Biometrics 66, 327–335
June 2010
DOI: 10.1111/j.1541-0420.2009.01308.x

Joint Inference on HIV Viral Dynamics and Immune Suppression in Presence of Measurement Errors

L. Wu[1,*] W. Liu,[2] and X. J. Hu[3]

[1]Department of Statistics, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada
[2]Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada
[3]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
[*]*email:* lang@stat.ubc.ca

Summary: In an attempt to provide a tool to assess antiretroviral therapy and to monitor disease progression, this article studies association of human immunodeficiency virus (HIV) viral suppression and immune restoration. The data from a recent acquired immune deficiency syndrome (AIDS) study are used for illustration. We jointly model HIV viral dynamics and time to decrease in CD4/CD8 ratio in the presence of CD4 process with measurement errors, and estimate the model parameters simultaneously via a method based on a Laplace approximation and the commonly used Monte Carlo EM algorithm. The approaches and many of the points presented apply generally.

Key words: Laplace approximation; Longitudinal data; Mixed-effects; Nonlinear models; Time-to-event.

1. Introduction

Human immunodeficiency virus (HIV) infection results in a progressive destruction of immune function, which may be indicated by a decrease of CD4 T-cells, an increase of CD8 T-cells, and a decrease in the ratio of CD4 to CD8 (Stevens et al., 2006). The CD4/CD8 ratio recently has become a tool for assessing the relative condition of HIV subjects. It, along with CD4 count, provides a way of gauging the progression from HIV to acquired immune deficiency syndrome (AIDS) for prognostic purposes. After HIV-infected subjects start an anti-HIV treatment, their viral loads typically decline in the initial period, often with a roughly biphasic exponential decay pattern, and then some subjects may experience viral rebound later in the study, which is possibly due to drug resistance and other factors. The relationship between HIV viral suppression and immune restoration has received great attention in AIDS research (Henry, Tebas, and Lane, 2006). However, little has been published about statistical analysis for the particular association of HIV viral dynamics with time trend in the CD4/CD8 ratio.

This article studies the association of viral decay with CD4/CD8 time trend by jointly modeling HIV viral dynamics, time to decrease in CD4/CD8, and CD4 process. The research was motivated by a recent AIDS Clinical Trials Group (ACTG) study, the data of which are used in Ding and Wu (2001), among others, to demonstrate that the initial viral decay rate reflects the efficacy of an anti-HIV treatment. We plotted the trajectories of the study subjects' HIV viral load, CD4/CD8 ratio, and CD4 count. The plots confirm that HIV viral load is negatively correlated with both CD4 count and the ratio CD4/CD8. In addition, we observed the following: (i) HIV viral rebound seems closely associated with time to decrease in CD4/CD8; and (ii) compared with the CD4 process, the CD4/CD8 ratio over time looks less wavering and its trend seems more closely associated with the viral dynamics.

Figure 1 presents the trajectories of four randomly selected study subjects. While the study data indicate a close association of the longitudinal viral load, CD4 count, and ratio CD4/CD8, they evidence large variation in the association across subjects and within each subject over time. The observations led us to model each of the three processes and link the three models with random effects, which characterize the underlying individual-specific effects. The specific models for this dataset are described in Section 3.

Joint modeling of longitudinal data and time-to-event data has recently received much attention in the literature (e.g., Henderson, Diggle, and Dobson, 2002; Guo and Carlin, 2004; Wu, Hu, and Wu, 2008). See Tsiatis and Davidian (2004) for a comprehensive overview. A major challenge with joint modeling is the associated intensive computation in the inference, and in some cases it can even be computationally infeasible. In the presence of measurement errors, missing data, and nonlinear longitudinal models, the computational problem becomes much worse. Wu et al. (2008) considered joint modeling of a nonlinear mixed-effects (NLME) model and the Cox proportional hazards model with missing data in the longitudinal responses. They applied the Monte Carlo EM algorithm (MCEM) for the joint likelihood inference, and experienced intensive computing.

The new contributions of this article are as follows. We study the association of viral dynamics with immune suppression through both CD4 count and the trend of CD4/CD8 ratio, which may provide new scientific insights into the association. From a statistical methodological point of view, while

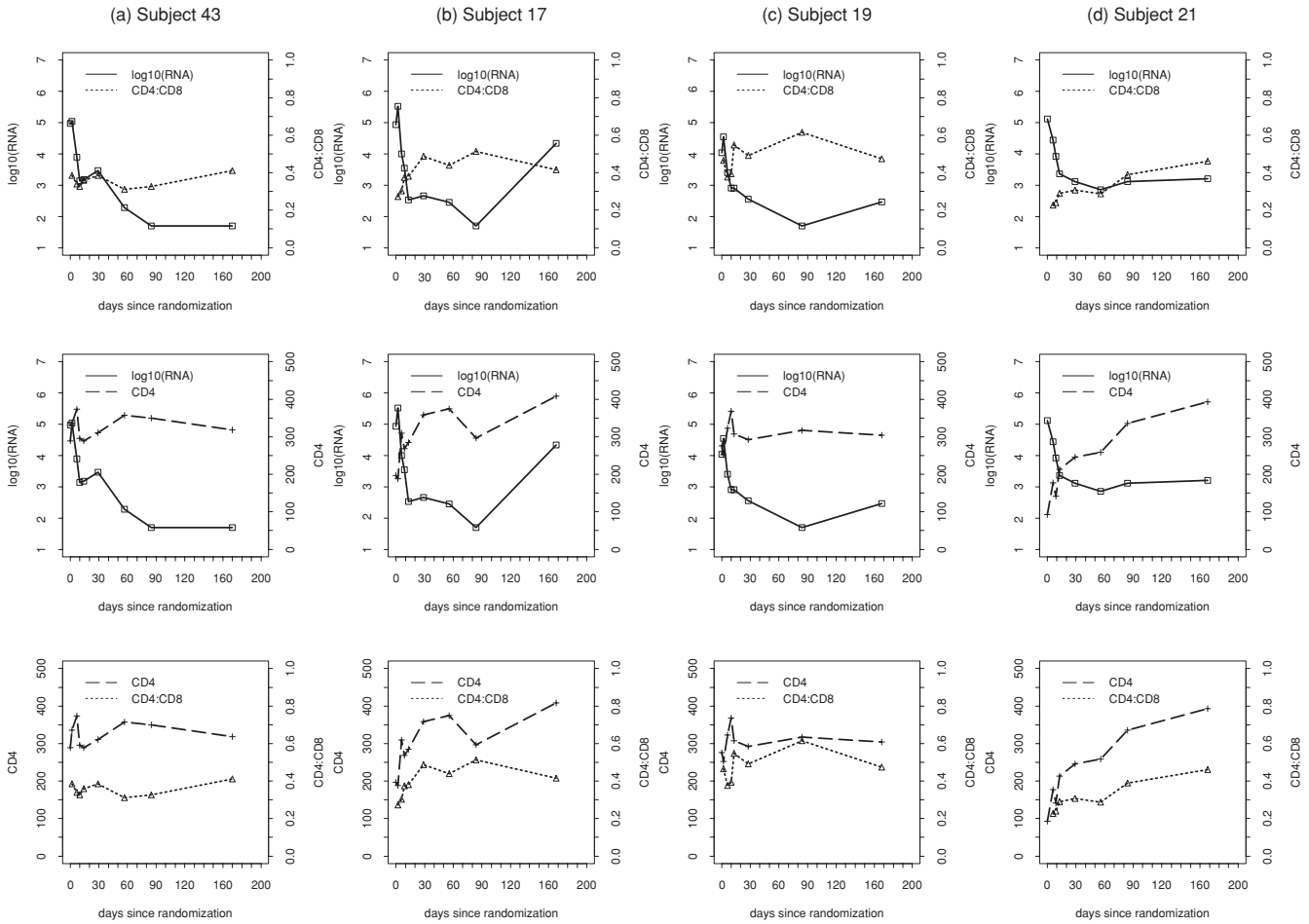

Figure 1. Viral load, CD4, and CD4/CD8 trajectories of four randomly selected subjects.

applying the well-established NLME model for the viral dynamics, we consider an empirical nonparametric mixed-effects model for CD4 count, which is viewed as a covariate process of the viral load, to incorporate *measurement errors* as well as potential missing data. The modeling for CD4 process also makes it feasible to relate the CD4 count as a time-dependent covariate to the time-to-event of interest, which is modeled via a parametric survival model. In addition, we propose a computationally efficient method based on a Laplace approximation, referred to as the saddle-point approximation by physicists, to address the computing challenge in a joint likelihood inference. The method offers a big computational advantage over the computationally intensive methods appearing in the joint model literature. Finally, the three models (the NLME model for viral dynamics, the nonparametric mixed model for CD4 process, and the parametric event-time model for CD4/CD8 decline) are linked through the random effects that characterize the underlying individual-specific longitudinal processes. The methods and many of the points discussed have broader applications, not just limited to the specific scientific problem discussed.

The rest of the article is organized as follows. Section 2 describes the joint models and the associated estimation procedures in general forms so that they can be considered in other applications. After a brief review of the commonly used approach in the joint likelihood inferences, the MCEM algorithm, we present a computationally efficient approximate likelihood approach based on a Laplace approximation. Section 3 describes the specific models and reports an analysis of the AIDS data that motivated this research. Section 4 presents a simulation study to examine the finite sample performances of the approaches used in the data analysis. We provide some discussion and final remarks in Section 5.

2. Joint Likelihood Inference

2.1 *Joint Modeling*

In a study with N independent subjects, let y_{ij} be the response of subject i at time $t_{ij}, j = 1, \ldots, n_i$ and $i = 1, \ldots, N$. Denote subject i's time to the event of interest by T_i. Given z_{ij}^*, the "true" (but unobservable) covariate value at time t_{ij}, we consider an NLME model for the longitudinal response process

$$y_{ij} = g(\boldsymbol{\beta}, t_{ij}, z_{ij}^*, \mathbf{b}_i) + e_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, N, \tag{1}$$

where $g(\cdot)$ is a nonlinear function, e_{ij} are within subject random errors, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^T$ is a vector of fixed effects,

and $\mathbf{b}_i = (b_{i1}, \ldots, b_{id})^T$ is a vector of random effects. We assume that $\mathbf{b}_i$ and $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})^T$ are independent, and $\mathbf{e}_i$ i.i.d. $\sim N(\mathbf{0}, \nu^2 I_i)$ and $\mathbf{b}_i$ i.i.d. $\sim N(\mathbf{0}, B)$, where ν is an unknown parameter, I_i is the identity matrix, and B is an unknown covariance matrix.

For the covariate process, we adopt a flexible empirical nonparametric mixed-effects model to address measurement errors

$$z_i(t) = r(t) + h_i(t) + \xi_i(t) \equiv z_i^*(t) + \xi_i(t), \quad i = 1, \ldots, N, \tag{2}$$

where $z_i^*(t) = r(t) + h_i(t)$ are the true but unobservable covariate values at time t, $r(t)$ and $h_i(t)$ are unknown nonparametric smooth fixed-effects and random-effects functions, respectively, and $\xi_i(t) \sim N(0, \sigma^2)$. The random smooth function $h_i(t)$ is introduced to incorporate the large interindividual variation in the covariate process, while the fixed smooth function $r(t)$ represents population average of the covariate process. We assume that $h_i(t)$ is the realization of a zero-mean stochastic process.

We approximate the nonparametric functions $r(t)$ and $h_i(t)$ by linear combinations of the natural cubic spline basis functions with percentile-based knots (Green and Solverman, 1994), $\mathbf{\Psi}_p(t) = [\psi_0(t), \psi_1(t), \ldots, \psi_{p-1}(t)]^T$ and $\mathbf{\Phi}_q(t) = [\phi_0(t), \phi_1(t), \ldots, \phi_{q-1}(t)]^T$. That is,

$$\begin{aligned} r(t) &\approx r_p(t) = \sum_{k=0}^{p-1} \alpha_k \, \psi_k(t) = \mathbf{\Psi}_p(t)^T \, \boldsymbol{\alpha}, \\ h_i(t) &\approx h_{iq}(t) = \sum_{k=0}^{q-1} a_{ik} \, \phi_k(t) = \mathbf{\Phi}_q(t)^T \, \mathbf{a}_i, \end{aligned} \tag{3}$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})^T$ is a $p \times 1$ vector of fixed effects and $\mathbf{a}_i = (a_{i0}, \ldots, a_{i,q-1})^T$ is a $q \times 1$ vector of random effects with $\mathbf{a}_i$ i.i.d. $\sim N(0, A)$. Appropriate values of p and q can be determined via the standard model selection criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) values (e.g., Rice and Wu, 2001). Approximating $r(t)$ and $h_i(t)$ by $r_p(t)$ and $h_{iq}(t)$, respectively, we can then approximate the covariate model (2) by the following linear mixed-effects (LME) model

$$z_i(t) \approx \mathbf{\Psi}_p(t)^T \, \boldsymbol{\alpha} + \mathbf{\Phi}_q(t)^T \, \mathbf{a}_i + \xi_i(t) \approx z_i^*(t) + \xi_i(t). \tag{4}$$

We assume that $\mathbf{a}_i, \mathbf{b}_i, \mathbf{e}_i$, and ξ_i are mutually independent. Note that, since the LME model allows for unbalanced response data, this covariate model can incorporate missing data in the time-dependent covariates when the missing is at random, in addition to addressing covariate measurement errors.

The event time T_i is likely related to the longitudinal response and covariate processes. This association is of much interest in many practical situations. We specify the association by assuming that, conditional on the random effects $\mathbf{a}_i$ in the covariate model (4) and the random effects $\mathbf{b}_i$ in the response model (1), the event time $T_i \sim f(t \,|\, \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \lambda)$ with unknown parameters $\boldsymbol{\gamma}$ and λ. That is, we assume that the event time is related to the longitudinal processes through the random effects that characterize the individual-specific effects. Particularly, we consider the following mixed-effects event time model:

$$\log(T_i) = \gamma_0 + \boldsymbol{\gamma}_1^T \, \mathbf{a}_i + \boldsymbol{\gamma}_2^T \, \mathbf{b}_i + \epsilon_i, \quad i = 1, 2, \ldots, N, \tag{5}$$

where the coefficients $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T)^T$, and the random errors ϵ_i's are i.i.d. and follow a parametric distribution with mean 0 and the other parameters λ such as $N(0, \lambda^2)$. We assume that ϵ_i's are independent of $\mathbf{a}_i$ and $\mathbf{b}_i$. Model (5) may be a good choice when the event times are thought to depend on individual-specific longitudinal trajectories, such as initial slopes and intercepts, or summaries of the longitudinal trajectories, and it is closely related to so-called shared parameter models (Wu and Carroll, 1988; DeGruttola and Tu, 1994). An alternative model is to relate the event times directly to the true but unobservable covariates. We will discuss this issue with some length in Section 5.

In some practical situations, the event times cannot be observed but only known as being contained in some time intervals, i.e., being *interval-censored*. In the AIDS study mentioned in Section 1, for example, given that CD4 and CD8 were collected at a finite number of times, we can only know that the time to the occurrence of the first CD4/CD8 decrease of a subject is between two data collection time points. The observed event time data are then $\{(w_i, v_i], i = 1, \ldots, N\}$, where $(w_i, v_i]$ is the smallest observed interval containing T_i. We take w_i as his latest time in the study and $v_i = \infty$ if subject i did not experience the event of interest during the whole study period.

2.2 *Estimation of Model Parameters*

In a longitudinal study, such as the AIDS study described in Section 1, the longitudinal response, the time-to-event, and the covariate processes are usually connected physically or biologically. As discussed in Section 1, we can model them jointly through the shared random effects, which is often biologically meaningful. Thus, statistical inferences on all the model parameters need to be made simultaneously; otherwise, if the parameters in each of the three models are estimated separately, the underlying association may not be fully addressed or captured. Moreover, if the shared parameters in one model are substituted by their estimates from another model, the uncertainty due to the estimation is not incorporated, and may lead to under-estimation of standard deviations. In this section, we consider joint likelihood inferences to estimate all the parameters simultaneously.

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, \nu, A, B, \boldsymbol{\gamma}, \lambda)$ be the collection of all model parameters, and use $f(\cdot)$ for a generic density function. With the available data, $\{(\mathbf{y}_i, \mathbf{z}_i, w_i, v_i) : i = 1, 2, \ldots, N\}$, the joint likelihood function of $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v}) = \prod_{i=1}^{N} \left[\iint L_i(\boldsymbol{\theta} \,|\, \mathbf{y}_i, \mathbf{z}_i, w_i, v_i; \mathbf{a}_i, \mathbf{b}_i) d\mathbf{a}_i d\mathbf{b}_i \right], \tag{6}$$

where $L_i(\boldsymbol{\theta} \,|\, \mathbf{y}_i, \mathbf{z}_i, w_i, v_i; \mathbf{a}_i, \mathbf{b}_i)$ is $f(\mathbf{y}_i \,|\, \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) f(\mathbf{b}_i; B)$ $f(\mathbf{z}_i \,|\, \mathbf{a}_i; \boldsymbol{\alpha}, \sigma) f(\mathbf{a}_i; A) f^*(w_i, v_i \,|\, \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \lambda)$ with $f^*(w_i, v_i \,|\, \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \lambda) = \int_{w_i}^{v_i} f(t \,|\, \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \lambda) \, dt$. It is in general hard to conduct inferences based on (6) directly, due to the high-dimensional and intractable integral. Numerical integration methods such as the Gauss–Hermite quadrature can be computationally very intensive or even infeasible when the

dimension of the random effects $(\mathbf{a}_i, \mathbf{b}_i)$ is not small (say, higher than 2). Another approach is to use the EM algorithm to obtain the maximum-likelihood estimate (MLE) of $\boldsymbol{\theta}$, by treating the unobservable random effects $(\mathbf{a}_i, \mathbf{b}_i)$ as "missing values," as follows.

The EM algorithm iterates between an E-step, which computes the conditional expectation of the "complete data" log-likelihood given the available data and current parameter estimates, and an M-step, which maximizes the conditional expectation from the E-step to update the parameter estimates. Specifically, in this application the "complete data" are $\{(\mathbf{y}_i, \mathbf{z}_i, w_i, v_i, \mathbf{a}_i, \mathbf{b}_i),\ i = 1, 2, \ldots, N\}$, so the E-step evaluates

$$Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)}) = E[l(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v}; \mathbf{a}, \mathbf{b}) \,|\, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v}; \boldsymbol{\theta}^{(t)}],$$

where $\boldsymbol{\theta}^{(t)}$ is the estimate of $\boldsymbol{\theta}$ at the tth EM iteration, $l(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v}; \mathbf{a}, \mathbf{b}) = \sum_{i=1}^{N} l_i(\boldsymbol{\theta}; \mathbf{a}_i, \mathbf{b}_i)$, and $l_i(\boldsymbol{\theta}; \mathbf{a}_i, \mathbf{b}_i) = \log L_i(\boldsymbol{\theta} \,|\, \mathbf{y}_i, \mathbf{z}_i, w_i, v_i; \mathbf{a}_i, \mathbf{b}_i)$ is the complete-data log-likelihood for subject i. The M-step maximizes $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)})$ to produce an updated estimate $\boldsymbol{\theta}^{(t+1)}$ for $t = 0, 1, 2, \ldots$ till convergence. We may obtain an initial estimate $\boldsymbol{\theta}^{(0)}$ by, for example, separately estimating the model parameters in each of the three models. Available computer programs for optimizing a nonlinear function can be used to realize the M-step.

The E-step is, however, hard to compute in the current application, because of the required calculation of the multidimensional integral. The MCEM algorithm, a well-known variation of the EM algorithm, approximates each conditional expectation term in $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)})$ at the $(t+1)$th EM iteration with the sample mean of $l_i(\boldsymbol{\theta}; \mathbf{a}_i, \mathbf{b}_i)$ based on a large sample of $(\mathbf{a}_i, \mathbf{b}_i)$ generated from the conditional distributions $f(\mathbf{a}_i,\ \mathbf{b}_i \,|\, \mathbf{y}_i,\ \mathbf{z}_i, w_i, v_i;\ \boldsymbol{\theta}^{(t)})$. This Monte Carlo sampling in the E-step is often accomplished via a Markov chain Monte Carlo method such as the Gibbs sampler, along with rejection sampling methods. Thus, the convergence of the MCEM involves convergence of both the Gibbs sampler within *each* EM iteration and the global convergence of the EM algorithm. See Ibrahim et al. (2005) and Wu et al. (2008) for applications of the approach in similar problems. Details of the MCEM in the current application are provided in Web Appendix 1.

The implementation of the MCEM algorithm in this application is rather computationally intensive and may even exhibit convergence problems. It requires the convergence of the Gibbs sampler within each EM iteration as well as the global convergence of the EM algorithm. In some situations, especially when the dimension of the random effects $(\mathbf{a}_i, \mathbf{b}_i)$ is high, the convergences of the Gibbs sampler and the EM algorithm are difficult to check and the MCEM algorithm may lead to undesirable outcomes. Alternative approaches are therefore in big demand. Computationally more efficient approximate methods may be obtained based on Taylor or Laplace approximations. We present below a method based on a Laplace approximation. It approximates the observed-data joint likelihood $L(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v})$ using a first-order Laplace approximation, which has an analytic expression and is thus computationally more efficient than the MCEM since Monte Carlo sampling is no longer needed.

Let $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) = \{(\tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i), i = 1, 2, \ldots, N\}$ be the collection of the solutions to the following equations with a fixed $\boldsymbol{\theta}$,

$$\partial l_i(\boldsymbol{\theta}; \mathbf{a}_i, \mathbf{b}_i)/\partial(\mathbf{a}_i, \mathbf{b}_i) = \mathbf{0}, \quad i = 1, 2, \ldots, N. \tag{7}$$

Following Lee, Nelder, and Pawitan (2006), we obtain the first-order Laplace approximation to the observed-data joint log-likelihood $l(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v})$ based on

$$\tilde{l}\big(\boldsymbol{\theta}; \tilde{\mathbf{a}}, \tilde{\mathbf{b}}\big) = l(\boldsymbol{\theta}\big|\mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v}; \tilde{\mathbf{a}}, \tilde{\mathbf{b}}) - \frac{1}{2} \log \det\Big\{ -\frac{1}{2\pi} \frac{\partial^2 l(\boldsymbol{\theta}\big|\mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{v}; \mathbf{a}, \mathbf{b})}{\partial(\mathbf{a}, \mathbf{b}) \partial(\mathbf{a}, \mathbf{b})^T} \Big\} \Big|_{(\mathbf{a}, \mathbf{b}) = (\tilde{\mathbf{a}}, \tilde{\mathbf{b}})}, \tag{8}$$

where $\det\{A\}$ is the determinant of matrix A. Estimates obtained by maximizing $\tilde{l}(\boldsymbol{\theta}; \tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ are thus approximate MLEs. Lee et al. (2006) provided some detailed discussion about this approach in the case of complete-data generalized linear mixed models (GLMM). The ideas can be applied in the current models.

An algorithm for obtaining an estimator of $\boldsymbol{\theta}$ based on the foregoing Laplace approximation is as follows. Let the estimate of $\boldsymbol{\theta}$ in the lth iteration be $\boldsymbol{\theta}^{(l)}$. For $l = 0, 1, 2, \ldots,$

Step 1. Solve equations (7) with $\boldsymbol{\theta} = \boldsymbol{\theta}^{(l)}$ to obtain the random effect estimates $(\tilde{\mathbf{a}}^{(l)}, \tilde{\mathbf{b}}^{(l)})$,

Step 2. Maximize the approximate joint log-likelihood $\tilde{l}(\boldsymbol{\theta}; \tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ in (8) with $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) = (\tilde{\mathbf{a}}^{(l)}, \tilde{\mathbf{b}}^{(l)})$ to obtain a new estimate $\boldsymbol{\theta}^{(l+1)}$.

Repeat the two steps until the sequence of $\{\boldsymbol{\theta}^{(l)} : l = 0, 1, 2, \ldots\}$ converges. Denote the limit of the sequence by $\hat{\boldsymbol{\theta}}$, which is an approximate MLE of $\boldsymbol{\theta}$. There are many optimization procedures available in standard softwares, such as *ms()* and *optim()* in `Splus` or `R`, which need only the original functions rather than their derivatives.

We may interpret $(\tilde{\mathbf{a}}^{(l)}, \tilde{\mathbf{b}}^{(l)})$ as empirical Bayesian estimates of the random effects in the lth iteration. The standard error of the resulting estimator $\hat{\boldsymbol{\theta}}$ can be computed using

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \left[-\frac{\partial^2 \tilde{l}(\boldsymbol{\theta}; \tilde{\mathbf{a}}^L, \tilde{\mathbf{b}}^L)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]^{-1}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}},$$

where $(\tilde{\mathbf{a}}^L, \tilde{\mathbf{b}}^L)$ is the random effects estimate obtained in the last iteration. This matrix is often a routine output in many optimization procedures, such as *optim()* in `R`.

In simpler settings, the performance of such a Laplace approximation to the likelihood functions is usually excellent with a continuous response. See Pinheiro and Bates (1995) for simulation results for complete-data NLME models and Joe (2008) for complete-data GLMMs and an extensive discussion. The current setting is more complex, but the general approach remains essentially the same. In Section 4, we conduct a simulation study to examine the performance of the approximate method in the current setting. We denote the approximate method by Laplace approximation (LAAP) in the rest of the article. It appears that the LAAP approach, compared with the MCEM algorithm, works well in our application.

3. Analysis of AIDS Data

3.1 *Data and Models*

The AIDS study mentioned in Section 1 had 46 HIV-infected subjects who started an anti-HIV treatment 10 days after entry. Viral loads were measured in HIV RNA and collected along with CD4 and CD8 counts on study days 0, 2, 7, 10, 14, 21, 28 and study weeks 8, 12, 24, and 48. For simplicity, we imputed the viral loads below the assay's lower detection limit 100 copies/ml with a half of the limit. A total of 361 viral load measures were recorded, with the number of viral load measures for each subject ranging from 4 to 10. The missing rate of the associated CD4 and CD8 measures was 20% (72 out of 361). The objective of the analysis was to study the association of the viral decay rates with the time to decrease in CD4/CD8 in the presence of CD4 counts with measurement errors. We considered the joint modeling of the three processes to characterize the association as well as to incorporate measurement errors and missing data in CD4 (the covariate process in the analysis).

In the initial period during an anti-HIV treatment, the following biphasic exponential viral decay model (a two-compartment model) has been shown to provide a good fit in many studies with a strong biological justification and interpretation (Wu and Ding, 1999)

$$y_{ij} = \log_{10}(P_{1i}e^{-\lambda_{1ij}t_{ij}} + P_{2i}e^{-\lambda_{2ij}t_{ij}}) + e_{ij}, \qquad (9)$$

where y_{ij} is the $\log_{10}$-transformation of the HIV viral load measurement for subject i at time t_{ij}, P_{1i} and P_{2i} are individual-specific baseline viral load values, λ_{1ij} and λ_{2ij} are the individual-specific first-phase and second-phase viral decay rates, respectively, and e_{ij} is the random error for within-subject measurements. We assume further that

$$\log(P_{1i}) = \beta_1 + b_{i1}, \;\; \lambda_{1ij} = \beta_2 + b_{i2},$$
$$\log(P_{2i}) = \beta_3 + b_{i3}, \;\; \lambda_{2ij} = \beta_4 + \beta_5 z^*_{ij} + b_{i4},$$

where z^*_{ij} is the true CD4 count at time t_{ij} (which is unobservable but might be estimated), β_j's are population parameters, and b_{ik}'s are random effects that characterize between-subject variations. See Liu and Wu (2007).

CD4 measures are known with nonnegligible errors. Ignoring covariate measurement errors can lead to severely misleading results in a statistical inference. Replications or validation data are in general needed to address measurement errors (Carroll et al., 2006). With CD4 measures collected over time from the AIDS study, we may model the CD4 process to partially address the measurement errors. See Wu (2002) for an example of modeling the process parametrically. However, the CD4 trajectories are often complicated, and there is no well-established model for the CD4 process. We, thus, model the CD4 process empirically using a nonparametric mixed-effects model, which is flexible and works well for complex longitudinal data. Based on AIC and BIC model selection criteria, we obtained the following nonparametric mixed-effects CD4 model for the current dataset:

$$z_{ik} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})\psi_1(u_{ik}) + (\alpha_3 + a_{i3})\psi_2(u_{ik}) + \xi_{ik}, \qquad (10)$$

where z_{ik} is the observed CD4 value at time u_{ik}, $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are two basis functions given in Section 2, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ are population parameters (fixed-effects), $\mathbf{a}_i = (a_{i1}, a_{i2}, a_{i3})^T$ are random effects, and ξ_{ik} is the measurement error at time u_{ik}. We may view $z^*_{ik} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})\,\psi_1(u_{ik}) + (\alpha_3 + a_{i3})\,\psi_2(u_{ik})$ as the true (but unobservable) CD4 value at time u_{ik}.

Let T_i be subject i's time to his first decline in the CD4/CD8 ratio. We are interested in the association of the time to the immune suppression with the individual-specific initial viral decay rates and the true CD4 trajectory, which are characterized by the random effects (or individual effects) in the viral load and CD4 models. We may view the (unobservable) random effects as error-free covariates in time-to-event models. The associated inferences can be computationally challenging when joint models consist of (nonlinear) longitudinal models and the semiparametric Cox proportional hazards model for time-to-event, a commonly used failure time model (Tsiatis and Davidian, 2004; Wu et al., 2008), especially when the event times are interval censored. To focus on the primary issues discussed in the article, we consider the following parametric model of time-to-event:

$$\log(T_i) = \gamma_0 + \gamma_1 a_{i1} + \gamma_2 a_{i2} + \gamma_3 b_{i2} + \gamma_4 b_{i4} + \epsilon_i, \qquad (11)$$

where γ_j's are parameters and ϵ_i is the random error. In Model (11), the random effects b_{i2} and b_{i4} represent individual variations in viral decay rates so they may be predictive for event times, while b_{i1} and b_{i3} in the viral dynamic model represent variations in the baseline viral loads, which seem not highly predictive for event times, so they are excluded from the model to reduce number of parameters. The random effects a_{i1} and a_{i2} capture the main features of individual CD4 trajectories, so they are included in the model. Model (11) is a parametric accelerated failure time (AFT) model with random effects and seems to fit the current dataset reasonably well (see Web Figure 1). It offers the advantages of easy interpretation and robustness against neglected covariates (Hougaard, 1999). An alternative model is to link the event times to the unobserved true CD4 values z^*_{ik}, but model (11) can offer some advantages. See discussion about it in Section 5.

For the nonparametric mixed-effects CD4 model, we used linear combinations of natural cubic splines with percentile-based knots to approximate $r(t)$ and $h_i(t)$. We set $\psi_0(t) = \phi_0(t) \equiv 1$ and took the same natural cubic splines in the approximations (3) with $q \leqslant p$ in order to decrease the dimension of random effects. AIC and BIC criteria were used to determine the best values of p and q, leading to CD4 model (10). In addition, we standardized the CD4 counts and rescaled the original time (in days) into $[0, 1]$ in the analysis, in order to avoid too small or large estimates as they may be unstable. Since each of the study subjects started the treatment 10 days after his entry, the time to the first CD4/CD8 decrease after day 10 was taken as the time-to-event of interest. Given the CD4 and CD8 evaluation scheme, the times to the CD4/CD8 decrease in the analysis were interval-censored. The median of the observed interval length was 27 days, while seven subjects in the study did not experience the CD4/CD8 decrease. In practice, the right limits of the intervals are often used to approximate the event times with interval censored data to simplify the data analysis. The relatively wide time intervals

Table 1
Estimates (standard errors) of the parameters in models (9), (10), and (11)

Method	β_1	β_2	β_3	β_4	β_5	α_1	α_2	α_3	γ_0	γ_1	γ_2	γ_3	γ_4
Naive	11.65	60.59	6.42	−0.44	1.20	−0.45	2.08	0.09	−1.86	−0.10	0.05	0.02	−0.01
	(0.18)	(3.31)	(0.27)	(0.75)	(0.39)	(0.12)	(0.23)	(0.17)	(0.09)	(0.13)	(0.10)	(0.01)	(0.03)
MCEM	11.66	63.1	6.48	−1.37	3.14	−0.45	2.17	0.01	−1.69	0.04	0.06	0.01	0.01
	(0.19)	(4.06)	(0.32)	(0.86)	(0.63)	(0.14)	(0.29)	(0.23)	(0.13)	(0.25)	(0.15)	(0.01)	(0.05)
LAAP	11.64	63.23	6.47	−1.43	3.18	−0.45	2.18	0.01	−1.68	0.04	0.06	0.01	0.01
	(0.21)	(4.01)	(0.34)	(0.85)	(0.63)	(0.15)	(0.28)	(0.23)	(0.13)	(0.24)	(0.15)	(0.01)	(0.05)

Estimates of the variance-covariance parameters in models (9), (10), and (11):
MCEM: $\hat{\nu} = 0.36,\ \hat{\sigma} = 0.51,\ \hat{\lambda} = 0.58,\ \widehat{\text{diag}(A)} = (0.52, 1.55, 0.36),\ \widehat{\text{diag}(B)} = (1.02, 111.18, 2.10, 16.18)$.
LAPP: $\hat{\nu} = 0.35,\ \hat{\sigma} = 0.52,\ \hat{\lambda} = 0.57,\ \widehat{\text{diag}(A)} = (0.50, 1.52, 0.39),\ \widehat{\text{diag}(B)} = (1.07, 111.31, 1.97, 15.91)$.

here indicate that such approximation may not work well with the current AIDS data.

3.2 *Estimation and Interpretation*

The MCEM algorithm and the LAAP-based approach were applied to estimate the parameters in the three joint models for the viral dynamics, the CD4 process, and the time to CD4/CD8 decrease. The estimates, together with their estimated standard errors, are shown in Table 1. The two sets of parameter estimates seem quite close, while the LAAP approach took much less time to implement.

The estimates of the parameters in the CD4 model (10) suggest a significant overall increase in CD4 after the treatment, and both individual variation and measurement error in CD4 appear substantial. The two sets of estimates for the individual-specific initial viral decay rate $\lambda_{1ij} = \beta_2 + b_{i2}$ in the viral dynamics model (9) indicate that the overall first-phase viral decay is significant with a substantial variation across subjects; the estimates of the population parameter β_2 (standard errors) are 63.10 (4.06) (MCEM) and 63.23 (4.01) (LAAP), and the estimates of the variances of the random effect (individual effect) b_{i2} are 111.18 (MCEM) and 111.31 (LAAP). The second-phase viral decay rate (i.e., $\lambda_{2ij} = \beta_4 + \beta_5 z^*_{ij} + b_{i4}$) appears positively associated with true CD4 value (z^*_{ij}) over time; the estimates of β_5 (standard errors) are 3.14 (0.63) (MCEM) and 3.15 (0.63) (LAAP). This is consistent with the findings reported in Ding and Wu (2001) and Wu et al. (2008). Figure 2 shows 20 realizations of the viral load process generated from each of the fitted viral dynamics models along with the observed trajectories of the four randomly selected subjects whose data are presented in Figure 1.

Different from what was anticipated, the estimates for the parameters in the event time model (11) do not directly show that the time to CD4/CD8 decrease is highly associated with either the two viral decay rates or the CD4 changing rates over time. To further explore this, we generated 20 sets of the 46 study subjects' times to CD4/CD8 decrease and times to HIV RNA rebound from the fitted event time model and viral dynamics model based on the LAAP method, respectively. The corresponding 20 "lowess" curves, nonparametric estimates of the association of the event times, and the rebound times are plotted in Figure 3, along with the linear regression line. Here, we used the left limits of the observed intervals to approximate the event times. Figure 3 reveals a positive association of the event time with the viral rebound. We also estimated the survivor function of the event time by applying the Turnbull estimator, a nonparametric estimator. The two parametric estimates are close to each other and quite consistent with the nonparametric estimate.

For comparison, we also employed three "naive" approaches to estimate the model parameters: (i) using only baseline CD4 in models (9) and (11), (ii) estimating the parameters in the three models separately, and (iii) ignoring the measurement errors in CD4 and using the observed CD4 in the viral dynamics model (9). The first naive approach could not capture the dynamics association of the viral load and CD4 processes. Leaving alone its relatively large standard errors with the parameter estimates, the second naive approach obtained a negative correlation of CD4 with the second-phase viral decay, which is counter-scientific. The parameter estimates of the third naive approach are presented in Table 1, labeled by "Naive." The difference of naive estimates and MCEM/LAAP estimates, due to whether or not ignoring potential CD4 measurement errors, indicates that CD4 measurement errors cannot be ignored in the analysis.

4. A Simulation Study

In this simulation study, we evaluate the performances of the proposed methods (MCEM and LAAP), compare them with the naive method (Naive), and conduct a sensitivity analysis. The response NLME model and the event time model are the same as those in the real-data example, but with equal-spaced observation times between 0 and 1. The true values of model parameters used in the simulations are given in the tables for simulation results. The covariate model is carefully chosen to mimic the observed CD4 trajectories in the real dataset

$$z_{ij} = (-1.37 + 0.5w_{i1}) + (2.17 + 0.6w_{i2}) \sin(1.9t_{ij} + 0.39 + 0.2w_{i3}) + \xi_{ij}, \quad (12)$$

where w_{i1}, w_{i2}, and w_{i3} are independent and follow $N(0,1)$. We generated 100 independent datasets based on the above models with $N = 100$ and $n_i = 5, 10$. To conduct a sensitivity analysis for the time-to-event model, we consider the following three distributions for error ϵ_i in event time model (11): $\epsilon_i \sim N(0, 0.2^2)$, $\epsilon_i \sim Gumbel(0,1)$, and $\epsilon_i \sim Logistic(0,1)$. To summarize simulation results, we compute: EST (average of estimates), ASE (average of standard errors from each simulation), SSE (simulated standard errors), BIAS (the percent relative bias defined by $(\hat{\beta}_j - \beta_j)/|\beta_j| \times 100\%$), MSE (the

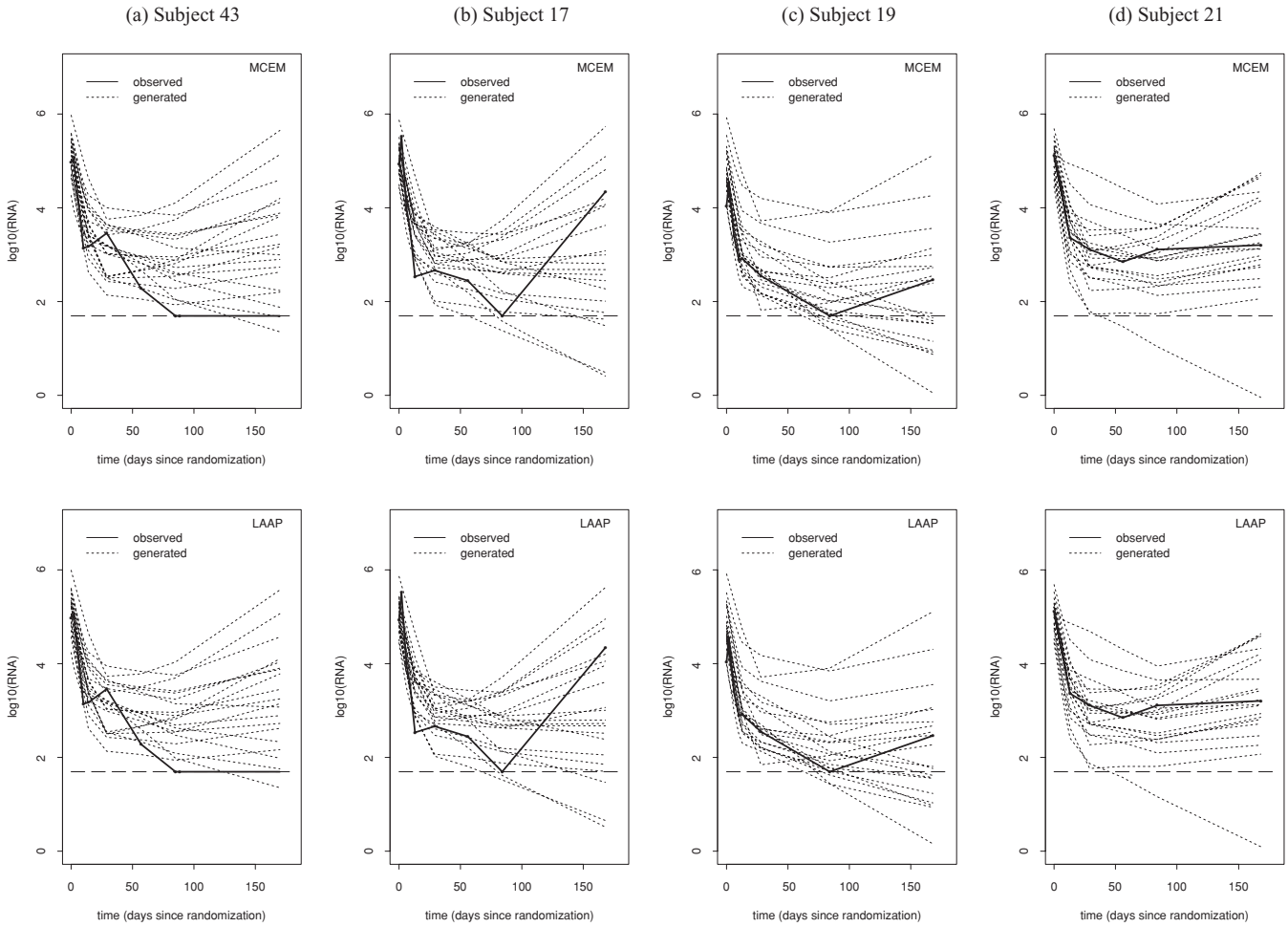

Figure 2. Observed and generated viral load trajectories of four randomly selected subjects.

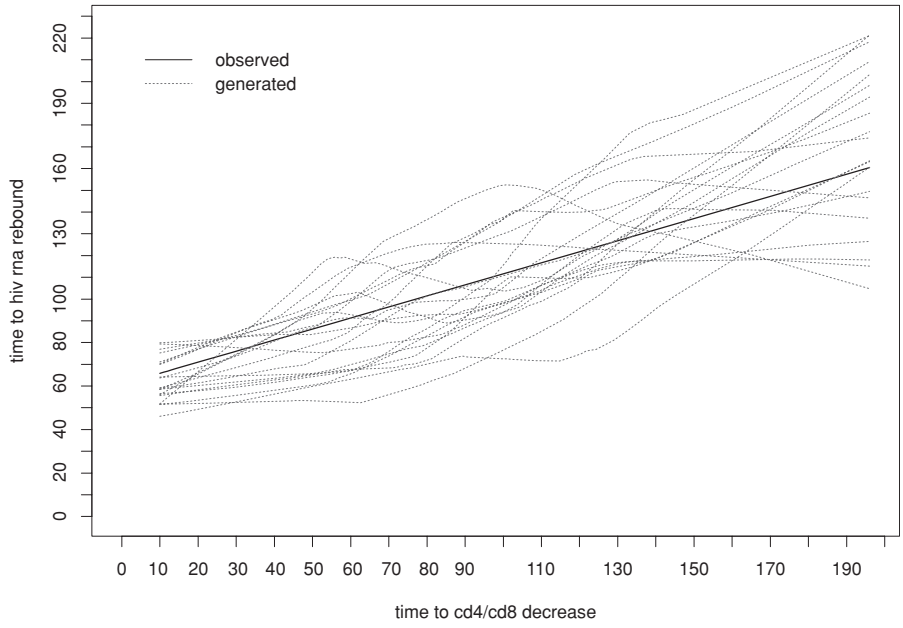

Figure 3. Association of times to CD4/CD8 decrease and viral load rebound.

Table 2
Simulation results based on 100 *repetitions with* $n_i = 10$ *and* $\epsilon_i \sim N(0, 0.2^2)$

Method	Estimate	β_1 12	β_2 63	β_3 6	β_4 −2	β_5 3	γ_0 −1	γ_1 0.4	γ_2 0.6	γ_3 0.3	γ_4 0.2
Naive	EST	12.02	59.27	5.87	−1.34	1.35	−0.86	0.28	0.44	0.22	0.15
	ASE	0.14	1.43	0.21	0.35	0.09	0.04	0.06	0.05	0.02	0.02
	SSE	0.14	1.47	0.21	0.39	0.11	0.04	0.07	0.06	0.02	0.02
	BIAS	0.16	−5.93	−2.15	33.18	−55.03	14.20	−29.28	−26.18	−27.46	−25.99
	MSE	1.20	6.35	4.16	37.53	55.11	14.80	33.04	27.38	27.95	27.89
	CRI	0.97	0.30	0.88	0.51	0.01	0.06	0.54	0.12	0.01	0.27
MCEM	EST	12.02	62.99	5.99	−2.01	3.00	−1.02	0.40	0.60	0.29	0.20
	ASE	0.14	1.25	0.20	0.29	0.14	0.14	0.13	0.10	0.05	0.07
	SSE	0.14	1.22	0.21	0.29	0.11	0.15	0.11	0.09	0.05	0.08
	BIAS	0.12	−0.01	−0.22	−0.73	−0.01	−2.40	−1.04	−0.07	−2.25	−1.00
	MSE	1.21	1.98	3.23	14.52	4.60	13.85	32.07	16.67	15.86	33.46
	CRI	0.95	0.95	0.94	0.96	0.98	1.00	0.95	0.96	0.99	0.97
LAAP	EST	12.03	62.74	5.98	−2.04	3.01	−1.03	0.38	0.59	0.29	0.20
	ASE	0.14	1.27	0.20	0.29	0.13	0.14	0.14	0.11	0.05	0.08
	SSE	0.14	1.24	0.21	0.28	0.11	0.12	0.15	0.09	0.05	0.08
	BIAS	0.23	−0.41	−0.35	−1.89	0.37	−2.52	−3.83	−1.89	−4.56	−1.77
	MSE	1.18	2.06	3.35	14.64	4.35	14.32	35.36	18.41	17.00	40.04
	CRI	0.96	0.94	0.96	0.95	0.95	0.97	0.96	0.94	0.95	0.94

percent relative root MSE defined by $\sqrt{(\hat{\beta}_j - \beta_j)^2 + \text{ASE}^2}/|\beta_j| \times 100\%)$, and CRI (coverage rates of 95% confidence intervals).

The MCEM, the LAAP, and the third naive approach were applied to estimate the model parameters using each of the generated datasets. The results based on 100 repetitions with $n_i = 10$ are presented in Table 2. Both MCEM and LAAP perform well, in terms of either bias or MSE or coverage rates, and their estimates are close to the true values. The MCEM approach appears slightly more efficient than the LAAP approach. The naive method performs the worst based on any of the criteria. The estimates obtained by the naive method are similar to the MCEM and LAAP estimates for the β parameters in the viral dynamic model (9), but show a trend of under-estimating the γ parameters, i.e., the covariate effects in the time-to-event model (11).

Sensitivity analyses with $\epsilon_i \sim Gumbel(0, 1)$ and $\epsilon_i \sim Logistic(0, 1)$ and simulation results with $n_i = 5$ are reported in Web Tables 1–3. It appears that both MCEM and LAAP are quite robust to the choice of the ϵ_i distributions in the current simulation settings. The LAAP performs reasonably well with few repeated measurements (i.e., $n_i = 5$).

5. Discussion

The computation associated with the joint inferences in joint models of longitudinal data and survival data can be extremely intensive and may lead to convergence problems (Tsiatis and Davidian, 2004; Wu et al., 2008). This is especially the case when nonlinear and semiparametric/nonparametric models are used and missing data or measurement errors are present. The LAAP-based method presented in this article offers a favorable alternative. It approximates the multidimensional integral in the observed-data joint likelihood by an analytic expression. Following Vonesh (1996) and Zeng and Cai (2005), we may establish the consistency and asymptotical normality of the resulting estimators when both the sample size and the number of within-individual measurements go to infinite.

Note that in this article the longitudinal and survival models are linked through shared random effects. The advantages of such an approach are that the link between the models is made clear, a computationally efficient approximate method can be easily implemented, and model interpretation may be reasonable in some applications. However, in some situations this formulation could become problematic, such as difficulty in interpretation, especially if the random effects structures are more complex. Recently, Tseng, Hsieh, and Wang (2005) considered an alternative approach for joint inference of AFT models and time-dependent covariates.

The performance of the LAAP approach with discrete responses such as binary responses may be less satisfactory since Laplace approximations may not be satisfactory with discrete function (Joe, 2008). For example, if the longitudinal responses and/or covariates are discrete and are modeled using GLMMs, a joint model approach similar to the one in this article may not perform well. In this case, a second-order Laplace approximation may be needed and may improve the approximation. Performances of methods based on Laplace approximations depend on both the sample size and the number of within-individual measurements (Vonesh, 1996).

As pointed out by the referees, an alternative event-time model is to link the event times to the unobserved true time-dependent covariates rather than the random effects in the assumed covariate model. Such a model is easy to interpret. However, in many practical situations such as the AIDS study that partly motivated this research, each of the covariate processes is not fully observed but only at a few time points, which along with the unobservable true covariate processes causes difficulties in implementing the model. Moreover, the current event-time model (5) can offer the following advantages: (i) the random effects in the covariate model summarize

the history of the covariate process, with individual-specific intercepts and rates of change, and the summary quantities are likely better predictors than the covariates at several particular times; (ii) the link between the three models is made clear by the shared random effects; (iii) it is easy to implement. In addition, when it is desirable, we can evaluate the effect of the true covariate process to the event time based on the fitted AFT model in combination of the covariate process model.

For joint inference, parameter (or model) identifiability can be an important but difficult problem when many model parameters must be estimated simultaneously. Thus, we prefer parsimonious models that contain fewer parameters, especially for models of secondary interest such as the covariate model whose parameters may be viewed as nuisance parameters. Fortunately, with nonlinear models such as the ones considered in this article, parameter identifiability is often less of a concern than with linear models (Carroll et al., 2006). In practice, if the models are not identifiable, the EM algorithm would diverge quickly. In the application considered in this article, the EM algorithm converged without problems and we did not observe potential identifiability problems.

We have focused on time-dependent covariates with measurement errors and missing data. To address measurement errors in time-independent covariates, validation data or replicates are needed (Carroll et al., 2006). Moreover, the Laplace approximation may become somewhat more complicated if the ranges of the covariate values are not the whole space due to the nature of Laplace approximation. When missing covariates are nonignorable, the missing data mechanism needs to be modeled and incorporated in the likelihood, sensitivity analyses should be performed, and then the proposed work can be extended straightforwardly.

6. Supplementary Materials

Web Appendix 1, Figure 1, and Tables 1–3 referenced in Sections 2.2, 3.1, and 4 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

The authors thank two referees, an associate editor, and an editor for their helpful comments and suggestions. The research was partially supported by grants from the U.S. National Institute of Allergy and Infectious Diseases and the Natural Sciences and Engineering Research Council of Canada.

References

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. 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**(1), 13–29.

Green, P. J. and Solverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

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.

Henry, W. K., Tebas, P., and Lane, H. C. (2006). Explaining, predicting, and treating HIV-associated CD4 cell loss. *Journal of the American Medical Association* **12,** 1523–1525.

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics* **55,** 13–22.

Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis* **52,** 5066–5074.

Ibrahim, J. G., Chen, M. H., Lipsitz, S. R., and Herring, A. H. (2005). Missing data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association* **100,** 332–346.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Boca Raton, Florida: Chapman & Hall/CRC.

Liu, W. and Wu, L. (2007). Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics* **63,** 342–350.

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

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed-effects models for unequally sampled noisy curves. *Biometrics* **57,** 253–259.

Stevens, R. A., Lempicki, R. A., Natarajan, V., Higgins, J., Adelsberger, J. W., and Metcalf, J. A. (2006). General immunologic evaluation of patients with immunodeficiency virus infection. In *Manual of Molecular and Clinical Laboratory Immunology*, 7th edition, B. Detrick, R. G. Hamilton, J. D. Folds, et al. (eds), 848–861. Washington, DC: ASM Press.

Tseng, Y. K., Hsieh, F., and Wang, J. L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika* **92,** 587–603.

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

Vonesh, E. F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* **83,** 447–452.

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, 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, L., Hu, X. J., and Wu, H. (2008). Joint inference for nonlinear mixed-effects models and time-to-event at the presence of missing data. *Biostatistics* **9,** 308–320.

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

Zeng, D. and Cai, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics* **33,** 2132–2163.

Received September 2008. *Revised April* 2009. *Accepted May* 2009.