The Annals of Applied Statistics
2019, Vol. 13, No. 4, 2140–2157
https://doi.org/10.1214/19-AOAS1274

JOINT MODEL OF ACCELERATED FAILURE TIME AND MECHANISTIC NONLINEAR MODEL FOR CENSORED COVARIATES, WITH APPLICATION IN HIV/AIDS[1]

BY HONGBIN ZHANG AND LANG WU

City University of New York and University of British Columbia

For a time-to-event outcome with censored time-varying covariates, a joint Cox model with a linear mixed effects model is the standard modeling approach. In some applications such as AIDS studies, *mechanistic* nonlinear models are available for some covariate process such as viral load during anti-HIV treatments, derived from the underlying data-generation mechanisms and disease progression. Such a mechanistic nonlinear covariate model may provide better-predicted values when the covariates are left censored or mismeasured. When the focus is on the impact of the time-varying covariate process on the survival outcome, an accelerated failure time (AFT) model provides an excellent alternative to the Cox proportional hazard model since an AFT model is formulated to allow the influence of the outcome by the entire covariate process. In this article, we consider a nonlinear mixed effects model for the censored covariates in an AFT model, implemented using a Monte Carlo EM algorithm, under the framework of a joint model for simultaneous inference. We apply the joint model to an HIV/AIDS data to gain insights for assessing the association between viral load and immunological restoration during antiretroviral therapy. Simulation is conducted to compare model performance when the covariate model and the survival model are misspecified.

1. Introduction. Human immunodeficiency virus (HIV) infection results in progressive destruction of immune function, which may be indicated by a decrease of CD4 T-cells and an increase of CD8 T-cells. Successful treatments such as antiretroviral therapy (ART) would typically suppress the amount of virus, as measured by viral loads (RNA copies per milliliter of blood plasma), with partial reconstitution of the immunologic perturbations. However, despite many advances in drug development for suppressing HIV, the immune dynamics after initiating an antiretroviral therapy remain not fully understood (Deeks et al. (2000), Lederman,

Received May 2018; revised April 2019.

[1]Supported in part by the City University of New York High-Performance Computing Center, College of Staten Island, funded in part by the City and State of New York, City University of New York Research Foundation and National Science Foundation Grants CNS-0958379, CNS-0855217 and ACI-112611; also partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant No. 22R80742.

Key words and phrases. Censored data, HIV/AIDS, mechanistic model, nonlinear mixed effects model, survival data.

Connick and Landy (1998), Palmisano and Yella (2011)). The CD4:CD8 ratio recently has become a tool for assessing the condition of HIV subjects. Compared to research focusing on the CD4 and CD8 processes separately, other research shows that the CD4:CD8 ratio process may be more closely associated with viral dynamics: the time until first decline in this ratio has been shown to be more closely associated with HIV viral rebound (see, e.g., Wu, Liu and Hu (2010) and Huang, Hu and Dagne (2014)). Here we further investigate this claim. Wu, Liu and Hu (2010) investigated such a relationship with a shared parameter joint modeling approach where individual random effects are used to capture the characteristics of viral load trajectories. In this paper, we model the error-prone viral load as a time-dependent regressor for event outcome to directly model their relationship for better interpretation. In particular, we aim to model the event times to first ratio decline (FRD) as a function of ART received, the longitudinal viral loads and their interaction for a better understanding of the interplay.

One of the main challenges in the analyses is that viral loads may be measured with errors and may be left censored due to a lower detection limit. In survival models, when a time-dependent covariate is left censored or measured with errors, a usual approach is to assume a linear or generalized linear mixed effects covariate model, and assume that the model holds for the unobserved censored values (e.g., Bernhardt, Wang and Zhang (2014), Sattar and Sinha (2019)). The inference is then based on a joint model for the covariate and the response. However, such an approach may lead to biased results if the assumed covariate model is misspecified, especially for censored values. Nonlinear models, on the other hand, are typically *scientific* or *mechanistic* models since they are usually derived based on the underlying data-generation mechanisms, not just based on observed data. Although such nonlinear models may not always be available, in many applications nonlinear models can be obtained, such as in HIV viral dynamics (Wu, Zhao and Liang (2004)). These mechanistic models are derived based on the underlying viral elimination and production mechanisms and have been confirmed to fit the observed data well (Wu and Ding (1999)). These mechanistic models are usually *nonlinear* models since they are solutions of a set of differential equations which describe the underlying data-generation mechanisms. Therefore, these mechanistic models should hold for censored viral loads as well. In this paper, along with the line of our recent development (Zhang, Wong and Wu (2018), Zhang and Wu (2018)), we employed a mechanistic nonlinear mixed effects model (NLME) for the covariate process. For the time to *FRD*, however, we consider the accelerated failure time (AFT) models as an alternative to the Cox proportional hazard model (Zhang and Wu (2018)). AFT models relax the proportional hazard assumption and allow the entire viral load trajectory, including the early decay phase as well as the portion censored by the detection limit, to influence the immune restoration, leading to a possibly more biological plausible joint model.

To our knowledge, a joint model of NLME and AFT model with censored covariates has not been studied in the literature yet, especially using a mechanistic nonlinear model to address censored and mismeasured covariates in the AFT

model. Tseng, Hsieh and Wang (2005) considered joint linear mixed effects (LME) and AFT model, without censoring in covariates. Wu, Liu and Hu (2010) studied a joint NLME and AFT model based on a shared parameter model, again without censoring in covariates. Riopoulos (2012) reviewed commonly used joint models for longitudinal and survival data. There has been extensive research on joint models in more recent years, but none seems to consider the approaches described in this article.

The innovations in the current article are as follows: (a) Instead of a shared parameter joint model where the random effects from the NLME model are used as "covariate," here we use viral loads as time-varying covariates directly in the AFT model to allow more natural interpretation. This extension is nontrivial because (i) the parameter estimation becomes more challenging since the baseline hazard of the AFT model involves unknown parameters from the covariate model, and (ii) the covariate values may be left censored or truncated due to the lower detection limit. (b) In the current article, we focus on addressing left censored or truncated time-dependent covariates in the joint model. This is important because in HIV studies left censored viral loads reflect "viral suppression" which has important clinical implications. (c) The computation becomes even more challenging in the current model setting than that in a shared parameter model where the "covariates" random effects are time-independent without censoring, and a simpler log-linear representation of the AFT model can be used.

We present the joint NLME and AFT model with covariate censoring in Section 2. In Section 3, we describe the Monte Carlo EM algorithm. In Section 4, we analyze an AIDS dataset which motivated our research. In Section 5, we conduct a simulation study to assess the joint model's performance and compare with results from other misspecified models. We conclude the article with some discussions in Section 6.

2. The joint NLME and AFT model. In this section, we present the models in general forms to enable generalizations in different settings. We consider a joint model for longitudinal and survival data, where the longitudinal data are treated as covariates with some covariates possibly being censored and measured with errors. For simplicity of presentation, we focus on the key covariate with left censoring and suppress other covariates. Suppose that the key covariate is repeatedly measured over time, subject to left censoring and measurement errors. We also assume that a scientific or mechanistic nonlinear model can be obtained for the covariate process. We use this nonlinear mechanistic covariate model to "predict" the unobserved censored or mismeasured covariate values. The repeated measures may serve as "replicates" to estimate magnitudes of measurement errors, assuming the covariate values change smoothly over time. Note that there are two types of censoring here: the longitudinal covariate values may be *left censored*, and the event times may be *right censored*.

Let the covariate value for individual i at time t_{ij} be x_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, n_i$, subject to censoring and measurement errors. Let x_{ij}^* be the corresponding true but unobserved covariate value. We consider the following NLME model, which may be viewed as a classical measurement error model:

$$x_{ij} = g(t_{ij}, \boldsymbol{\alpha}, \boldsymbol{a}_i) + e_{ij} \equiv x_{ij}^* + e_{ij}, \tag{2.1}$$

where $g(\cdot)$ is a known *nonlinear* function, $\boldsymbol{\alpha}$ is a vector of parameters (fixed effects), $\boldsymbol{a}_i$ is a random effects vector, and e_{ij} is measurement error. We assume that $\boldsymbol{a}_i \sim N(\mathbf{0}, A)$, and $e_{ij}|\boldsymbol{a}_i$ i.i.d. $\sim N(0, \sigma^2)$, that is, e_{ij} is independent of e_{ik}, $j \neq k$, where σ^2 represents the magnitude of the measurement error. To quantify the covariate effect history, we recast model (2.1) in continuous time, $x_i(t) = g(t, \boldsymbol{\alpha}, \boldsymbol{a}_i) + e_i(t) \equiv x_i^*(t) + e_i(t)$ where $x_i(t)$, $g(t, \boldsymbol{\alpha}, \boldsymbol{a}_i)$, and $e_i(t)$ are the covariate value, known nonlinear function, and measurement error at time t.

Suppose that there is a lower detection limit d for the covariate such that the value of x_{ij} with $x_{ij} < d$ cannot be observed. In the presence of left censoring in the covariate, the covariate $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{in_i})$ can be written as $(\boldsymbol{x}_{oi}, \boldsymbol{x}_{ci})$, where $\boldsymbol{x}_{oi} = (x_{oi,1}, \ldots, x_{oi,n_{oi}})$ represents the portion of observed covariate values for individual i and $\boldsymbol{x}_{ci} = (x_{ci,1}, \ldots, x_{ci,n_{ci}})$ represents the left censored portion, and $n_i = n_{oi} + n_{ci}$.

For the survival data, we consider the usual right censoring. We define T_i to be the minimum of the event time T_i^* and the right censoring time C_i, and define Δ_i to be the event indicator which takes value "1" if an event occurs before the censoring time and "0" otherwise. We aim to associate $x_i^*(t)$, the true and unobserved values of the longitudinal covariate process, with the event outcome. The commonly used Cox model (Cox (1972)) can be written as $h_i(t; X_i^*(t)) = h_0(t) \exp(x_i^*(t)\beta)$, where $X_i^*(t) = \{x_i^*(s) : 0 \leq s \leq t\}$ is the covariate history up to time t, and $h_0(t)$ the baseline hazard function. When covariates are constant, that is, time-fixed, Cox model reduces to a proportional hazard model where the proportionality assumption is hard to meet in practice. For the time-varying covariate, although Cox model is no longer proportional in hazard, however, the baseline hazard is still assumed to be independent with the covariate. Alternatively, Cox and Oakes (1984) presented a so-called AFT model (see also Miller (1976)) in which an individual with covariate $x_i^*(\cdot)$ and survival time T_i^* uses up her life at the rate of $e^{-x_i^*(t)\beta}$ relative to a "baseline time" T_{0i} in the relationship $T_{0i} = \int_0^{T_i^*} e^{x_i^*(s)\beta}\, ds$. The model is succinctly written as

$$S_i\big(t|x_i^*, \beta\big) = S_{0i}\big(\bar{c}_i(t)t\big), \qquad \bar{c}_i(t) = \frac{1}{t}\int_0^t e^{x_i^*(s)\beta}\, ds,$$

where $S_{0i}(\cdot)$ is the baseline survival function. Here, we can interpret $\bar{c}_i(t)$ as the average value of an acceleration factor $e^{x_i^*(s)\beta}$ over $s \in [0, t)$. The model can be equivalently specified through the hazard function as

$$h_i\big(t; X_i^*(t)\big) = h_0\big(\bar{c}_i(t)t\big) \exp\big(x_i^*(t)\beta\big). \tag{2.2}$$

One can understand this model through the survivor function $S_i(t|x_i^*, \beta) = S_{0i}(\bar{c}_i(t)t)$. So if we have standardized the covariate process, suppose that we have an individual with $\bar{c}(t) \equiv 1$, for all t, so that we can regard them as a baseline individual with survivor function S_{0i}. Then we can compare this individual to one with $\bar{c}(t) = 2$, say, for a value of t of interest. Then the prospects of living at least t units of time for this individual are the same as the prospects for the baseline individual to live at least $2t$ units. This model then allows survival prospects to slide forwards or backwards relative to a hypothetical baseline individual. It is not possible with the Cox model to make such a simple comparison with a baseline individual. In addition, we assume that the time-varying covariate process is internal or endogenous (see, e.g., Kalbfleisch and Prentice (2002)).

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \sigma^2, A, \boldsymbol{h}_0, \boldsymbol{\beta})$ be the collection of all unknown parameters in model (2.1) and (2.2), and let $f(\cdot)$ denote a generic density function and $f(X|Y)$ denote a conditional density of X given Y. Under the assumptions of noninformative right censoring of event times and conditional independence, the "joint likelihood" for individual i can be written as

$$
\begin{aligned}
L_{\text{obs}}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \int_{-\infty}^{d} \big[& f(T_i, \Delta_i | \boldsymbol{a}_i; \boldsymbol{\alpha}, \boldsymbol{h}_0, \boldsymbol{\beta}) \\
& \times f\big(\boldsymbol{x}_{oi} | \boldsymbol{a}_i; \boldsymbol{\alpha}, \sigma^2\big) f\big(\boldsymbol{x}_{ci} | \boldsymbol{a}_i; \boldsymbol{\alpha}, \sigma^2\big) f(\boldsymbol{a}_i | A) \big] \, d\boldsymbol{x}_{ci} \, d\boldsymbol{a}_i,
\end{aligned} \tag{2.3}
$$

where the integration $\int$ may be multidimensional. In (2.3), for the survival part, we have

$$
\begin{aligned}
f(T_i, \Delta_i | \boldsymbol{a}_i; \boldsymbol{\alpha}, \boldsymbol{h}_0, \boldsymbol{\beta}) = & \big(h_i\big(T_i | X_i^*(T_i); \boldsymbol{\alpha}, \boldsymbol{h}_0, \boldsymbol{\beta}\big)\big)^{\Delta_i} \\
& \times S\big(T_i | X_i^*(T_i); \boldsymbol{\alpha}, \boldsymbol{h}_0, \boldsymbol{\beta}\big).
\end{aligned} \tag{2.4}
$$

Note that, in the above covariate model for x_{ij}'s, we have assumed that the censored (unobserved) values of x_{ij} continue to follow the parametric distribution assumed for the observed x_{ij} values. This assumption should be reasonable since such a nonlinear mechanistic model should hold for both observed and censored covariate values. When considering the censored data as nonignorable missing data, our approach may provide better "predictions" of the unobserved values, implied by the EM algorithm described in the next section, than those from the commonly used linear covariate models which may not hold for censored values.

3. Parameter estimation and inference. For statistical inference of joint models of longitudinal and survival data, the key implementation difficulty is that the integral in the likelihood $L_{\text{obs}}(\theta)$ is typically quite intractable, due to the nested integrals because of the embedded nonlinear covariate in the AFT structure. When the dimension of the censored portion and random effects, that is, $\dim(\boldsymbol{x}_{ci}, \boldsymbol{a}_i)$, is not low, numerical methods such as Gaussian quadrature can be computationally very intensive and may offer nonconvergence. We therefore offer a Monte Carlo EM algorithm.

The EM algorithm is a standard approach for likelihood estimation in the presence of missing data. When the E-step is highly complicated, Monte Carlo methods can be used to approximate the expectation, leading to a MCEM algorithm. In our case, by treating the censored values $\boldsymbol{x}_{\mathrm{cen},i}$ and random effects $\boldsymbol{a}_i$ as "missing data," we have "complete data" as $\{(\boldsymbol{x}_{ci}, \boldsymbol{x}_{oi}, T_i, \Delta_i, \boldsymbol{a}_i) = (\boldsymbol{x}_i, T_i, \Delta_i, \boldsymbol{a}_i), i = 1, \ldots, n\}$, and the "complete-data" log-likelihood function for individual i can be expressed as

$$
\begin{aligned}
l_c^i(\boldsymbol{\theta}) = {} & \log f(T_i, \Delta_i | \boldsymbol{a}_i; \boldsymbol{\alpha}, \boldsymbol{h}_0, \boldsymbol{\beta}) + \log f\big(\boldsymbol{x}_{ci} | \boldsymbol{a}_i; \boldsymbol{\alpha}, \sigma^2\big) \\
& + \log f\big(\boldsymbol{x}_{oi} | \boldsymbol{a}_i; \boldsymbol{\alpha}, \sigma^2\big) + \log f(\boldsymbol{a}_i; A).
\end{aligned} \tag{3.1}
$$

The EM algorithm iterates between an E-step and a M-step until convergence. Let $\boldsymbol{\theta}^{(t)}$ be the parameter estimates from the tth EM iteration. The E-step for individual i at the $(t+1)$th EM iteration can be expressed as

$$
\begin{aligned}
Q_i(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) = {} & \int \int \big[\log f(T_i, \Delta_i | \boldsymbol{a}_i; \boldsymbol{\alpha}, \boldsymbol{h}_0, \boldsymbol{\beta}) \\
& + \log f\big(\boldsymbol{x}_{ci} | \boldsymbol{a}_i; \boldsymbol{\alpha}, \sigma^2\big) + \log f\big(\boldsymbol{x}_{oi} | \boldsymbol{a}_i; \boldsymbol{\alpha}, \sigma^2\big) \\
& + \log f(\boldsymbol{a}_i; A)\big] \times f(\boldsymbol{x}_{ci}, \boldsymbol{a}_i | \boldsymbol{x}_{oi}, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})\, d\boldsymbol{x}_{ci}\, d\boldsymbol{a}_i.
\end{aligned} \tag{3.2}
$$

The above E-step again involves an intractable integration. However, because expression (3.2) is an expectation with respect to $f(\boldsymbol{x}_{ci}, \boldsymbol{a}_i | \boldsymbol{x}_{oi}, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})$, and it can be evaluated using the MCEM algorithm (Ibrahim, Lipsitz and Chen (1999), Wei and Tanner (1990)). Specifically, we may use the Gibbs sampler (Gelfand and Smith (1990)) to generate many samples from $f(\boldsymbol{x}_{ci}, \boldsymbol{a}_i | \boldsymbol{x}_{oi}, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})$ by iteratively sampling from the full conditionals $[\boldsymbol{x}_{ci} | \boldsymbol{x}_{oi}, T_i, \Delta_i, \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}]$, and $[\boldsymbol{a}_i | \boldsymbol{x}_i, T_i, \Delta_i; \boldsymbol{\theta}^{(t)}]$ based on the following results:

$$
f(\boldsymbol{x}_{ci} | \boldsymbol{x}_{oi}, T_i, \Delta_i, \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}) \propto f(\boldsymbol{x}_i | \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}),
$$

$$
f(\boldsymbol{a}_i | \boldsymbol{x}_i, T_i, \Delta_i; \boldsymbol{\theta}^{(t)}) \propto f(\boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}) \cdot f(\boldsymbol{x}_i | \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}) \cdot f(T_i, \Delta_i | \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}).
$$

Monte Carlo samples from each of the above full conditionals can be generated using multivariate rejection sampling methods (see the Appendix for details). After generating large random samples from the conditional distribution $f(\boldsymbol{x}_{ci}, \boldsymbol{a}_i | \boldsymbol{x}_{oi}, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})$, we can then approximate the expectation $Q_i(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$ in the E-step by its empirical mean, with "missing data" replaced by simulated values. Then the M-step, which maximizes $\sum_{i=1}^n Q_i(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$, is like a complete-data maximization, so complete-data optimization procedures such as the Newton–Raphson method may be used to update the parameter estimates. At convergence, we obtain the MLE of $\boldsymbol{\theta}$ or a possibly local maxima. We may try different starting values to roughly check if an MLE (i.e., global maxima) is obtained.

To obtain the variance-covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$, we consider the following approximate formula in McLachlan and Krishnan (1997). Denote the score

function of the complete-data likelihood by $S_c^{(i)} = \partial l_c^{(i)}/\partial \boldsymbol{\theta}$. Then an approximate formula for the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \left[\sum_{i=1}^{n} \mathrm{E}\big(S_c^{(i)}|\boldsymbol{x}_{oi}, T_i, \Delta_i; \hat{\boldsymbol{\theta}}\big)\mathrm{E}\big(S_c^{(i)}|\boldsymbol{x}_{oi}, T_i, \Delta_i; \hat{\boldsymbol{\theta}}\big)^T\right]^{-1},$$

where the expectation can be approximated by Monte Carlo empirical means.

4. Data analysis. In this section, we analyze the ACTG 388 dataset (Fischl, Ribaudo and Collier (2003)). ACTG 388 study was a randomized, open-label study comparing two different 4-drug regimens (efavirenz and nelfinavir) with a standard 3-drug regimen for 517 subjects with no or limited previous experience with antiretroviral therapy (ART) who had a CD4 cell count ≤ 200 cells/mm^3 or a plasma HIV-1 RNA level $\geq 80{,}000$ copies/mL at screening. The plasma HIV-1 RNA (viral load) is repeatedly quantified at weeks 0, 4, 8, 16, and every eight weeks until week 72. Out of a total of 517 patients, 65 subjects were excluded from this study as RNA assay results are not available for these subjects. The CD4 and CD8 cell counts were also measured throughout the study on a similar scheme. The event times of interest here are the occurrences of *FRD* (first CD4:CD8 ratio declining) for which we treat the censored event times as right censoring for simplicity, although strictly speaking, they are interval-censored. To see if this simple assumption of right censoring is reasonable, we created three datasets where the event times were assumed to occur at left-ends, middle and right-ends of the time intervals. No major differences were found, and the results hereafter are based on the dataset where the right-ends of the intervals were treated as the times of the event. Also, we removed 60 subjects as they had less than three viral load measurements. The final data has 392 individuals where 123 (31%) are in the "standard" 3-drugs treatment arm, 140 (36%) in the "efavirenz" arm, 129 (33%) in the "nelfinavir" arm.

Figure 1 displays viral load trajectories (up to the time point of *FRD*) for the three treatment arms. The event *FRD* occurred to most individuals. The viral load detection limit in this study is 25 copies/mL (or 1.39 in $\log_{10}$ scale). In total, 15% of viral loads are left censored. In Figure 1, the censored viral loads are replaced by half of the detection limit. We use $\log_{10}$ transformation on viral load to stabilize the variation of the measurement errors, to make the normality assumption of the errors more reasonable, and to speed up the estimation algorithm. To avoid too small or too large parameter estimates, which may be unstable, we also re-scaled the original time (in days) so that the new timescale is between 0 and 1.

For the observed viral load data x_{ij}, which is subject to measurement error and left censoring, we consider the following mechanistic NLME model based on Wu

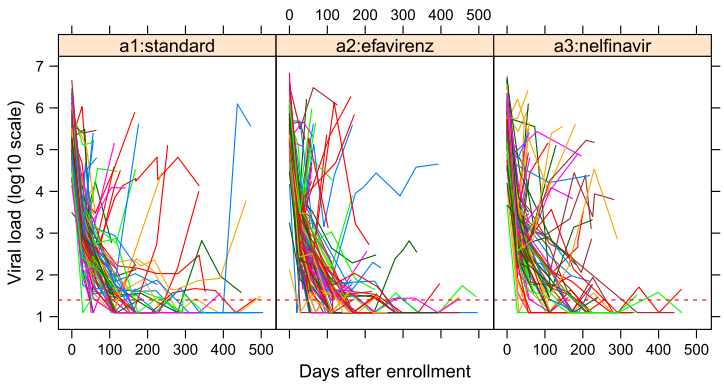

FIG. 1. *Individual viral loads stratified by the three treatment arms. The lower detection limit for viral loads is* $\log 10(25)(=1.39)$, *and censored data below the detection limit are replaced with half of the limit* (*i.e.*, $\log 10(25/2)=(1.10)$) *in the plots.*

and Ding (1999):

$$
\begin{aligned}
&x_{ij} = \log_{10}\bigl\{e^{\alpha_{1i}-\alpha_{2i}*t_{ij}} + e^{\alpha_{3i}-(\alpha_{4i}+\alpha_{5i}*t_{ij})*t_{ij}}\bigr\} + e_{ij} \equiv x^*_{ij} + e_{ij},\\
&\alpha_{1i} = \alpha_1 + a_{1i}, \qquad \alpha_{2i} = \alpha_2 + a_{2i},\\
&\alpha_{3i} = \alpha_3 + a_{3i}, \qquad \alpha_{4i} = \alpha_4 + a_{4i}, \qquad \alpha_{5i} = \alpha_5 + a_{5i},\\
&\boldsymbol{a}_i = (a_{1i}, a_{2i}, a_{3i}, a_{4i}, a_{5i})^T \sim N(\mathbf{0}, A), \qquad e_{ij}|\boldsymbol{a}_i \sim N\bigl(0,\sigma^2\bigr),
\end{aligned}
\tag{4.1}
$$

where x_{ij} is the observed error-prone viral load ($\log_{10}$ transformed) for individual i at time j, x^*_{ij} is the corresponding (unobserved) true viral load, a_{ki}'s are random effects, e_{ij} represents measurement error, parameter α_{2i} represents the first-phase viral decay rate and $\alpha_{4i}+\alpha_{5i}t_{ij}$ represents the time-varying second-phase viral change rate, parameters α_{1i} and α_{3i} are the amounts of virus produced and cleared from productively infected cells and long-lived and/or latently infected cells respectively.

Our primary goal in this analysis is to determine if and how the time to *FRD* may be related to entire viral load process before the event and how treatments modify the relationship. We consider the following AFT model for the time to *FRD*:

$$
h_i(t) = h_0\bigl[\psi\bigl\{X^*(t);\boldsymbol{\beta}\bigr\}\bigr]\psi'\bigl\{X^*(t);\boldsymbol{\beta}\bigr\},
\tag{4.2}
$$

where $\psi(X^*(t);\boldsymbol{\beta}) = \int_0^t e^{\beta_1 z_{1i}+\beta_2 z_{2i}+\beta_3 x_i^*(s)+\beta_4 z_{1i}x_i^*(s)+\beta_5 z_{2i}x_i^*(s)}\,ds$, and ψ' is the first derivative of ψ, $x_i^*(s)$ is the true viral load value at time s subject to left censoring, z_{1i} and z_{2i} are dummy variables for the three treatment arms such that $z_{1i}=0$ & $z_{2i}=0$ for the "standard" treatment arm, $z_{1i}=1$ & $z_{2i}=0$ for the "efavirenz" arm, and $z_{1i}=0$ & $z_{2i}=1$ for the "nelfinavir" arm, and β_j's are the corresponding parameters.

Handling the AFT structure in the joint modeling setting is more difficult since the baseline hazard function $h_0(\cdot)$ involves unknown quantities $(\boldsymbol{a}_i, \boldsymbol{\alpha}, \boldsymbol{\beta})$. One cannot use the point mass function with masses assigned to all uncensored survival times for the baseline hazard function h_0 as in standard semi-parametric Cox model implementation. To circumvent this, Tseng, Hsieh and Wang (2005) assumed the baseline hazard function h_0 to be a step function, while more recently, Tseng and Yang (2016) propose a kernel-smooth function which requires an arbitrary choice on the number of the knots. Here we consider the parametric Weibull distribution for the survival data since it provides a reasonable fit to the data and is also easier to handle. In this case, we have $h_0(t) = \lambda\gamma t^{\gamma-1}$, where λ is the location parameter and γ is the shape parameter. For the integration in the baseline hazard function, we use a numerical method called Gauss–Kronrod rule (Press et al. (2007)).

In our implementation, the E-step in the MCEM method is the most time consuming portion of the entire estimation procedure. Consider sampling from $f(\boldsymbol{x}_{ci}|\boldsymbol{x}_{oi}, T_i, \Delta_i, \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)})$. Since $f(\boldsymbol{x}_{ci}|\boldsymbol{x}_{oi}, T_i, \Delta_i, \boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}) \propto f(\boldsymbol{x}_i|\boldsymbol{a}_i; \boldsymbol{\theta}^{(t)})$, we could get the samples directly from the multivariate normal distribution $f(\boldsymbol{x}_i|\boldsymbol{a}_i; \boldsymbol{\theta}^{(t)})$ by rejecting those that do not fit the observed censoring patterns. However, such a procedure may have an arbitrarily low yields and be time consuming. In the implementation, we take an alternative approach by using Gibbs samplers again. Given an initial value of $\boldsymbol{x}_i$, new values of $\boldsymbol{x}_i$ are generated by iteratively sampling from univariate conditional distributions (Breslaw (1994)). To get the initials values, we first fit the NLME covariate model separately, ignoring the survival data. In particular, we use the popular iterative linearization method of Lindstrom and Bates (1990), in which the left-censored response (covariate) values are handled in the working linear mixed effect model (LME) step within each iteration based on the R package "lmec" (Vaida and Liu (2009)). We then obtain the estimates of the survival model's parameters with the true covariate values and random effects substituted by their estimates from the previous step.

We start the MCEM estimation procedure with $k_0 = 100$ Monte-Carlo samples and increase the Monte-Carlo sample size as the number of iteration t increases: $k_{t+1} = k_t + k_t/c$ with $c = 5$ (see Booth and Hobert (1999)). We assess the convergence of the Gibbs sampler by examining time series plots and sample autocorrelation function plots. We notice that the Gibbs sampler converges quickly and the autocorrelations between successive generated samples are negligible after lag 30 (for the random effects) and 10 for the censored covariate. Therefore, we discard the first 300 samples as the burn-in, and then we take one sample from every 30 simulated samples to obtain "independent" samples. Convergence of the EM algorithm was considered to be achieved when the maximum percentage change of all estimates was less than 0.01 in two consecutive iterations.

For comparison purpose, we also consider three alternative joint models for the data, hoping to gain additional insights into the analysis. The first joint model is "LME + AFT," where an LME model is applied to the covariate data and an

AFT model on survival data. The second joint model is "NLME + Cox," where an NLME model is fitted to the covariate data and a Cox model is assumed for the survival data. The third model is "LME + Cox," which is the standard joint model in the literature. For the LME model, we consider the following LME model (quadratic polynomial with random coefficients) based on AIC/BIC criteria $x_{ij} = (\alpha_1 + a_{1i}) + (\alpha_2 + a_{2i})t_{ij} + (\alpha_3 + a_{3i})t_{ij}^2 + e_{ij}$, where a_{1i}, a_{2i}, a_{3i} are the random effects for intercept, slope and quadratic terms and e_{ij} is the error term for measurement error. Again we assume $\boldsymbol{a}_i = (a_{1i}, a_{2i}, a_{3i})^T \sim N(\mathbf{0}, A)$ and $e_{ij} \sim N(0, \sigma^2)$.

Table 1 shows the results of the population parameters estimates for both the covariate model and the survival model under the four modeling approaches. Across the models, the covariate parameters are all estimated with high precision and are all significant. When the NLME model is used, for example, in the "NLME + AFT" and "NLME + Cox" joint models, the magnitude of measurement error was estimated to be smaller, comparing to the LME joint models, for example, "LME + AFT" and "LME + Cox" joint models. For the survival response model, all four models produced a significant (and positive) estimate of the key parameter β_3, which reflects the effects of the longitudinal (true) viral load on the risk of *FRD*.

TABLE 1

Data analysis results of ACTG 388: *estimates* (*est*), *standard errors* (*se*), *and p-values* (*pv*)

	NLME + AFT			LME + AFT			NLME + Cox			LME + Cox		
Model	est	se	pv	est	se	pv	est	se	pv	est	se	pv
α_1	6.40	0.18	<0.001	4.81	0.06	<0.001	6.23	0.15	<0.001	4.86	0.06	<0.001
α_2	5.38	0.74	<0.001	−20.81	0.69	<0.001	4.60	0.64	<0.001	−21.74	0.66	<0.001
α_3	12.56	0.16	<0.001	25.93	1.00	<0.001	12.53	0.15	<0.001	28.13	1.01	<0.001
α_4	122.86	6.96	<0.001				121.65	7.54	<0.001			
α_5	−112.63	7.55	<0.001				−103.24	9.90	<0.001			
σ	0.66	0.01	<0.001	0.97	0.02	<0.001	0.71	0.01	<0.001	0.94	0.01	<0.001
$\log(\lambda)$	0.93	0.10	<0.001	0.91	0.10	<0.001	1.07	0.21	<0.001	1.37	0.17	<0.001
$\log(\gamma)$	0.39	0.11	<0.001	0.54	0.09	<0.001	0.44	0.08	<0.001	0.48	0.07	<0.001
β_1	0.24	0.17	0.185	0.01	0.32	0.998	0.35	0.33	0.352	0.02	0.20	0.873
β_2	0.47	0.24	0.051	0.20	0.21	0.377	0.44	0.31	0.182	0.22	0.24	0.386
β_3	0.18	0.03	<0.001	0.22	0.02	<0.001	0.27	0.10	0.004	0.23	0.07	0.001
β_4	−0.11	0.10	0.663	−0.05	0.06	0.409	−0.10	0.14	0.445	0.02	0.09	0.822
β_5	−0.06	0.08	0.454	−0.01	0.08	0.861	−0.06	0.14	0.643	−0.01	0.11	0.974

Estimates of random effects covariance matrix parameters:

$$A_{\text{NLME + AFT}} = \begin{pmatrix} 3.31 & 1.79 & 0.16 & 0.68 & 0.55 \\ & 26.93 & -0.16 & 4.59 & 1.81 \\ & & 1.10 & 0.10 & 0.23 \\ & & & 21.62 & -0.40 \\ & & & & 18.92 \end{pmatrix}, \quad A_{\text{LME + AFT}} = \begin{pmatrix} 0.04 & 0.17 & 0.88 \\ & 28.93 & -55.24 \\ & & 145.99 \end{pmatrix},$$

$$A_{\text{NLME + Cox}} = \begin{pmatrix} 2.43 & 3.38 & 0.19 & 0.45 & -0.28 \\ & 61.62 & -0.40 & -9.35 & -6.96 \\ & & 0.56 & 0.39 & -0.01 \\ & & & 51.49 & 0.01 \\ & & & & 46.57 \end{pmatrix}, \quad A_{\text{LME + Cox}} = \begin{pmatrix} 0.04 & 0.16 & 1.31 \\ & 36.16 & -76.05 \\ & & 227.28 \end{pmatrix}.$$

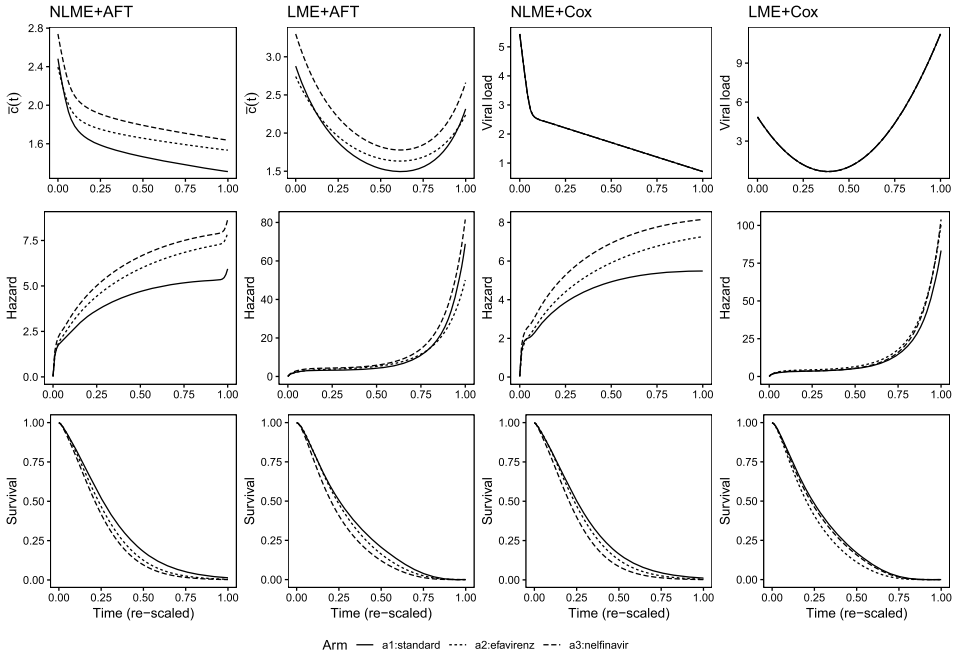

FIG. 2. *Plots for typical individuals' $\bar{c}(t)$ functions, hazard functions, and survival functions based on four joint models (NLME + AFT, LME + AFT, NLME + Cox and LME + Cox) stratified by three treatment arms. As no $\bar{c}(t)$ function is available for non-AFT models (NLE + Cox and LME + Cox), viral load trajectories for the typical individual are displayed.*

However, for the parameter β_2 which measures the effect of treatment "nelfinavir," the proposed "NLME + AFT" joint model gives a nearly significant estimate at 5% level (none of its interaction terms are significant), while all other models fail to detect this significance. This is a new finding from the proposed model.

To gain better understanding of the results, we checked several related functions for a "typical individual" for whom the random effects in the covariate model are all set at zero values as opposed to integrating out the random effects, see Figure 2. For AFT models, we plotted the function $\bar{c}(t)$ (i.e., the average value of the acceleration factor) for each treatment arm. Note that for the current application, we have $\bar{c}(t) = \frac{1}{t}\int_0^t \exp(\beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 x_i^*(s) + \beta_4 z_{1i} x_i^*(s) + \beta_5 z_{2i} x_i^*(s))\, ds$. For Cox models, we only plotted the viral load trajectory for the "typical individual." As a comparison, the corresponding treatment stratified hazard functions and survival functions are also displayed. For the leftmost panel in Figure 2 for the "NLME + AFT" model, we see that treatments lead to different $\bar{c}(t)$ profiles. The "nelfinavir" treatment has the highest level of profile, with the "efavirenz" treatment in the middle and the "standard" 3-drug treatment the lowest. The order of magnitude of the treatment effects is preserved in the hazard functions and the survival functions. For the "LME + AFT" model, hazard and survival functions seem

not distinguishable among the three treatments, especially between the "efavirenz" arm and the "standard" arm. Notice that, for the "LME + Cox" model, the hazard function and survival function are even more severely collided.

In the preceding data analysis, we have considered different models for longitudinal data and survival data respectively. Standard model diagnostics such as residual plots show that these models fit the observed data reasonably well. However, the four joint models offer different estimations of key parameters, both the magnitudes of the estimates and the corresponding p-values, as discussed above. Moreover, for the longitudinal data, the NLME model has the advantage of being biologically justified so that it may predict the left-truncated viral loads better than the empirical LME model. For the survival data, the AFT model provides alternative biological interpretations than the Cox model, and it may also be viewed as a tool for sensitivity analysis, that is, how sensitive the parameter estimates are to the survival model specifications.

5. A simulation study. In this section, we evaluate the proposed model and method when there are different amounts of left-censoring in the time-varying covariate, and we also assess model misspecifications through simulation. We generate the data based on the "NLME + AFT" joint model, and then we conduct data analysis using the "NLME + AFT" joint model and the misspecified "LME + AFT," "NLME + Cox" and "LME + Cox" joint models. The models and their true parameter values in the simulations are chosen to be the same (or similar) to those in the real data analysis. This setting allows us to validate the analysis results in the previous section, in addition to the evaluation of model performance.

We first generate random effects for each individual, $\boldsymbol{a}_i \sim N(\mathbf{0}, A)$, $i = 1, \ldots, n$, and then we simulate the true covariate values from the two-compartment model (4.1). We then generate the event time for each individual based on model (4.2). In the simulation, as we mimic ACTG 388 which ends at week 72 (504 days), event times greater than 504 are treated as right censored. The resulting proportion of right-censored event times is about 5%, which is comparable with the real data. For each individual, we only preserve the portion of the longitudinal data up to his/her event time. Left censoring of the longitudinal data is created at the last step. We consider left-censored longitudinal data with censoring rate at 15% or 30% respectively. We choose the sample size to be the same as the real data, that is, $n = 392$.

Table 2 summarizes the simulation results based on 100 simulation repetitions. We present averaged parameter estimates (Est), averaged standard errors (SE), standard deviations (SD) of the estimated parameter from the simulated samples, percent biases (Bias%) (percentage on the change of the difference between the estimated parameter and the true value relative to the true value) and coverage rate (CR) of the 95% confidence intervals of the parameters. We see that the joint model "NLME + AFT" produces the least biased estimates and coverage probabilities closest to the nominal level 95%. For the "LME + AFT" joint model where

TABLE 2
Simulation results (*NA*: *NLME + AFT*; *LA*: *LME + AFT*; *NC*: *NLME + Cox*; *LC*: *LME + Cox*)

Model	True	α_1 6.40	α_2 5.38	α_3 12.56	α_4 122.86	α_5 −112.63	log(λ) 0.93	log(γ) 0.39	β_1 0.24	β_2 0.47	β_3 0.18	β_4 −0.11	β_5 −0.06
						15% left-censored							
NA	Est	6.42	5.47	12.68	124.96	−113.44	1.06	0.35	0.24	0.33	0.10	−0.09	−0.04
	SE	0.18	0.56	0.14	4.46	5.24	0.08	0.08	0.22	0.24	0.04	0.09	0.08
	SD	0.11	0.85	0.10	4.23	8.98	0.20	0.08	0.17	0.15	0.08	0.06	0.05
	Bias%	1.94	20.33	0.93	4.96	4.27	15.62	−10.81	−7.35	−21.00	−14.29	−24.49	−17.01
	CR	0.95	0.92	0.95	0.93	0.93	0.89	0.94	0.95	0.95	0.85	0.95	0.95
LA	Est	-	-	-	-	-	0.55	0.47	0.21	0.61	0.28	−0.09	−0.11
	SE	-	-	-	-	-	0.31	0.06	0.26	0.27	0.07	0.09	0.09
	SD	-	-	-	-	-	0.35	0.07	0.28	0.29	0.08	0.09	0.09
	Bias%	-	-	-	-	-	−21.05	19.29	−18.61	19.24	24.92	−18.79	44.02
	CR	-	-	-	-	-	0.76	0.74	0.91	0.87	0.74	0.95	0.89
NC	Est	6.50	5.71	12.61	121.66	−112.28	1.32	0.31	0.07	0.53	0.04	−0.07	−0.02
	SE	0.17	0.69	0.11	3.61	4.05	0.17	0.05	0.23	0.24	0.05	0.06	0.07
	SE	0.38	1.17	0.12	9.44	10.89	0.34	0.07	0.24	0.23	0.07	0.08	0.07
	Bias%	1.58	6.19	0.41	−0.97	−0.31	22.40	−20.85	−32.70	12.79	−19.28	−20.78	−33.23
	CR	0.81	0.86	0.88	0.59	0.61	0.86	0.81	0.84	0.94	0.83	0.95	0.72
LC	Est	-	-	-	-	-	0.17	0.44	0.16	0.59	0.15	−0.04	−0.11
	SE	-	-	-	-	-	0.22	0.05	0.32	0.33	0.09	0.12	0.12
	SD	-	-	-	-	-	0.25	0.05	0.35	0.38	0.10	0.12	0.13
	Bias%	-	-	-	-	-	−41.50	71.43	68.68	89.89	64.18	60.27	−91.89
	CR	-	-	-	-	-	0.40	0.60	0.94	0.61	0.41	0.88	0.40
						30% left-censored							
NA	Est	6.23	5.08	12.42	125.55	−116.79	1.21	0.32	0.19	0.38	0.08	−0.07	−0.04
	SE	0.21	0.60	0.15	6.67	6.72	0.11	0.09	0.22	0.27	0.05	0.09	0.10
	SD	0.24	0.76	0.14	12.45	8.98	0.16	0.11	0.29	0.23	0.09	0.12	0.10
	Bias%	−3.45	−31.97	−0.88	10.12	7.73	21.44	−17.27	−23.55	−45.65	−21.08	−35.91	−27.22
	CR	0.90	0.88	0.95	0.92	0.94	0.85	0.92	0.93	0.93	0.86	0.92	0.95
LA	Est	-	-	-	-	-	0.13	0.48	0.27	0.74	0.35	−0.10	−0.15
	SE	-	-	-	-	-	0.41	0.06	0.32	0.32	0.08	0.11	0.11
	SD	-	-	-	-	-	0.50	0.08	0.36	0.36	0.10	0.12	0.11
	Bias%	-	-	-	-	-	−56.99	71.77	95.18	49.61	92.75	87.66	−97.35
	CR	-	-	-	-	-	0.20	0.30	0.86	0.77	0.42	0.84	0.52
NC	Est	6.07	2.82	12.59	115.11	−107.19	1.31	0.30	0.03	0.49	0.02	−0.01	−0.01
	SE	0.16	0.48	0.11	3.41	3.92	0.18	0.05	0.24	0.25	0.04	0.06	0.06
	SD	0.40	1.13	0.12	9.51	10.19	0.50	0.07	0.22	0.25	0.07	0.07	0.08
	Bias%	−5.08	−47.56	0.20	−6.31	−4.83	21.30	−23.30	−50.07	4.41	−49.57	−47.41	−41.52
	CR	0.54	0.05	0.92	0.49	0.58	0.57	0.58	0.77	0.94	0.57	0.51	0.68
LC	Est	-	-	-	-	-	0.51	0.45	0.25	1.05	0.47	−0.11	−0.21
	SE	-	-	-	-	-	0.27	0.05	0.38	0.39	0.10	0.14	0.14
	SD	-	-	-	-	-	0.33	0.05	0.44	0.47	0.13	0.15	0.16
	Bias%	-	-	-	-	-	−48.33	65.01	92.22	118.64	171.53	178.25	−110.97
	CR	-	-	-	-	-	0.10	0.20	0.87	0.47	0.39	0.85	0.44

the longitudinal covariate model is misspecified, we only show the estimation results for the survival model part. As expected, the misspecified model performs worse than the NLME + AFT joint model. For example, when 15% of longitudinal data are left censored, the coverage rate of the confidence intervals are less than 80% for parameters log(λ), log(γ) and β_3, and the coverage drops below 50% when there are 30% censoring in the longitudinal covariate data. Comparing to the "LME + AFT" model, the consequences of the misspecified survival model in the "NLME + Cox" model is less severe. This is not surprising since the fo-

TABLE 3
Fitting NLME + AFT model under different data-generating mechanism

	LME + AFT		NLME + Cox		LME + COX	
Parameter	Bias%	CR	Bias%	CR	Bias%	CR
$\log(\lambda)$	41.37	0.39	−0.29	0.88	22.41	0.44
$\log(\gamma)$	−39.11	0.64	32.28	0.25	−53.72	0.62
β_1	−78.60	0.72	−54.91	0.92	99.12	0.67
β_2	15.45	0.88	−55.74	0.86	−45.45	0.72
β_3	−77.00	0.42	−81.23	0.23	−173.29	0.43
β_4	−43.09	0.70	3.20	0.95	81.01	0.39
β_5	29.44	0.89	−31.23	0.91	162.32	0.57

cus is on the left-censored longitudinal covariate data and Weibull model is also proportional hazards. The misspecified joint model "LME + Cox" represents the worst case since both the covariate model and the survival model are misspecified. We see that the coverage rates for "LME + Cox" can be quite low, for example, only 10% for $\log(\lambda)$ and the percent bias can be more than 100 when the covariate left-censoring rate is at 30%.

Although the simulation results are somewhat expected, they do give us some confidence about the data analysis results in the previous section and some idea about how much worse a misspecified model may perform.

We have conducted another simulation study where we simulate the data under the scenario that the time-varying covariate follows an LME model and the survival response follows a Cox model. Then we fit an NLME + AFT joint model to check the performance of this misspecified joint model. The simulation results are presented in Table 3, where we show the estimates of the parameters from the survival response model, under 15% left censoring of the time-varying covariate data. As expected, the NLME + AFT joint model does not perform as well when it is misspecified. Thus, it is important to correctly specify the joint model. In practice, since the true models are not known, we should choose models that fit observed data reasonably well *and* have reasonable biological justifications or interpretations.

6. Discussion. In some applications, AFT models may be more appealing than Cox models since AFT models do not rely on the proportional hazard assumption and allow for the entire covariate process to impact the disease process. Moreover, AFT models have an attractive interpretation of "speed up" or "slow down" of the disease process, which seems reasonable in HIV studies. Of course, Cox models may be more popular in practice and have their advantages. The main purpose of this article is to provide an alternative modeling approach, hoping to gain new insights into the scientific problem.

When the time-varying covariate is left censored due to a detection limit, understanding the underlying data-generating process is important since it may provide better "predicted values" for the censored data. Our simulation results show that serious biases may occur if the models are misspecified. On the other hand, both the mechanistic NLME model and the empirical LME model may be misspecified for the unobserved censored values since neither can be tested or verified based on the observed data. However, the NLME model is derived based on reasonable biological arguments, and it fits observed data well, while the LME model has no biological justifications. Thus, although the NLME model may still be misspecified, it should be more desirable than the LME model.

The biological relationship between HIV viral dynamics and immunological restoration is a complicated issue. In this paper, we study the possible relationship between the risk of *FRD* and the viral load process before the event. Since viral loads may be left censored and measured with errors, appropriately addressing censoring and measurement errors allows us to more accurately estimate the magnitude and significance of the association between the risk intensity of immunological prognosis and viral load. We attempted several model formulations in the data analysis and obtained consistent conclusions on the strong association between immune prognosis and viral load process. Based on the AFT model with the average accelerated factor ($\bar{c}(t)$), we establish a natural connection between the covariate process and the quantities (e.g., hazard function, survival function) of the time-to-event outcome. Also, based on the proposed "NLME + AFT" joint model, we find that the treatment "nelfinavir" is significantly associated with the event risk, which is unavailable based on other joint models.

The joint models in our implementation are parametric models, including the Cox sub-model where a Weibull baseline function is assumed. Typically Cox models are semiparametric. We choose parametric models because they fit the data reasonably well and the computation is simpler. As a reviewer pointed out, a semiparametric Cox model would mitigate some model misspecification, the computation also becomes more challenging in the current settings.

A major challenge for the joint model in this article is computation, since the baseline hazard function of the AFT model involves unknown parameters from the covariates model as well as censoring in the covariates. A similar "NLME + AFT" joint model was considered in Wu, Liu and Hu (2010), but they considered a shared parameter joint model, which is computationally much simpler since the time-independent random effects from the covariate model replace the time-varying covariates so the AFT model can be written in a log-linear form. Moreover, they did not consider censoring in the covariates. The Monte Carlo EM algorithm in this article can be computationally quite intensive. Alternative approaches include approximate inferences based on Laplace approximations or linearization methods as reviewed in Wu (2010). We may also consider multiple imputation methods where the NLME covariate model is used to generate imputed values for the censored covariates. This approach may be computationally simpler

since software is available for multiple imputations. We are currently investigating approximate methods and multiple imputation methods.

APPENDIX: MULTIVARIATE REJECTION SAMPLING ALGORITHM

Sampling from the full conditional of the random effects can be accomplished by a multivariate rejection algorithm. If the density functions are log-concave in the appropriate parameters, the adaptive rejection algorithm of Gilks and Wild (1992) may be used, as in Ibrahim, Lipsitz and Chen (1999). However, for joint modelling with survival and NLME models, some densities may not be log-concave. In such cases, the multivariate rejection sampling method Geweke (1996) may be used to obtain the desired samples. Booth and Hobert (1999) discussed such a method in the context of complete-data generalized linear mixed models, which can be extended to our models as follows.

Consider sampling from $f(\boldsymbol{a}_i|\boldsymbol{x}_i, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})$. Let $f^*(\boldsymbol{a}_i) = f(\boldsymbol{x}_i|\boldsymbol{a}_i; \boldsymbol{\theta}^{(t)}) \cdot f(T_i, \Delta_i|\boldsymbol{a}_i; \boldsymbol{\theta}^{(t)})$, and $\xi = \sup_{\mathbf{u}}\{f^*(\mathbf{u})\}$. A random sample from $f(\boldsymbol{a}_i|\boldsymbol{x}_i, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})$ can be obtained as follows:

Step 1: sample $\boldsymbol{a}_i^*$ from $f(\boldsymbol{a}_i; \boldsymbol{\theta}^{(t)})$, and independently, sample w from the uniform(0, 1) distribution;

Step 2: if $w \le f^*(\boldsymbol{a}_i^*)/\xi$, then accept $\boldsymbol{a}_i^*$ as a sample point from $f(\boldsymbol{a}_i|\boldsymbol{x}_i, T_i, \Delta_i; \boldsymbol{\theta}^{(t)})$, otherwise, go back to step 1 and continue.

Acknowledgments. The authors thank the Editor, Associate Editor and the anonymous reviewers for their very constructive and helpful comments.

REFERENCES

BERNHARDT, P. W., WANG, H. J. and ZHANG, D. (2014). Flexible modeling of survival data with covariates subject to detection limits via multiple imputation. *Comput. Statist. Data Anal.* **69** 81–91. MR3146878

BOOTH, J. G. and HOBERT, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. Roy. Statist. Soc. Ser. B* **61** 265–285.

BRESLAW, J. A. (1994). Random sampling from a truncated multivariate normal distribution. *Appl. Math. Lett.* **7** 1–6. MR1349883

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data*. *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. MR0751780

DEEKS, S., BARBOUR, J., MARTIN, J. et al. (2000). Sustained CD4+ T cell response after virologic failure of protease inhibitor-based regimens in patients with human immunodeficiency virus infection. *J. Infect. Dis.* **181** 946–953.

FISCHL, M., RIBAUDO, H. and COLLIER, A. (2003). A randomized trial of 2 different 4-drug antiretroviral regimens versus a 3-drug regimen, in advanced human immunodeficiency virus disease. *J. Infect. Dis.* **188** 625–634.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

GEWEKE, J. (1996). *Handbook of Computational Economics*. North-Holland, Amsterdam.

GILKS, W. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41** 337–348.

HUANG, Y., HU, X. J. and DAGNE, G. A. (2014). Jointly modeling time-to-event and longitudinal data: A Bayesian approach. *Stat. Methods Appl.* **23** 95–121. MR3171402

IBRAHIM, J. G., LIPSITZ, S. R. and CHEN, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 173–190. MR1664045

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, New York. MR1924807

LEDERMAN, M. M., CONNICK, E. and LANDY, A. (1998). Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir: Results of AIDS clinical trials group protocol 315. *Journal of Infectious Disease* **178** 79–79.

LINDSTROM, M. J. and BATES, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46** 673–687. MR1085815

MCLACHLAN, G. J. and KRISHNAN, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York. MR1417721

MILLER, R. G. (1976). *Survival Analysis*, 2nd ed. Wiley Classics Library, New York.

PALMISANO, L. and YELLA, L. (2011). A brief history of antiretroviral therapy of HIV infection: Success and challenges. *Ann* 1*st Super Sanita* **47** 44–48.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and FLANNERY, B. P. (2007). *Numerical Recipes*: *The Art of Scientific Computing*, 3rd ed. Cambridge Univ. Press, New York.

RIOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*. Chapman & Hall, London.

SATTAR, A. and SINHA, S. K. (2019). Joint modeling of longitudinal and survival data with a covariate subject to a limit of detection. *Stat. Methods Med. Res.* **28** 486–502. MR3903754

TSENG, Y.-K., HSIEH, F. and WANG, J.-L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika* **92** 587–603. MR2202648

TSENG, Y. K. and YANG, Y. F. (2016). A kernel smooth approach for joint modeling of accelerated failure time and longitudinal data. *Comm. Statist. Simulation Comput.* **45** 1240–1248. MR3486657

VAIDA, F. and LIU, L. (2009). Fast implementation for normal mixed effects models with censored response. *J. Comput. Graph. Statist.* **18** 797–817. MR2750442

WEI, G. C. and TANNER, M. A. (1990). A Monte-Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *J. Amer. Statist. Assoc.* **85** 699–704.

WU, L. (2010). *Mixed Effects Models for Complex Data*. Chapman & Hall, London.

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

WU, 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. MR2758812

WU, H., ZHAO, C. and LIANG, H. (2004). Comparison of linear, nonlinear and semiparametric mixed-effects models for estimating HIV dynamic parameters. *Biom. J.* **46** 233–245. MR2099362

ZHANG, H., WONG, H. and WU, L. (2018). A mechanistic nonlinear model for censored and mismeasured covariates in longitudinal models, with application in AIDS studies. *Stat. Med.* **37** 167–178. MR3738069

ZHANG, H. and WU, L. (2018). A non-linear model for censored and mismeasured time varying covariates in survival models, with applications in human immunodeficiency virus and acquired immune deficiency syndrome studies. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 1437–1450. MR3873715

DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS
INSTITUTE FOR IMPLEMENTATION SCIENCE
IN POPULATION HEALTH
GRADUATE SCHOOL OF PUBLIC HEALTH
AND HEALTH POLICY
CITY UNIVERSITY OF NEW YORK
55 WEST 125TH STREET
NEW YORK, NEW YORK 10027
USA
E-MAIL: hongbin.zhang@sph.cuny.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA V6T 1Z4
CANADA
E-MAIL: lang@stat.ubc.ca