Mixture Joint Models for Event Time and Longitudinal Data With Multiple Features

Yangxin Huang, Getachew A. Dagne, and Jeong-Gun Park

ABSTRACT
It often happens in longitudinal studies that repeated measurements of markers are observed with various data features of a heterogeneous population comprising of several subclasses, left-censoring due to a limit of detection (LOD) and covariates measured with error. Moreover, repeatedly measured markers in time may be associated with a time-to-event of interest. Inferential procedures may become very complicated when one analyzes data with these features together. This article explores a finite mixture of hierarchical joint models of event times and longitudinal measures with an attempt to alleviate departures from homogeneous characteristics, tailor observations below LOD as missing values, mediate accuracy from measurement error in covariate and overcome shortages of confidence in specifying a parametric time-to-event model with a nonparametric distribution. The Bayesian joint modeling is employed to not only estimate all parameters in mixture of joint models, but also evaluate probabilities of class membership. A real data example is analyzed to demonstrate the methodology by jointly modeling the viral dynamics and the time to decrease in CD4/CD8 ratio in the presence of CD4 cell counts with measurement error and the analytic results are reported by comparing potential models for various scenarios. Supplementary materials for this article are available online.

1. Introduction

Many studies aim at exploring the relationship between a time-to-event outcome and a longitudinal marker. For example, the relationship of HIV viral suppression and immune restoration after a treatment has received great attention in HIV/AIDS research (Fischl et al. 2003). One is often interested in simultaneously studying the HIV dynamics and immune restoration, which may be characterized by the time to CD4/CD8 ratio decline (Pawitan and Self 1993). Relatively little work has been published about statistical analysis for the particular association of HIV viral dynamics and time trend in CD4/CD8 ratio in the presence of CD4 covariate process. The research was motivated by an AIDS Clinical Trials Group study 388 (ACTG388, Fischl et al. 2003) to understand within-subject patterns of change in HIV-1 RNA copies (also referred to as viral load) or CD4 cell count, and to study the relationship of features of viral load and CD4 profiles with time to decrease in CD4/CD8 ratio.

Joint analysis of event times and longitudinal measures is an active area of biostatistics and statistics research. There have been a considerable number of statistical approaches in the literature. Researchers may often confront the task of developing inference, where longitudinal outcomes of interest may follow heterogeneous (not homogeneous) characteristics, suffer from left-censoring due to a limit of detection (LOD), and measure with substantial error. Modeling such data has many challenges due to the following issues of inherent data features. First, in the literature, most studies of longitudinal modeling assume that all subjects come from a homogeneous population where large between- and within-individual variations were accommodated by random-effects and/or time-varying covariates in the models. These typical between- and within-individual variations are shown in Figure 1(a), viral load trajectory profiles of three representative patients in ACTG388 (see Section 2 for details of this study and data description); these viral load trajectories can be roughly classified into three classes, which are biologically and clinically interpretable. We, therefore, can reasonably assume that these patients are from a population which consists of three relatively distinct classes and, thus, we consider a finite mixture of nonlinear mixed-effects (NLME) models for such dataset. Second, the outcome of a longitudinal study may be subject to LOD because of the low sensitivity of current standard assays (Perelson et al. 1997). For example, for the ACTG388 study, which was designed to collect data on every individual at each assessment, the response (viral load) measurements may be subject to left-censoring due to an LOD. It can be seen from Figure 1(a) that for some patients their viral loads are below LOD. The proportion of data censored may not be trivial, so failure to account for censoring in the analysis may result in significant biases in the estimates of the fixed-effects and variance components (Hughes 1999; Lachos et al. 2011). Third, measurement error in covariates is another typical feature of longitudinal data and ignoring this phenomenon may result in unreliable statistical inference. This is the usual case in the longitudinal studies; for instance, CD4 cell counts are often measured with substantial measurement error (Wu 2002; Liu and Wu 2007; Huang and Dagne 2011). Finally, it is a commonplace in clinical and public health research that longitudinal studies consist of repeated measurements on continuous variables, an
observation on a possibly censored time-to-event and additional covariate information collected for each subject. Interest often focuses on interrelationships among these variables (Tsaiatis and Davidian 2004). A joint model that links the event time to these longitudinal measures, which can also incorporate information about left-censoring due to LOD and measurement error in covariates, is becoming increasingly important in many applications.

Although joint modeling of both longitudinal and event time data has received a great deal of attention in the statistical literature in recent years, the majority of the statistical literature for the joint modeling of longitudinal and/or event time data has focused on the development of models that aim at capturing only specific aspects of the motivating studies. These specific data features considered in the literature include longitudinal data with, but are not limited to, censored response (Hughes 1999; Wu 2002; Thiebaut et al. 2005; Lachos et al. 2011), mixture response (Muthén and Shedden 1999; Pauler and Laird 2000; Lu and Huang 2014), covariate measurement error (Wu 2002; Liu and Wu 2007; Huang and Dagne 2011; Y. Liu, and Wu 2011), and longitudinal-survival (or event time) data (Brown and Ibrahim 2003; Tsaiatis and Davidian 2004; Wu, Liu, and Hu 2010; Guedj, Rodolphe Thiebaut, and Commenges 2010; Huang, Dagne, and Wu 2011; Mbogningabc, Bleakleyab, and Lavielleab 2014). However, there is relatively few studies on simultaneous inference for longitudinal data with features of heterogeneity and left-censoring due to LOD in response and measurement error in covariate as well as event time data incorporated. It is not clear how heterogeneity, left-censoring, covariate measurement error and event time of data may interact and simultaneously influence inferential procedures. Statistical inference and analysis complicate dramatically when all of these issues are present. In this article, we develop a mixture of mixed-effects joint model to investigate the effects on inference when all of these typical data features exist.

We employ a Bayesian framework for a mixture of nonlinear mixed-effects joint (MNLMEJ) models for longitudinal response and time-to-event in the presence of covariates with measurement error to accommodate a large class of data structures with various features. In particular, we consider the MNLMEJ model with the following components to quantify these data features: (i) a mixture of NLME models for viral load response process in connection with a Tobit model (Tobin 1958) to tailor observations below LOD; (ii) nonparametric mixed-effects model for CD4 covariate process to modulate accuracy from measurement error; and (iii) an accelerated failure time (AFT) model for time to decline of CD4/CD8 ratio using the nonparametric Dirichlet process (DP) prior as a distribution (Ferguson 1973; Ishwaran and Zarepour 2000), which is linked to longitudinal models through random effects, providing further bridge between time-to-event and longitudinal observations. To the best of our knowledge, there has been relatively little work done to address the simultaneous impact of longitudinal data with heterogeneity, left-censoring due to LOD, measurement error in covariate and event time by joint modeling the three processes above based on the Bayesian MNLMEJ model. It is noted that some models published in the literature (Huang and Dagne 2011; Huang, Dagne, and Wu 2011; Lu and Huang 2014) can be treated as special cases of MNLMEJ model proposed in this article when our model is tailored in this way; see detailed discussion in Appendix A. Approaching the mixture joint modeling problem from the Bayesian perspective is more natural and straightforward. It avoids many complicated approximations required by the frequentist approach. Yet, with noninformative priors, we can still get maximum likelihood-type estimates.

The remainder of this article evolves as follows. Section 2 describes the dataset that motivated this research and discusses the specific MNLMEJ models, where mixture models are formulated by three nonlinear mean functions of different mixture components for viral load response with left-censoring, CD4 covariate process with measurement errors and AFT model for time to decline of CD4/CD8 ratio. Simultaneous Bayesian inferential procedure for estimating both parameters and probabilities of class membership are presented in Section 3. In Section 4, we demonstrate to apply the proposed methodologies to an AIDS dataset described in Section 2 and report the analytic results. Finally, we conclude the article with discussion in Section 5.

2. Motivating Dataset and a Mixture of Joint Models for HIV Dynamics

2.1. Motivating Dataset

The dataset that motivated this research is from ACTG388 study (Fischl et al. 2003). ACTG388 study was a randomized,
open-label study comparing 2 different 4-drug regimens with a standard 3-drug regimen for 517 subjects with no or limited previous experience with antiretroviral treatment (ART) who had a CD4 cell count ≤ 200 cells/mm³ or a plasma HIV-1 RNA level ≥ 80,000 copies/mL at screening. The planned study duration was 72 weeks, which was subsequently increased to 96 weeks beyond the enrollment of the last subject. The plasma HIV-1 RNA (viral load) is repeatedly quantified at weeks 0, 4, 8, 16, and every 8 weeks until the last patient on study. The number of viral load measurements for each individual varies from 2 to 18. Out of total 517 patients, 65 subjects had no viral load measurements because all RNA assay results are not available for these subjects and 25 subjects had less than three viral load measurements (which is not suitable for classification) because of permanently discontinued study prior to study completion with various reasons, which were excluded from this study. The remaining 427 subjects whose viral load measurements vary from 3 to 18 were included in data analysis. Thirty-four percent of viral load observations were measured below the LOD of 25 copies/mL in our data. The CD4 and CD8 cell counts were also measured throughout study on a similar scheme. The detailed data description can be found in Fischl et al. (2003). The log_{10} transformation of viral load and square-root of CD4 were used in the analysis to make data more symmetric, to stabilize the variation of the measurement errors and to speed up estimation algorithm. In addition, to avoid too small or large estimates which may be unstable, we rescaled the original time (in days) so that the timescale is between 0 and 1.

Since viral load is an important marker for assessing virologic response of an ART regimen, our objective is to develop a mixture of joint modeling for characterizing population-level and individual-level viral load trajectories. As discussed in Section 1, viral load trajectory profiles of three representative patients displayed in Figure 1(a) can be roughly classified into three classes in terms of the effect of treatment on viral load responses: for class 1, patient’s viral loads decrease rapidly and constantly in a short-term period (dotted line with solid circle sign) because some patients withdrew too early to be clustered into either class 2 (with viral load suppression or without viral load rebound) or class 3 (with viral load rebound); it is noted that missing data due to early dropouts of patients occur in the response (viral load) measurements and are missing at random so that the missing data mechanism can be ignored in the analysis. For class 2, patient’s viral loads decrease at the beginning and then stay stable at a low level (solid line with circle sign). The classes 1 and 2 with suppression of plasma HIV-1 RNA levels indicate that the treatment can be thought successful without serious clinical problems arose, suggesting a confirmed virologic response. While for class 3 patient’s viral loads decrease at the beginning, but they experience viral load increase, which results in viral load rebound eventually (dashed line with plus sign), implying a virologic failure. Along with these observations, we can reasonably assume that patients from a population consist of three relatively homogeneous classes.

Figure 1 presents the trajectories of HIV viral load, CD4 cell count, and CD4/CD8 ratio of three representative subjects from ACTG388 study (Fischl et al. 2003). While the data indicate a close association of the longitudinal viral load, CD4 cell count, and CD4/CD8 ratio, the data show a large variation in the association across subjects and over time within each subject. This observation together with the published researches in the area (Wu and Ding 1999; Liu and Wu 2007; Wu, Liu, and Hu 2010; Huang, Dagne, and Wu 2011; Lu and Huang 2014) led us to propose the following MNLMEJ models for HIV dynamics in connection with three outcome variables of concern. Because our objective in this study is HIV viral dynamics, we choose viral load as a response variable and consider CD4 cell count as a covariate (Wu 2002; Liu and Wu 2007; Wu, Liu, and Hu 2010; Lachos et al. 2011) to develop MNLMEJ models below.

### 2.2. Measurement Error Models

Denote the number of subjects by \( n \) and the number of measurements on the \( i \)th subject by \( n_i \). Let \( z_{ij} = (z_{i1}, \ldots, z_{in_i})^T \), where \( z_{ij} \) is time-varying CD4 covariate for the \( i \)th subject at time \( t_{ij} \) (\( i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, n_i \)). As is evident from Figure 1(a), the interpatient variations in viral load appear to be large and these variations change over time. Previous studies suggest that the interpatient variation in viral load may be partially quantified by time-varying CD4 cell count (Wu 2002; Liu and Wu 2007; Huang and Dagne 2011). The covariate (CD4 cell counts) often have substantial measurement errors and ignoring these errors may lead to misleading results (Carroll et al. 2006). With CD4 measures collected over time, we may model the CD4 process to partially address the measurement errors (Wu 2002). 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 CD4 data as follows:

\[
\begin{align*}
    z_{ij} &= w(t_{ij}) + h(t_{ij}) + \epsilon_{ij} \\
    &= (z_{ij1}, \ldots, z_{ijn_i})^T \\
    \epsilon_{ij} &\sim N_1(0, \sigma_1^2 I_{n_i}),
\end{align*}
\]

where \( w(t_{ij}) \) and \( h(t_{ij}) \) are unknown nonparametric smooth fixed-effect and random-effect functions, respectively, and \( \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T \) follows a multivariate normal distribution. \( z_{ij} = w(t_{ij}) + h(t_{ij}) \) are the true (but unobservable) covariate values at time \( t_{ij} \). The fixed smooth function \( w(t) \) represents population average of the covariate process, while the random smooth function \( h(t) \) is introduced to incorporate the large inter-individual variation in the covariate process. We assume that \( h(t) \) is the realization of a zero-mean stochastic process.

Model (1) is more flexible than parametric mixed-effects models. To fit model (1), we apply a regression spline method to \( w(t) \) and \( h(t) \). The working principle is briefly described as follows and more details can be found in literature (Davidian and Giltinan 1995; Wu and Zhang 2006). The main idea of regression spline is to approximate \( w(t) \) and \( h(t) \) by using a linear combination of spline basis functions. For instance, \( w(t) \) and \( h(t) \) can be approximated by a linear combination of basis
functions $\Psi_p(t) = \{\psi_0(t), \psi_1(t), \ldots, \psi_{p-1}(t)\}$ and $\Phi_q(t) = \{\phi_0(t), \phi_1(t), \ldots, \phi_{q-1}(t)\}$, respectively. That is,

$$w(t) \approx w_p(t) = \sum_{l=0}^{p-1} \alpha_l \psi_l(t) = \Psi_p(t)^T \alpha,$$

$$h_i(t) \approx h_{iq}(t) = \sum_{l=0}^{q-1} \alpha_l \phi_l(t) = \Phi_q(t)^T \alpha_l,$$

(2)

where $\alpha = (\alpha_0, \ldots, \alpha_{p-1})^T$ is a $p \times 1$ vector of fixed-effects and $\alpha_l = (\alpha_0, \ldots, \alpha_{q-1})^T$ ($q \leq p$) is a $q \times 1$ vector of random-effects with $\alpha_l \sim N_q(0, \Sigma_q)$. Based on the assumption of $h_i(t)$, we can regard $\alpha_l$ as iid realizations of a zero-mean random vector. Substituting $w(t)$ and $h_i(t)$ by their approximations $w_p(t)$ and $h_{iq}(t)$, we can approximate model (1) by the following linear mixed-effects (LME) model.

$$z_{ij} \approx \Psi_p(t_{ij})^T \alpha + \Phi_q(t_{ij})^T \alpha_l + e_{ij} \approx z_{ij}^* + \epsilon_{ij},$$

$$e_{ij} \sim N_n(0, \sigma_i^2 I_n),$$

(3)

For this model, we consider natural cubic splines bases with the percentile-based knots. To select an optimal degree of regression spline and numbers of knots, that is, optimal sizes of $p$ and $q$, the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) is often applied (Davidian and Giltinan 1995). Following the study in Wu, Liu, and Hu (2010), we set $\psi_0(t) = \phi_0(t) = 1$ and take the same natural cubic splines in the approximation (2) with $q \leq p$ (to limit the dimension of random effects). The numbers of knots (see equation (2) in detail), $p$ and $q$, are determined based on the AIC/BIC criteria. For the ACTG388 data, the AIC/BIC values are evaluated for various models with $(p, q) = ((1,1), (2,1), (2,2), (3,1), (3,2), (3,3))$ which was found that the model with $(p, q) = (3,3)$ has the smallest AIC/BIC values being 5733.5/5749.8. We, thus, adopted the following CD4 data model:

$$z_{ij} \approx (\alpha_0 + \alpha_1 t_{ij}) \psi_1(t_{ij}) + (\alpha_1 + \alpha_2) \psi_2(t_{ij}) + e_{ij} \approx z_{ij}^* + \epsilon_{ij},$$

(4)

where $z_{ij}$ is the observed $\sqrt{CD4}$ value at time $t_{ij}$, $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are two basis functions described in Equation (2), $\alpha = (\alpha_0, \alpha_1, \alpha_2)^T$ is a vector of population (fixed-effects) parameters, $\alpha_l = (\alpha_0, \alpha_1, \alpha_2)^T$ is a vector of random-effects.

### 2.3. A Mixture of NLMet Tobit Models for HIV Dynamics

Let $y_{ij}$ be the value of viral load response for the individual $i$ at time $t_{ij}$ ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n_i$). To introduce the Tobit model to deal with observations below LOD in our mixture joint modeling framework, denote the observed value $y_{ij}$ by $(q_{ij}, d_{ij})$, where $d_{ij}$ is the censoring indicator and $q_{ij}$ is the latent response variable. The latent $d_{ij}$ is observed, as $y_{ij}$, if and only if $y_{ij} > \rho$ (a known constant LOD). When $q_{ij}$ is observed we have $d_{ij} = 0$. Otherwise we have $d_{ij} = 1$. Let $y_i = (y_{i1}, \ldots, y_{in_i})^T$, $q_i = (q_{i1}, \ldots, q_{in_i})^T$ and $d_i = (d_{i1}, \ldots, d_{in_i})^T$. Finite mixture models for longitudinal studies were considered in the literature (Muthén and Shedden 1999; Pauer and Laird 2000), where the latent classes corresponding to the mixture components and cluster individuals may provide a better inference. However, most finite mixture models for longitudinal data are currently based on linear (polynomial) (Muthén and Shedden 1999) or piecewise linear (Pauer and Laird 2000) mean functions. The partial reason is that the computation can be conveniently carried out because the likelihood function of a model based on these “linear” mean functions has a closed form (Muthén and Shedden 1999). However, in practice, most longitudinal trajectories appear to be nonlinear patterns. When a mixture model is extended to incorporate nonlinear mean functions which will be conducted in this article, inferential procedures complicate dramatically because a closed form of likelihood function no longer exists.

We assume that there are $K$ plausible nonlinear trajectory classes with mean functions $g_k(\cdot)$ ($k = 1, \ldots, K$), which are known to be specified (in our case conducted $K = 3$). The true trajectory mean function of the $i$th subject might be $g_k(\cdot)$ with unknown probability $\pi_k = P(c_i = k)$ which satisfies $\sum_{k=1}^{K} \pi_k = 1$, where $c$ is a latent indicator. For the response process with left-censoring, an NLMet model for $i$th subject, given $c_i = k$, can be formulated by

$$(y_i|c_i = k) = g_k(t_i, A_k \beta_i) + e_i, \quad e_i \sim N_n(0, \sigma^2_i I_n),$$

(5)

where $t_i = (t_{i1}, \ldots, t_{im})^T$, the vector of random errors $e_i = (e_{i1}, \ldots, e_{im})^T$ follows a multivariate normal distribution with mean zero and unknown variance parameter $\sigma^2_i$, and $\beta_i = (\beta_{ij}, j = 1, 2, \ldots, n_i)$ with $\beta_i = (\beta_{ij}, \ldots, \beta_{ij})^T$ being an $s$-dimensional vector of individual parameters for the $i$th subject, which is expressed by

$$\beta_{ij} = Z_{ij} \beta + X b_i,$$

(6)

where $\beta = (\beta_1, \ldots, \beta_s)^T$ is a $r$-dimensional vector of population parameters, $Z_{ij}$ ($s \times r$) is a design matrix including time-independent and/or time-varying covariates such as CD4 cell count, $X$ is an $s \times m$ matrix indicator to determine how random effects are specified, and the $m$-dimensional vector of random-effects $b_i = (b_{i1}, \ldots, b_{im})^T \sim N_m(0, \Sigma_b)$ with an unknown variance–covariance matrix $\Sigma_b$ ($m \times m$). We assume one of the elements in the $Z_{ij}$, $z_{ij}$ in model (3), is a summary of true (but unobservable) time-varying covariate value at time $t_{ij}$. It is noted that $r \geq s \geq m$; in particular, for real-data application below, it is seen that $r = 6, s = 5$ and $m = 4$. $g_k(t_i, A_k \beta_i) = (g_{k1}(t_i, A_k \beta_i), \ldots, g_{k6}(t_{im}, A_k \beta_{i6}))^T$, and $A_k$ ($k = 1, \ldots, K$) is known $s \times s$ square-matrix indicator, of which diagonal elements are either 0 or 1 and off-diagonal elements are all 0. $A_k$ is introduced because the mean functions of $y_{ij}$ ($i = 1, \ldots, n$; $j = 1, \ldots, n_i$), specified by the nonlinear functions $g_k(\cdot)$, may only involve different subsets of $\beta_i$. By introducing $A_k$, $A_k \beta_i$ will set unrelated elements of $\beta_i$ to 0 in the $k$th trajectory class, respectively. We will illustrate the choice of $A_k$ and specify nonlinear mean functions $g_k(\cdot)$ in the HIV dynamic application below.

Similar to discussion in Pauer and Laird (2000) and Lu and Huang (2014), model (5) can be specified conditionally and marginally, respectively, by

$$(y_i|c_i = k) \sim N_n(g_k(t_i, A_k \beta_i), \sigma^2_i I_n),$$

(7)

$$y_i \sim \sum_{k=1}^{K} \pi_k N_n(g_k(t_i, A_k \beta_i), \sigma^2_i I_n).$$

(8)
Thus, Equation (8) along with the Tobit formulation forms the finite mixture of NLME Tobit models for response variable with missing values due to LOD in the presence of measurement error model (3). In (8), the vector of mixture probabilities \( \pi = (\pi_1, \ldots, \pi_K)^T \) can be also viewed as the mixture weights of all plausible components within the finite mixture model framework. Model (8) is identifiable, as long as each of the component models is identifiable and distinguishable from each other; when the component models are identifiable but not distinguishable from each other, some constraints may be required to make model (8) identifiable (Titterington, Smith, and Makov 1985).

For the viral load response in HIV dynamics, based on discussion in Appendix A, we consider one- and two-compartment models with constant decay rate(s) for trajectory classes 1 and 2, respectively, and a two-compartment model with a time-varying decay rate in the second compartment for trajectory class 3 described above. Toward this end, the mean functions of \( K = 3 \) components in the mixture model are specified by

1. One-compartment model with a constant decay rate for class 1 trajectory
   \[
   g_1(t_{ij}, A_1, \beta_{1ij}) = \log_{10}(e^{p_{u-\lambda_{1ij}}}),
   \]  
   (9)

2. Two-compartment model with constant decay rates for class 2 trajectory
   \[
   g_2(t_{ij}, A_2, \beta_{ij}) = \log_{10}(e^{p_{u-\lambda_{1ij}} + e^{p_{u-\lambda_{2ij}}}}),
   \]  
   (10)

3. Two-compartment model with constant and time-varying decay rates for class 3 trajectory
   \[
   g_3(t_{ij}, A_3, \beta_{ij}) = \log_{10}(e^{p_{u-\lambda_{1ij}} + e^{p_{u-\lambda_{ij}}}}). \]  
   (11)

In (9)–(11),

\[
\begin{align*}
\beta_{1i} &= \beta_1 + b_{1i}, \\
\beta_{2i} &= \lambda_{1i} + \beta_2 + \beta_3 z_{0i} + b_{2i}, \\
\beta_{3i} &= \beta_3 + b_{3i}, \\
\beta_{5ij} &= \lambda_{2ij} + \beta_5 + \beta_6 z_{ij}^* + b_{4i}, \\
\beta_{ij} &= (\beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{5ij})^T, \\
\beta &= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)^T, \\
b_i &= (b_{1i}, \ldots, b_{4i})^T, \\
A_1 &= \text{diag}(1, 1, 0, 0, 0), \\
A_2 &= \text{diag}(1, 1, 1, 1, 0), \\
A_3 &= \text{diag}(1, 1, 1, 0, 1),
\end{align*}
\]

\[
Z_{ij} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & z_{0i} & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & z_{ij}^*
\end{pmatrix}
\quad \text{and} \quad
X = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0
\end{pmatrix},
\]  
(12)

where \( z_{0i} \) is the baseline \( \sqrt{CD4} \) and \( E(z_{ij}) = z_{ij}^* \) is true (but unobservable) value of \( \sqrt{CD4} \) at time \( t_{ij} \) defined in (4). The decay rate of the second compartment in (11), \( \beta_{5ij} \), is time-varying due to \( z_{ij}^* \), but other parameters in \( \beta_{ij} \) are time independent. As mentioned previously, the mean functions in different components may involve different subsets of \( \beta_{ij} \); for example, \( g_1(\cdot) \) only involves parameters \( \beta_{1ij} \) and \( \beta_{2ij} \), and \( g_2(\cdot) \) and \( g_3(\cdot) \) share the same parameters \( \beta_{1ij}, \beta_{2ij}, \) and \( \beta_{3i} \) but have different second-phase decay rate, \( \beta_{4ij} \) and \( \beta_{5ij} \), respectively. The diagonal indicator matrices, \( A_1, A_2, \) and \( A_3 \), determine which elements of \( \beta_{ij} \) are involved and set other unrelated parameters to be 0 in the mean functions, \( g_1(\cdot), g_2(\cdot), \) and \( g_3(\cdot), \) respectively. It is noted that (11) is a natural extension of (10) to consider a time-varying decay rate for capturing viral rebound in class 3 trajectories. With this mixture clustering, our mixture modeling can be used to estimate probabilities of class membership which is either viral rebound eventually, suggesting virologic failure (class 3) or viral decrease continuously, indicating a confirmed shorter-term virologic suppression (class 1) and longer-term virologic suppression (class 2).

### 2.4. Time-to-Event Models with an Unspecified Distribution

Various accelerated failure time (AFT) models for time-to-event were investigated in the literature (Tsatis and Davidian 2004; Wu, Liu, and Hu 2010). However, the common assumption of distributions for model errors is log-normal and this assumption may lack the robustness and/or may violate the agreement with observed data. Thus, statistical inference and analysis with log-normal assumption may lead to misleading results. We consider AFT models with error term to have a nonparametric prior distribution, which follows a Dirichlet process (DP) prior (Ferguson 1973; Ishwaran and Zarepour 2000). The event time \( \tau_i \) is likely related to the longitudinal response and covariate processes. We specify the association by assuming that, conditional on the random-effects \( a_i \) and \( b_i \), the event time is related to the longitudinal processes through the random effects that characterize the individual-specific effects (Wu, Liu, and Hu 2010; Huang, Dagane, and Wu 2011).

Assume that \( \tau_i \) is time to the first decline in the CD4/CD8 ratio for the \( i \)th subject \((i = 1, 2, \ldots, n)\). We are interested in the association of time to immune suppression of the individual-specific initial viral decay rates and the true CD4 trajectory, which are characterized by the random effects in the viral load response and CD4 covariate models. We may view the (unobservable) random effects as error-free covariates in time-to-event models. The associated inference 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 (Tsatis and Davidian 2004), especially when the event times are interval censored. To focus on the primary interests discussed in the article, we consider the following AFT model for time to first decline of the CD4/CD8 ratio.

\[
\begin{align*}
\ln(\tau_i) &= Y^T \tau_i + \epsilon_i, \\
\epsilon_i &\sim G(\cdot), \\
G(\cdot) &\sim \text{DP}(\eta G_0), \\
G_0 &\sim N(\xi_0, \sigma^2_0), \\
\eta &\sim \Gamma(\eta_{10}, \eta_{20}).
\end{align*}
\]  
(13)
the other parameter of the DP prior, $\eta$, reflects the degree of closeness of $G(\cdot)$ to the prior mean $G_0$. Large values of $\eta$ make $G(\cdot)$ very close to $G_0$, while small values of $\eta$ allow $G(\cdot)$ to deviate from $G_0$. This nonparametric analysis of specifying the distribution of $\varepsilon$ is robust to misspecification of the distributional assumptions about the model error. To implement the DP prior $DP(\eta G_0)$, a popular way for specifying the DP prior is the stick-breaking prior representation; see Appendix A for detailed discussion.

In model (13), the random-effects $b_{2i}$ and $b_{4i}$ represent individual variations in the first- and second-phase viral decay rates, respectively, so they may be predictive for event times. While $b_{3i}$ and $b_{1i}$ represent variations in the baseline viral loads, they do not appear to be highly predictive of event time, so they are excluded from the model to reduce the number of parameters. Model (13) offers the following advantages: (i) the random effects in the covariate model summarize the history of the covariate process and the summary quantities are likely better predictors than the covariates at several particular times; (ii) the random effects in the response model summarize individual variations in the first- and second-phase viral decay rates to predict time to immune suppression; (iii) the link among the three models is made clear by the shared random effects; (iv) the model error is assumed to be more flexible and unspecified distribution that has DP prior; (v) it is easy to implement.

Model (13) may be a good choice when the event times are thought to depend on individual-specific longitudinal trajectories, such as initial intercepts and slopes, or summaries of the longitudinal trajectories, and it is closely related to so-called shared parameter models (DeGruttola and Tu 1994). As pointed out in the studies by Tsatis and Davidian (2004) and Wu, Liu, and Hu (2010), in some practical situations, the event times cannot be observed but are only known as being contained in some time intervals, that is, being interval-censored. In the AIDS study mentioned previously, for example, given that CD4 and CD8 were collected at a finite number of times, we can only know that the time to the first occurrence of CD4/CD8 decrease in a subject is between two data collection time points. The observed event time data are then $(u_i, v_i), i = 1, \ldots, n$, where $(u_i, v_i)$ is the smallest observed interval containing $\zeta_i$. We take $u_i$ as the subject’s latest time in the study and $v_i = \infty$ if the $i$th subject did not experience the event of interest during the whole study period.

3. Simultaneous Bayesian Inferential Approach

In a longitudinal study, such as the AIDS study described previously, the longitudinal response, the covariate processes, and the time-to-event are usually connected physically or biologically. Models (3), (6), and (8) in connection with the time-to-event (13) define the MNLMEJ models considered in this article. They can be jointly modeled through the shared random effects. Statistical inference on all of the model parameters can then be made simultaneously. Although a simultaneous inferential method based on a joint likelihood for the response, covariate, and event time data may be favorable, the computation associated with the simultaneous likelihood inference in the MNLMEJ models can be extremely complex and may lead to convergence problems; in some cases it can even be computationally infeasible (Wu 2002; Tsatis and Davidian 2004; Thiebaut et al. 2005; Liu and Wu 2007). Here, we propose a simultaneous Bayesian inferential method via Markov chain Monte Carlo (MCMC) procedure to estimate both probabilities of class membership and all model parameters. The MNLMEJ models-based Bayesian inferential approach may pave a way to balance the computational burdens and convergence problems for such complex model setting.

Let $\theta = (\alpha, \beta, \gamma, \sigma^2_0, \Sigma_{\alpha}, \Sigma_{\beta})$ be the collection of unknown parameters in the MNLMEJ models except for the mixture weight $\pi$ in (8). Under the Bayesian framework, we next need to specify prior distributions for all the unknown parameters in the MNLMEJ models as follows:

\[
\begin{align*}
\alpha & \sim N_1 (\tau_1, \Lambda_1), \\
\beta & \sim N_4 (\tau_2, \Lambda_2), \\
\gamma & \sim N_6 (\tau_3, \Lambda_3), \\
\sigma^2_0 & \sim IG (\omega_1, \omega_2), \\
\sigma^2 & \sim IG (\omega_3, \omega_4), \\
\Sigma_{\alpha} & \sim IW (\Omega_1, v_1), \\
\Sigma_{\beta} & \sim IW (\Omega_2, v_2),
\end{align*}
\]

(14) where the mutually independent normal ($N$), inverse gamma (IG), and inverse Wishart (IW) prior distributions are chosen to facilitate computations. The super-parameter matrices $\Lambda_1, \Lambda_2, \Lambda_3, \Omega_1$, and $\Omega_2$ can be assumed to be diagonal for convenient implementation. By its definition, the latent indicating variables $c_i (i = 1, \ldots, n)$ follow a Categorical distribution (Cat)

\[
c_i \overset{iid}{\sim} Cat((1, 2, 3), (\pi_1, \pi_2, \pi_3)),
\]

(15) in which $\pi = (\pi_1, \pi_2, \pi_3)^T$ follows a Dirichlet distribution (Diebolt and Robert 1994),

\[
\pi \sim Dir (\phi_1, \phi_2, \phi_3).
\]

(16) Under the umbrella of the MNLMEJ models (3), (6), (8), and (13), the MCMC procedure consists of the following two iterative steps:

(i) Sampling class membership indicators $c_i (i = 1, \ldots, n)$, conditional on population parameters, $\theta$, and individual random-effects, $a_i$ and $b_i$.

Generate $c_i (i = 1, \ldots, n)$ from

\[
P(c_i = k | a_i, b_i, \theta, y_i) = \frac{\pi_k f(y_i | a_i, b_i, c_i = k, \theta)}{\sum_{m=1}^{3} \pi_m f(y_i | a_i, b_i, c_i = m, \theta)},
\]

(17) where $f(y_i | a_i, b_i, c_i = k, \theta)$ ($k = 1, 2, 3$) is a conditional density function of $y_i$ based on (7). Then, the probability $\pi$ can be updated from the following distribution for next iteration:

\[
(\pi | num_1, num_2, num_3) \sim Dir (\phi_1 + num_1, \phi_2 + num_2, \phi_3 + num_3).
\]

(18) where $num_k = \sum_{i=1}^{n} I(c_i = k), (k = 1, 2, 3)$, in which $I(\cdot)$ is an indicator function.

(ii) Sampling parameters $\theta$, and individual random-effects $a_i$ and $b_i$, conditional on class membership indicator $c = (c_1, \ldots, c_n)^T$.

It can be shown, conditional on $c_i$ determined in step (i), that $z_i$ and $y_i$ (in the presence of left-censoring) with respective
random effects \( a_i \) and \( b_i \) in conjunction with the time-to-event model (13) can be hierarchically formulated as
\[
\begin{align*}
z_{ij} | a_i \sim N_n \left( z_{ij}^*, \sigma_1^2 I_n \right), & \quad \tau_i \sim N_1 \left( 0, \Sigma_{\tau} \right), \\
y_{ij} | a_i, b_i, c_i \sim N_n \left( g_c \left( t_i, A_i, \beta \right), \sigma_2^2 I_n \right), & \quad b_i \sim N_1 \left( 0, \Sigma_{b} \right), \\
\tau_i | \gamma_i \sim G \left( \ln(t_i) - \gamma_i^{\top} \right), & \quad (19)
\end{align*}
\]
where \( z_{ij}^* = (z_{ij1}, \ldots, z_{ijn})^\top \).

Let \( f(\cdot) \), \( F(\cdot) \), and \( h(\cdot) \) denote a probability density function (pdf), cumulative density function (cdf), and prior density function, respectively. Conditional on the random variables and some unknown parameters, the specification of the Tobit model, a detectable measurement \( y_{ij} \) contributes \( f(y_{ij} | b_i, a_i, c_i) \), whereas a nondetectable measurement contributes \( F(\rho | b_i, a_i, c_i) \equiv P(y_{ij} < \rho | b_i, a_i, c_i) \) in the likelihood. We assume that \( a_i, \beta, \gamma, \alpha_i, \sigma_1^2, \Sigma_{\tau}, \Sigma_{b} \) are independent of each other, that is, \( h(\theta) = h(\alpha) h(\beta) h(\gamma) h(\alpha_i^2) h(\sigma_1^2) h(\Sigma_{\tau}) h(\Sigma_{b}) \).

After we specify the models for the observed data denoted by \( \mathcal{Y} = \{ y_{ij}, x_i, z_{ij}, \tau_i, u_i, v_i \} \), \( i = 1, \ldots, n \) and the prior distributions for the unknown model parameters, we can make statistical inference for the parameters based on their posterior distributions under Bayesian framework. Thus, the joint posterior density of \( \theta \) based on the observed data \( \mathcal{Y} \) and classification indicator \( c \) can be given by
\[
f(\theta|\mathcal{Y}, c) \propto \prod_{i=1}^n \int L_y f(z_{ij}|a_i) F^{\tau}(u_i, v_i | \tau_i) \\
\times f(a_i) f(b_i) da_i db_i \ h(\theta), \quad (20)
\]
where \( F^{\tau}(u_i, v_i | \tau_i) = G(\ln(u_i) - \gamma_i^{\top} \tau_i) - G(\ln(u_i) - \gamma_i^{\top} \tau_i); \)
\( L_y = \prod_{i=1}^n f(y_{ij} | b_i, a_i, c_i)^{1-d_i} F(\rho | b_i, a_i, c_i)^{d_i} \) is the likelihood for the observed response data in which \( d_i \) is the censoring indicator such that \( y_{ij} = q_{ij} \) is observed if \( d_i = 0 \) and \( y_{ij} \) is left-censored (i.e., treated as missing) if \( d_i = 1 \).

In general, the integrals in (20) are of high dimension and do not have a closed form. Analytic approximations to the integrals may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of \( \theta \) based on the observed data and class membership. As an alternative, the MCMC procedure can be used to sample population parameters, \( \theta \), and random effects, \( a_i \) and \( b_i \), as well as latent class indicator \( c_i \) \( (i = 1, \ldots, n) \), from conditional posterior distributions, based on (20), using the Gibbs sampler along with the Metropolis-Hastings (M-H) algorithm (Huang, Liu, and Wu 2006). Steps (i) and (ii) considered in the MCMC procedure above are repeated alternatively in iterations of MCMC procedure until convergence is reached. An important advantage of the above representations based on the hierarchical models is that they are easily implemented using the freely available WinBUGS software (Lunn et al. 2000) interacted with a function called bugs in a package R2WinBUGS of R. Note that when WinBUGS software is used to implement our modeling approach, it is not necessary to explicitly specify the full conditional posterior distributions or proportional functions of the density functions of full conditional posterior distributions for parameters to be estimated. Although their derivations are straightforward by working the complete joint posterior, some cumbersome algebra will be involved (Gelfand et al. 1990). As examples, the full conditional posterior distributions of representative parameters are presented in Appendix B.

\section{Analysis of AIDS Clinical Data}

\subsection{Model Implementation}

We conducted the following scenarios of modeling comparison. First, we proposed the joint modeling (JM) approach based on the NLME model (denoted as Model I) in Section 3. Here, we estimated the model parameters by using the “naive” method (denoted by NM), which ignores measurement error in CD4 covariate. That is, the “naive” method uses only the observed CD4 value \( z_{ij} \) rather than expected (unobservable) CD4 value \( E(z_{ij}) = z_{ij}^* \) in the mixture model for viral load response (denoted by Model II). We used it as a comparison to the JM approach. This comparison attempted to investigate how the measurement errors in CD4 contribute to modeling results. Second, we further investigated a commonly used NLME joint model (Model III, ignoring data feature of heterogeneous population) where the mean function is specified by (11) alone. This NLME model has been widely applied to analyze viral load data (Wu 2002; Wu and Zhang 2006; Liu and Wu 2007; Huang, Chen, and Yan 2012) in significantly advancing the understanding of pathogenesis of HIV infection and the assessment of effectiveness of ART. We compared Model III with Model I proposed in this article to explore how heterogeneous feature influences modeling results.

To carry out the Bayesian inference, we took weakly informative prior distributions for the parameters which are detailed in Appendix C. The MCMC sampler was implemented using WinBUGS software (Lunn et al. 2000) interacted with R2WinBUGS of R, and the program code for mixture joint model is available in Appendix D. When the MCMC procedure was applied to the actual clinical data, convergence of the generated samples was assessed using standard tools within WinBUGS software such as trace plots and Gelman–Rubin (GR) diagnostics (Gelman and Rubin 1992). Figure I in Appendix C shows the trace plots, autocorrelation, and dynamic version of GR diagnostic plots based on Model I for the representative parameters \( \alpha_2, \beta_2 \) and \( \sigma_2^2 \). We observed from the plots of trace (top panel), autocorrelation (middle panel) and GR diagnostics (bottom panel) that the algorithm has approached convergence. When these criteria suggested the convergence of chains, we proposed that, after an initial number of 50,000 burn-in iterations of three chains of length 100,000, every 50th MCMC sample was retained from the next 50,000 for each chain. Thus, we obtained a total of 3,000 samples of targeted posterior distributions of the unknown parameters for statistical inference. The computational burden for fitting such mixture of highly nonlinear joint models via MCMC procedure was reasonable. For example, to fit NLMEJ model using JM approach, it took about 9 hr on a Windows PC with Intel Core i7-2600 CPU @ 6.80 GHz and RAM of 16 GB.

\subsection{Results of Data Analysis}

The Bayesian joint modeling approach in conjunction with the three models for the viral load response, the CD4 covariate and
the time to first decline of CD4/CD8 ratio with different scenarios was used to fit the data. From the model fitting results, we have seen that, in general, all the models provided a reasonably good fit to the observed viral load data above LOD for most patients in our study, although the fitting for a few patients was not completely satisfactory due to unusual data fluctuation patterns for these patients, particularly for Models II and III. As an example, the three representative individual estimates of viral load trajectories based on Models I (solid lines), II (dashed lines), and III (dotted lines) are displayed in Figure 2. The following findings are observed from joint modeling. The estimated individual trajectories for Model I fit the originally observed values above LOD more closely than those for Models II and III which are comparable. The predicted values where the viral load observations are below LOD appear quite different between Model I and the other two models (see detailed discussion and Figure 3).

To assess the goodness of fit of the proposed models, the diagnosis plots of the fitted values versus the observed values (top panel) and normal Q–Q plots (bottom panel) from Models I, II, and III are presented in Figure II in Appendix C. It was seen from the top panel that Model I provides a better fit to the observed viral load data, compared with Models II and III which perform similarly. This finding can also be explained by examining the Q–Q plots of the residuals (bottom panel) that all plots show the existence of outliers, but it is clearly seen that Model I has few outliers and thus, provides a better goodness of fit to the data than Models II and III.

![Figure 2](image-url)

**Figure 2.** Individual fitted curves of viral load for three representative patients based on Models I (solid lines), II (dashed lines), and Model III (dotted lines). The observed values are indicated by circle “•”. The horizontal lines represent LOD at $1.398 = \log_{10}(25)$.

![Figure 3](image-url)

**Figure 3.** Histograms of inaccurate raw data (a) below LOD (dash line), and predicted values of viral load below LOD based on Models I (b), II (c), and III (d).
The population posterior mean (PM), corresponding standard deviation (SD), and 95% credible interval (CI) for fixed-effects parameters based on Models I, II, and III with two methods are presented in Table 1 and 2. The following findings are observed for estimated results of parameters.

In the mixture of NLME response model with the three components (9)–(11), the findings, particularly for the fixed-effects ($\beta_5$, $\beta_6$), which are parameters related to the second-phase viral decay rate, show that these estimates are different from zero for the three Models since the 95% credible intervals do not contain zero. Nevertheless, for the estimate of the coefficient of CD4 covariate $\beta_6$, although its estimate based on Model I (0.098) is much smaller than the counterpart based on Models II and III (0.245 and 0.157), its estimate is always significantly positive. This means that CD4 has a significantly positive effect on the first-phase viral decay rate. The fixed-effects ($\beta_2$, $\beta_3$), which are parameters of the first-phase viral decay rate, show that the estimate of $\beta_3$, the coefficient of baseline CD4 count, is significantly positive, indicating that the baseline CD4 has positive effect on the first-phase viral decay rate.

For parameter estimates of the CD4 covariate model (4), the estimates of the coefficients based on the three models show slight differences. The estimated linear coefficient $\alpha_1$ is significantly different from zero with a positive value, suggesting that there is a positive relation between CD4 cell count and measurement time. This result may be explained by the fact that CD4 cell count may increase during treatment, and in turn, indicates an overall improvement of AIDS progression in this AIDS trial study.

The estimates of parameters in the AFT model (13) do not directly show that the time to CD4/CD8 decrease is highly associated with either the two viral decay rates (due to insignificant estimates of $\gamma_1$ and $\gamma_2$) or the CD4 changes (due to insignificant estimates of $\gamma_1$ and $\gamma_2$) over time, which is different from what was anticipated. This finding is consistent with the study by Wu, Liu, and Hu (2010) in which further explanations are provided.

For comparison, we employed the NM to estimate the model parameters presented in Table 1 and 2 using the observed CD4 and ignoring the CD4 measurement error in the mixture of NLME response model with the three components (9)–(11). It is seen from estimates of ($\beta_2$, $\beta_3$), which depend on whether or not to ignore potential CD4 measurement error, that the differences between the naive estimates ($-3.157$, 0.098) of ($\beta_2$, $\beta_3$) and the joint modeling estimates ($-5.693$, 0.245), indicate that CD4 measurement error cannot be ignored in the analysis.

We further investigated a commonly used NLME joint model (Model III, where data feature of heterogeneous population is ignored). We compared Model III with Model I to explore how heterogeneous feature influences modeling results. We found the important differences in the estimates of parameters ($\beta_5$, $\beta_6$), which are associated with the second-phase viral decay rate. It is noted that one advantage of Model I over Model III is its flexibility to provide not only estimates of all model parameters, but also evaluate class membership probabilities at both population and individual levels, which is helpful for clinicians to develop individualized treatment.

### Table 1. Summary of estimated posterior mean (PM) of population (fixed-effects) parameters, the corresponding standard deviation (SD), and lower limit ($L_{CI}$) and upper limit ($U_{CI}$) of 95% equal-tail credible interval (CI).

| Method | Model | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\sigma_1^2$ | $\sigma_2^2$ |
|--------|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|
| JM I   | PM    | 11.36     | 103.8     | 1.854     | 2.604     | -3.157    | 0.098     | 13.53     | 8.712     | 68.27      | 4.871      | 0.940      |
|       | $L_{CI}$ | 11.04     | 92.09     | 0.307     | 1.740     | -4.836    | 0.027     | 13.00     | 7.997     | 49.93      | 4.677      | 0.872      |
|       | $U_{CI}$ | 11.62     | 115.7     | 3.303     | 3.527     | -0.892    | 0.184     | 13.98     | 9.536     | 86.99      | 5.058      | 1.023      |
| SD    | 0.145  | 6.075     | 0.720     | 0.510     | 1.059     | 0.052     | 0.271     | 0.426     | 9.405     | 0.097      | 0.043      |

### Table 2. Summary of estimated posterior mean (PM) of population (fixed-effects) parameters, the corresponding standard deviation (SD), and lower limit ($L_{CI}$) and upper limit ($U_{CI}$) of 95% equal-tail credible interval (CI) as well as DIC and EPD values based on the joint models.

| Method | Model | $\gamma_0$ | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\gamma_4$ | $\gamma_5$ | DIC | EPD |
|--------|-------|------------|------------|------------|------------|------------|------------|-----|-----|
| JM I   | PM    | -2.783     | -0.155     | -0.265     | 1.251      | 0.054      | 0.026      | 10368.9 | 1.675 |
|       | $L_{CI}$ | -2.870     | -0.493     | -0.864     | -0.502     | -0.478     | -0.036     | 12972.8 | 1.932 |
|       | $U_{CI}$ | -2.706     | 0.750      | 0.219      | 2.458      | 1.127      | 0.230      | 14099.1 | 2.637 |
| SD    | 0.043  | 0.335      | 0.375      | 0.735      | 0.326      | 0.065      |            |      |     |

| NM II  | PM    | -2.802     | -0.073     | -0.110     | -0.050     | -0.020     | -0.005     | 12972.8 | 1.932 |
|       | $L_{CI}$ | -2.890     | -0.401     | -1.389     | -2.068     | -0.246     | -0.034     | 14099.1 | 2.637 |
|       | $U_{CI}$ | -2.715     | 0.270      | 0.766      | 1.085      | 0.026      | 0.026      | 14099.1 | 2.637 |
| SD    | 0.045  | 0.158      | 0.599      | 0.778      | 0.064      | 0.015      |            |      |     |

| JM III | PM    | -2.785     | -0.112     | -0.177     | 0.485      | 0.001      | -0.003     | 14099.1 | 2.637 |
|        | $L_{CI}$ | -2.871     | -0.581     | -0.943     | -1.611     | -0.004     | -0.027     | 14099.1 | 2.637 |
|        | $U_{CI}$ | -2.701     | 0.307      | 0.271      | 1.446      | 0.004      | 0.019      | 14099.1 | 2.637 |
| SD    | 0.044  | 0.199      | 0.322      | 0.763      | 0.003      | 0.012      |            |      |     |
We now focus on the lower end of the distribution of the viral load due to LOD, where the Tobit model is offered to mediate the observations below LOD as missing values. As it was mentioned in Section 1, the current assay techniques for quantifying HIV-RNA viral load may not give accurate readings below LOD. In our analysis, we treated those inaccurate observed viral loads as missing values and predict them using the proposed the (mixture) joint models in connection with the Tobit model. Note that the main advantage of our proposed Tobit model is its ability to predict the true viral load values (log_{10} scale) below LOD based on a latent variable approach. The fitted results of these models for values below LOD are depicted in Figure 3, where the histograms show the observed distributions but inaccurate values (Figure 3(a)) below LOD and the predicted values below LOD under Models I, II, and III shown in Figure 3(b)–(d). The vertical-dashed line denotes the LOD value at 1.398. It can be seen from the histograms that most observed values are piled up at the range (0.6, 1.4) in the upper-left histogram (Figure 3(a)), whereas for Models I, II, and III the predicted values of the unobserved viral load below LOD are spread out as expected (see Figure 3(b)–(d)). However, Models II and III produce the predicted values exceeded the LOD much more than Model I does. In addition, when we compare the three models in terms of their distributions in predicting viral loads below LOD, we can see that Model I gives more plausible values ranged within (−0.2, 1.5) than Models II and III do in the sense that the distribution is closely fits the lower part of the whole distribution of the predicted viral load values based on Model I as expected, implying that Model I is the better model. This finding is also confirmed by the results (see Table 2 for details) from other criteria such as deviance information criterion (DIC; Spiegelhalter et al. 2002) and expected predictive deviance (EPD; Gelman et al. 2003) discussed in Appendix D. In summary, our results may suggest that it is important to consider heterogeneous and/or covariate measurement error data features-based the mixture of joint models to achieve reliable results. Based on these findings, we will further report our results in detail only for the best Model I below.

As mentioned in Section 1, one of the primary objectives in this analysis was to cluster all individuals’ membership into three classes based on viral load trajectories. Based on the mixture joint modeling, we are able to obtain a summary of class membership at both the population and individual levels. At population level, the MCMC procedure yields samples from the posterior distribution of \((\pi_1, \pi_2, \pi_3)\) in (18), the population proportion of individuals in each class. The estimates of population proportion and associated 95% CI of \((\pi_1, \pi_2, \pi_3)\) for three classes are 5.29% (3.15%, 7.47%), 57.13% (54.34%, 59.79%), and 37.58% (34.83%, 40.43%), respectively. It can be seen that class 2 (decrease and maintain stable) has the largest proportion 57.13%, followed by class 3 (decrease and rebound later) with proportion 37.58%, and then class 1 (decrease all the time) with the lowest proportion 5.29%. Thus, out of 427 patients, the patterns of changing viral load of 23, 244, and 160 patients followed classes 1, 2, and 3, respectively. This indicates that a confirmed shorter-term (class 1 due to early dropout) and longer-term (class 2) virologic responses were observed in a total of 62.42% of the patients in classes 1 and 2.

At individual level, the posterior probability of individual \(i\) belonging to the \(k\) (\(k = 1, 2, 3\)) class, \(p_{ik} = E[I(c_i = k)]\), can be approximated by \(\frac{1}{M} \sum_{m=1}^{M} I(c^{(m)}_i = k)\), where \(c_i^{(m)}\) is class membership of individual \(i\) drawn from the posterior distribution (17) in the \(m\)th MCMC iteration \((m = 1, \ldots, M)\) and \(M\) is a total number of posterior sample iterations \((M = 3000\) in this study). Barplot shown in Figure 4 displays the probabilities for the selected 20 individuals (from the 1st to the 20th subjects). The probability corresponding to individual patient who is classified as either viral load rebound or not may help physicians to refine treatment strategy and to identify the reason of viral load rebound for such individual patient. As examples, for the three patients shown in Figure 1 and 2, the patient 112 belongs to class 1 because the viral load decreases constantly in an early short-term period with probability 87.5%; the viral load of the patient 17 decreases and then maintain stable, and thus, this patient belongs to class 2 with probability 91.1%; and finally, the patient 7 is in class 3 (indicating viral load rebound) with probability 83.7%.

5. Concluding Discussion

We present a Bayesian joint modeling for three (response, covariate, and time-to-event) processes linked through the random effects that characterize the underlying individual-specific longitudinal processes. We consider the MNLMEJ models for survival-longitudinal data with multiple features. The approach is applied to jointly model HIV dynamics and time to decrease
in CD4/CD8 ratio in the presence of the CD4 covariate process with measurement error to provide a tool to assess ART and to monitor disease progression. Among the models considered in the application, Model I for longitudinal data with features of heterogeneity, missing due to LOD in response and measurement error in covariate was found to be favorable. One advantage of mixture joint modeling approach-based models proposed in this article is its flexibility to study simultaneous impact of various data characteristics (heterogeneity, non-linearity, missingness due to LOD, and incompleteness).

Another advantage of our mixture joint modeling approach is to provide not only all model parameter estimates, but also model-based probabilistic clustering to obtain class membership probabilities at both population and individual levels. In addition, the proposed mixture joint modeling approach can be easily implemented using the publicly available WinBUGS interacted with R. This makes our models and methods quite powerful and accessible to practicing statisticians in the field.

The estimated results of fixed effects presented in Table 1 based on Model I indicate that the first-phase decay rate, and the second-phase decay rate without and with time-varying CD4 covariate may be approximated by \( \hat{\lambda}_1 = 103.8 + 1.854z_{0i} \), \( \hat{\lambda}_2 = -3.157 \), and \( \hat{\lambda}_2(t) = -3.157 + 0.098(-13.53 + 8.712 \psi(t) + 68.27\psi_2(t)) \), respectively, where \( z_{0i} \) is the baseline \( \sqrt{CD4} \) value, \( \psi(t) \) and \( \psi_2(t) \) are two basis functions given in Section 2.1. Thus, the population viral load processes of three classes may be approximated by \( \hat{V}_1(t) = \exp[11.36 - \hat{\lambda}_1 t] \), \( \hat{V}_2(t) = \exp[11.36 - \hat{\lambda}_2 t] + \exp[2.604 - \hat{\lambda}_2 t] \), and \( \hat{V}_3(t) = \exp[11.36 - \hat{\lambda}_3 t] + \exp[2.604 - \hat{\lambda}_2 t] \). Since the first-phase viral decay rate, \( \lambda_1 \), is significantly associated with the baseline CD4 (due to significant estimate of \( \beta_1 \)) and the second-phase viral decay rate \( \lambda_2^* \) in the component three is significantly associated with the time-varying CD4 values (due to significant estimate of \( \beta_3 \)), this suggests that the viral load change \( V(t) \) may be significantly associated with both the baseline CD4 and time-varying CD4 values. This simple approximation considered here may provide a rough guidance and point to further research even though the true association described above may be more complicated. The analysis results suggest that for patients with viral load rebound, the CD4 process at time \( t \) has a significantly positive effect on the second-phase viral decay rate; this finding confirms the fact that the CD4 covariate is a more significant predictor on a viral decay rate during late stage and more rapid increase in CD4 cell count may be associated with faster viral decay in late stage, which, in turn, may result in a viral load suppression. In addition, the estimated results of the parameters in the time-to-event model (13) based on (best) Model I suggest that the time to first decline of CD4/CD8 ratio is not highly associated with either the two viral decay rates or the CD4 changes over time. This finding is consistent with that in the study by Wu, Liu, and Hu (2010) in which further explanations are provided.

To examine the sensitivity of parameter estimates to the prior distributions, we conducted a limited sensitivity analysis using the uniform prior distribution \( U(0, 100) \) for standard deviation scales \( \sigma_1 \) and \( \sigma_2 \) (Gelman 2006) instead of inverse gamma prior distribution (used in this article) for variance parameters \( \sigma_1^2 \) and \( \sigma_2^2 \); we refitted data based on Model I. We found from the results (not shown here) of sensitivity analysis that the conclusions with different prior distributions are in agreement. Thus, the findings from our analysis remain unchanged.

For inference of mixture modeling, parameter (or model) identifiability can be an important but difficult problem when a large number of model parameters must be estimated simultaneously. We must ensure each component model to be identifiable to make whole mixture model identifiable. Thus, we assume \( \lambda_{1i} > \lambda_{2i} \) in (10) and \( \lambda_{1i} > \lambda_{2ij}^* \) in (11) (Wu and Ding 1999) to make sure that equations (10) and (11) are identifiable. In practice, if the models are not identifiable, the MCMC algorithm would diverge quickly. In the application considered in this article, the MCMC algorithm was converged without problems and we did not observe potential identifiability problems. It is noted that the mixture components, \( g_k(\cdot) \) in the finite mixture model may have the same family of densities but differ only in specific values of parameters such as in their means, or have completely different functional forms with parameters of different dimensions and meanings across the sub-models (Pauder and Laird 2000), which is the case adopted in this article. Because of complexity of longitudinal studies, an advantage of choosing components correspond to different densities with distinguished mean functions \( g_k(\cdot) \), which are known and pre-specified, rather than with the same general family in the finite mixture models is that identifiability problem can be avoided (Pauder and Laird 2000).

As one of the referees pointed out, instead of the JM method we may adopt a more measured approach such as so-called two-step (TS) method (Higgins, Davidian, and Giltinan 1997). However, a lot of studies compared JM method and TS method in the literature (Wu 2002, 2004; Huang, Chen, and Yan 2012). The results from both real applications and simulation studies indicated that, in general, there were differences in the estimates and the standard deviations from TS method is smaller than those obtained by the JM method. This is because the usual TS method ignores the variability due to estimating the parameters in the first step. So, the standard deviation produced by TS method may be underestimated. The simulation studies also showed that estimates in JM method have smaller biases than those in the TS method. The estimates from the TS method may also be biased, and thus the JM method produces less-biased estimates and more-reliable standard errors.

The final three issues related to this application are noted as follows. (i) The original motivation of this mixture joint modeling was to cluster all patients into two classes with or without viral load rebound, which is of main interest from a clinical prospective. But another class was needed (class 1) to capture some patients who withdrew too early to be clustered into either class 2 with viral load suppression or class 3 with viral load failure. Thus, the number of components in this analysis was determined empirically based on the viral load trajectory patterns and clinical interpretability. Note that class 1 is referred as a confirmed “short-term virologic response” due to early dropout which may be informative dropout and may not indicate virologic suppression in long-term treatment. Toward this end, we may consider that dropouts are likely to be informative or non-ignorable missing in the sense that the probability of dropouts (missing data) may depend on the missing values. Thus, we can modify the mixture joint models conducted in this article to cluster all patients into two classes with virologic suppression.
and failure, respectively. (ii) Although the normal distribution assumption in both response and covariate models may be appropriate (histograms of viral load (in log scale) and CD4 (in square-root) not shown here) for this specific data analysis, the mixture joint models in this article can be extended to assume model errors with skew distributions as considered by Bandyopadhyay et al. (2012); Huang and Dagne (2011); Huang, Dagne, and Wu (2011); Lachos et al. (2011) if data exhibit nonnormality; however, the computational burden and other issues may be involved. (iii) The purpose of this article is to demonstrate the proposed MNLMEJ models using a real data analysis, although a limited simulation study may be conducted to evaluate the performance of the proposed models and method. However, since this article investigated different scenarios based on relatively complex model specifications focusing on real data analysis, the complex nature considered in this article, especially the three model components involved, will pose some challenges for such simulation studies with intensive computational burden which requires additional significant efforts. We are actively investigating these important research problems and hope that the relevant technical specifications are met and longitudinal measurements are assumed to arise from two or more identifiable subclasses within a population.

**Supplementary Material**

Appendices of the online supplementary materials referenced in Section 1, Section 2.3, Section 2.4, Section 3, Section 4.1, and Section 4.2 in detail are available under the article information link at the Statistics in Biopharmaceutical Research Web site.

**Acknowledgments**

The authors gratefully acknowledge the Editor, an Associate Editor, and two anonymous referees for their insightful comments and constructive suggestions that led to a marked improvement of the article.

**References**

Bandyopadhyay et al. (2012), “Skew-Normal/Independent Linear Mixed Models for Censored Responses With Applications to HIV Viral Loads,” *Biometrical Journal*, 54, 405–425. [205]

Brown, E. R., and Ibrahim, J. G. (2003), “A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data,” *Biometrics*, 59, 221–228. [195]

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), London: Chapman and Hall. [196]

Davidian, M., and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman and Hall. [196,197]

DeGruttola, V., and Tù, X. M. (1994), “Modeling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time,” *Biometrics*, 50, 1003–1014. [199]

Diebolt, J., and Robert, C. (1994), “Estimation of Finite Mixture Distributions by Bayesian Sampling,” *Journal of the Royal Statistical Society*, Series B, 56, 363–375. [199]

Ferguson, T. S. (1973), “A Bayesian Analysis of Some Nonparametric Problems,” *Annals of Statistics*, 1, 209–230. [195,198]

Fischl, M. A., Ribaudo, H. J., Collier, A. C., et al. (2003), “A Randomized Trial of 2 Different 4-Drug Antiretroviral Regimens Versus a 3-Drug Regimen, in Advanced Human Immunodeficiency Virus Disease,” *Journal of Infectious Disease*, 188, 625–634. [194,195,196]

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990), “Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling,” *Journal of the American Statistical Association*, 85, 972–985. [200]

Gelman, A. (2006), “Prior Distributions for Variance Parameters in Hierarchical Models,” *Bayesian Analysis*, 1, 515–533. [204]

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), *Bayesian Data Analysis*, London: Chapman and Hall. [203]

Gelman, A., and Rubin, D. (1992), “Inference From Iterative Simulation Using Multiple Sequences,” *Statistical Science*, 7, 457–511. [200]

Guedj, J., Rodolphe Thiebaut, R., and Commenges, D. (2010), “Joint Modeling of the Clinical Progression and of the Biomarkers’ Dynamics Using a Mechanistic Model,” *Biometrics*, 67, 59–66. [195]

Higgins, K. M., Davidian, M., and Giltinan, D. M. (1997), “A Two-Step Method to Measurement Error in Time-Dependent Covariates in Nonlinear Mixed Effects Models, With Application to IGF-I Pharmacokinetics,” *Journal of the American Statistical Association*, 92, 436–448. [204]

Huang, Y., Liu, D., and Wu, H. (2006), “Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System,” *Biometrics*, 62, 413–423. [200]

Huang, Y., and Dagne, G. A. (2011), “A Bayesian Approach to Joint Mixed-Effects Models With a Skew-Normal Distribution and Measurement Errors in Covariates,” *Biometrics*, 67, 260–269. [194,195,196,205]

Huang, Y., Dagne, G. A., and Wu, L. (2011), “Bayesian Inference on Joint Models of HIV Dynamics for Time-to-Event and Longitudinal Data With Skewness and Covariate Measurement Errors,” *Statistics in Medicine*, 30, 2930–2946. [195,196,198,205]

Huang, Y., Chen, J., and Yan, C. (2012), “Mixed-Effects Joint Models With Skew-Normal Distribution for HIV Dynamic Response With Missing and Mismeasured Time-Varying Covariate,” *International Journal of Biostatistics*, 8, Article 34. [200,204]

Hughes, J. P. (1999), “Mixed Effects Models With Censored Data With Applications to HIV RNA Levels,” *Biometrics*, 55, 625–629. [194,195]

Ishwaran, H., and Zarepour, M. (2000), “Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models,” *Biometrika*, 87, 371–390. [195,198]

Lachos, V. H., Bandyopadhyay, D., and Dey, D. K. (2011), “Linear and Nonlinear Mixed-Effects Models for Censored HIV Viral Loads Using Normal/Independent Distributions,” *Biometrics*, 67, 1594–1604. [194,195,196,205]

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. [194,195,196,199,200]

Lu, X., and Huang, Y. (2010a), “Bayesian Analysis of Nonlinear Mixed-Effects Mixture Models for Longitudinal Data With Heterogeneity and Skewness,” *Statistics in Medicine*, 33, 2830–2849. [194,195,197]

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), “WinBUGS—A Bayesian Modelling Framework: Concepts, Structure, and Extensibility,” *Statistics and Computing*, 10, 325–337. [200]

Mbogningab, C., Bleakleyab, K., and Lavieilleb, M. (2014), “Joint Modelling of Longitudinal and Repeated Time-to-Event Data Using Nonlinear Mixed-Effects Models and the Stochastic Approximation Expectation-Maximization Algorithm,” *Journal of Statistical Computation and Simulation*, DOI:10.1080/00949655.2013.878938 . [195]

Muthén, B., and Shedden, K. (1999), “Finite Mixture Modeling With Mixture Outcomes Using the EM Algorithm,” *Biometrics*, 55, 463–469. [195,197]

Pauker, D., and Laird, N. M. (2000), “A Mixture Model for Longitudinal Data With Application to Assessment of Noncompliance,” *Biometrics*, 56, 464–472. [195,197,204]

Pawitan, Y., and Self, S. (1993), “Modeling Disease Marker Processes in AIDS,” *Journal of the American Statistical Association*, 88, 719–726. [194]
Perelson, A. S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D. D. (1997), "Decay Characteristics of HIV-1-Infected Compartments During Combination Therapy," Nature, 387, 188–191. [194]

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," Journal of the Royal Statistical Society, Series B, 64, 583–639. [203]

Thiebaut, R., Jacqmin-Gadda, H., Babiker, A., and Commenges, D. (2005), "Joint Modelling of Bivariate Longitudinal Data With Informative Dropout and Left-Censoring, With Application to the Evolution of CD4+ Cell Count and HIV RNA Viral Load in Response to Treatment of HIV Infection," Statistics in Medicine, 24, 65–82. [195,199]

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985), Statistical Analysis of Finite Mixture Distributions, Chichester, UK: Wiley. [198]

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26, 24–36. [195]

Tsiatis, A. A., and Davidian, M. (2004), "An Overview of Joint Modeling of Longitudinal and Time-to-Event Data," Statistical Sinica, 14, 809–834. [195,198,199]

Wu, H., and Ding, A. A. (1999), "Population HIV-1 Dynamics in vivo: Applicable Models and Inferential Tools for Virological Data From AIDS Clinical Trials," Biometrics, 55, 410–418. [196,204]

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. [194,195,196,199,200,204]

——— (2004), "Simultaneous Inference for Longitudinal Data With Detection Limits and Covariates Measured With Errors, With Application to AIDS Studies," Journal of the Statistics in Medicine, 23, 1715–1731. [204]

Wu, H., and Zhang, J. T. (2006), Nonparametric Regression Methods for Longitudinal Data Analysis, Hoboken, New Jersey: Wiley. [196,200]

Wu, L., Liu, W., and Hu, X. J. (2010), "Joint Inference on HIV Viral Dynamics and Immune Suppression in Presence of Measurement Errors," Biometrics, 66, 327–335. [195,196,197,198,199,202,204]

Yi, G. Y., Liu, W., and Wu, L. (2011), "Simultaneous Inference and Bias Analysis for Longitudinal Data With Covariate Measurement Error and Missing Responses," Biometrics, 67, 67–75. [195]