 cells/mL. These women were followed up every 6 months for up to a total of 12 visits. The response variable of interest is whether or not each subject was hospitalized since the previous visit (about 6 months). PI use at each visit was recorded, which was defined as having taken PIs at any time during the interval. In addition, we have data on a wide variety of variables, including age, race, history of drug use, baseline HIV-1 RNA level, and study site. Table 3 presents the proportion of hospitalizations and PI use at each visit, along with the number of subjects still remaining in the study. For example, 48% of the women had a hospitalization between the baseline and second visit. Some things to note. First, several women dropped out at each visit, with only 21 of 126 women with data at every visit. Also, note that PIs were not available

Table 3

Proportion of patients with a hospitalization and proportion of protease inhibitor (PI) use at each interval

Characteristic	Interval between HERS visit numbers											
	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13
Subjects	126	115	109	95	78	69	67	61	55	44	33	21
Hospitalization	0.48	0.43	0.47	0.39	0.38	0.29	0.22	0.18	0.13	0.20	0.18	0.19
PI use	0	0	0.01	0.02	0.14	0.26	0.39	0.59	0.58	0.59	0.64	0.67

at the start of the study; virtually no one used PIs prior to visit 5. Once PIs were introduced as a method of treatment, PI use in this population increased steadily over time. The proportion of hospitalization declines over time. Note that over time three things are happening: we were losing subjects to follow-up, the observed proportion of hospitalizations was going down, and PI use was going up. A naive analysis ignoring dropouts would lead one to conclude that PI use reduced the probability of hospitalization. However, it is not hard to imagine that less healthy patients may be more likely to drop out from the study. Therefore, that could be the cause of the observed decline in the proportion of hospitalizations. In particular, the probability of dropout at visit k may depend on whether or not the subject was hospitalized between visits $k - 1$ and k , which would be unobserved for subjects who dropped out at that time. In addition, the subject's PI status would also be unobserved at that time. We therefore use the methods described in this article to account for possible nonignorable missing data, along with a missing time-varying covariate.

For the analysis, we defined the response Y_{ik} to take value of 1 if subject i was hospitalized between visits $k - 1$ and k , and take value of 0 otherwise. The covariates S_{ik} included: a vector of ones (intercept); age at baseline (years); an indicator of whether or not the subject had a history of injection drug use; baseline HIV-1 RNA level (copies/mL), which was categorized as <500 (referent), 500–5000, 5000–30,000, and >30,000; race, which was categorized as African American, Hispanic or “other,” and white (referent); indicators for study

site; and time itself (entered into the model as the visit number). There was one time-varying covariate that was subject to dropout-related missingness, and that was PI use. Therefore, T_{ik} was an indicator for PI use between visits $k - 1$ and k . We assumed that Y_i given covariates and random intercept b_i followed a Bernoulli distribution, with mean modeled by (1), where $g(\cdot)$ was the logit link. Next we modeled the probability of dropout using (2). We included indicators of HIV-1 RNA level in the model, as these were important predictors of dropout. Finally, we modeled PI use using the transition model (3), where $h(\cdot)$ was the logit link. After conditioning on past PI use, other covariates were not significant predictors of PI use, and therefore we only included past PI use in the transition model.

The data were analyzed assuming several plausible values of α_2 (the parameter linking the current value of Y to the probability of dropout), ranging from -1 to 1 . When α_2 was positive, this implied that dropout is more likely if the subject had a recent hospitalization. A recent hospitalization was unrelated to dropout if $\alpha_2 = 0$ (the MAR case). If a hospitalization decreased the likelihood of dropout, then $\alpha_2 < 0$. We believe α_2 was likely to be positive, but considered a range of values that included negative numbers. The results from the analyses are given in Tables 4 and 5. For each value of α_2 ($-1, -0.5, 0, 0.5, 1$), there was a significant decline in the risk of hospitalization for PI uses, after controlling for the other covariates. The PI effect was largest (in magnitude; odds ratio = 0.38), assuming $\alpha_2 = 1$, i.e., assuming a recent hospitalization greatly increased the chances of dropping out. However, the

Table 4
Parameter estimates and estimated standard errors of the parameters of primary interest from a sensitivity analysis of the hospitalization data

Variable	$\alpha_2 = -1$	$\alpha_2 = -0.5$	$\alpha_2 = 0$	$\alpha_2 = 0.5$	$\alpha_2 = 1$
Intercept	0.66 (0.48)	0.60 (0.49)	0.51 (0.51)	0.40 (0.51)	0.29 (0.51)
Visit	-0.16 (0.03)	-0.14 (0.03)	-0.13 (0.03)	-0.11 (0.04)	-0.09 (0.04)
Drug use	0.33 (0.23)	0.33 (0.24)	0.32 (0.24)	0.31 (0.25)	0.30 (0.25)
Age	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
African American	0.51 (0.35)	0.54 (0.36)	0.57 (0.37)	0.59 (0.37)	0.61 (0.38)
Hispanic	0.71 (0.38)	0.75 (0.39)	0.79 (0.40)	0.81 (0.40)	0.82 (0.40)
HIV RNA > 30,000	-1.18 (0.46)	-1.15 (0.48)	-1.11 (0.49)	-1.06 (0.49)	-0.99 (0.50)
HIV RNA 5000–30,000	-0.77 (0.47)	-0.72 (0.48)	-0.66 (0.49)	-0.58 (0.50)	-0.50 (0.50)
HIV RNA 500–5000	-0.51 (0.46)	-0.43 (0.47)	-0.34 (0.48)	-0.23 (0.49)	-0.12 (0.50)
Site 1	-0.54 (0.33)	-0.55 (0.33)	-0.56 (0.34)	-0.55 (0.35)	-0.55 (0.35)
Site 2	-0.10 (0.37)	-0.10 (0.38)	-0.10 (0.39)	-0.10 (0.39)	-0.09 (0.40)
Site 3	-0.56 (0.39)	-0.55 (0.40)	-0.58 (0.41)	-0.58 (0.42)	-0.58 (0.42)
PI use	-0.76 (0.29)	-0.79 (0.29)	-0.84 (0.30)	-0.89 (0.30)	-0.96 (0.30)

Table 5

Parameter estimates and estimated standard errors of the parameters of secondary interest from a sensitivity analysis of the hospitalization data

Parameter	$\alpha_2 = -1$	$\alpha_2 = -0.5$	$\alpha_2 = 0$	$\alpha_2 = 0.5$	$\alpha_2 = 1$
α_{11} HIV RNA > 30,000	1.32 (0.64)	1.40 (0.64)	1.51 (0.64)	1.64 (0.65)	1.79 (0.66)
α_{12} HIV RNA 5000–30,000	1.62 (0.64)	1.70 (0.64)	1.80 (0.64)	1.90 (0.65)	2.02 (0.66)
α_{13} HIV RNA 500–5000	1.98 (0.65)	2.05 (0.64)	2.13 (0.65)	2.22 (0.65)	2.33 (0.66)
α_{14} Y_{k-1}	0.88 (0.23)	0.81 (0.23)	0.72 (0.23)	0.64 (0.23)	0.53 (0.23)
λ_0	-2.22 (0.14)	-2.22 (0.14)	-2.23 (0.14)	-2.24 (0.14)	-2.25 (0.14)
λ_1	4.05 (0.27)	4.05 (0.27)	4.05 (0.27)	4.06 (0.27)	4.06 (0.27)
θ	0.52 (0.18)	0.55 (0.19)	0.59 (0.20)	0.62 (0.21)	0.65 (0.21)

time effect (coefficient of “visit”) was smallest (in magnitude) when $\alpha_2 = 1$. This is not surprising, because if less healthy patients were dropping out, the observed decline in hospitalizations over time that we saw in Table 3 may be in part due to this fact.

We next compared the results from the proposed full model to that from the naive approaches. The baseline approach was not an option here as all subjects had PI equal to 0 at baseline. The results assuming IM are given in Tables 4 and 5 under the $\alpha_2 = 0$ heading. For the LOCF approach, it was possible to fit the model for each value of α_2 ; those are reported in Table 4. To keep the discussion concise, however, we instead focused on the case where $\alpha_2 = 1$ (which we believe is a plausible value). The estimated effect of PI use in that case was -0.57 with an estimated standard error of 0.30. Thus, the LOCF method estimated a smaller PI effect (in magnitude) than did the proposed method or IM approach, and the effect was only marginally significant. The other parameter estimates were very similar across the three approaches.

7. Discussion

In this article, we considered longitudinal data with informative dropouts and missing time-varying covariates. This problem is common in practice, because at the time of dropout, time-varying covariates are often missing in addition to the missing outcome. The existing statistical models for informative dropouts generally require covariates to be completely observed. To apply these models in the presence of time-varying covariates, one has to make the covariate data complete. The naive approaches to fill in missing covariates include the baseline approach, the LOCF approach, and the IM approach. As an alternative to the naive approach, we propose in this article a selection–transition model to allow for missing time-varying covariates at the time of dropout for longitudinal data with informative dropouts.

Both the bias analysis and simulations demonstrated that when there are nonignorable dropouts, the naive approaches considered here may perform poorly in terms of bias. In finite samples, the naive approaches often have a higher MSE

than do estimators from a correctly specified model proposed in this manuscript. Both the baseline and IM approaches performed poorly in most of the situations we considered. Not surprisingly, the LOCF approach performed reasonably well in situations where within-subject variance of the time-varying covariate is small (i.e., the covariate is relatively stable over time within subjects), but performed poorly if the time-varying covariate fluctuates substantially within a subject.

For simplicity, we assumed the selection model (2) was not a function of missing time-varying covariates $T_{i,k}$ at the time of dropout. It is possible in practice that the dropout probability might depend on the missing time-varying covariates $T_{i,k}$ at the time of dropout in addition to the missing outcome. One can easily extend the selection model (2) to allow for dependence of the dropout probability on the missing covariates $T_{i,k}$. In addition, we assumed that each time-varying covariate T_{ikl} , conditional on past values of that covariate, was independent of $T_{ikl'}$, for $l \neq l'$. This assumption could be weakened by specifying the joint likelihood for these covariates using a factorization similar to that advocated by Ibrahim et al. (1999) and Stubbendick and Ibrahim (2003).

The proposed methodology does require extensive modeling assumptions, including specification of the dropout mechanism and partial specification of the covariate distributions. An area in need of further research is sensitivity of the results to a misspecified dropout and/or covariate distribution. Clearly, there are trade-offs between protection against bias and increased variance by fitting flexible models with large numbers of parameters for the covariate and dropout distributions. A possible direction for future research is to develop semiparametric models for the covariate distributions. An alternative, using the methods proposed here, would be to carry out a sensitivity analysis, where the data are reanalyzed under several different models for the covariates. The purpose of such an analysis would be to determine whether inference about the parameters of primary interest is affected by these modeling assumptions.

ACKNOWLEDGEMENTS

We thank the editor, associate editor, and referees for their insightful comments, which led to an improved version of the manuscript. This research was supported in part by NIH grants R01-CA76404, R01-AI50505, and P30-AI42853 (Lifespan/Tufts/Brown Center for AIDS Research). Data from the HERS were collected under CDCP grant U64-CCU10675.

REFERENCES

- Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed models likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 265–285.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.
- Diggle, P. and Kenward, M. G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–94.
- Ibrahim, J. G., Lipsitz, S. R., and Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society, Series B* **61**, 173–190.
- Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551–564.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.
- Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226–233.
- Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339.
- Roy, J. and Lin, X. (2002). Analysis of multivariate longitudinal outcomes with non-ignorable dropouts and missing covariates: Changes in methadone treatment practices. *Journal of the American Statistical Association* **97**, 40–52.
- Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussions). *Journal of the American Statistical Association* **94**, 1096–1120.
- Smith, D. K., Warren, D. L., Vlahov, D., Schuman, P., Stein, M. D., Greenberg, B. L., and Holmberg, S. D. (1997). Design and baseline participant characteristics of human immunodeficiency virus epidemiology research (HER) study: A prospective cohort study of human immunodeficiency virus infection in US women. *American Journal of Epidemiology* **146**, 459–469.
- Stubbendick, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* **59**, 1140–1150.
- Tashima, K. T., Hogan, J. W., Gardner, L. I., Korkontzelou, C., Schoenbaum, E. E., Schuman, P., Rompalo, A., and Carpenter, C. C. J. (2001). A longitudinal analysis of hospitalization and emergency department use among human immunodeficiency virus-infected women reporting protease inhibitor use. *Clinical Infectious Diseases* **33**, 2055–2060.
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics* **57**, 7–14.
- Wei, G. C. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704.

Received May 2004. Revised September 2004.

Accepted November 2004.