Biostatistics (2008), **9**, 2, *pp.* 308–320
doi:10.1093/biostatistics/kxm029
Advance Access publication on August 29, 2007

Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data

LANG WU*

Department of Statistics, 333-6356 Agricultural road, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada
lang@stat.ubc.ca

X. JOAN HU

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

HULIN WU

Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA

Summary

In many longitudinal studies, the individual characteristics associated with the repeated measures may be possible covariates of the time to an event of interest, and thus, it is desirable to model the time-to-event process and the longitudinal process jointly. Statistical analyses may be further complicated in such studies with missing data such as informative dropouts. This article considers a nonlinear mixed-effects model for the longitudinal process and the Cox proportional hazards model for the time-to-event process. We provide a method for simultaneous likelihood inference on the 2 models and allow for nonignorable data missing. The approach is illustrated with a recent AIDS study by jointly modeling HIV viral dynamics and time to viral rebound.

Keywords: EM algorithm; longitudinal data; proportional hazards model; shared parameter model.

1. Introduction

In many longitudinal studies, we need to model a longitudinal process and a time-to-event process simultaneously. For example, in AIDS studies, we are often interested in modeling the HIV viral dynamics in the early period of an anti-HIV treatment, and in the meantime, we are also interested in the relationship between the individual-specific characteristics of the viral load process in the early period and a long-term antiviral response such as the time to viral rebound (or viral suppression or death). An important question is then to check if patients with a faster initial viral decay rate may have earlier viral rebound later in

*To whom correspondence should be addressed.

the study. Nonlinear mixed-effects (NLME) models are very useful in many longitudinal studies because these models are based on the underlying mechanisms which generate the data (Davidian and Giltinan, 1995). For time-to-event data, Cox proportional hazards model is often used. In these studies, missing data are common since subjects may drop out early for various reasons such as toxicities or side effects and data may be missing at scheduled times. The missing data may be nonignorable (or informative) in the sense that the missingness may be related to unobserved values. For example, the dropout or missing data process may be related to the initial (unobservable) true viral decay rates. Thus, analyses of longitudinal data often involve methods for missing data. In this article, we consider a "joint" likelihood method for an NLME model and a survival model, incorporating missing data in the time-varying responses and baseline covariates.

1.1 *The data set and models*

Our research is motivated from an AIDS study (Wu *and others*, 2004). The study consists of 115 subjects enrolled in an AIDS clinical trial. The viral load is repeatedly measured over time during an anti-HIV treatment. Some covariate measurements are also available. Here, we focus on the first 90-day data after start of the treatment because data after 3 months may be complicated by long-term clinical factors. The number of viral load measurements in the first 90 days varies from 3 to 10 (with a mean of 9 and a standard deviation of 1.4), and the typical measurement times are day 0, 1, 2, 7, 10, 14, and every 1 or 2 weeks (some measurements are made near the scheduled times). Twenty-four subjects drop out early and 3 subjects have missing CD4 values at baseline. Figure 1 shows viral load trajectories for 6 randomly selected patients from the study. We see that patients' viral loads after initiating antiviral treatment declined in the early period, and then, some patients' viral loads rebounded in the later period. This is probably because HIV virus was sensitive to the antiviral treatment in the initial period but developed drug resistance after the initial period. There is a substantial variation between patients, and some patients do not experience viral rebound during the study period. Some patients with faster initial viral decays appear to have earlier viral rebound. Ding and Wu (2001) show that the initial viral decay rate may reflect the efficacy/potency of the anti-HIV treatment. Therefore, it is important to study if the initial viral decay rate is predictive for time to viral rebound.

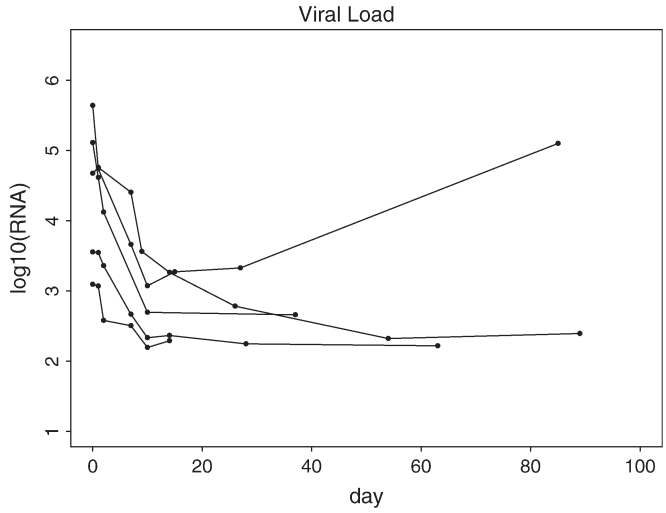

Fig. 1. Profiles of viral load (RNA) for 6 randomly selected patients.

Based on some biological arguments, Wu and Ding (1999) derived the following nonlinear exponential decay model with individual-specific parameters for HIV viral dynamics (see also Wu, 2002):

$$y_{ij} = \log_{10}(P_{1i}\, e^{-\lambda_{1i} t_{ij}} + P_{2i}\, e^{-\lambda_{2i} t_{ij}}) + e_{ij}, \tag{1.1}$$

$$\log(P_{1i}) = \beta_1 + b_{1i}, \quad \lambda_{1i} = \beta_2 + \beta_3 \mathrm{CD4}_i + b_{2i},$$

$$\log(P_{2i}) = \beta_4 + b_{3i}, \quad \lambda_{2i} = \beta_5 + b_{4i}, \quad i = 1, 2, \ldots, N, \ j = 1, 2, \ldots, n, \tag{1.2}$$

where y_{ij} is the $\log_{10}$-transformation of the viral load measurement for the ith patient at jth time point, λ_{1i} and λ_{2i} represent individual-specific first and second phases of viral decay rates, respectively, P_{1i} and P_{2i} are individual-specific baseline values, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_5)^T$ are population parameters (fixed effects), e_{ij} represent within-individual errors, b_{ki} are random effects, and N is the number of patients and n is the number of viral load measurements for patient i. In models (1.1) and (1.2), baseline CD4 values are introduced to partially explain the between-individual variation in the initial viral decay rate λ_{1i} (Wu, 2002) and the random effects $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})^T$ represent individual characteristics of the viral load trajectories. The exponential decay rates λ_{1i} and λ_{2i} can be interpreted as turnover rates of productively infected cells and long-lived and/or latently infected cells, respectively. We assume that $\lambda_{1i} > \lambda_{2i}$. Parameters P_{1i} and P_{2i} represent baseline viral loads (in $\log_{10}$-scale). Model (1.1) is a 2-compartment model which is a simplification from a multi-compartment model under some assumptions (see Wu and Ding, 1999, for details). We assume that e_{ij} i.i.d. $\sim N(0, R_i)$ and are independent of $\mathbf{b}_i \sim N(\mathbf{0}, D)$.

1.2 *Outline*

In this article, we consider a joint likelihood method for the viral dynamic model (1.1) and (1.2) and a proportional hazards model for time to viral rebound, incorporating informative dropouts and missing covariates, and estimate all model parameters simultaneously. We present our method in a more general framework—joint likelihood inference for a NLME model and a proportional hazards model at the presence of missing data—so that our method may be applied to other problems. The missing time-varying responses in the NLME model are allowed to be nonignorable for dropout patients. For covariates, we focus on missing baseline covariates with an ignorable missing mechanism (or missing at random) and ignorable measurement errors in the covariates. We extend our method to missing time-varying covariates and mismeasured covariates in Section 5, with technical details outlined in Appendix B. The random effects in the NLME model, which represent individual-specific characteristics of the longitudinal process, are used as possible error-free "covariates" for the proportional hazards model and the missing response model. A Monte Carlo expectation maximization (EM) algorithm is used for likelihood estimation.

Joint modeling of longitudinal data and survival data has been studied in the literature (e.g. DeGruttola and Tu, 1994; Wulfsohn and Tsiatis, 1997; Henderson *and others*, 2002; Guo and Carlin, 2004). Tsiatis and Davidian (2004) provide a very nice review. These methods often consider linear (mixed) models for the longitudinal process and focus on complete data or ignorably missing data cases. Here, we consider NLME models for the longitudinal process and incorporate nonignorably missing data (or informative dropouts).

In Section 2, we describe the models for longitudinal data and time-to-event data, as well as the models for the missing data mechanism. In Section 3, we describe the Monte Carlo EM algorithm for inference, with computational details given in Appendix A. A real-data example is presented in Section 4. In Section 5, we conclude the article and discuss an extension of our method to time-dependent covariates with measurement errors, with technical details given in Appendix B.

2. NOTATION AND MODELS

2.1 *Notation*

Suppose that there are N individuals. Let y_{ij} be the response value for individual i at time t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n$, and let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in})^T$. We assume that the measurement schedules are fixed and common to all individuals. We consider a response missing if its value is not observed at or near the scheduled time point. Let $\mathbf{z}_i$ be the collection of time-independent (baseline) covariates for individual i. We write $\mathbf{y}_i = (\mathbf{y}_{i,\text{mis}}, \mathbf{y}_{i,\text{obs}})$, where $\mathbf{y}_{i,\text{mis}}$ is the collection of missing responses for dropout patients and $\mathbf{y}_{i,\text{obs}}$ is the collection of observed responses, and similarly, we write $\mathbf{z}_i = (\mathbf{z}_{i,\text{mis}}, \mathbf{z}_{i,\text{obs}})$. Let $\mathbf{s}_i = (s_{i1}, \ldots, s_{in})^T$ be a vector of missing response indicators such that $s_{ij} = 1$ if y_{ij} is missing and 0 otherwise. We allow for arbitrary patterns of missing data, that is, we do not restrict to monotone missing data patterns, which assume that once a subject leaves the study, return is not possible. Let $\mathbf{r}_i = (r_{i1}, \ldots, r_{im})^T$ be the vector of an "event" indicator with individual i: $r_{ij} = 1$ or 0 if the event has happened or not by time t_{ij}. We assume that $r_{i1} = 0$ for all i. If patient i drops out between time t_{ik} and $t_{i,k+1}$ and does not return to study later, then $s_{i1} = \cdots = s_{ik} = 0$, $s_{ij} = 1$ for $j > k$.

For individual i, let T_i be the time to an event (or the duration time until an event occurs) and assume $P(T_i < \infty) = 1$. In the AIDS example, T_i is the time to the first viral rebound in the study, which is usually not observable but confirmed in practice by an observed rise of viral load after the initial decay, say, 2 consecutive increases in viral load. Specifically, if there is $k \leqslant m$ such that k is the smallest number with $y_{i,k-1} < y_{ik}$ and $y_{i,k-1} < y_{i,k+1}$, in practice we take $r_{i1} = \cdots = r_{i,k-1} = 0$ and $r_{ik} = \cdots = r_{im} = 1$, that is, $t_{i,k-1} < T_i \leqslant t_{ik}$ (the rebound takes place during the time period $(t_{i,k-1}, t_{ik}]$). If there is no such a number k, we view $r_{ij} = 0$ for $j = 1, \ldots, m$ and thus $T_i > t_{im}$. This type of event time data structure is referred to as interval-censored event times (see, e.g. Lawless, 2003).

2.2 *Models for longitudinal and time-to-event data*

For the longitudinal process, we consider the following general NLME model (Davidian and Giltinan, 1995):

$$y_{ij} = g(t_{ij}, \boldsymbol{\beta}_i) + e_{ij}, \quad \mathbf{e}_i \sim N(\mathbf{0}, R_i), \tag{2.1}$$

$$\boldsymbol{\beta}_i = \mathbf{h}(\mathbf{z}_i, \boldsymbol{\beta}) + B_i \mathbf{b}_i, \quad \mathbf{b}_i \text{ i.i.d.} \sim N(\mathbf{0}, D), \quad j = 1, \ldots, n, \ i = 1, \ldots, N, \tag{2.2}$$

where $g(\cdot)$ is a nonlinear function, $\mathbf{e}_i = (e_{i1}, \ldots, e_{in})^T$ are measurement errors, $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{is})^T$ is a vector of individual-specific regression parameters, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^T$ is a vector of population parameters, $\mathbf{h}(\cdot)$ is a s-dimensional vector-valued function, B_i is an incidence matrix of 0's and 1's, $\mathbf{b}_i = (b_{i1}, \ldots, b_{is})^T$ is a vector of random effects and is independent of $\mathbf{e}_i$, R_i is the unknown within-individual covariance matrix which contains distinct parameters $\boldsymbol{\sigma}$, and D is an unstructured covariance matrix. If there are no missing data, the probability density for $\mathbf{y}_i$ can be written as

$$f(\mathbf{y}_i | \mathbf{z}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}, D) = \int f(\mathbf{y}_i | \mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}) f(\mathbf{b}_i | D) \mathrm{d}\mathbf{b}_i. \tag{2.3}$$

For the time-to-event process, we assume that the distribution of T_i depends on the random effects $\mathbf{b}_i$ which represent individual-specific longitudinal processes. For example, in AIDS studies, patients with faster (or slower) initial viral decays may be more likely to have an earlier viral rebound, so the time to viral rebound T_i depends on the random effect associated with initial viral decays. We therefore consider a frailty model for T_i which is linked to the NLME model (2.1)–(2.2) through the random effects $\mathbf{b}_i$. Specifically, we assume that the conditional hazard rate of T_i at time t_i is

$$\lambda(t_i | \mathbf{z}_i, \mathbf{b}_i) = \lambda_0(t_i) \exp(\boldsymbol{\gamma}_1^T \mathbf{z}_i + \boldsymbol{\gamma}_2^T \mathbf{b}_i), \tag{2.4}$$

where $\lambda_0(t_i)$ is an unspecified baseline hazard function and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are unknown parameters linking baseline covariates $\mathbf{z}_i$ and random effects $\mathbf{b}_i$ to the conditional hazard rate, respectively. For the above survival model, we assume that all individuals have the same set of measurement schedules, although for the longitudinal model we allow the response measurement schedules to be somewhat different across individuals or ignorable missing responses for subjects who do not drop out.

Let

$$\begin{aligned} p_{ik} &= P(r_{ik} = 1 | r_{il} = 0, 0 \leqslant l < k; \mathbf{z}_i, \mathbf{b}_i) \\ &= 1 - P(T_i \geqslant t_{ik} | T_i \geqslant t_{i,k-1}; \mathbf{z}_i, \mathbf{b}_i), \quad k = 1, 2, \ldots, m. \end{aligned} \tag{2.5}$$

Then, we have

$$p_{ik} = 1 - \exp[-\exp(\gamma_{0k} + \boldsymbol{\gamma}_1^T \mathbf{z}_i + \boldsymbol{\gamma}_2^T \mathbf{b}_i)] \tag{2.6}$$

or $\log(-\log(1 - p_{ik})) = \gamma_{0k} + \boldsymbol{\gamma}_1^T \mathbf{z}_i + \boldsymbol{\gamma}_2^T \mathbf{b}_i$, where $\gamma_{0k} = \log \int_{t_{i,k-1}}^{t_{ik}} \lambda_0(u) \mathrm{d}u$, $k = 1, \ldots, m$. Given the current observation mechanism, we need only to deal with the finite number of parameters γ_{0k} instead of the unknown function $\lambda_0(t)$ in the likelihood estimation. Denote the vector that its components are all the distinct $\gamma_{01}, \ldots, \gamma_{0m}$ for $i = 1, \ldots, N$ by $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. Note that

$$f(\mathbf{r}_i | \mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}) = \prod_{k=1}^{m} f(r_{ik} | r_{il}, 0 \leqslant l < k; \mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}), \tag{2.7}$$

where

$$f(r_{ik} | r_{il} = 0, 0 \leqslant l < k; \mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}) = p_{ik}^{r_{ik}} (1 - p_{ik})^{1 - r_{ik}},$$

and r_{ik} equals 0 before an event and 1 after an event. With $l_i = \max(t_{ij}: r_{ij} = 0)$ and $u_i = \min(t_{il}: r_{il} = 1)$, (2.7) can be written as

$$\begin{aligned} &P(l_i < T_i \leqslant u_i | \mathbf{z}_i, \mathbf{b}_i) \\ &\quad = \exp\left(-\int_0^{l_i} \lambda_0(t) \mathrm{d}t \exp(\boldsymbol{\gamma}_1 \mathbf{z}_i + \boldsymbol{\gamma}_2 \mathbf{b}_i)\right) \left\{1 - \exp\left(-\int_{l_i}^{u_i} \lambda_0(t) \mathrm{d}t \exp(\boldsymbol{\gamma}_1 \mathbf{z}_i + \boldsymbol{\gamma}_2 \mathbf{b}_i)\right)\right\}, \end{aligned}$$

which may reduce some computing and simplify the presentation. Here, $u_i = \infty$ if $r_{il} = 0$ for $l = 1, \ldots, m$.

2.3 *Missing data model*

When there are informative dropouts (or nonignorable missing longitudinal responses), the missing data mechanism must be taken into account for valid likelihood inference, but the missing data mechanism can be ignored in likelihood inference if the missing data are ignorable in the sense of missing at random or missing completely at random (Little, 1995). As noted in Section 1, we assume a missing response model which allows the missing probability to possibly depend on the unobservable random effects $\mathbf{b}_i$. Such a missing data model is related to the shared parameter models or random effect–based dropouts (Wu and Carroll, 1988; DeGruttola and Tu, 1994; Little, 1995; Follmann and Wu, 1995; Ten Have *and others*, 1998). In other words, the missingness depends on both $\mathbf{y}_{\mathrm{mis},i}$ and $\mathbf{y}_{\mathrm{obs},i}$ through the random effects $\mathbf{b}_i$. For such missing responses, a model specifying the missing response mechanism must be incorporated in likelihood inference. The probability of missing responses at the time t_{ij} may also depend on the missing status at the previous time point $t_{i,j-1}$. We assume that the missing baseline covariates are missing at

random (or ignorable) in the sense that the missingness may be related to the observed data but not the missing values, so we do not need to specify a missing covariate mechanism. However, we need to make a distributional assumption for the incompletely observed covariates for likelihood inference. No distributional assumption is needed for covariates without missing data.

Based on the above arguments, as an example, we may consider the following model for the missing responses:

$$\text{logit}(P(s_{ij} = 1|s_{i,j-1}, \mathbf{b}_i, \boldsymbol{\phi})) = \phi_0 + \phi_1 s_{i,j-1} + \boldsymbol{\phi}_2^T \mathbf{b}_i, \tag{2.8}$$

$$f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}) = f(s_{i1}|\mathbf{b}_i, \boldsymbol{\phi}) \prod_{j=2}^{n} f(s_{ij}|s_{i,j-1}, \mathbf{b}_i, \boldsymbol{\phi}), \tag{2.9}$$

where the parameters $\boldsymbol{\phi}$ may be viewed as nuisance parameters and are usually not of inferential interest. In the above missing data model, the probability of missing data (or dropout) is related to the individual characteristics (i.e. random effects $\mathbf{b}_i$) of the longitudinal process, which appears to be reasonable for AIDS studies as noted earlier. The probability of missing data may also depend on covariates $\mathbf{z}_i$, but we should avoid to build a too complicated missing data model if $\mathbf{z}_i$ are not highly predictive of the missingness.

3. Inference based on joint likelihood

3.1 *The joint likelihood*

In this section, we consider simultaneous likelihood inference for all parameters based on the joint likelihood of the observed data $\{(\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{r}_i, \mathbf{s}_i), i = 1, 2, \ldots, N\}$. Let $f(\cdot)$ denote a generic density function and $f(y|x)$ denote the conditional distribution of y given x. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\phi}, D)$ denote the collection of all unknown parameters. We assume that $\mathbf{y}_i$ and $\mathbf{r}_i$ are conditionally independent given the random effects $\mathbf{b}_i$, that is, $\mathbf{r}_i$ depends on $\mathbf{y}_i$ through the random effects $\mathbf{b}_i$. In AIDS studies, this implies that the time to viral rebound depends on viral load trajectories through individual-specific (error-free) initial viral levels and viral decay rates. Based on the motivation discussed in Section 2.3, we also assume that $f(\mathbf{s}_i|\mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\phi}) = f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi})$. Thus, we have

$$f(\mathbf{y}_i, \mathbf{r}_i, \mathbf{s}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}) f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}) f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}).$$

The joint likelihood for the "observed" data can then be written as

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left[\int\int\int f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}) f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}) f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}) f(\mathbf{z}_i|\boldsymbol{\alpha}) f(\mathbf{b}_i|D) \text{d}\mathbf{y}_{i,\text{mis}} \, \text{d}\mathbf{z}_{i,\text{mis}} \, \text{d}\mathbf{b}_i \right],$$

where $f(\mathbf{z}_i|\boldsymbol{\alpha})$ is the assumed distribution for the incompletely observed covariates $\mathbf{z}_i$ with unknown parameters $\boldsymbol{\alpha}$. Maximum likelihood estimates (MLEs) of all parameters $\boldsymbol{\theta}$ can be obtained by maximizing the observed data likelihood $L_o(\boldsymbol{\theta})$. However, the observed data likelihood $L_o(\boldsymbol{\theta})$ may be difficult to evaluate because it involves an intractable and high-dimensional integral. In the following, we use a Monte Carlo EM algorithm to obtain the MLEs.

3.2 *A Monte Carlo EM algorithm*

The EM algorithm is a widely used method for finding MLEs in the presence of missing data. It iterates between an E-step and an M-step: the E-step computes the conditional expectation of the complete data log-likelihood given the observed data, and the M-step gives updated parameter estimates by maximizing

the conditional expectation in the E-step. Iterating between an E-step and an M-step until convergence leads to MLEs. When the conditional expectation in the E-step is difficult to evaluate analytically, Monte Carlo approximations may be used, which leads to a Monte Carlo EM algorithm (Wei and Tanner, 1990). For the current problem, if we treat the unobservable random effects $\mathbf{b}_i$ as additional "missing data," we can write the "complete data" as $\{(\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \mathbf{s}_i, \mathbf{b}_i), i = 1, 2, \ldots, N\}$. Thus, the complete data log-likelihood for individual i can be written as

$$l_c^{(i)}(\boldsymbol{\theta}) = \log f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}) + \log f(\mathbf{z}_i|\boldsymbol{\alpha}) + \log f(\mathbf{b}_i|D) + \log f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}) + \log f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}).$$

The E-step at the tth iteration of the EM algorithm for individual i can then be written as

$$Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \iiint \{\log f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}) + \log f(\mathbf{z}_i|\boldsymbol{\alpha}) + \log f(\mathbf{b}_i|D) + \log f(\mathbf{r}_i|\mathbf{b}_i, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}) + \log f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi})\} f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)}) \mathrm{d}\mathbf{y}_{i,\text{mis}}\, \mathrm{d}\mathbf{z}_{i,\text{mis}}\, \mathrm{d}\mathbf{b}_i.$$

Since it is difficult to evaluate the integral $Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ analytically, we approximate the integral by Monte Carlo methods as follows.

Since $Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ is a (conditional) expectation with respect to the density $f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\, \mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$, we may approximate Q_i by its empirical mean obtained by simulating many samples from the conditional density $f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$ and then replacing the expectation by an empirical mean. To generate random samples from the conditional density $f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$, we may use the Gibbs sampler method (Gelfand and Smith, 1990) along with rejection sampling methods. See Appendix A for details.

Suppose that $\{(\tilde{\mathbf{y}}_{i,\text{mis}}^{(1)}, \tilde{\mathbf{z}}_{i,\text{mis}}^{(1)}, \tilde{\mathbf{b}}_i^{(1)}), \ldots, (\tilde{\mathbf{y}}_{i,\text{mis}}^{(m_t)}, \tilde{\mathbf{z}}_{i,\text{mis}}^{(m_t)}, \tilde{\mathbf{b}}_i^{(m_t)})\}$ is a random sample of size m_t generated from $f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$. The E-step of the Monte Carlo EM algorithm at the $(t+1)$th iteration can be approximated as follows:

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N} Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \approx \sum_{i=1}^{N} \Bigg\{ \frac{1}{m_t} \sum_{j=1}^{m_t} [&\log f(\mathbf{y}_{i,\text{obs}}, \tilde{\mathbf{y}}_{i,\text{mis}}^{(j)}|\mathbf{z}_{i,\text{obs}}, \tilde{\mathbf{z}}_{i,\text{mis}}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) \\ &+ \log f(\mathbf{z}_{i,\text{obs}}, \tilde{\mathbf{z}}_{i,\text{mis}}^{(j)}|\boldsymbol{\alpha}) + \log f(\tilde{\mathbf{b}}_i^{(j)}|D) \\ &+ \log f(\mathbf{r}_i|\mathbf{z}_{i,\text{obs}}, \tilde{\mathbf{z}}_{i,\text{mis}}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \boldsymbol{\gamma}) + \log f(\mathbf{s}_i|\tilde{\mathbf{b}}_i^{(j)}, \boldsymbol{\phi})] \Bigg\}. \end{aligned} \tag{3.1}$$

The above approximation can be made arbitrary accurate by increasing m_t. The M-step of the Monte Carlo EM algorithm is then to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, which is just like a complete data maximization, so standard optimization procedures for complete data models such as the Newton–Raphson method can be used to obtain the updated parameters $\boldsymbol{\theta}^{(t+1)}$. If we assume that the parameters in each term of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ are distinct, we can maximize each term of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ separately using standard methods for linear, nonlinear, and logistic regression models. The convergence of the Monte Carlo EM algorithm is discussed in Appendix A.

The variance–covariance matrix of $\boldsymbol{\theta}$ can be approximated as follows: At the convergence of EM, let $S_{ij}(\hat{\boldsymbol{\theta}}) = \partial l(\boldsymbol{\theta}|\mathbf{y}_{\text{obs},i}, \tilde{\mathbf{y}}_{\text{mis},i}^{(j)}, \mathbf{z}_{\text{obs},i}, \tilde{\mathbf{z}}_{\text{mis},i}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \mathbf{r}_i, \mathbf{s}_i)/\partial\boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. We have

$$I(\hat{\boldsymbol{\theta}}) \approx \sum_{i=1}^{N} \sum_{j=1}^{m_t} \frac{1}{m_t} S_{ij}(\hat{\boldsymbol{\theta}}) S_{ij}^T(\hat{\boldsymbol{\theta}}).$$

The approximate asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is $I^{-1}(\hat{\boldsymbol{\theta}})$.

4. Data analysis and simulation

4.1 *Data analysis*

In this section, we analyze the data set described in Section 1 based on the proposed method. In the analysis, we assume that $R_i = \sigma^2 I$. One of our objectives is to test if the time to rebound T_i depends on baseline CD4 values and the random effects b_{i1}, b_{i2}, b_{i3}, so we consider the following model for T_i:

$$\lambda(t_i|\mathbf{z}_i, \mathbf{b}_i) = \lambda_0(t_i)\exp(\gamma_1 z_{i1} + \gamma_2 b_{i1} + \gamma_3 b_{i2} + \gamma_4 b_{i3}), \tag{4.1}$$

where $z_{i1} = \text{CD4}_i$. Since baseline CD4 contains some missing data, we need to make a distributional assumption for baseline CD4 values and assume $z_i \sim N(\alpha_1, \alpha_2)$. We also assume that the missing CD4 values are ignorable (i.e. missing at random). This assumption should not be too restrictive here since the missing rate in CD4 is low. There are also some dropouts, which lead to missing values in $\mathbf{y}_i$, and the missingness may be informative as discussed in Section 1. Note that if data are collected close to the scheduled time points but not exactly at the scheduled time points, the data at the scheduled time points are not viewed as missing data here. We assume the following simple model for the missing response mechanism:

$$f(\mathbf{s}_i|\mathbf{y}_i, \mathbf{b}_i, \mathbf{z}_i, \boldsymbol{\phi}) = \prod_{j=1}^{n} P(s_{ij} = 1|\boldsymbol{\phi}, \mathbf{b}_i)^{s_{ij}}(1 - P(s_{ij} = 1|\boldsymbol{\phi}, \mathbf{b}_i))^{1-s_{ij}}, \tag{4.2}$$

$$\log\left(\frac{P(s_{ij} = 1|\boldsymbol{\phi}, \mathbf{b}_i)}{1 - P(s_{ij} = 1|\boldsymbol{\phi}, \mathbf{b}_i)}\right) = \phi_0 + \phi_1 b_{1i} + \phi_2 b_{2i} + \phi_3 b_{3i} + \phi_4 b_{4i}, \quad i = 1, 2, \ldots, N, \tag{4.3}$$

where the missingness probability of the responses may depend on the random effects which characterize individual differences of the viral load trajectories. Although we can assume more complicated models for the missing response mechanism and include other covariates in (4.3), we should avoid building a too complicated missing response model since large number of nuisance parameters may lead to poor precision if these covariates are not the main focus and are not highly significant.

We use the likelihood method described in Section 3 to obtain the parameter estimates. The starting values for the EM algorithm are obtained by fitting the models separately and ignoring all missing data. The convergence of the Gibbs sampler is based on visual inspection of autocorrelation plots. We used 400 burn-ins for the Gibbs sampling. The EM is considered converged if the maximum difference in consecutive parameter estimates is less than 1%. The number of Monte Carlo samples m_t is increased at each iteration until convergence.

Table 1 shows the resulting parameter estimates and the associated standard errors. The p-values are computed based on Wald-type tests. The estimate of β_3 indicates that higher initial CD4 values are associated with faster initial viral decay (p-value $= 0.001$). The estimate of β_5 (negative sign, not significant) reflects the viral rebound at later stage. We see that the missing responses (or dropouts) depend on initial viral decay rate ϕ_2 (p-value $= 0.001$) and initial viral load level ϕ_3 (p-value $= 0.008$). Time to viral rebound depends on initial CD4 values γ_1 (p-value < 0.001) so that smaller baseline CD4 values are associated with earlier viral rebound. However, the viral rebound time does not appear to be significantly associated with initial viral decay rate. Although the viral decay rates may reflect the efficacy of the treatment (Ding and Wu, 2001), the effect of viral decay rates on the viral rebound is not big enough. This is probably because many other factors also contribute to the viral rebound. These results may be useful for future AIDS studies.

Table 1. *Estimates of model parameters*

Model	Estimates of the NLME model parameters in (1.1)									
Parameter	β_1		β_2		β_3		β_4		β_5	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Estimate	10.32	0.15	53.98	9.17	14.01	4.37	6.02	0.22	–0.55	0.34
p-value	<0.001		<0.001		0.001		<0.001		0.10	
Model	Estimates of the dropout model parameters in (4.3)									
Parameter	ϕ_0		ϕ_1		ϕ_2		ϕ_3		ϕ_4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Estimate	3.00	0.006	0.017	0.019	–0.057	0.018	–0.056	0.021	–0.056	0.061
p-value	<0.001		0.37		0.001		0.008		0.36	
Model	Estimates of the survival model parameters in (4.1)									
Parameter	γ_1		γ_2		γ_3		γ_4			
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE		
Estimate	–1.47	0.06	0.041	0.046	–0.012	4.20	0.001	3.81		
p-value	<0.001		0.37		0.99		0.99			

Estimates of the diagonal elements of matrix D is (0.103, 0.104, 0.083, 0.010). Estimates of the off-diagonal elements are all less than 0.004. SE, standard error.

4.2 *Simulation*

We conducted a simulation study to evaluate the proposed method and compare the method with a naive method where missing data are ignored. All the models used in the simulation are the same as those in the example. The true values of the model parameters are $\boldsymbol{\beta} = (12, 40, 8, 8, 2)$, $\boldsymbol{\alpha} = (0, 1)$, $\boldsymbol{\phi} = (3, 0.04, 0.08, -0.01, -0.1)$, $\boldsymbol{\gamma} = (-1.5, 0.03, 0.01, 0.02)$, $\sigma = 0.2$, $D = \text{diag}(1, 0.5, 1, 0.01)$, $t = (2, 4, 6, 8, 10, 12, 15, 19, 23, 28)/28$, and $N = 100$. We generated roughly 20% missing values and rebounds in the responses based on the assumed models. The covariate is assumed to be completely observed. The simulation was repeated 200 times, and the resulting estimates are averaged.

For each estimate, we computed percent bias and percent mean square error (MSE) as follows (say, for β_j): percent bias = $100 \times (\hat{\beta}_j - \beta_j)/\beta_j$ and percent MSE = $100 \times \sqrt{\text{bias}^2 + \text{variance}}$, where $\hat{\beta}_j$ is the average of all estimates for β_j from simulation. The biases and MSEs are then averaged over all simulations. Table 2 shows the simulation results. We see that the proposed method (joint model) performs well in terms of both bias and MSE, and it is better than the naive method. The proposed method gives approximately unbiased estimates and reasonable MSEs, so the method is feasible and reasonable. The purpose of this small simulation study is to preliminary check the feasibility and performance of the proposed method. Due to space limitation, a more comprehensive and thorough simulation study will be reported separately.

5. DISCUSSION

An alternative approach for the problem discussed in this article, if there are no missing data, is the so-called 2-step method: in the first step, we estimate the parameters and random effects based on the NLME model alone, and then in the second step, we substitute the estimates from the first step into the survival model to obtain estimates of parameters in the survival model. It is well known in the joint model literature (e.g. Tsiatis and Davidian, 2004), however, that such a 2-step method ignores the variability in estimation of the parameters in the first step, so may lead to underestimation of the variability of

Table 2. *Simulation results*

Method	Estimates of the parameters in model (1.1)									
	$\beta_1 = 12$		$\beta_2 = 40$		$\beta_3 = 8$		$\beta_4 = 8$		$\beta_5 = 2$	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Joint model	11.75	0.30	35.62	4.35	6.84	1.46	7.97	0.12	1.97	0.12
Naive method	11.87	0.35	33.74	3.54	6.91	1.99	7.95	0.36	1.95	0.33
Method	Percent biases and percent MSEs									
	$\beta_1 = 12$		$\beta_2 = 40$		$\beta_3 = 8$		$\beta_4 = 8$		$\beta_5 = 2$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Joint model	−2	3	−11	15	−14	23	0	2	−2	7
Naive method	−1	3	−16	18	−14	28	−1	5	−3	17

SE, standard error.

parameter estimates in the second step. The joint model method described in this article, on the other hand, incorporates all variability, so it should lead to reliable estimates and standard errors, as demonstrated in other cases in the literature (Tsiatis and Davidian, 2004). Moreover, the proposed method incorporates both missing covariates and missing responses. Note that the missing data models cannot be tested based on the observed data, so the choices of the missing data models should be based on practical reasonability. Normally, we should try different missing data models to see if the results for the main model parameters are robust against the missing data models.

The likelihood method based on a Monte Carlo EM algorithm as described in the article can be computationally very intensive. To reduce computational burden, we may consider approximation methods which are based on Laplace approximations or Taylor expansions about the random effects to linearize the nonlinear models (so sampling the random effects in the E-step may be avoided). Detailed implementation of these approximate methods are under investigation.

The method presented in this article can be extended to time-dependent covariates where the covariates may be measured with errors or may be missing. In practice, some covariates may be measured with substantial errors, and the time-varying covariates may also be missing due to different measurement schedules from the response measurements or other problems. For example, in AIDS studies, CD4 count is often measured with substantial errors and may have measurement schedules different from the viral load measurement schedules. To address covariate measurement errors or missing data, we may model the time-dependent covariates empirically using multivariate linear mixed-effects (LME) models. Then, a joint likelihood method can be developed in a similar way. An outline of the approach is presented in Appendix B.

Acknowledgments

The authors are grateful to the associate editor and 3 referees for helpful comments and Shijun Song for programming assistance. *Conflict of Interest:* None declared.

Funding

Canada Natural Sciences and Engineering Research Council (RGPIN238677) to L.W.; United States National Institutes of Health (R01-AI56695) and Canada Natural Sciences and Engineering Research Council (RGPIN262823) to X.J.H.; United States National Institutes of Health (AI052765, AI055290) to H.W.

APPENDIX A: COMPUTATIONAL DETAILS, SAMPLING METHODS, AND CONVERGENCE

For the Monte Carlo EM algorithm in Section 3, to generate random samples from the conditional density $f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$, we may use the Gibbs sampler method by iteratively sampling from the full conditionals $f(\mathbf{y}_{i,\text{mis}}|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$, $f(\mathbf{z}_{i,\text{mis}}|\mathbf{z}_{i,\text{obs}}, \mathbf{y}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$, and $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$ in turn until the resulting Markov chain converges. To sample these full conditionals, note that

$$f(\mathbf{y}_{i,\text{mis}}|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_i, \mathbf{b}_i, \mathbf{r}_i, \mathbf{s}_i, \boldsymbol{\theta}^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}^{(t)}, \boldsymbol{\sigma}^{(t)}), \tag{A.1}$$

$$f(\mathbf{z}_{i,\text{mis}}|\mathbf{z}_{i,\text{obs}}, \mathbf{y}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}^{(t)}, \boldsymbol{\sigma}^{(t)}) f(\mathbf{z}_i|\boldsymbol{\alpha}) f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}^{(t)}), \tag{A.2}$$

$$\begin{aligned} f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)}) &\propto f(\mathbf{b}_i|D^{(t)}) f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}^{(t)}, \boldsymbol{\sigma}^{(t)}) \\ &\quad \times f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}^{(t)}) f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}^{(t)}). \end{aligned} \tag{A.3}$$

Thus, rejection sampling methods such as a multivariate rejection method may be used to sample from each of the full conditionals (see next paragraph). Iteratively sampling from each of the full conditionals in turn for a burn-in period, we obtain a sample from $f(\mathbf{y}_{i,\text{mis}}, \mathbf{z}_{i,\text{mis}}, \mathbf{b}_i|\mathbf{y}_{i,\text{obs}}, \mathbf{z}_{i,\text{obs}}, \mathbf{s}_i, \mathbf{r}_i, \boldsymbol{\theta}^{(t)})$.

Sampling from the distributions (A.1)–(A.3) can be accomplished by rejection sampling methods as follows: If the appropriate densities on the right-hand sides of (A.1)–(A.3) are log-concave, the adaptive rejection algorithm of Gilks and Wild (1992) may be used. If some densities are not log-concave, we may consider the multivariate rejection sampling method. For example, suppose that we want to generate random samples from $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \mathbf{s}_i, \boldsymbol{\theta}^{(t)})$ in (A.3). Let $h(\mathbf{b}_i) = f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}^{(t)}, \boldsymbol{\sigma}^{(t)}) f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\gamma}^{(t)})$ $f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}^{(t)})$ and $\tau = \sup_{\mathbf{b}}\{h(\mathbf{b})\}$. A random sample from $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \mathbf{s}_i, \boldsymbol{\psi}^{(t)})$ can be obtained as follows. Step 1: sample $\mathbf{b}_i^*$ from $f(\mathbf{b}_i|D^{(t)})$, and independently, sample w from the uniform(0,1) distribution; Step 2: if $w \leqslant h(\mathbf{b}_i^*)/\tau$, then accept $\mathbf{b}_i^*$; otherwise, go to Step 1. Samples from the other 2 full conditionals can be obtained in a similar way. Therefore, the E-step of the Monte Carlo EM method can be accomplished by the Gibbs sampler method combined with the rejection sampling methods. To assess the convergence of the Gibbs sampler, we may use standard graphical tools such as time-series plots and autocorrelations and determine the burn-in or warm-up iterations of the Gibbs sampler based on some preliminary draws.

The above rejection sampling methods may be slow when the dimension of $(\mathbf{y}_{\text{mis},i}, \mathbf{z}_{\text{mis},i}, \mathbf{b}_i)$ is large. In this case, we may use importance sampling methods. The importance function can be chosen to be a multivariate normal density or a multivariate student's t density whose mean and variance match the mode and curvature of $f(\mathbf{y}_{\text{mis},i}, \mathbf{z}_{\text{mis},i}, \mathbf{b}_i|\mathbf{y}_{\text{obs},i}, \mathbf{z}_{\text{obs},i}, \mathbf{r}_i, \mathbf{s}_i, \boldsymbol{\theta}^{(t)})$. Other sampling methods include Metropolis–Hastings or Markov chain methods.

To implement the E-step of the Monte Carlo EM algorithm, we should choose the numbers of Monte Carlo samples m_t. Generally, larger values of m_t will result in more exact approximation in the E-step but the computation will be slower. To ensure convergence of the Monte Carlo EM algorithm, we should increase m_t as the number t of EM iterations increases. For Monte Carlo EM algorithms, the incomplete data log-likelihood is not guaranteed to increase at each iteration due to Monte Carlo error at the E-step. However, under suitable regularity conditions, Monte Carlo EM algorithms still converge to the MLE (Fort and Moulines, 2003). For sufficiently large values of m_t, the Monte Carlo EM algorithm would inherit the properties of the exact versions, such as the likelihood-increasing properties of EM.

APPENDIX B: TIME-DEPENDENT COVARIATES WITH MEASUREMENT ERRORS OR MISSING DATA

In this section, we outline an approach which extends the proposed method to time-dependent covariates with measurement errors or missing data. Let z_{ikl} be the observed value and z^*_{ikl} be the (unobservable) "true" value of covariate k for the ith individual at time $u_{il}, i = 1, \ldots, N, k = 1, \ldots, v, l = 1, \ldots, m$. We focus on the case where z^*_{ikl} is the current true covariate value, but our method can be extended to the case where z^*_{ikl} is a summary of the true covariate history up to time u_{il}. We allow the covariate measurement times u_{il} to differ from the response measurement times t_{ij}, so we allow (ignorable) missing data in the covariates.

Let $\mathbf{z}_i = (\mathbf{z}_{i1}^T, \ldots, \mathbf{z}_{im}^T)^T$, where $\mathbf{z}_{il} = (z_{i1l}, \ldots, z_{ivl})^T, l = 1, \ldots, m$. Following Shah *and others* (1997), we consider the following multivariate LME model to empirically describe the covariate processes:

$$\mathbf{z}_{il} = U_{il}\boldsymbol{\alpha} + V_{il}\mathbf{a}_i + \boldsymbol{\epsilon}_{il} \equiv \mathbf{z}^*_{il} + \boldsymbol{\epsilon}_{il}, \quad i = 1, \ldots, N, \ \ l = 1, \ldots, m, \tag{B.1}$$

where U_{il} and V_{il} are design matrices, $\boldsymbol{\alpha}$ and $\mathbf{a}_i$ are unknown population (fixed-effects) and individual-specific (random-effects) parameter vectors, and $\boldsymbol{\epsilon}_{il}$ are the random measurement errors for the ith individual at the time u_{il}. The true (unobservable) covariate values are assumed to be $\mathbf{z}^*_{il} = U_{il}\boldsymbol{\alpha} + V_{il}\mathbf{a}_i$. We also assume that $\mathbf{a}_i$ i.i.d. $\sim N(\mathbf{0}, A)$, $\boldsymbol{\epsilon}_{il}$ i.i.d. $\sim N(0, R)$, and $\mathbf{a}_i$ and $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_{i1}^T, \ldots, \boldsymbol{\epsilon}_{im}^T)^T$ are independent, where A and R are unknown and unstructured covariance matrices. We further assume that $\boldsymbol{\epsilon}_i$ and $\mathbf{a}_i$ are independent of $\mathbf{e}_i$ and $\mathbf{b}_i$ in the response model. Models such as (B.1) may be interpreted as a covariate measurement error model (Carroll *and others*, 1995). To allow for missing data in the time-varying covariates, we recast model (B.1) in continuous time:

$$\mathbf{z}_i(t) = U_i(t)\boldsymbol{\alpha} + V_i(t)\mathbf{a}_i + \boldsymbol{\epsilon}_i(t), \quad i = 1, \ldots, N,$$

where $\mathbf{z}_i(t)$, $U_i(t)$, $V_i(t)$, and $\boldsymbol{\epsilon}_i(t)$ are the covariate values, design matrices, and measurement errors at the time t, respectively. At the response measurement time t_{ij}, the possibly unobserved true covariate values can be viewed as $\mathbf{z}^*_{ij} = U_{ij}\boldsymbol{\alpha} + V_{ij}\mathbf{a}_i$, where $U_{ij} = U_i(t_{ij})$ and $V_{ij} = V_i(t_{ij})$.

When the covariates are measured with errors, we assume that the response and the time-to-event distributions $f(\mathbf{y}_i|\mathbf{a}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma})$ and $f(\mathbf{r}_i|\mathbf{a}_i, \mathbf{b}_i, \boldsymbol{\gamma})$ may depend on the unobserved true covariate values rather than the observed mismeasured covariate values, that is, the distributions of $\mathbf{y}_i$ and $\mathbf{r}_i$ may depend on the random effects $\mathbf{a}_i$ and $\mathbf{b}_i$. The observed data log-likelihood can thus be written as

$$L^*_{\mathrm{o}}(\boldsymbol{\theta}) = \prod_{i=1}^{N}\left[\int\!\!\int f(\mathbf{y}_i|\mathbf{a}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\sigma}) f(\mathbf{b}_i|D) f(\mathbf{z}_i|\mathbf{a}_i, \boldsymbol{\alpha}) f(\mathbf{a}_i|A) f(\mathbf{r}_i|\mathbf{a}_i, \mathbf{b}_i, \boldsymbol{\gamma}) f(\mathbf{s}_i|\mathbf{b}_i, \boldsymbol{\phi}) \mathrm{d}\mathbf{y}_{i,\mathrm{mis}}\, \mathrm{d}\mathbf{a}_i\, \mathrm{d}\mathbf{b}_i\right].$$

Then, a Monte Carlo EM algorithm similar to that in Section 3 can be used to obtain the MLEs of all unknown parameters. A main modification needed in the E-step is to sample the random effects $\mathbf{a}_i$ instead of $\mathbf{z}_{\mathrm{mis},i}$, which can again be accomplished by Gibbs sampler combined with rejection sampling methods in a similar way.

REFERENCES

CARROLL, R. J., RUPPERT, D. AND STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman & Hall.

DAVIDIAN, M. AND GILTINAN, D. M. (1995). *Nonlinear Models for Repeated Measurements Data*. London: Chapman & Hall.

DEGRUTTOLA, V. AND TU, X. M. (1994). Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

DING, A. AND WU, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics* **2**, 13–29.

FOLLMANN, D. AND WU, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 15–168.

FORT, G. AND MOULINES, E. (2003). Convergence of the Monte-Carlo EM for curved exponential families. *Annals of Statistics* **31**, 1220–1259.

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.

GUO, X. AND CARLIN, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* **58**, 1–9.

HENDERSON, R., DIGGLE, P. J. AND DOBSON, A. (2002). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

LAWLESS, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition, Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons.

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

SHAH, A., LAIRD, N. AND SCHOENFELD, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* **92**, 775–779.

TEN HAVE, T. R., PULKSTENIS, E., KUNSELMAN, A. AND LANDIS, J. R. (1998). Mixed effects logistics regression models for longitudinal binary response data with informative dropout. *Biometrics* **54**, 367–383.

TSIATIS, A. A. AND DAVIDIAN, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica* **14**, 793–818.

WEI, G. C. AND TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association* **85**, 699–704.

WU, H. AND DING, A. (1999). Population HIV-1 dynamics in vivo: application models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410–418.

WU, H., LATHEY, J., RUAN, P., DOUGLAS, S. D., SPECTOR, S. A., LINDSEY, J., HOUGHES, M. D., RUDY, B. J., FLYNN, P. M. AND PACTG 381 TEAM (2004). Relationship of plasma HIV-1 RNA dynamics to baseline factors and virological responses to highly active antiretroviral therapy in adolescents (aged 12–22 years) infected through high-risk behavior. *Journal of Infectious Diseases* **189**, 593–601.

WU, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association* **97**, 955–964.

WU, M. C. AND CARROLL, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **55**, 410–418.

WULFSOHN, M. S. AND TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

[*Received April 28, 2006; first revision June 20, 2007; second revision July 19, 2007; accepted for publication July 19, 2007*]