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

Jason Roy
Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, U.S.A.
email: jason_roy@urmc.rochester.edu

and

Xihong Lin
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109 U.S.A.
email: xlin@sph.umich.edu

May 10, 2004

SUMMARY. We consider estimation in generalized linear mixed models for longitudinal discrete and continuous 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 paper, 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; (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. We allow mixed continuous and discrete time-varying covariates and assume they follow generalized transition models with different link functions. The EM algorithm is used for inference in the proposed model. Data from a longitudinal study of HIV-infected women are used to illustrate the methodology.

KEY WORDS: Asymptotic bias; EM algorithm; Generalized Linear Mixed Models; Missing data; Mixed continuous and discrete covariates; Random effects; Sensitivity analysis; Transition model.

1 Introduction

Subjects often drop out in longitudinal studies prior to completion of the study. Dropouts may be non-ignorable, 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 (Diggle and Kenward 1994; Little 1995) and semiparametric (Rotnitzky et al. 1998) approaches. It is commonly assumed in the longitudinal dropout literature that 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.

For example, Tashima et al. (2001) reported results from a longitudinal study of human immunodeficiency virus (HIV) infected women. This was a sub-study of the HIV Epidemiological Research Study (HERS) (Smith et al. 1997), where interest was in determining whether use of protease inhibitors (PIs) (yes/no) reduced the number of hospitalizations and emergency department visits. The HERS recruited 1310 participants from 4 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 these women were more likely to have been hospitalized in the past six months. A further complication is PI use changes over time and has the same pattern of missingness as the response. We are therefore interested in models that account for the missing binary response and covariates that are missing not at random.

Roy and Lin (2002) proposed a model for dealing with both dropout-related missing response and time-varying covariates in a multivariate, latent variable setting. They restricted in their

model to normally distributed outcomes and continuous time-varying covariates. However, it is common to have a discrete response and/or time-varying covariates of discrete or mixed type, e. g., the HERS data. Further, although several naive methods have been proposed in the literature to handle missing time-varying covariates, such as the last observation carried forward method, little is known about their performance, particularly in comparison to more complicated modeling strategies such as the one proposed here.

In this paper, we consider the situation where dropouts are informative and both the outcome and time-varying covariates are missing at the time of dropout. We first 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 last observation carried forward; (3) assuming dropouts are ignorable. Our asymptotic bias analysis shows these naive approaches yield substantially biased estimators of model parameters.

We next propose a model that allows for non-ignorable dropouts as well as missing covariates. Unlike Roy and Lin (2002), who assumed both the outcome and the time-varying covariates are continuous and follow linear models, we relax this strong assumption in this paper and allow the outcome to be continuous or discrete and follow generalized linear mixed models, and allow the dropout-related missing time-varying covariates to be continuous, discrete or mixed and follow generalized transition models with possibly different link functions. Non-ignorable dropouts are modeled using a selection model. In view of multi-dimensional integration required for a full likelihood analysis, a Monte Carlo EM algorithm (Wei and Tanner 1990) is developed for inference. Because the observed data cannot distinguish between missing at random and informative missingness (Little, 1995), we proceed with a sensitivity analysis (Rotnitzky, et al. 1998; Verbeke, et al., 2001).

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 provides an asymptotic bias analysis of the use of three naive approaches to handle missing

covariates. Section 4 describes the EM algorithm for inference in the proposed model. Simulation studies are carried out in Section 5, comparing the proposed methods to the naive approaches in finite samples for both 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 for K time points, but due to dropouts the ith unit is sampled at $K_i \leq K$ time points. 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 $i = 1, \cdots, n$ and time points $k = 1, \cdots, K_i$. The variable $R_i = k$ if unit i dropped out at timepoint k. We distinguish between two types of covariates: those that are known 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 Z_{ik} are completely observed.

We assume a generalized linear mixed model (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 following the quasilikelihood with means μ_{ik} and variances $\phi_y a_{ik}^{-1} v_y(\mu_{ik})$, where ϕ_y is a scale parameter, a_{ik} is a weight and $v_y(\cdot)$ is a variance function. The conditional mean is related to the covariates through the following generalized linear model:

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

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 following $N\{0, D(\theta)\}$, and 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:

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

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 $\{S_{i(k)}, T_{i(k)}, Y_{i(k)}\}$ up to time point k, where (k) denotes the history up to time point k. Since 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.

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}. Likelihood-based inference therefore requires the specification of an additional model for these missing covariates.

We allow the p_2 time-varying covariates T_{ik} to be continuous, discrete and mixed, and propose generalized transition models with possibly different links to model missing T_{ik}. For simplicity, we here assume the T_{ik} are fully observed at the baseline. Specifically, we assume the lth time-varying covariate T_{ikl} follows a quasilikelihood distribution with mean $\nu_{ikl} = \mathrm{E}(T_{ikl})$ and variance ${a_{Tl}}^{-1}\phi_{Tl}v_{Tl}(\nu_{ikl})$, where a_{Tl} is a weight, ϕ_{Tl} is a scale parameter and $v_{Tl}(\cdot)$ is a variance function, specific to the lth covariate T_l. Let $\phi_T = (\phi_{T1}, \cdots, \phi_{Tp_2})^T$. We assume T_{ikl} follows a first-order generalized linear transitional model ($l = 1, \cdots, p_2;\ k = 2, \cdots, K$) with possibly different link functions

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

where $h_l(\cdot)$ is a known lth covariate-specific monotone link function and $\lambda_l = (\lambda_{0l}, \lambda_{1l}, \lambda_{2l}^T)^T$ is an unknown parameter vector. Model (3) hence allows missing mixed discrete and continuous time-varying covariates. We assume T_{ikl} is independent of $T_{ikl'}$ for ($l \neq l'$) conditional on the history $T_{i(k-1)}$ and $S_{i(k)}$. However, no assumption is made 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. Model (3) can be easily extended to allow T_{ikl} to depend on the other time-varying covariates $T_{ikl'}$ $(l' \neq l)$.

Let $Y_{obs,i} = (Y_{i1}, \cdots, Y_{iK_i})^T$ denote the observed values of Y_i, with T_i, $S_{obs,i}$ and $Z_{obs,i}$ defined similarly. For the dropout units, denote by $Y_{mis,i} = Y_{i,K_i+1}$ and $T_{mis,i} = T_{i,K_i+1}$ the missing observations. Then define $Y_i = Y_{obs,i}$ if $K_i = K$ (i.e., unit i did not dropout) and $Y_i = (Y_{obs,i}^T, Y_{mis,i})^T$ otherwise (i.e., unit i drops out at time point $K_i + 1$). Define T_i similarly. Let $S_i = (S_{i1}, \cdots, S_{i,K_i+1})^T$ 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 likelihood $L(Y_{obs,i}|T_{obs,i}, S_{obs,i}, Z_{obs,i}; \beta, \theta)$, as in Breslow and Clayton (1993).

When $\alpha_2 \neq 0$, the missing data mechanism is non-ignorable, and inference needs to be based on the joint integrated quasilikelihood

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

where $\lambda = (\lambda_1^T, \cdots, \lambda_{p_2}^T)^T$, $\alpha = (\alpha_0^T, \alpha_1^T)^T$, 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. The loglikelihoods in the integrand in (4) are

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

where $\ell(\cdot) = lnL(\cdot)$, $d(y, \mu) = -2 \int_y^\mu \{\phi a_{ik}^{-1} v(u)\}^{-1} (y - u) du$ is the deviance function and $d_l(t, \nu)$ is defined similarly for T_{ikl}, 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 one of the three naive approaches: (1) using the baseline measures throughout by assuming the time-varying covariates have not changed since baseline (§ 3.2); (2) carrying forward the last observation, i.e., assuming their values at the time of dropout are the same as the previous values (§ 3.3); (3) ignoring missing data completely by assuming missing at random (§ 3.4). These naive approaches enable one to fit models using the existing approaches that assume covariates are completely observed. For example, Diggle and Kenward's method (1994) could be applied using (1) and (2); a standard GLMM could be applied using (3). 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

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + b_i + \epsilon_{ik},\ k = 1, \cdots, K, \tag{5}$$

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 linear transition model to account for missing X_{ik}

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

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

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

where P_{ik} is defined in § 2.

The naive model assumes

$$Y_{ik} = \beta_{0,naive} + \beta_{1,naive} X^*_{ik} + b_i + \epsilon_{ik}, \tag{8}$$

where $b_i \sim N(0, \theta_{naive})$ and $\epsilon_{ik} \sim N(0, \tau^2_{naive})$, 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.

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

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

where the expectation is taken under the true models (5)-(7). The solution $\{\beta_{naive}, \xi_{naive}\}$ 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. We provide below detailed asymptotic bias calculations for the three naive approaches.

3.2 The baseline approach

The simplest naive approach to fill in the missing covariate values is the baseline approach by setting $X^*_{ik} = X_{i1}$, for $k = 1, \cdots, K$. The naive loglikelihood under this approach is

$$\begin{aligned} \ell_{naive}(Y_i, R_i) &= \sum_{k=2}^{K} I[R_i = k] \ln \int \left\{ \prod_{j=2}^{2 \vee (k-1)} (1 - P_{ij}) \right\}^{I(k>2)} P_{ik} L(Y_{i1}, \cdots, Y_{ik} | X^*_i) dY_{ik} \\ &\quad + I[R_i = K+1] \ln \left\{ L(Y_i | X^*_i) \prod_{j=2}^{K} (1 - P_{ij}) \right\}. \end{aligned}$$

Since P_{ik} for $k = 2, \cdots, K$, is free of γ, some calculations show that, apart from a constant,

$$\begin{aligned} \ell_{naive}(Y_i, R_i) &= \ell(Y_{i1} | X_{i1}) + \sum_{k=2}^{K} I[R_i = k] \ln \int P_{ik} L(Y_{i2}, \cdots, Y_{ik} | Y_{i1}, X^*_i) dY_{ik} \\ &\quad + I[R_i = K+1] \ell(Y_i | Y_{i1}, X^*_i). \end{aligned} \tag{10}$$

Solving equation (9) is relatively straightforward. For illustration consider the simple case where $K = 2$. The solution for general K is similar. Suppressing the subscript i, some calculations show that the asymptotic limits of the naive estimates $\{\beta_{naive}, \xi_{naive}\}$ solve

$$\mathrm{E}\left\{\frac{\widetilde{\mathrm{X}}_1(\mathrm{Y}_1 - \widetilde{\mathrm{X}}_1^{\mathrm{T}}\beta_{\mathrm{naive}})}{\theta_{\mathrm{naive}} + \tau^2_{\mathrm{naive}}}\right\} + \Pr(\mathrm{R}=2)\mathrm{E}\left\{\frac{\partial}{\partial\beta_{\mathrm{naive}}}\log\int \mathrm{L}(\mathrm{Y}_2|\mathrm{Y}_1, X^*)\mathrm{P}_2\mathrm{dY}_2\middle|\mathrm{R}=2\right\}$$
$$+\Pr(\mathrm{R}=3)\mathrm{E}\left[\frac{\partial}{\partial\beta_{\mathrm{naive}}}\ell(\mathrm{Y}_2|\mathrm{Y}_1, X^*)\middle|\mathrm{R}=3\right] = 0, \tag{11}$$

$$\frac{1}{2}\mathrm{E}\left\{\frac{(Y_1 - \widetilde{X}_1^T\beta_{naive})^2}{(\theta_{naive} + \tau^2_{naive})^2} - \frac{1}{\theta_{naive} + \tau^2_{naive}}\right\} 1_2 + \Pr(\mathrm{R}=2)\mathrm{E}\left\{\frac{\partial}{\partial\xi_{naive}}\log\int L(Y_2|Y_1, X^*)P_2dY_2\middle|R=2\right\}$$
$$+\Pr(\mathrm{R}=3)\mathrm{E}\left[\frac{\partial}{\partial\xi_{naive}}\ell(Y_2|Y_1, X^*)\middle|R=3\right] = 0, \tag{12}$$

where $\widetilde{X}_1 = (1, X_1)^T$, $X^* = (X_1, X_1)^T$, $1_2 = (1,1)^T$ and the expectations are taken under the true models (5)-(7) with respect to the true distributions of (X_1, X_2, Y_1, Y_2). See Appendix for details.

3.3 The last observation carried forward approach

The LOCF approach fills in the missing covariate data by carrying forward the last observed value by setting $X^*_{ik} = X_{ik}$ for $k = 1, \cdots, K_i$, and $X^*_{i,K_i+1} = X_{i,K_i}$ if $R_i < K+1$. For units that did not drop out, this approach uses all of the covariate information. For units that did drop out, it assumes the values of the time-varying covariates at the time of dropout are the same as those at the previous time point. One hence expects it would perform better than the baseline approach.

The naive loglikelihood under the LOCF approach takes the same form as (10) with X^*_i defined in the this section. The asymptotic limits of the naive estimators $\{\beta_{naive}, \xi_{naive}\}$, for the special case where $K = 2$, simultaneously solve equations taking the same form as (11) and (12) except that $X^* = (X_1, X_1)$ in the second term and $X^* = (X_1, X_2)$ in the third term. Hence the first two terms in each equation are the same as they were for the baseline approach, while the third term is different. The equations are similar for general K. See Appendix for details.

3.4 The ignorable missing data approach

The easiest approach is to ignore missing data by assuming missing data are ignorable (IM). Unlike the baseline and LOCF approaches, the bias here arises from ignoring both missing X's and missing

Y's, and the dropout model (7) is not needed. The resulting naive loglikelihood for unit i is

$$\ell_{naive}(Y_i|X_i) = \ell(Y_{i1}|X_{i1}) + \sum_{k=2}^{K} I(R_i = k+1)\ell(Y_{i2}, \cdots, Y_{ik}|Y_{i1}, X_i).$$

Setting the expectation of the naive score function to zero and suppressing the subscript i, the asymptotic limits of the naive estimators $\{\beta_{naive}, \xi_{naive}\}$, for the illustrative $K = 2$ case, solve

$$\mathrm{E}\{(\theta_{\mathrm{naive}} + \tau^2_{\mathrm{naive}})^{-1}\widetilde{\mathrm{X}}_1(\mathrm{Y}_1 - \widetilde{\mathrm{X}}_1^{\mathrm{T}}\beta_{\mathrm{naive}})\} + \Pr(\mathrm{R}=3)\mathrm{E}\left\{\frac{\partial}{\partial\beta_{\mathrm{naive}}}\ell(\mathrm{Y}_2|\mathrm{Y}_1, X)\Big|\mathrm{R}=3\right\} = 0, \quad (13)$$

$$\frac{1}{2}\mathrm{E}\left\{\frac{(Y_1 - \widetilde{X}_1^T\beta_{naive})^2}{(\theta_{naive} + \tau^2_{naive})^2} - \frac{1}{\theta_{naive} + \tau^2_{naive}}\right\}1_2 + \Pr(\mathrm{R}=3)\mathrm{E}\left\{\frac{\partial}{\partial\xi_{naive}}\ell(Y_2|Y_1, X)\Big|R=3\right\} = 0, \quad (14)$$

where $X = (X_1, X_2)^T$, and the expectations are taken under the true models (5)-(7) with respect to the true distributions of (X_1, X_2, Y_1, Y_2). Note that equations (13) and (14) are the same as equations (11) and (12) for LOCF, except with the second terms missing. That is because the second terms in (11) and (12) correspond to missing data, which is ignored in the IM analysis.

3.5 The results

solutions. We numerically calculated the asymptotic biases of the naive estimators by simultaneously solving equations (11) and (12) for the baseline and LOCF approaches, and equations (13) and (14) for the IM approach. 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 parameter values 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 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.

For the ease of interpretability, we plot in Figure 1 the relative bias plots for the three appraoches a function of the dropout probability instead of α_2, for $(\lambda_0, \lambda_1) = (0.5, 0.5)$, which assumes the value of X at the two time points are positively correlated. When the marginal probability of dropout

is zero, the LOCF and IM approaches yield unbiased parameters, since all of the data are used. However, the baseline approach still yield biased estimates even when there are no dropouts. As the marginal probability of dropout goes to one, 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 dropout 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 approach 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.

We also calculated bias curves under other scenarios. When the value of X at time points 1 and 2 were assumed to be 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. When X_2 is equal to X_1 plus random noise (i.e., $(\lambda_0, \lambda_2) = (0, 1)$), except for τ^2, the magnitude of the biases is smaller than in Figure 1. The bias plot of τ^2 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 parameter estimates. 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.

4 Estimation and Inference

The asymptotic bias results in § 3 suggest that a statistical model needs to be developed to account for account for missing time-varying covariates in the presence of informative dropouts. We propose such a model using the selection-transition model (1)-(3). Statistical inference for model (1)-(3) is

challenged by the possibly high dimensional integrals in the joint quasilikelihood (4). For Gaussian outcome Y_i and Gaussian time-varying covariates T_{ikl} the quasilikelihood (4) can be simplified and only involves one dimensional integral (Roy and Lin, 2002). However, for non-Gaussian data the dimension of integration of the joint quasilikelihood (4) is $q+1+p_2$, which may be quite large. 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 $\{Y_{obs,i}, R_i, T_{obs,i}, S_i\}$. The complete data loglikelihood is $\ell_c(\Omega) = \sum_{i=1}^n \{\ell(Y_i|S_i, T_i, b_i; \beta, \phi) + \ell(T_i|T_{i1}, S_i; \lambda, \sigma) + \ell(b_i; \theta) + \ell(R_i|Y_i, T_{obs,i}, S_{obs,i}; \alpha)\}$, where each of the terms was defined in §2.

Let $\Omega^{(k)}$ denote the estimate of Ω at the kth iteration. The M-step updates β by solving

$$\sum_{i=1}^{n} \sum_{k=1}^{(K_i+1)\wedge K} \mathrm{E}\left\{ X_{ik}\mu'_{ik}V_{ik}^{-1}(Y_{ik}-\mu_{ik}) \middle| Y_{obs,i}, T_{obs,i}, S_i, R_i; \Omega^{(k)} \right\} = 0, \tag{15}$$

where $\mu'_{ik} = 1/g'(\mu_{ik})$ and $V_{ik} = \phi_y a_{ik}^{-1} v_y(\mu_{ik})$. For linear mixed models, (15) has a closed form; for GLMMs for non-normal data, (15) needs to be solved iteratively using the Fisher-scoring algorithm. The estimate of ϕ_y is updated similarly. 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 $D^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \mathrm{E}\{b_i b_i^T | Y_{obs,i}, T_{obs,i}, S_i, R_i; \Omega^{(k)}\}$. The parameters in the dropout model $\alpha = (\alpha_{02}, \cdots, \alpha_{0K}, \alpha_1^T)^T$ are updated iteratively using the Newton-Raphson algorithm

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

where

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

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

and $\Delta_{ik} = \{\delta_k, H_{i(k-1)}^T\}^T$ and δ_k is a $(K-1)\times 1$ vector of zeros except that the kth element is 1.

The E-step involves taking expectations of functions of the 'missing data' conditional on the observed data and take the form $\mathrm{E}\{r(b_i, Y_{mis,i}, T_{mis,i})|Y_{obs,i}, T_{obs,i}, S_i, R_i\}$, where r is a function of the missing data, e.g., $r(b_i, Y_{mis,i}, T_{mis,i}) = b_i b_i^T$. For subjects not dropping out, the expectation is

$$\mathrm{E}\{r(b_i)|Y_i, T_i, R_i\} = \frac{\int h(b_i)L(Y_i|T_i, S_i)L(R_i|Y_i, T_i, S_i)L(T_i)L(b_i)db_i}{\int L(Y_i|T_i, S_i)L(R_i|Y_i, T_i, S_i)L(T_i)L(b_i)db_i},$$

where the integrals are q-dimensional and the Gaussian quadrature can be used to evaluate the integral. For subjects that did drop out of the study, the expectations take the form

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

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, or components of T that are binary or normal, they could be integrated over directly. A Monte Carlo approximation (Wei and Tanner 1990; Booth and Hobert 1999) can be used as an alternative for all or part of the integrals. The method of Louis (1982) is used for obtaining standard errors estimates.

5 Simulations

We carry out two simulation studies – one for continuous data and one for binary data. The goal are to evaluate the finite sample performance of the EM algorithm for the proposed model, and the finite sample bias of the three naive methods for handling missing covariates. We expand the setting of the asymptotic bias by considering more than two time points in our simulation.

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. One time-varying covariate T_{ik} was generated. We assumed the probability P_{ik} for $k = 2, 3$,

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

Each simulated data set was analyzed using four methods – the EM-based MLE method and three naive methods discussed in Section 3. For each configuration, 1000 replications were done.

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 $e_{ik} \sim N(0, 1)$. The response was then $Y_{ik} = \beta_0 + \beta_s k + \beta_T T_{ik} + \varepsilon_{ik}$, where b_i and ε_i were independently distributed as $N(0, 1)$.

We set $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 1$, $\alpha_{02} = \alpha_{03} = 0.5$ 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. The results are given in Table 1. The proposed EM-based MLE method 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 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 performs 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 50 percent.

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 0.5, then generated T_{ik} from a Bernoulli distribution with success probability $\text{logit}^{-1}(\lambda_0 + \lambda_1 T_{i,k-1})$, for $k = 2, 3$. 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 $b_i \sim N(0, 1)$.

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 to change the probability of dropout and the dependence of dropout on the current value of the response. We also considered several values of λ to assess how the methods performed as serial correlation in T varied. Table 2 shows the results. The proposed EM-based MLE yielded parameters with little bias ($< 3\%$). The bias substantially decrease if we increased sample size $n = 500$ (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, unlike the continuous case, the LOCF approach performed fairly well in these simulations, with $< 10\%$ bias for β_S and β_T in all cases.

6 Application

We next analyzed a subset of the HERS data described in the Introduction. We were interested in whether or not the use of protease inhibitors (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, and subjects with CD4 cell counts < 200 have the most advanced disease, and are prime candidates for the therapy.

At baseline, 126 women from HERS had CD4 cell count < 200 cells/mL. These women were followed up every six 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. Other covariates include 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. A large number of women dropped during the study period, with only 21 of 126 women remaining at the end of study. PIs were not available 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 and reached 67% by the end of the study. The proportion of hospitalization declined over time from 48% to 19% on the other hand. 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 and cause the observed decline in the proportion of hospitalizations. In view of the potential informative dropouts and the missing PI status at the time of dropout, we used the methods proposed in Section 4 to account for possible non-ignorable missing data with a missing time-varying covariate.

The response Y_{ik} took value of 1 if subject i was hospitalized between visits $k-1$ and k, and 0 otherwise. The covariates S_{ik} included: 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), $500 - 5,000$, $5,000 - 30,000$ and $> 30,000$; race, which was categorized as African American, Hispanic or "other" and white (reference); indicators for study site; and time, entered into the model as the visit number. PI use was the only one time-varying covariate that was subject to dropout-related missingness. Therefore, T_{ik} was an indicator for PI use between visits $k-1$ and k. We assumed that Y_i given covariates followed a logistic random intercept model. 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 a logistic transition model (3). After conditioning on past PI use, other covariates were not significant predictors of PI use, and therefore we only included past PI use in (3).

We varied α_2 in the dropout model (2) from -1 to 1. When α_2 was positive, dropout is more likely

if the subject had a recent hospitalization; $\alpha_2 = 0$ means a recent hospitalization was unrelated to dropout (the MAR case); $\alpha_2 < 0$ implies a hospitalization decreased the likelihood of dropout. 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 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.

7 Discussion

In this paper 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 fill the missing covariates naively using the baseline approach, the last observation carried forward approach, and the ignorable missingness approach. Our asymptotic bias analysis shows that these naive approaches often yield biased estimates of the model parameters. We propose in this paper a selection-transition model to allow for missing mixed discrete and continuous time-varying covariates at the time of dropout for longitudinal discrete and continuous outcomes with informative dropouts.

For simplicity, we assumed the selection model (2) did not depend on missing time-varying covariates $T_{i,k}$ at the time of dropout. 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 to the missing

outcome. 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, Lipsitz and Chen (1999).

Acknowledgment

The work of Roy was supported by NIH grants R01-AI50505 and P30-AI42853 (Lifespan/Tufts/Brown Center for AIDS research). Data from HERS were collected under CDCP grant U64-CCU10675. The work of Lin was supported by NIH grant CA76404.

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.

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.

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.

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.

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.

Appendix: Details in bias analysis

To obtain the asymptotic limits of the naive estimators, we need to solve equations (11) and (12) for the baseline and LOCF approaches, and equations (13) and (14) for the IM approach, for $(\beta_{naive}, \xi_{naive})$ as a function of the true values (β, ξ). It can be easily shown that the first terms in equations (11)-(14) have closed form

$$\mathrm{E}(\tilde{X}_1(Y_1 - \tilde{X}_1^T\beta_{naive})) = \begin{bmatrix} \beta_0 - \beta_{0,naive} + \mu_1(\beta_1 - \beta_{1,naive}) \\ \mu_1(\beta_0 - \beta_{0,naive}) + (\mu_1^2 + \delta)(\beta_1 - \beta_{1,naive}) \end{bmatrix}$$

$$\mathrm{E}(Y_1 - \tilde{X}_1^T \beta_{naive})^2 \quad = \quad \{\beta_0 - \beta_{0,naive} + \mu_1(\beta_1 - \beta_{1,naive})\}^2 + \delta(\beta_1 - \beta_{1,naive})^2 + \theta + \tau^2.$$

The second term in (11) does not have a closed form and needs to be evaluated numerically

$$\mathrm{E}\Big[\frac{\partial}{\partial \beta_{naive}} \log \int L(Y_2|Y_1, X^*) P_2 dY_2 \Big| R = 2\Big] =$$
$$\int\int \Big[\frac{\partial}{\partial \beta_{naive}} \log \int L(Y_2|Y_1, X^* = X_1 1) P_2 dY_2\Big] L(Y_1, X_1|R = 2) dY_1 dX_1,$$

where $[Y_2|Y_1, X^* = X_1 1]$ is distributed as

$$N\{\tilde{X}_1^T \beta_{naive} + \theta_{naive}(\theta_{naive} + \tau^2_{naive})^{-1}(Y_1 - \tilde{X}_1^T \beta_{naive}), \theta_{naive} + \tau^2_{naive} - (\theta_{naive} + \tau^2_{naive})^{-1}\theta^2_{naive}\}$$

and

$$L(Y_1, X_1|R = 2) = \frac{L(Y_1, X_1) \int L(R = 2|Y_1, Y_2) L(Y_2|Y_1, X_1) dY_2}{P(R = 2)}.$$

The second term in equations (11) and (12) can therefore be evaluated by finding the corresponding partial derivatives numerically, and evaluating the integrals using Gauss-Hermite quadrature.

The third term in equations (11) and (12) can be evaluated numerically in a similar fashion. The results will differ depending on whether the baseline or LOCF approach is used. If the problem of missing covariates is addressed using baseline measures only, then we have

$$\mathrm{E}\left[\frac{\partial}{\partial \beta_{naive}} \ell(Y_2|Y_1, X^*) \Big| R = 3\right] = \int \frac{\partial}{\partial \beta_{naive}} \ell(Y_2|Y_1, X^* = X_1 1) L(Y_1, Y_2, X_1|R = 3) dY_2 dY_1 dX_1.$$

Calculation of this expectation is the same as in the LOCF approach, except here $\ell(Y_2|Y_1, X^* = X_1 1)$ now becomes $\ell(Y_2|Y_1, X^* = (X_1, X_2)^T)$, $L(Y_1, Y_2, X_1|R = 3)$ becomes $L(Y_1, Y_2, X_1, X_2|R = 3)$, and integration is over X_2 as well. Calculations of the second term of equations (13) and (14) are the same as calculating the third term in equation (11) and (12) for LOCF.

Table 1

Results from simulation study with continuous response. Percentage bias 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.

		Method							
		Baseline		LOCF		IM		Proposed	
α_2	(λ_0, λ_1)	β_s	β_T	β_s	β_T	β_s	β_T	β_s	β_T
0.5	(0.5,0.5)	72.8	-30.9	1.6	-2.9	-15.3	-1.7	-0.4	0.0
	(0,1)	-12.2	4.2	-0.1	0.1	-12.8	-1.5	-1.0	0.1
	(1,0)	104.2	-51.7	5.1	-5.0	-15.9	-2.4	0.2	-0.6
1.0	(0.5,0.5)	60.4	-24.0	8.8	-10.9	-46.6	-5.5	0.1	-0.5
	(0,1)	-42.0	3.0	-1.9	-0.1	-25.4	-0.2	-0.8	-0.4
	(1,0)	108.3	-40.5	30.6	-20.9	-51.4	-8.6	-1.4	-0.5

Table 2

Results from simulation study with binary response. Percentage bias 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.

		Method							
		Baseline		LOCF		IM		Proposed	
α_2	(λ_0, λ_1)	β_s	β_T	β_s	β_T	β_s	β_T	β_s	β_T
0.5	(-0.5,1)	2.4	-43.7	3.8	4.5	-9.7	-42.4	1.6	1.3
	(-1.5,3)	4.7	-16.2	3.2	3.0	-8.0	-21.3	0.9	1.6
	(0,0)	6.5	-40.1	6.5	4.5	-11.4	-51.8	1.1	1.2
1.0	(-0.5,1)	8.2	-32.0	4.2	4.1	-24.7	-80.2	1.0	1.5
	(-1.5,3)	10.3	-13.2	4.3	3.6	-21.2	-19.7	1.2	2.2
	(0,0)	8.1	-37.9	7.2	8.9	-26.8	-46.7	0.6	1.6

Table 3

Percentage of patients with a hospitalization and percentage of protease inhibitor (PI) use at each interval

	Interval between HERS visit numbers											
Characteristic	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

Table 4

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

Variable	$\alpha_2 = -1$	$\alpha_2 = -0.5$	$\alpha_2 = 0$	$\alpha_2 = 0.5$	$\alpha_2 = 1$
Intercept	0.66	0.60	0.51	0.40	0.29
	(0.48)	(0.49)	(0.51)	(0.51)	(0.51)
Visit	-0.16	-0.14	-0.13	-0.11	-0.09
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Drug use	0.33	0.33	0.32	0.31	0.30
	(0.23)	(0.24)	(0.24)	(0.25)	(0.25)
Age	0.01	0.01	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
African American	0.51	0.54	0.57	0.59	0.61
	(0.35)	(0.36)	(0.37)	(0.37)	(0.38)
Hispanic	0.71	0.75	0.79	0.81	0.82
	(0.38)	(0.39)	(0.40)	(0.40)	(0.40)
HIV RNA $> 30,000$	-1.18	-1.15	-1.11	-1.06	-0.99
	(0.46)	(0.48)	(0.49)	(0.49)	(0.50)
HIV RNA 5000-30,000	-0.77	-0.72	-0.66	-0.58	-0.50
	(0.47)	(0.48)	(0.49)	(0.50)	(0.50)
HIV RNA 500-5000	-0.51	-0.43	-0.34	-0.23	-0.12
	(0.46)	(0.47)	(0.48)	(0.49)	(0.50)
site1	-0.54	-0.55	-0.56	-0.55	-0.55
	(0.33)	(0.33)	(0.34)	(0.35)	(0.35)
site2	-0.10	-0.10	-0.10	-0.10	-0.09
	(0.37)	(0.38)	(0.39)	(0.39)	(0.40)
site3	-0.56	-0.55	-0.58	-0.58	-0.58
	(0.39)	(0.40)	(0.41)	(0.42)	(0.42)
PI use	-0.76	-0.79	-0.84	-0.89	-0.96
	(0.29)	(0.29)	(0.30)	(0.30)	(0.30)

Table 5

Parameter estimates and estimated standard errors of the parameters of secondary interest from a sensitivity analysis 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 > 30000	1.32	1.40	1.51	1.64	1.79
	(0.64)	(0.64)	(0.64)	(0.65)	(0.66)
α_{12} HIV RNA 5000-30000	1.62	1.70	1.80	1.90	2.02
	(0.64)	(0.64)	(0.64)	(0.65)	(0.66)
α_{13} HIV RNA 500-5000	1.98	2.05	2.13	2.22	2.33
	(0.65)	(0.64)	(0.65)	(0.65)	(0.66)
α_{14} Y_{k-1}	0.88	0.81	0.72	0.64	0.53
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
λ_0	-2.22	-2.22	-2.23	-2.24	-2.25
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
λ_1	4.05	4.05	4.05	4.06	4.06
	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)
θ	0.52	0.55	0.59	0.62	0.65
	(0.18)	(0.19)	(0.20)	(0.21)	(0.21)

List of Illustrations