Simultaneous Inference for Longitudinal Data with Detection Limits and Covariates Measured with Errors, with Application to AIDS Studies

Lang Wu

Department of Statistics, University of British Columbia,

Vancouver, B.C., V6T 1Z2, Canada

Email: lang@stat.ubc.ca

December, 2001

Abstract

In AIDS studies such as HIV viral dynamics, statistical inference is often complicated because the viral load measurements may subject to left censoring due to a detection limit and time-varying covariates such as CD4 counts may be measured with substantial errors. Mixed-effects models are often used to model the response and the covariate processes in these studies. We propose a unified approach which addresses the censoring and measurement errors simultaneously. We estimate the model parameters by a Monte-Carlo EM algorithm via the Gibbs sampler. A simulation study is conducted to compare the proposed method with the usual two-step method and a naive method. We find that the proposed method produces approximately unbiased estimates with more reliable standard errors. A real dataset from an AIDS study is analyzed using the proposed method.

KEY WORDS: EM algorithm, Gibbs sampler, HIV, Mixed-effects model.

1. INTRODUCTION

In AIDS studies, researchers have recently shown great interest in modeling the viral load (plasma HIV-1 RNA copies) trajectory after initiation of a potent antiretroviral treatment (Ho, *et al.*, 1995; Perelson, *et al.*, 1996; Wu and Ding, 1999). These viral dynamic models provide good understanding of the pathogenesis of HIV infection and evaluation of antiretroviral therapies. Linear and nonlinear mixed-effects models have been used in these studies to account for inter-patient and intra-patient variations. The inter-patient variations typically receive great attention and may be partially explained by time-varying covariates such as CD4 counts (the number of CD4+ T lymphocytes per volume of blood) (Wu, *et al.*, 1999). Statistical analyses in these studies are often complicated because the viral load measurements may subject to left censoring due to a detection limit (i.e., a certain threshold below which viral loads are not quantifiable) and the covariates such as CD4 are often measured with substantial errors or measured at time points different from the viral load measurements (i.e., missing covariates). In this article, we consider an unified approach to address censoring and measurement errors (or missing covariates) simultaneously.

Our research was motivated by the AIDS Clinical Trials Group (ACTG) Study 315. In this study, fifty-three HIV-1 infected patients were treated with a potent antiretroviral regimen which consists of protease inhibitor (PI) and reverse transcriptase inhibitor (RTI) drugs. Viral load is repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, and weeks 8, 12, 24, and 48 after initiation of treatment. Covariates such as CD4 counts were also measured throughout the study. In this study, the viral load detectable limit is 100 copies per milliliter, and there are 40 out of 361 (11%) of all viral load measurements are below the detection limit. Figure 1 shows the measurements of HIV viral load and CD4 for four randomly selected patients from ACTG Protocol 315. Previous studies (e.g., Wu, *et al.*, 1999) show that the rate of change in viral load may be associated with change in CD4 counts. It is important to study the patterns of virological response to treatment, and to understand the mechanism of the inter-patient variation in order to make clinical decisions and provide individualized treatments. Since CD4 counts are typically measured with substantial errors, statistical analyses must take the measurement error into account. Even for covariates measured with little errors, missing data can be common in such studies since the covariates measurement times may be different from the response measurement times, as is evident in Figure 1.

When viral loads are below the detection limit, a common practice is to impute the censored values by the detection limit or half the detection limit. Such ad-hoc methods may produce biased results (Hughes, 1999; Hughes, 2000). Linear mixed-effects models with censored responses were studied in Pettitt (1986) and Hughes (1999). Ibrahim, *et al* (2001) recently consider missing response in generalized linear models when the missing data mechanism is non-ignorable. There is extensive literature on measurement errors (Carroll, *et al*, 1995). When covariates are measured with errors, a common approach is the so-called two-step method: the first step estimates the "true" covariate values based on some covariate models; at the second step the covariates in the response model are substituted by their estimates from the first step. This two-step method, however, may underestimate the standard errors since it fails to take the covariate estimation into account (Higgins, Davidian, and Giltinan, 1997). Wang and Taylor (2001) recently consider a joint model for analysis of longitudinal data and event time data.

In this article we propose to model the covariate and the response processes jointly, incorporating censoring in the response and measurement errors or missing data in the covariates simultaneously. We consider linear mixed-effects models for both the covariate and the response processes, and estimate the model parameters using a Monte-Carlo EM algorithm. Section 2 presents the models, likelihoods, and a Monte-Carlo EM algorithm. Section 3 presents a real data example. A simulation study is conducted in Section 4. In Section 5, we discuss a more general case for longitudinal data with non-ignorable missing response. We conclude the article in Section 6.

2. A JOINT MODEL

2.1. The Models

We consider longitudinal data with left censoring, but our method can be easily extended to "doubly censored" case. Let y_{ij} be the value of the response Y for individual i at time t_{ij}, subject to left censoring, $i = 1, \ldots, n;\ j = 1, \ldots, n_i$. The observed value of y_{ij} is (q_{ij}, c_{ij}), where c_{ij} is the censoring indicator such that y_{ij} is observed if $c_{ij} = 0$ and y_{ij} is left censored if $c_{ij} = 1$, i.e.,

$$y_{ij} = q_{ij} \text{ if } c_{ij} = 0, \qquad y_{ij} \leq d \text{ if } c_{ij} = 1, \tag{1}$$

where d is some known constant (e.g., the lower detection limit). For simplicity, we first consider a single time-varying covariate Z. Extensions to more than one covariates will be discussed in Section 2.3. Let z_{ik} be the covariate value for individual i at time u_{ik}, $i = 1, \cdots, n; \quad k = 1, \cdots, m_i$. Note that we allow the covariate observation times u_{ik} to differ from the scheduled response measurement times t_{ij}. Thus we allow missing data in the covariate. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T, \mathbf{z}_i = (z_{i1}, \ldots, z_{im_i})^T, \mathbf{q}_i = (q_{i1}, \ldots, q_{in_i})^T$, and $\mathbf{c}_i = (c_{i1}, \ldots, c_{in_i})^T$. The observed data are $\{(\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i), \ i = 1, \ldots, n\}$.

We model the covariate process by the following linear mixed effects model

$$\begin{aligned} z_{ik} &= \mathbf{u}_{ik}^T \boldsymbol{\alpha}_i + \epsilon_{ik}, \qquad (2) \\ \boldsymbol{\alpha}_i &\sim N(\boldsymbol{\alpha}, A), \qquad \epsilon_{ik} \sim N(0, \sigma^2), \qquad i = 1, \cdots, n; \quad k = 1, \cdots, m_i, \end{aligned}$$

where $\mathbf{u}_{ik}$ is a $p \times 1$ design vector, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \cdots, \alpha_{ip})^T$ is a $p \times 1$ vector of unknown individual-specific parameters, ϵ_{ik} is the random measurement error for individual i at time u_{ik}, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^T$ and A are the $p \times 1$ unknown population mean vector and $p \times p$ covariance matrix respectively, and σ^2 is the unknown within-individual variance. We assume that $\epsilon_{i1}, \ldots, \epsilon_{im_i}$ are conditional independent given $\boldsymbol{\alpha}_i$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ is independent of $\boldsymbol{\alpha}_i$. In the absence of theoretical rational, a reasonable approach for the covariate process is to adopt an empirical, parametric representation that captures the main features of the covariate path, such as a low-order polynomial. In this case, we have $\mathbf{u}_{ik} = (1, u_{ik}, \ldots, u_{ik}^{p-1})^T$. Thus, for example, a quadratic function corresponds to $p = 3$. Note that model (2) may be interpreted as a covariate measurement error model (Carroll, *et al*, 1995; Wang, *et al*, 2000).

In the covariate model (2), $z^*_{ik} \equiv \mathbf{u}_{ik}^T \boldsymbol{\alpha}_i$ represents the true but unobservable covariate value for individual i at time u_{ik}. At any time point, the trajectory of the covariate up to that time could be summarized by the vector $\boldsymbol{\alpha}_i$. For example, if we fit a straight line to the covariate process, i.e., $\mathbf{u}_{ik} = (1, u_{ik})^T$, then $\boldsymbol{\alpha}_i$ represents the covariate initial value and slope (rate of change). When the covariates z_{ij} are measured with errors, we may be interested in the relationship between the response y_{ij} and the true but unobservable covariate values z^*_{ij} or the past history or summaries of the covariate values such as the covariate initial value and rate of change. We assume that $\mathbf{z}_i$ is a surrogate for $\boldsymbol{\alpha}_i$ such that the distribution of $(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i)$ is that of $(\mathbf{y}_i|\boldsymbol{\alpha}_i)$, independent of $\mathbf{z}_i$. To allow for missing data in the covariate, we recast the covariate model (2) in continuous time as follows

$$z_i(t) = \mathbf{u}_i(t)^T \boldsymbol{\alpha}_i + \epsilon_i(t), \qquad i = 1, \cdots, n, \tag{3}$$

where, for individual i, $z_i(t), \mathbf{u}_i(t)$, and $\epsilon_i(t)$ are the covariate value, design vector, and measurement error at time t respectively. At response measurement time t_{ij}, the "true" covariate values are $z^*_{ij} = \mathbf{u}_{ij}^T \boldsymbol{\alpha}_i$, where $\mathbf{u}_{ij} = \mathbf{u}_i(t_{ij})$.

Let $h(\boldsymbol{\alpha}_i)$ be a known linear function which summarizes the time-varying covariate and may depend on other covariates such as time. We consider the following linear mixed-effects model with time-dependent parameters for the response process,

$$\begin{aligned} y_{ij} &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_{ij} + e_{ij}, \\ \boldsymbol{\beta}_{ij} &= \mathbf{d}(h(\boldsymbol{\alpha}_i), \boldsymbol{\beta}, \mathbf{b}_i), \qquad i = 1, \cdots, n; \quad j = 1, \cdots, n_i, \end{aligned} \tag{4}$$

where $\mathbf{x}_{ij}$ is a $s \times 1$ design vector, $\boldsymbol{\beta}_{ij} = (\beta_{ij1}, \cdots, \beta_{ijs})^T$ is a $(s \times 1)$ vector of individual-specific and time-dependent regression parameters, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_r)^T$ is a $(r \times 1)$ vector of fixed effects, $\mathbf{d}(\cdot)$ is a s-dimensional vector-valued linear function, $\mathbf{e}_i = (e_{i1}, \cdots, e_{in_i})^T$, and $\mathbf{b}_i = (b_{i1}, \cdots, b_{is})^T$ is a vector of random effects. Thus in the response model (4) we link the time-dependent parameters $\boldsymbol{\beta}_{ij}$ to the covariate through the function $h(\boldsymbol{\alpha}_i)$ which summarizes the covariate process. We assume that

$$\mathbf{e}_i | \boldsymbol{\beta}_{ij} \sim N(\mathbf{0}, \delta^2 I), \qquad \mathbf{b}_i \text{ i.i.d.} \sim N(\mathbf{0}, B),$$

where δ^2 is the unknown within individual variance, I is the identity matrix, and the $s \times s$ matrix B quantifies the random inter-individual covariance. We also assume that $\mathbf{e}_i$ is independent of $\boldsymbol{\epsilon}_i$, and $\mathbf{b}_i$ is independent of $\mathbf{e}_i$ and $\boldsymbol{\alpha}_i$. A common form for the linear function $\mathbf{d}(\cdot)$ is $\mathbf{d}(h(\boldsymbol{\alpha}_i), \boldsymbol{\beta}, \mathbf{b}_i) = D_{ij}\boldsymbol{\beta} + V_i \mathbf{b}_i$, where the matrix D_{ij} depends on $\boldsymbol{\alpha}_i$ and other covariates such as time and V_i is a matrix of 0's and 1's. Note that, when $h(\boldsymbol{\alpha}_i) = z_{ij}^* \equiv \mathbf{u}_{ij}^T \boldsymbol{\alpha}_i$, we relate the parameters $\boldsymbol{\beta}_{ij}$ to the true but unobservable covariate values z_{ij}^*. This special case will be our focus in the real data analysis section.

A simple approach to the above problem, which is often called the *two-step* method, is to replace the unobserved $\boldsymbol{\alpha}_i$ in the response model (4) by their estimates from the covariate model (2), such as the individual empirical Bayes estimates. This method, however, may underestimate the standard errors for the response parameters since it fails to take the covariate estimation into account. Here we consider a simultaneous inference for models (2) and (4) based on a joint likelihood for the covariate and the response data. Thus our method not only incorporates the uncertainty from estimating the covariate process, but it also uses all data to estimate both the covariate and the response models. Our method can also accommodate censoring in the responses.

Let $f_X(\cdot)$ denote the density function for a random variable X and $F_X(\cdot)$ denote the corresponding cumulative density function (cdf). The observed value of X is denoted by x. Conditioning on the random effects $\boldsymbol{\alpha}_i$ and $\mathbf{b}_i$, a detectable measurement y_{ij} contributes $f_Y(y_{ij}|z_{ij}, \boldsymbol{\alpha}_i, \mathbf{b}_i)$ in the likelihood, whereas a non-detectable measurement contributes $F_Y(d\,|\,z_{ij}, \boldsymbol{\alpha}_i, \mathbf{b}_i) \equiv P(Y_{ij} < d\,|\,z_{ij}, \boldsymbol{\alpha}_i, \mathbf{b}_i)$ in the likelihood. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \delta^2, A, B)$ be the collection of all unknown parameters. The likelihood for the *observed data* $\{(\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i),\ i = 1, \ldots, n\}$ can be written as

$$
\begin{aligned}
L_o(\boldsymbol{\theta}) \;=\; \prod_{i=1}^{n} \Bigg[\iint & \left\{ \prod_{j=1}^{n_i} \left(f_Y(y_{ij}|z_{ij}, \boldsymbol{\alpha}_i, \mathbf{b}_i)\right)^{1-c_{ij}} \left(F_Y(d|z_{ij}, \boldsymbol{\alpha}_i, \mathbf{b}_i)\right)^{c_{ij}} \right\} \\
& \times \left\{ \prod_{k=1}^{m_i} f_Z(z_{ik}|\boldsymbol{\alpha}_i) \right\} f(\boldsymbol{\alpha}_i) f(\mathbf{b}_i) d\boldsymbol{\alpha}_i d\mathbf{b}_i \Bigg], \qquad (5)
\end{aligned}
$$

where $(Y_{ij}|z_{ij}, \boldsymbol{\alpha}_i, \mathbf{b}_i) \sim N(\mathbf{x}_{ij}^T \boldsymbol{\beta}_{ij}, \delta^2) \equiv N(\mathbf{x}_{ij}^T \mathbf{d}(h(\boldsymbol{\alpha}_i), \boldsymbol{\beta}, \mathbf{b}_i), \delta^2)$ and $(Z_{ik}|\boldsymbol{\alpha}_i) \sim N(\mathbf{u}_{ik}^T \boldsymbol{\alpha}_i, \sigma^2)$. In the next subsection, we estimate $\boldsymbol{\theta}$ by an EM algorithm (Dempster, *et al*, 1977).

2.2. Estimation by a Monte-Carlo EM Algorithm

Laird and Ware (1982) used the EM algorithm for estimation in standard linear mixed-effects models. In this section, we propose a Monte-Carlo EM algorithm to find the MLEs of the parameters in the covariate model and the response model simultaneously, incorporating censoring in the response.

Treating the non-detectable components in $\mathbf{y}_i$ and the unobservable random effects $\boldsymbol{\alpha}_i$ and $\mathbf{b}_i$ as missing data, we have *complete data* $(\mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i)$ for individual i. The complete-data log-likelihood can then be written as

$$
l_c(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n} l_c^{(i)}(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i)
$$

$$= \sum_{i=1}^{n} \left[logf_Y(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i) + logf_Z(\mathbf{z}_i|\boldsymbol{\alpha}_i) + logf(\boldsymbol{\alpha}_i) + logf(\mathbf{b}_i) \right]. \qquad (6)$$

We can write the E-step for the ith observation at the $(k+1)$st EM iteration as

$$Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = E\left\{ l_c^{(i)}(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i)|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)} \right\}.$$

To carry out the E-step, it is straightforward to show that we only need to compute the following conditional expectations of the sufficient statistics:

$$\begin{array}{lll} E(\mathbf{y}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}), & E(\boldsymbol{\alpha}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i, \boldsymbol{\theta}^{(k)}), & E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i, \boldsymbol{\theta}^{(k)}), \\ E(\mathbf{b}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i, \boldsymbol{\theta}^{(k)}), & E(\mathbf{b}_i\mathbf{b}_i^T|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i, \boldsymbol{\theta}^{(k)}). & \qquad (7) \end{array}$$

To compute these quantities, we note that

$$E(\boldsymbol{\alpha}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) = \int_{R_i} E(\boldsymbol{\alpha}_i|\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) f(\mathbf{y}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) d\mathbf{y}_i, \qquad (8)$$

$$E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) = \int_{R_i} E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T|\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) f(\mathbf{y}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) d\mathbf{y}_i, \qquad (9)$$

where R_i is the set of $\mathbf{y}_i$ consistent with the observed data $(\mathbf{q}_i, \mathbf{c}_i)$. Similar expressions hold for $E(\mathbf{b}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)})$ and $E(\mathbf{b}_i\mathbf{b}_i^T|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)})$.

For simple cases, the integrals in (8) and (9) may be evaluated by numerical quadrature, but this approach becomes intractable for high dimension integrals. Following Hughes (1999), here we consider a Monte Carlo approach. We use the Gibbs sampler (Gelfand and Smith, 1990) to generate $\mathbf{y}_i$ from $f(\mathbf{y}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)})$, and then approximate the integrals by empirical means. Specifically, suppose that $\mathbf{y}_i^{(l)}$, $l = 1, \cdots, M$, are generated from $f(\mathbf{y}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)})$, where M is the number of Monte Carlo samples. Then we have

$$E(\boldsymbol{\alpha}_i|\mathbf{q}_i, \mathbf{c}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}) \approx \frac{1}{M}\sum_{l=1}^{M} E(\boldsymbol{\alpha}_i|\mathbf{y}_i^{(l)}, \mathbf{z}_i; \boldsymbol{\theta}^{(k)}),$$

$$E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T|\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)}) \approx \frac{1}{M}\sum_{l=1}^{M} E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T|\mathbf{y}_i^{(l)},\mathbf{z}_i;\boldsymbol{\theta}^{(k)}). \tag{10}$$

Quantities $E(\mathbf{b}_i|\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)})$ and $E(\mathbf{b}_i\mathbf{b}_i^T|\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)})$ can be computed similarly. These computations involve evaluating the conditional expectations: $E(\boldsymbol{\alpha}_i|\mathbf{y}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)})$, $E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T|\mathbf{y}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)})$, $E(\mathbf{b}_i|\mathbf{y}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)})$, and $E(\mathbf{b}_i\mathbf{b}_i^T|\mathbf{y}_i,\mathbf{z}_i;\boldsymbol{\theta}^{(k)})$. To obtain these conditional expectations, note that $(\boldsymbol{\alpha}_i,\mathbf{y}_i,\mathbf{z}_i)$ and $(\mathbf{b}_i,\mathbf{y}_i,\mathbf{z}_i)$ follow multivariate normal distributions. Therefore the conditional expectations can be obtained using well-known results for multivariate normal distributions.

The M-step of the EM algorithm is like a complete-data maximization and is straightforward for linear mixed-effects models. To obtain the asymptotic covariance matrix of $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$, we may apply the formula of Louis (1982) at convergence. Let $\hat{\boldsymbol{\theta}}$ denote the estimate of $\boldsymbol{\theta}$ at convergence of EM. Let

$$\begin{aligned} S_c(\hat{\boldsymbol{\theta}}) &= \partial l_c(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \\ \mathcal{I}(\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\hat{\boldsymbol{\theta}}) &= E_{\boldsymbol{\theta}}\left(\partial^2 l_c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T \mid \mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i\right) \mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \end{aligned} \tag{11}$$

Then, the observed information matrix is given by

$$I(\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\hat{\boldsymbol{\theta}}) = \mathcal{I}(\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\hat{\boldsymbol{\theta}}) - E_{\boldsymbol{\theta}}\left(S_c(\hat{\boldsymbol{\theta}})S_c^T(\hat{\boldsymbol{\theta}}) \mid \mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i\right).$$

The estimate of the asymptotic covariance matrix of $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ is the upper $(p+r)\times(p+r)$ block of $I^{-1}(\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\hat{\boldsymbol{\theta}})$. It is straightforward to show that $I(\mathbf{q}_i,\mathbf{c}_i,\mathbf{z}_i;\hat{\boldsymbol{\theta}})$ can be computed using the conditional expectations in (7).

Model selections can be performed using the likelihood ratio test or AIC or BIC criteria since the observed data likelihoods for various models can be obtained at convergence. Predicted values for censored responses can be obtained based on the fitted model.

2.3. More Than One Covariates

When there are more than one covariates, a simple approach would be to fit each covariate process separately, and then incorporate the covariates information in the response model as in the previous section. The estimation procedure can be modified in a straightforward way. Sometimes, however, we may be able to model all or some covariates jointly using a multivariate mixed-effects model. This approach is described below.

Let z_{ikl} be the value of covariate l for individual i at time u_{ikl}, measured with errors. Let $\mathbf{w}_{ik}$ be the collection of other covariates measured without errors. Then, a multivariate linear mixed effects model for the covariates measured with errors can be written as

$$z_{ikl} = \mathbf{u}_{ikl}\boldsymbol{\alpha}_{il} + \epsilon_{ikl}, \quad i = 1, \cdots, n; \;\; k = 1, \cdots, m_{ik}; \;\; l = 1, \cdots, L, \tag{12}$$

where $\mathbf{u}_{ikl}$ are design vectors, $\boldsymbol{\alpha}_{il}$ are individual-specific parameters, and ϵ_{ikl} are error terms. We may assume that $(\boldsymbol{\alpha}_{i1}^T, \cdots, \boldsymbol{\alpha}_{iL}^T)^T$ follows a multivariate normal distribution with mean vector $\boldsymbol{\alpha}$ and covariance matrix A and are i.i.d across i, the errors ϵ_{ikl} are i.i.d across i and k with $\epsilon_{ikl} \sim N(0, \sigma_l^2)$, and ϵ_{ikl} are independent of the $\boldsymbol{\alpha}_{il}$. Note that in the above model we allow for the possibility that not all covariates are measured at all study visits since the observation times u_{ikl} are allowed to depend on individual i and covariate l. The second stage model for the response process can then be written as $\boldsymbol{\beta}_{ij} = \mathbf{d}(h(\boldsymbol{\alpha}_{i1}, \cdots, \boldsymbol{\alpha}_{iL}, \mathbf{w}_{ij}), \boldsymbol{\beta}, \mathbf{b}_i)$. The estimation procedure described in the previous sections can then be extended in a straightforward way.

3. EXAMPLE

We now analyze the ACTG 315 data based on the proposed method. For this dataset, as is evident from Figure 1, the inter-patient variations in viral load appear to be large and these variations appear to change over time. Previous studies suggest that the inter-patient variation in viral load may be partially explained by time-varying CD4 counts (Wu *et al.*, 1999). To model the CD4 process, we consider linear mixed effects model (2) with linear ($p = 2$), quadratic ($p = 3$), and cubic ($p = 4$) polynomials. The AIC (BIC) values corresponding to each model are 4102.8 (4126.1), 4012.5 (4070.9), and 4048.6 (4087.5) respectively. We thus adapt the quadratic polynomial as the model for the CD4 trajectory: $z_{ik} = \alpha_{i1} + \alpha_{i2} u_{ik} + \alpha_{i3} u_{ik}^2$. Standard diagnostic methods indicate that the quadratic polynomial CD4 model fits well.

For the initial stage of viral decay after treatment, a useful model is the following uni-exponential model (Ho, *et al*, 1995; Wei, *et al*, 1995)

$$V(t) = V(0)e^{-\lambda t}, \tag{13}$$

where $V(t)$ is the total virus at time t and λ is the rate of change in viral load. To model the entire viral load trajectory, one possible extension of model (13) is to allow λ to vary over time (Grossman, *et al*, 1999). A simple model for time-varying λ is $\lambda(t) = a + bt$. For HIV viral dynamic models, it is typical to take *log*-transformations of the raw viral loads in order to stabilize the variance and makes the raw data more normally distributed (Wu and Ding, 1999). After *log*-transformation of $V(t)$, together with the simple linear model for λ, we obtain the following quadratic linear mixed-effects model

$$y_{ij} = \beta_{i1} + \beta_{ij2} t_{ij} + \beta_{ij3} t_{ij}^2 + e_{ij}, \tag{14}$$

where $y_{ij} = log(V_i(t_{ij}))$. Since CD4 counts are measured with substantial errors, we assume that the individual-specific and time-varying parameters in (14) are related to the *true* CD4 values z_{ij}^* rather than the raw values z_{ij}. Standard model selection methods (AIC, BIC, and likelihood ratio test) favor the following second stage model

$$\beta_{i1} = \beta_1 + b_{i1}, \quad \beta_{ij2} = \beta_2 + \beta_3 z_{ij}^* + b_{i2}, \quad \beta_{ij3} = \beta_4 + \beta_5 z_{ij}^* + b_{i3}. \tag{15}$$

Usual model diagnostic techniques such as residual plots indicate that models (14) and (15) fit quite well.

We consider three methods for estimation: the proposed joint model method (JM), the two-step method (TS), and a "naive method" (NV) where the raw (observed) CD4 values rather than the true CD4 values are substituted in the second stage (15). Thus this naive method ignores measurement errors in CD4 values and treats the observed CD4 as the true values. Since very small/large estimates may be unstable, we standardized the CD4 values and re-scaled the original time t (in days) so that the new time scale is between 0 and 1. We assess the convergence of the Markov chains by examining time-series plots and sample autocorrelation functions, and find that the sampler typically burns in quite rapidly. The number of Monte Carlo samples to compute the expected sufficient statistics during the E-step is 500. The starting values for the EM algorithm were chosen to be the estimates from the two-step method.

The resulting estimates and standard errors are shown in Table 1. The joint model and the two-step method give similar estimates of the model parameters. However, the two-step method produces smaller standard errors, presumably because it fails to incorporate the uncertainty due to fitting the CD4 model. So the standard errors produced by the two-step method may be underestimated. For example, the estimated standard error for β_2 using the two-step method is about 55% smaller than that using the joint model. As will be demonstrated in the simulation study, the standard errors given by the joint model should be more reliable. Table 1 also indicates that the naive method, which ignores measurement errors, may produce severely biased estimates and may substantially underestimate the covariate effect. Thus it is important to take the measurement error into account. The analysis of ACTG 315 data suggests that CD4 may have a significant effect on the viral load trajectory after initiation of an antiretroviral treatment. Such effect may be severely underestimated if the measurement errors in CD4 are ignored and may be overestimated if the two-step method is used for estimation. The proposed joint model method may give most reliable results.

Finally, the fitted viral load model may provide prediction for non-detectable viral loads. Such prediction may be important in AIDS studies. For example, we may use the fitted viral load model to predict the "nadir value", the lowest level of viral load attained after initiation of therapy, and the "turning point", the time point when the viral load attains its minimum value, for each individual. These values may be associated with viral rebound based on previous studies. The nadir values and the turning points are typically difficult to estimate due to non-detectable viral load and sparseness of measurements. A well fitted viral load model may facilitate such estimation. The fitted viral load model may also be used to provide more accurate estimates of the viral rebound times, which may be used in survival analyses.

4. SIMULATION

We evaluate the three methods via a simulation study. The covariate model and the response model are chosen to be the same as that in the previous section. The "true" values of the model parameters are the same as those in Table 1. We generate random samples for the covariate model and the response model, and fit the respective models using the joint model method, the two-step method, and the naive method. We simulated 200 datasets and calculated the averages of the resulting estimates, together with the averages of their default standard errors (SE) and Monte-Carlo standard deviations (SD). We focus on estimation of the mean parameters. The number of burn-in steps in the Gibbs sampler and the number of Monte-Carlo samples in the E-steps are the same as those in the real data example.

The simulation results are presented in Table 2. Both the two-step method and the joint model method give approximately unbiased estimates for the model parameters. The naive method, however, may produce substantially biased estimates and tend to severely underestimate the covariate effects. The most noticeable difference between the two-step method and the joint model method is their estimated standard errors. For the response model, the joint model method produces larger standard errors which agree more closely with the Monte-Carlo standard deviations. Thus the standard errors produced by the joint model method may be more reliable, suggesting that more variability due to the estimation of the covariate process is incorporated in the joint model. The two-step method, on the other hand, fails to take covariate uncertainty into account and thus underestimate the standard errors. For the covariate model, notice that the joint model method produces smaller standard errors, indicating the increase in efficiency in using the joint model where response data were also used in estimating the covariate parameters. This is in contrast with the two-step method which only uses covariate data to estimate the parameters in the covariate model.

5. GENERALIZATION

Censored data due to a detection limit is a special case of non-ignorable missing data, i.e., the missing data may be related to the values being missing. The joint model approach described in the previous sections can be generalized to joint models for linear mixed-effects models with non-ignorable missing responses and mis-measured covariates. We briefly describe the approach as follows. Let $\mathbf{r}_i = (r_{i1}, \ldots, r_{in_i})^T$ be the vector of missing data indicators for individual i such that $r_{ij} = 1$ if y_{ij} is missing and $r_{ij} = 0$ if y_{ij} is observed, $j = 1, \ldots, n_i$. Let the conditional distribution of $\mathbf{r}_i$ given $(\mathbf{y}_i, \mathbf{z}_i)$ be $f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \phi)$, where we assume that $\mathbf{r}_i$ does not depend on the random effects $(\boldsymbol{\alpha}_i, \mathbf{b}_i)$ and ϕ are nuisance parameters. Denote all unknown parameters by $\boldsymbol{\eta} = (\boldsymbol{\theta}, \phi)$.

By treating the unobservable random effects $(\boldsymbol{\alpha}_i, \mathbf{b}_i)$ as additional missing data, we may write the "complete data" log-likelihood for individual i as follows

$$\begin{aligned} l(\boldsymbol{\eta};\ \mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \mathbf{r}_i) &= logf(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \boldsymbol{\theta}) + logf(\mathbf{z}_i|\boldsymbol{\alpha}_i, \boldsymbol{\theta}) + logf(\boldsymbol{\alpha}_i|\boldsymbol{\theta}) + \\ & \qquad logf(\mathbf{b}_i|\boldsymbol{\theta}) + logf(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \phi). \end{aligned} \tag{16}$$

Following Ibrahim, Chen, and Lipsitz (2001), we may then estimate the model parameters simultaneously using the Monte-Carlo EM algorithm. The E-step for individual i at the $(t+1)$st EM iteration can be written as

$$\begin{aligned} Q_i(\boldsymbol{\eta}|\boldsymbol{\eta}^{(t)}) &= E\left(l(\boldsymbol{\eta};\ \mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \mathbf{r}_i)|\mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)}\right) \\ &= \int\int\int \Big\{ logf(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \boldsymbol{\theta}) + logf(\mathbf{z}_i|\boldsymbol{\alpha}_i, \boldsymbol{\theta}) + logf(\boldsymbol{\alpha}_i|\boldsymbol{\theta}) + logf(\mathbf{b}_i|\boldsymbol{\theta}) \\ & \quad + logf(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \phi)\Big\} f(\mathbf{y}_{mis,i}, \boldsymbol{\alpha}_i, \mathbf{b}_i|\mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)}) d\mathbf{y}_{mis,i}\ d\boldsymbol{\alpha}_i\ d\mathbf{b}_i. \end{aligned} \tag{17}$$

The above integral generally does not have a closed form, but it can be approximated by the empirical mean based on samples generated from $f(\mathbf{y}_{mis,i}, \boldsymbol{\alpha}_i, \mathbf{b}_i|\mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)})$. Samples from $f(\mathbf{y}_{mis,i}, \boldsymbol{\alpha}_i, \mathbf{b}_i|\mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)})$ can be generated using the Gibbs sampler by sampling from the full conditionals $f(\mathbf{y}_{mis,i}|\mathbf{y}_{obs,i}, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)})$, $f(\boldsymbol{\alpha}_i|\mathbf{y}_{mis,i}, \mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)})$, and $f(\mathbf{b}_i|\mathbf{y}_{mis,i}, \mathbf{y}_{obs,i}, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)})$. Note that

$$f(\mathbf{y}_{mis,i}|\mathbf{y}_{obs,i}, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \boldsymbol{\eta}^{(t)}) f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\eta}^{(t)}), \tag{18}$$

$$f(\boldsymbol{\alpha}_i|\mathbf{y}_{mis,i}, \mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \boldsymbol{\eta}^{(t)}) f(\boldsymbol{\alpha}_i|\boldsymbol{\eta}^{(t)}), \tag{19}$$

$$f(\mathbf{b}_i|\mathbf{y}_{mis,i}, \mathbf{y}_{obs,i}, \mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\alpha}_i, \mathbf{b}_i, \boldsymbol{\eta}^{(t)}) f(\mathbf{b}_i|\boldsymbol{\eta}^{(t)}). \tag{20}$$

Thus, samples from the full conditionals can be generated using the adaptive rejection sampling of Gilks and Wild (1992). For the missing data model $f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \phi)$, we may choose any multinomial model such as a log-linear model. Ibrahim, *et al* (2001) recommend to model the joint distribution as a product of one-dimensional conditional distributions, which reduces the number of nuisance parameters. Note that for the censoring considered in the previous sections, we have $r_{ij} = I(y_{ij} < d)$. The M-step is then like a complete-data maximization and is therefore straightforward. Due to space limitation, the details are omitted here but are available from the author.

6. DISCUSSION

We have proposed a joint model approach for censored responses and covariates measured with errors in linear mixed-effects models. We model the covariate and the response processes simultaneously, and estimate the model parameters using a Monte-Carlo EM algorithm via the Gibbs sampler. The simulation study shows that the proposed joint model produces more reliable standard errors than the two-step method. The joint model also produces more accurate estimates for the covariate parameters since it uses all data for estimation. The naive method, which ignores measurement errors, may produce substantially biased estimates and may severely underestimate the true covariate effects.

The proposed method may be generalized to nonlinear models such as generalized linear mixed-effects models and nonlinear mixed-effects models. For nonlinear models, the E-step may involve integrations of nonlinear functions via Monte Carlo methods, while the M-step may use standard complete-data nonlinear model methods for maximization. For nonlinear models, however, the resulting Monte Carlo EM algorithm may be very slow. Moreover, there may be additional complication associated with common linearization procedures that may affect the estimated standard errors.

In the application, we consider a case where a linear mixed-effects model seems appropriate after the responses are transformed. This however may not always be feasible. In some cases, nonlinear mixed-effects models would be more appropriate. Note that common linearization procedures for nonlinear mixed-effects models often involve iteratively solving certain linear mixed-effects models. Thus inference for nonlinear models may incorporate techniques for linear models. Note also that, when sampling random effects in the Monte Carlo EM method, computational difficulties such as inefficiency or slow convergence may arise. For linear models, we may avoid sampling random effects by integrating them out in the E-step thus obtain a much more efficient algorithm. Such approach may also be incorporated into linearization procedures for nonlinear models. These problems are currently under investigation.

ACKNOWLEDGMENTS

I am grateful to Victor DeGruttola, Peter Gilbert, and Hulin Wu for very helpful comments and Kunling Wu for programming assistance. This research was supported by a Canada NSERC grant.

REFERENCES

Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*, London: Chapman and Hall.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood Estimation from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* **39**, 1-38.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.

Grossman, Z., Polis, M., Feinberg, *et al.* (1999). Ongoing HIV dissemination during HAART. *Nature Medicine* **5**, 1099-1103.

Higgins, K.M., Davidian, M., and Giltinan, D.M. (1997). A Two-Step Approach to Measurement Error in Time-Dependent Covariates in Nonlinear Mixed-Effects Models, With Application to IGF-I Pharmacokinetics. *Journal of the American Statistical Association* **92**, 436-448.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126.

Hughes, J.P. (1999). Mixed effects models with censored data with applications to HIV RNA levels. *Biometrics* **55**, 625-629.

Hughes, M.D. (2000). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. *Statistics in Medicine* **19**, 3171-3191.

Ibrahim, J.G., Chen, M.H., and Lipsitz, S.R. (2001). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable.

Biometrika **88**, 551-564.

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

Louis, T.A. (1982). Finding the observed information matrix when suing the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226-233.

Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., and Ho, D.D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586.

Pettitt, A.N. (1986). Censored observations, repeated measures and mixed effects models: an approach using the EM algorithm and normal errors. *Biometrika* **73**, 635-643.

Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-92.

Wang, C.Y., Wang, N., and Wang, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics* **56**, 487-495.

Wang, Y. and Taylor (2001). Jointly modeling longitudinal and event time data with application to Acquired Immunodeficiency Syndrome. *Journal of the American Statistical Association* **96**, 895-905.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., Saag, M.S., and Shaw, G.M. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117-122.

Wu, H. and Ding, A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410-418.

Wu, H., Kuritzkes, D.R., and McClernon, D.R. et al. (1999), Characterization of viral dynamics in human immunodeficiency virus type 1-infected patients treated with combination antiretroviral therapy: relationships to host factors, cellular restoration and virological endpoints, *Journal of Infectious Diseases* **179**, 799-807.

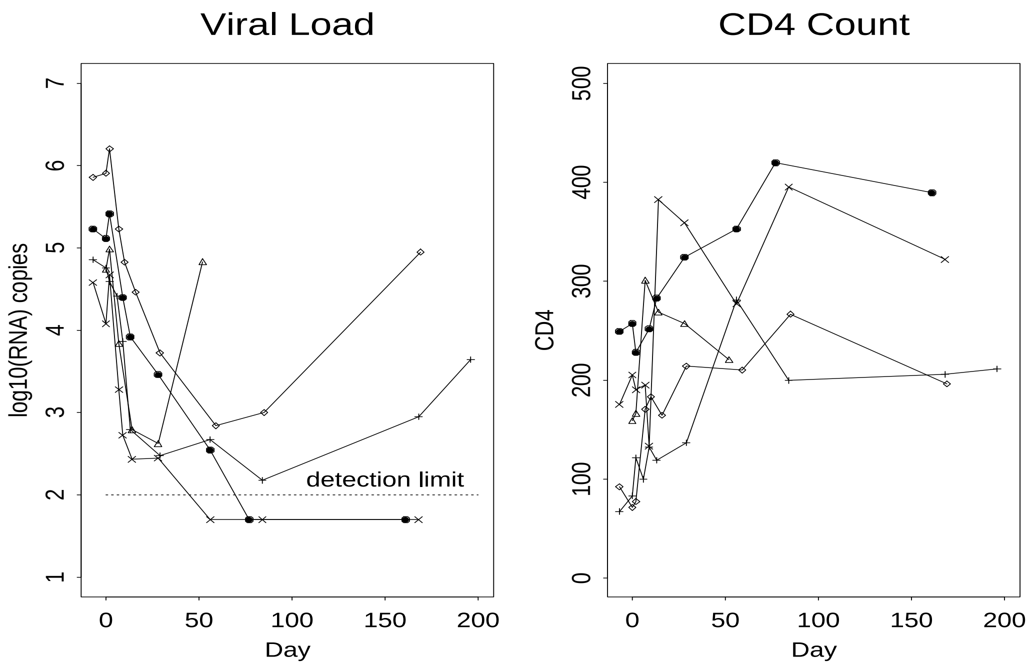

Figure 1: *Profiles of viral load and CD4 count for five randomly selected patients. The viral load detectable limit is log10(100) = 2. Viral loads below the detectable level are imputed by half the detection limit, which in* log_{10} *scale is* $log_{10}(50) = 1.69$.

Table 1. Estimation for the ACTG 315 data.

	Naive (NV)		Two-step (TS)		Joint Model (JM)	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
α_1	–	–	-0.42	0.11	-0.48	0.15
α_2	–	–	4.17	0.47	4.65	0.48
α_3	–	–	-3.74	0.49	-4.07	0.52
β_1	4.54	0.08	4.77	0.09	4.79	0.10
β_2	-9.50	0.54	-14.23	1.27	-14.38	2.32
β_3	0.07	0.42	8.14	0.80	8.05	0.88
β_4	9.97	0.76	14.50	1.33	14.54	2.23
β_5	-1.26	0.63	-9.27	0.95	-9.24	1.12

SE: standard error.

Estimates of Variance Components

$$\hat{\sigma}^2_{TS} = 0.27, \quad \hat{\sigma}^2_{JM} = 0.29; \qquad \hat{\delta}^2_{NV} = 0.30, \quad \hat{\delta}^2_{TS} = 0.18, \quad \hat{\delta}^2_{JM} = 0.19,$$

$$\hat{A}_{TS} = \begin{pmatrix} 0.52 & -0.14 & 0.09 \\ - & 4.05 & -2.83 \\ - & - & 1.98 \end{pmatrix}, \qquad \hat{A}_{JM} = \begin{pmatrix} 0.53 & -0.26 & 0.18 \\ - & 2.42 & -1.69 \\ - & - & 1.18 \end{pmatrix},$$

$$\hat{B}_{TS} = \begin{pmatrix} 0.32 & -1.09 & 0.99 \\ - & 58.39 & -56.49 \\ - & - & 58.41 \end{pmatrix}, \qquad \hat{B}_{JM} = \begin{pmatrix} 0.29 & -0.58 & 0.45 \\ - & 50.82 & -48.02 \\ - & - & 50.51 \end{pmatrix},$$

$$\hat{B}_{NV} = \begin{pmatrix} 0.23 & -0.20 & 0.16 \\ - & 3.76 & -3.86 \\ - & - & 6.37 \end{pmatrix}.$$

Table 2. Simulation results for comparing estimates.

		Naive			Two-Step			Joint Model		
Parameter	True Value	Estimate	SE	SD	Estimate	SE	SD	Estimate	SE	SD
α_1	-0.42	–	–	–	-0.40	0.11	0.11	-0.42	0.12	0.11
α_2	4.14	–	–	–	4.17	0.45	0.46	4.29	0.37	0.44
α_3	-3.72	–	–	–	-3.75	0.49	0.49	-3.83	0.28	0.48
β_1	4.80	4.57	0.09	0.09	4.72	0.09	0.10	4.72	0.10	0.10
β_2	-14.20	-8.76	1.02	1.00	-12.01	1.10	1.23	-12.05	1.28	1.23
β_3	8.10	2.87	0.47	0.56	8.15	0.77	1.09	8.11	0.92	1.08
β_4	14.50	9.29	1.12	1.14	12.42	1.18	1.32	12.43	1.38	1.32
β_5	-9.30	-3.64	0.66	0.80	-9.17	0.94	1.34	-9.14	1.15	1.33

SE is the mean of the estimated default standard errors. SD is the Monte Carlo standard deviation. 200 iterations.