Review Article

Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues

Lang Wu,[1] Wei Liu,[2] Grace Y. Yi,[3] and Yangxin Huang[4]

[1] *Department of Statistics, The University of British Columbia, Vancouver, BC, Canada V6T 1Z2*

[2] *Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3*

[3] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

[4] *Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, USA*

Correspondence should be addressed to Lang Wu, lang@stat.ubc.ca

In the past two decades, joint models of longitudinal and survival data have received much attention in the literature. These models are often desirable in the following situations: (i) survival models with measurement errors or missing data in time-dependent covariates, (ii) longitudinal models with informative dropouts, and (iii) a survival process and a longitudinal process are associated via latent variables. In these cases, separate inferences based on the longitudinal model and the survival model may lead to biased or inefficient results. In this paper, we provide a brief overview of joint models for longitudinal and survival data and commonly used methods, including the likelihood method and two-stage methods.

1. Introduction

Longitudinal data and survival data frequently arise together in practice. For example, in many medical studies, we often collect patients' information (e.g., blood pressures) repeatedly over time and we are also interested in the time to recovery or recurrence of a disease. Longitudinal data and survival data are often *associated* in some ways. The time to event may be associated with the longitudinal trajectories. *Separate* analyses of longitudinal data and survival data may lead to inefficient or biased results. Joint models of longitudinal and survival data, on the other hand, incorporate all information simultaneously and provide valid and efficient inferences.

Figure 1 shows a longitudinal dataset in which CD4 cell counts are measured repeatedly over time in an AIDS study. Here, the time to event could be time to viral rebound,
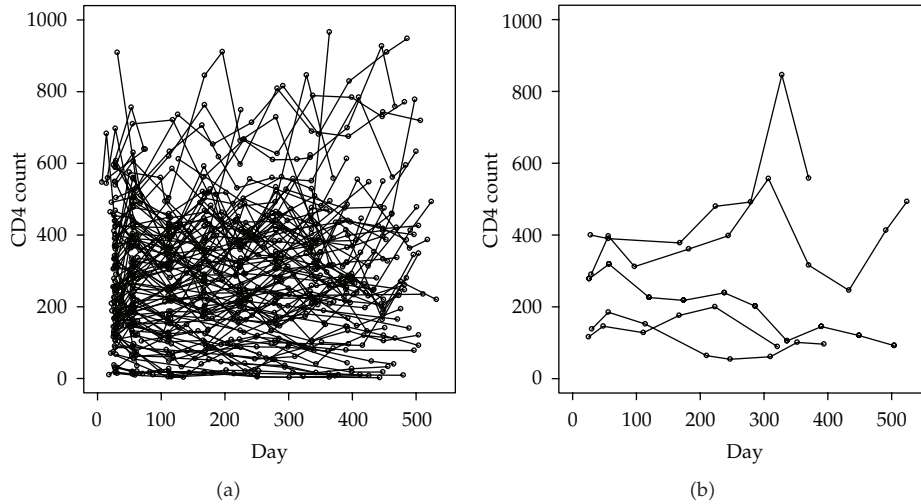
Figure 1: CD4 measurements over time. (a) All subjects. (b) Five randomly selected subjects.

time to dropout, or time to death, depending on the research objectives. Data analysis can mainly focus on either the longitudinal data or the survival data or both. When the analysis focuses on longitudinal data, we often need to address informative dropouts since dropouts are very common in longitudinal studies. When the analysis focuses on survival data, we often need to incorporate time-dependent covariates such as CD4 since the times to event may be associated with the covariate trajectories. Sometimes, the main interest may lie in the association between the longitudinal process and survival process. In any of these cases, joint models are required to feature correlated longitudinal and survival data.

Typically, joint models for longitudinal and survival data are required in the following situations:

   (i) survival models with measurement errors in time-dependent covariates;

  (ii) longitudinal models with informative dropouts;

 (iii) longitudinal and survival processes are governed by a common latent process;

 (iv) the use of external information for more efficient inference.

Joint models of longitudinal and survival data have attracted increasing attention over the last two decades. Tsiatis and Davidian [1] provided a nice overview of early work on joint models, including De Gruttola and Tu [2], Wulfsohn and Tsiatis [3], Henderson et al. [4], and Wang and Taylor [5], among others. More recent work includes Ding and Wang [6], Nathoo and Dean [7], Ye et al. [8], Albert and Shih [9], Jacqmin-Gadda et al. [10], Rizopoulos et al. [11], Wu et al. [12], Huang et al. [13], and Pan and Yi [14], among others. A typical model setting is to assume a mixed-effects model for the longitudinal data and a Cox model or an accelerated failure time (AFT) model for the survival data, with the two models sharing some random effects or variables. The likelihood method is often used, implemented by EM algorithms. Another common approach is based on two-stage methods, which are computationally simpler. Henderson et al. [4] allow different random effects in the longitudinal and survival models, but assume that the random effects are correlated. Bayesian methods have also been proposed [13, 15, 16].

Since the literature on joint models is quite extensive, it is difficult to review all references here. In this paper, we provide a brief review of the joint model literature. In Section 2, we describe a standard formulation of joint models. In Section 3, we review a commonly used two-stage method. In Section 4, we describe the standard likelihood method. In Section 5, we discuss some extensions of standard joint models. A real data example and a simulation study are presented in Section 6 to illustrate and evaluate the methods. We conclude the paper in Section 7 with discussion.

2. A Standard Formulation of a Joint Model

In this section, we consider a standard formulation of a joint model. In the literature, a typical setup is a survival model with measurement errors in time-dependent covariates, in which a linear mixed-effects (LME) model is often used to model time-dependent covariates to address covariate measurement errors and a Cox proportional hazards (PH) model is used for modelling the survival data. We focus on this setup to illustrate the basic ideas.

Consider a longitudinal study with n individuals in the sample. The objective is to model the time to an event of interest or survival time. Time-varying covariates are used in the survival model to partially explain the variation in the event times. Let s_i be the survival time for individual i, $i = 1, 2, \ldots, n$. Some individuals may not experience any events by the end of the study so their event times may be *right censored*. We assume that the censoring is random or *noninformative*. For individual i, let c_i be the censoring time, and let $\delta_i = I(s_i \leq c_i)$ be the censoring indicator such that $\delta_i = 0$ if the survival time for individual i is right censored and $\delta_i = 1$ otherwise. The observed survival data are $\{(t_i, \delta_i),\ i = 1, 2, \ldots, n\}$, where $t_i = \min(s_i, c_i)$.

In survival models, some time-dependent covariates may be measured with errors. For simplicity, we consider a single time-dependent covariate. Let z_{ij} be the *observed* covariate value for individual i at time u_{ij}, subject to measurement errors, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m_i$. Let the corresponding unobserved *true* covariate value be z_{ij}^*. Denote $\mathbf{z}_i = (z_{i1}, \ldots, z_{im_i})^T$, and $\mathbf{z}_i^* = (z_{i1}^*, \ldots, z_{im_i}^*)^T$. In many cases, the longitudinal covariate measurements are terminated at the event or censoring time t_i. For example, this is the case when the events are dropouts. In this case, we have $u_{im_i} \leq t_i$, and the covariate values after the event time t_i are all missing. Let $\mathbf{x}_i$ be covariates without measurement errors.

We consider the following Cox model for the survival data:

$$\lambda_i(t) = \lambda_0(t) \exp\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\right), \quad i = 1, \ldots, n, \tag{2.1}$$

where $\lambda_i(t)$ is the hazard function, $\lambda_0(t)$ is the unspecified baseline hazard function, and $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2^T)^T$ are unknown parameters. Survival model (2.1) assumes that the hazard function $\lambda_i(t)$ depends on the *unobserved true* covariate values $z_i^*(t)$, rather than the observed but mismeasured covariate value $z_i(t)$.

In Cox model (2.1), the time-dependent covariate value $z_i(t)$ should be available at any event time t. In practice, however, covariates are usually measured *intermittently* at times (say) $\{u_{ij},\ j = 1, 2, \ldots, m_i\}$ for individual i, with the measurement times possibly varying across individuals, leading to possible "missing covariates." Moreover, covariates may be measured with errors. Therefore, it is common to have both *missing data* and *measurement for errors* in time-dependent covariates, which must be addressed when conducting inference based on the Cox model (2.1). For simplicity, we assume that the missing covariate data are missing at random [17].

To address either missing data or measurement error or both, a standard approach is to model the time-dependent covariates. A common choice is the following LME model:

$$\mathbf{z}_i = U_i\boldsymbol{\alpha} + V_i\mathbf{a}_i + \boldsymbol{\epsilon}_i \equiv \mathbf{z}_i^* + \boldsymbol{\epsilon}_i, \quad i = 1,\ldots,n, \tag{2.2}$$

where U_i and V_i are known design matrices, $\boldsymbol{\alpha}$ is a vector of fixed-effects parameters, $\mathbf{a}_i$ is a vector of random effects, $\boldsymbol{\epsilon}_i = (\epsilon_{i1},\ldots,\epsilon_{im_i})^T$ is a vector of measurement errors, and the unobserved true covariates are $\mathbf{z}_i^* = U_i\boldsymbol{\alpha} + V_i\mathbf{a}_i$. Often, we assume that

$$\mathbf{a}_i \sim N(0, A), \qquad \boldsymbol{\epsilon}_i \sim N\left(0, \sigma^2 I\right), \tag{2.3}$$

and $\mathbf{a}_i$ and $\boldsymbol{\epsilon}_i$ are independent, where A is a covariance matrix, σ^2 is a parameter, and I is an identity matrix. Here, we focus on the case that the observations of the covariate process is truncated by the event time; that is, no covariate data are available after the event occurs (such as death or dropout).

Note that the survival model (2.1) and the longitudinal model (2.2) are *linked through the shared random effects* $\mathbf{a}_i$. In some applications, not necessarily in the measurement error context, the shared random effects may be viewed as a latent process that governs both the longitudinal process and the survival process. The shared random effects induce the *dependence* between the longitudinal and survival processes, and this dependence suggests the need of joint modelling.

There are two commonly used approaches for inference of joint models:

(i) two-stage methods,

(ii) likelihood methods.

In the following sections, we describe these two approaches in detail. Other approaches for joint modes have also been proposed, such as those based on estimating equations, but we omit them here for space consideration.

3. Two-Stage Methods

In the joint modelling literature, various two-stage methods have been proposed. A simple (naive) two-stage method is as follows.

Stage 1. Fit a LME model to the longitudinal covariate data, and estimate the missing or mismeasured covariates based on the fitted model.

Stage 2. Fit the survival model *separately*, with the missing or unobserved true covariate values substituted by their estimates from the first stage as if they were observed values and then proceed with the usual survival analysis.

Main advantages of the two-stage methods, including the modified two-stage methods as described below, are the simplicity and that they can be implemented with existing software. The limitation of those methods is that they may lead to biased inference for several reasons. First, in the estimation of the longitudinal covariate model parameters, the truncations resulted from the events are not incorporated. That is, the longitudinal covariate trajectories of subjects who experience an event may be different from those who do not

experience that event, so estimation of the parameters associated with the longitudinal covariate model in the first stage, based only on observed covariate data, may be biased. Second, the *uncertainty* of the estimation in the first stage is not incorporated in the second stage of the survival model estimation. Thus, standard errors of the parameter estimates of the survival model may be underestimated. Third, all information in the longitudinal process and the survival process is not fully combined in each model fitting to produce the most efficient estimates.

The bias in the estimation of the longitudinal model parameters caused by ignoring the informative truncations from the events may depend on the *strength of the association* between the longitudinal process and the survival process. The bias resulted from ignoring the estimation uncertainty in Stage 1 may depend on the *magnitude of measurement errors* in covariates. To address these issues, various modified two-stage methods have been proposed, leading to better two-stage methods.

Self and Pawitan [18] considered a two-stage method in which the least-square method was used to fit individual longitudinal covariate trajectories; the resulting estimates were used to impute the "true" covariate values in the survival models and inference was then based on the usual partial likelihood. Tsiatis et al. [19] considered an approximation to the hazard function and then using the approximation to construct the partial likelihood. They replaced the true covariate $z_i^*(t)$ by an empirical Bayes estimate of the conditional expectation $E(z_i^*(t) \mid z_i^H(t), \ t \le s_i)$, where $z_i^H(t) = \{z_i(u), \ u \le t\}$ is the covariate history up to time t. They obtained the empirical Bayes estimate from a standard fit of the LME model to the covariate data up to time t for all subjects still at risk at time t. Similar two-stage methods were also proposed in Bycott and Taylor [20] and Dafni and Tsiatis [21].

More recently, other two-stage methods have been developed in the literature. In the sequel, we review some of these recent methods. Following Prentice [22], we rewrite the survival model (2.1) as

$$\lambda_i(t; z_i(t), \mathbf{x}_i) = \lambda_0(t) E\left[\exp\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\right) \mid z_i(t), \mathbf{x}_i, \ t_i > t\right], \tag{3.1}$$

which involves an intractable conditional expectation. Following Dafni and Tsiatis [21] and Ye et al. [8], we approximate the above conditional expectation by

$$E\left[\exp\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\right) \mid z_i(t), \mathbf{x}_i, \ t_i > t\right]$$
$$\approx \exp\left[E\left(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2 \mid z_i(t), \mathbf{x}_i, \ t_i > t\right)\right]. \tag{3.2}$$

A two-stage method may then proceed as follows. In the first step, we estimate the conditional expectation $E(z_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2 \mid z_i(t), \mathbf{x}_i, \ t_i > t)$ by fitting the covariate model (2.2) to the observed longitudinal data and survival data. In the second step, we then substitute the conditional expectation (3.2) in (3.1) by its estimate from the first step and then we proceed with standard inference for the Cox model. Ye et al. [8] proposed two approaches for the first step, called *risk set regression calibration (RRC)* method and *ordinary regression calibration (ORC)* method, respectively. The idea is to fit the LME covariate model (2.2) to either the observed covariate data in the risk set or all observed covariate data.

Note that the bias resulted from the naive two-stage method is caused by the fact that the covariate trajectory is related to the length of followup. For example, subjects who drop

out early or die early may have different trajectories than those who stay in the study. Thus, much of the bias may be removed if we can recapture these missing covariate measurements due to truncation by incorporating the event time information. Albert and Shih [9] proposed to recapture the missing measurements by generating data from the conditional distribution of the covariate given the event time:

$$f(\mathbf{z}_i \mid s_i; \boldsymbol{\theta}) = \int f(\mathbf{z}_i \mid \mathbf{a}_i; \boldsymbol{\theta}) f(\mathbf{a}_i \mid s_i; \boldsymbol{\theta}) d\mathbf{a}_i, \tag{3.3}$$

where the covariate $\mathbf{z}_i$ and event time s_i are assumed to be conditionally independent given the random effects $\mathbf{a}_i$, and $\boldsymbol{\theta}$ contains all unknown parameters. They approximate the conditional density $f(\mathbf{z}_i \mid s_i; \boldsymbol{\theta})$ using a LME model, and then use standard software to simulate missing data from $f(\mathbf{z}_i \mid s_i; \boldsymbol{\theta})$. Once the missing measurements are simulated, the covariate model is then fitted to the "complete data," which are used in the second step. The procedure is iterated several times to incorporate the missing data uncertainty. Thus, the idea is similar to a multiple imputation method with nonignorable missing data. Such an approach may reduce the bias resulted from truncations.

To incorporate the estimation uncertainty in the first step, we may consider a *parametric bootstrap method* as follows.

Step 1. Generate covariate values based on the assumed covariate model, with the unknown parameters substituted by their estimates.

Step 2. Generate survival times from the fitted survival model.

Step 3. For each generated bootstrap dataset from Steps 1 and 2, fit the models using the two-stage method and obtain new parameter estimates.

Repeating the procedure B times (say, $B = 500$), we can obtain the standard errors for the fixed parameters from the sample covariance matrix across the B bootstrap datasets. This Bootstrap method may produce more reliable standard errors than the naive two-stage method if the assumed models are correct.

Two-stage methods have bearing with the *regression calibration* method in measurement error literature. Many of these two-stage methods may not completely remove biases. Moreover, they rely on certain assumptions and approximations. The validity of these assumptions and the accuracy of these approximations need to be further investigated.

4. Likelihood Methods

The likelihood method is perhaps the most widely used approach in the joint model literature. It provides a unified approach for inference, and it produces valid and the most efficient inference if the assumed models are correct. The likelihood method is based on the likelihood for both longitudinal data and survival data. However, since the likelihood function can be complicated, a main challenge for the likelihood method is computation.

4.1. The Likelihood

All the observed data are $\{(t_i, \delta_i, \mathbf{z}_i, \mathbf{x}_i), \ i = 1, 2, \ldots, n\}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma, A, \boldsymbol{\lambda}_0)$ denote the collection of all unknown parameters in the models, where $\boldsymbol{\lambda}_0 = \{\lambda_0(t_i), \ i = 1, 2, \ldots, n\}$.

We assume that the censoring of survival data and the assessment process of longitudinal measurements are noninformative. The (overall) likelihood for all the observed data is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f\left(t_i, \delta_i \mid \mathbf{z}_i^*, \lambda_0, \boldsymbol{\beta}\right) f\left(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2\right) f(\mathbf{a}_i \mid A) d\mathbf{a}_i, \tag{4.1}$$

where

$$f\left(t_i, \delta_i \mid \mathbf{z}_i^*, \lambda_0, \boldsymbol{\beta}\right) = \left[\lambda_0(t_i) \exp\left(z_i^*(t_i)\beta_1 + \mathbf{x}_i^T\beta_2\right)\right]^{\delta_i}$$

$$\times \exp\left[-\int_0^{t_i} \lambda_0(x) \exp\left(z_i^*(t_i)\beta_1 + \mathbf{x}_i^T\beta_2\right) dx\right],$$

$$f\left(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-m_i/2} \exp\left[-\frac{(\mathbf{z}_i - \mathbf{z}_i^*)^T (\mathbf{z}_i - \mathbf{z}_i^*)}{2\sigma^2}\right],$$

$$f(\mathbf{a}_i \mid A) = (2\pi|A|)^{-1/2} \exp\left[-\frac{(\mathbf{a}_i^T A^{-1}\mathbf{a}_i)}{2}\right].$$

$$\tag{4.2}$$

Parameter estimation can then be based on the observed-data likelihood $L(\boldsymbol{\theta})$ via the maximum likelihood method. Note that the baseline hazard $\lambda_0(t)$ in the Cox model is unspecified. It can be estimated using the nonparametric maximum likelihood method by assuming that $\lambda_0(t)$ takes discrete mass at each failure time t_i. Thus, the dimension of the parameter vector λ_0 is equal to the number of unique failure times. This converts the semiparametric Cox model to a parametric model, but it introduces a major challenge since standard asymptotic theory for the maximum likelihood estimators (MLEs) may not apply due to the infinitely dimensional nature of λ_0.

MLEs of the model parameters can either be obtained by a direct maximization of the observed data log likelihood or by using an EM algorithm. Since the observed data log likelihood involves an intractable integral, a direct maximization is often based on numerical integration techniques such as the Gaussian Hermite quadrature or Monte Carlo methods. These methods, however, can be quite computationally intensive if the dimension of the unobservable random effects $\mathbf{a}_i$ is not low. The EM algorithm is known for its stability and generality, so it is widely used for likelihood inference of joint models [1, 3, 11, 23]. Since the E-step of an EM algorithm still involves an intractable integral, Monte Carlo methods or Laplacian approximations are often used to approximate the conditional expectation in the E-step. In the M-step, the Newton-Raphson method is often used.

Hsieh et al. [24] noted that standard errors for estimators of the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)$ based on the Fisher information matrix may be problematic, because of the semiparametric nature of the joint model. They recommended a bootstrap method to obtain standard errors.

4.2. Computational Issues

A main challenge in the likelihood inference for joint models is the computational complexity, since numerical methods or Monte Carlo methods can be very computationally intensive

when the dimension of the random effects $\mathbf{a}_i$ is not small. Moreover, convergence of the EM algorithms can sometimes be an issue. Tsiatis and Davidian [1] and Tsiatis and Davidian [25] proposed an alternative approach, a so-called conditional score method, which makes no distributional assumption for the random effects $\mathbf{a}_i$. Their method treats $\mathbf{a}_i$ as "nuisance parameters" and conditions on an appropriate "sufficient statistic" for $\mathbf{a}_i$. Conditioning on this sufficient statistic would remove the dependence of the conditional distribution on the random effects $\mathbf{a}_i$. This approach leads to a set of unbiased estimating equations, similar to the usual partial likelihood score equations or generalized estimating equation (GEE). The resulting estimates are, under certain regularity conditions, consistent and asymptotically normal, although they may not be the most efficient. Moreover, their method is relatively simple to implement. Song et al. [26] considered an alternative approach to relax the normality assumption of $\mathbf{a}_i$. They assume that the random effects follow distributions in a class of smooth density family, including the standard normal density as a special case. They use the likelihood method for inference via an EM algorithm, but the computation is considerably more intensive than the conditional score method. Song and Wang [27] also proposed a local corrected score estimator and a local conditional score estimator, for which no distributional assumptions are needed for the underlying true covariates.

Approximate but computationally more efficient methods for joint models have also appeared in the literature, such as those based on Laplace approximations (e.g., [11, 12, 28]). When the dimension of the integration in the likelihood for joint models is high, the Laplace approximations offer considerably computational advantages over numerical or Monte Carlo methods. Note that the order of the Laplace approximation error is $O(m_i^{-1})$, which cannot be made arbitrarily accurate for a given dataset, where m_i is the number of within-individual measurements for individual i. Therefore, the Laplace approximation works well if the number of within-individual measurements is large. Approximate methods based on Taylor series approximations are similar to the Laplace approximation; that is, their performance improves as m_i increases.

Rizopoulos et al. [11] proposed to use the full exponential Laplace approximation in the E-step of the EM algorithm. Compared to the standard (first-order) Laplace approximation, the full exponential Laplace approximate method has approximation error of order $O(m_i^{-2})$ and requires a much smaller number of within-individual longitudinal measurements to produce reliable results. Lee et al. [28] suggested second-order Laplace approximations. However, these Laplace approximation methods cannot control the magnitude of the approximation errors, unlike Gaussian quadrature or Monte Carlo integration techniques.

5. Bayesian Methods

Bayesian joint models have also been studied by various authors, including Faucett and Thomas [29], Xu and Zeger [15], Wang and Taylor [5], Law et al. [30], Ibrahim et al. [16], and Huang et al. [13]. Joint models may contain many unknown parameters, which may lead to potential problems in inference. A main advantage of Bayesian methods is that they can borrow additional information from similar studies or from experts and incorporate this information in the current analysis, in the forms of prior distributions for the current model parameters. Thus, Bayesian methods can be very useful for inference of joint models.

For Bayesian joint models, the model parameters are assumed to follow some prior distributions, and inference is then based on the posterior distribution given the observed data. Let $\boldsymbol{\theta}$ denote the collection of unknown parameters in the joint model, and let $f(\boldsymbol{\theta})$

denote the prior distribution. Let $\mathfrak{D} = \{(t_i, \delta_i, \mathbf{z}_i, \mathbf{x}_i),\ i = 1, 2, \ldots, n\}$ denote all observed data. The joint *posterior distribution* for all unknown parameters $\boldsymbol{\theta}$ and random effects $\mathbf{a} = \{\mathbf{a}_i,\ i = 1, \ldots, n\}$ is then given by

$$f(\boldsymbol{\theta}, \mathbf{a} \mid \mathfrak{D}) \propto \prod_{i=1}^{n} \left[f(t_i, \delta_i \mid \mathbf{z}_i^*, \mathbf{x}_i, \boldsymbol{\theta}) f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\theta}) f(\mathbf{a}_i \mid A) \right] f(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0), \tag{5.1}$$

where $\boldsymbol{\theta}_0$ are known hyperparameters. Bayesian inference is then based on Monte Carlo samples drawn from the posterior distribution $f(\boldsymbol{\theta}, \mathbf{a} \mid \mathfrak{D})$ using an MCMC algorithm such as the Gibbs sampler. For example, the posterior means and variances of the parameters can be estimated based on these Monte Carlo samples, and Bayesian inference can then be based on these estimated posterior means and variances. This Monte Carlo sampling can be done using the publically available WinBUGS software [31], which is quite general, flexible, and easy to use.

Like other Bayesian methods, it is desirable to check if the final results are sensitive to the choices of prior distributions. Sometimes, in the absence of prior information, noninformative priors or flat priors may be desirable.

6. Other Joint Models

In the previous sections, we have focused on joint models based on a Cox model for right-censored survival data and a LME model for longitudinal data. Other models for survival data and longitudinal data can also be considered in joint models. For example, for survival data, we may consider accelerated failure time (AFT) models and models for interval censored data and models for recurrent events. For longitudinal data, nonlinear, generalized linear mixed models or semiparametric/nonparametric mixed models can be utilized. Although the different survival models and longitudinal models can be employed, basic ideas and approaches for inference remain essentially the same. In the following, we briefly review some of these joint models.

6.1. Joint Models Based on an LME Model and an AFT Model

In joint modelling of longitudinal and survival data, we can use the AFT model to feature survival data. Here, we focus on an AFT model with measurement errors in time-dependent covariates. For longitudinal data, we again consider LME models for simplicity. The description below is based on Tseng et al. [23]. A semiparametric AFT model can be written in a form similar to the Cox model:

$$h_i(t) = h_0 \left[\int_0^t \exp\{-z_i^*(u)\beta\} du \right] \exp\{-z_i^*(t)\beta\}, \tag{6.1}$$

where $h_i(t)$ is the hazard function of the ith individual at time t, $h_0(t)$ is the baseline hazard function, and $z_i^*(t)$ is the unobserved true covariate value at time t. For the observed measurements $z_i(t)$, we again consider the LME model (2.2).

Tseng et al. [23] proposed a likelihood method using an EM algorithm. The likelihood for all observed data is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(t_i, \delta_i \mid \mathbf{z}_i^*, h_0, \beta) f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2) f(\mathbf{a}_i \mid A) d\mathbf{a}_i, \tag{6.2}$$

where $f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma^2)$ and $f(\mathbf{a}_i \mid A)$ are the same as those for (4.1) and

$$f\left(t_i, \delta_i \mid \mathbf{z}_i^*, h_0, \boldsymbol{\beta}\right) = \left[h_0\left\{\phi(t_i; \boldsymbol{\theta}, \mathbf{a}_i)\right\} \frac{\partial \phi\left(t_i; \bar{\mathbf{z}}_i^*, \beta\right)}{\partial t_i}\right]^{\delta_i} \exp\left\{-\int_0^{\phi(t_i; \bar{\mathbf{z}}_i^*, \beta)} h_0(u)du\right\}, \qquad (6.3)$$

where $\bar{\mathbf{z}}_i^*$ denotes the covariate history and ϕ is a known function.

Handling the AFT structure in the joint modelling setting is more difficult than for the Cox model, since $f(t_i, \delta_i \mid \mathbf{z}_i^*, h_0, \boldsymbol{\beta})$ is more complicated and the baseline function $h_0\{\phi(t_i; \bar{\mathbf{z}}_i^*, \boldsymbol{\beta})\}$ involves unknown quantities $(\boldsymbol{\beta}, \boldsymbol{\alpha}, A, \mathbf{a}_i)$, while this is not the case in the Cox model. One cannot use the point mass function with masses assigned to all uncensored survival times t_i for the baseline hazard function h_0. In other words, in Cox models, the baseline hazard h_0 can be represented by a collection of parameters which are point masses, but this approach is not feasible for the AFT model because of its dependence on covariates via function $\phi(t_i; \bar{\mathbf{z}}_i^*, \boldsymbol{\beta})$. To circumvent this, Tseng et al. [23] assumed the baseline hazard function h_0 to be a step function, taking constant values between two consecutive failure times.

Tseng et al. [23] used an Monte Carlo EM algorithm to obtain the MLEs. The framework is similar to that in the previous section. They used a Monte Carlo method to approximate the conditional expectations in the E-step. The M-step involves more complicated computations due to the complicated baseline hazard h_0. To obtain the standard errors of the MLEs, the usual asymptotic formula based on Fisher information matrix may be questionable, so they used a bootstrap method.

6.2. Joint Models with Interval Censored Survival Data

In the previous sections, we have focused on right censored survival data and assume that either the *exact* survival times or censoring times are observed. In practice, however, we often cannot observe the exact survival nor censoring times, but we only know that events have occurred over certain time intervals. Such survival data are called *interval censored*. For simplicity, we assume that all individuals are assessed at the same times. Again, let S_i be the time to an event (survival time) for individual i, with observed value s_i. Let $\mathbf{r}_i = (r_{i1}, \ldots, r_{im})^T$ be the vector of event indicators such that $r_{ij} = 1$ if subject i has an event occurred from time t_{j-1} to time t_j, and let $r_{ij} = 0$ otherwise, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$. We assume that $r_{i1} = 0$ for all i. Let $p_{ij} = P(t_{j-1} \leq S_i < t_j)$, and let

$$\pi_{ij} = P\left(t_{j-1} \leq S_i < t_j \mid S_i \geq t_{j-1}\right) = 1 - P\left(S_i \geq t_j \mid S_i \geq t_{j-1}\right). \qquad (6.4)$$

Then, we have $p_{ij} = (1 - \pi_{i1})(1 - \pi_{i2}) \cdots (1 - \pi_{i,j-1})\pi_{ij}$. The probability function for the event indicator vector $\mathbf{r}_i$ can be written as

$$f(\mathbf{r}_i) = \prod_{j=1}^{m} p_{ij}^{r_{ij}} = \prod_{j=1}^{m} \pi_{ij}^{r_{ij}} \left(1 - \pi_{ij}\right)^{1-r_{ij}}, \qquad (6.5)$$

which is the probability function for a Bernoulli distribution. We can introduce observed error-prone covariate value $\mathbf{z}_i$, with true value $\mathbf{z}_i^*$, and assume

$$\log\left\{-\log(1 - \pi_{ij})\right\} = \boldsymbol{\beta}^T \mathbf{z}_i^* + \gamma_j, \qquad (6.6)$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$ are unknown parameters. Then, we can write the probability function of $\mathbf{r}_i$ as $f(\mathbf{r}_i \mid \mathbf{z}_i^*, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Alternatively, we can assume that $\mathbf{r}_i$ depends on $\mathbf{z}_i^*$ only through the random effects $\mathbf{a}_i$ and writing the probability function of $\mathbf{r}_i$ as $f(\mathbf{r}_i \mid \mathbf{a}_i, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

Let $f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma)$ be the conditional probability density function, given the random effects $\mathbf{a}_i$, and $f(\mathbf{a}_i \mid A)$ be the marginal probability density function for $\mathbf{a}_i$ with covariance matrix A. Let $\boldsymbol{\theta}$ denote the collection of all parameters in all models. Then, the *likelihood* for all the observed data can be written as

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^n \left[\int f(\mathbf{z}_i \mid \mathbf{a}_i, \boldsymbol{\alpha}, \sigma) f(\mathbf{r}_i \mid \mathbf{a}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) f(\mathbf{a}_i \mid A) d\mathbf{a}_i \right]. \tag{6.7}$$

MLE of parameters $\boldsymbol{\theta}$ can be obtained by maximizing the observed data likelihood $L_o(\boldsymbol{\theta})$. Because the observed-data likelihood $L_o(\boldsymbol{\theta})$ can be difficult to evaluate due to its involvement of an intractable and possibly high-dimensional integral, one may proceed with Monte Carlo EM algorithms or other computationally more efficient approximate methods.

6.3. GLMM and NLME Models for Longitudinal Data

We have focused on LME models for modelling the longitudinal data. Other models for longitudinal data can also be considered. For example, one may consider nonlinear mixed-effects (NLME) models for modelling the longitudinal data in joint models [12, 32]. NLME models are often mechanistic models in the sense that they are typically based on the underlying mechanisms which generate the data. On the other hand, LME models are typically empirical models; that is, they are usually used to approximately describe the observed data without considering possible data-generation mechanisms. Thus, NLME models may be scientifically more desirable if such models exist. Similarly, for nonnormal longitudinal data, generalized linear mixed models (GLMMs) can be considered, which are special nonlinear models but are essentially empirical models as well.

When the longitudinal models are nonlinear, the general ideas of the two-stage methods and likelihood methods for joint models can still be applied. The complication is that computation becomes more demanding, because of the nonlinearity of the longitudinal models.

6.4. Joint Models with Missing Data

For longitudinal data, missing values are very common. When missing data are nonignorable in the sense that the missingness probability may be related to the missing values or the random effects, the missing data process is often needed to be incorporated in inferential procedures in order to obtain valid results. For likelihood methods, it is straightforward to incorporate missing data mechanisms in joint model inference. However, the computation becomes even more challenging. Wu et al. [12, 32] considered the missing data problems for joint models, using Monte Carlo EM algorithms and Laplace approximations.

7. Example and Simulation

7.1. An Illustrating Example

As an illustration, we consider an AIDS dataset which includes 46 HIV infected patients receiving an anti-HIV treatment. Viral load (i.e., plasma HIV RNA) and CD4 cell count were

repeatedly measured during the treatment. The number of viral load measurements for each individual varies from 4 to 10. It is known that CD4 is measured with substantial errors. About 11% viral load measurements are below the detection limit after the initial period of viral decay. We call the viral load below the detection limit as "viral suppression." We wish to check if the initial CD4 trajectories are predictive for the time to viral suppression.

Let s_i be the time to viral suppression, that is, the time from the start of the treatment to the first scheduled time when the viral load drops below the detection limit. The viral suppression times for patients whose viral loads never dropped below detection limit may be regarded as right censored, with the study end time as the censoring time. We employ two models, the Cox model (2.1) or the AFT model (6.1), to feature the time to viral suppression, where $z_i^*(t)$ is the unobserved true CD4 cell count at time t for individual i. We do not consider additional covariates here.

To address the measurement error in the time-dependent covariate CD4 cell count, we use the LME model to model the CD4 trajectories:

$$\text{CD4}_{ij} = (\alpha_0 + a_{i1}) + (\alpha_1 + a_{i2}u_{ij}) + \epsilon_{ij}, \tag{7.1}$$

where the parameters, random effects and random errors, and the assumed distributions are the same as those described in the previous sections. The fixed parameters (α_0, α_1) are population intercept and slope of the CD4 process and (a_{i1}, a_{i2}) are individual variations from the population averages. To avoid very small (large) estimates, which may be unstable, we standardize the CD4 cell counts and rescale the original time t (in days) so that the new time scale is between 0 and 1. We estimate the model parameters using the joint model method and the two-stage method with/out bootstrap standard error correction. The number $B = 500$ of the bootstrap samples is taken. For the joint model method, we consider the Cox model (2.1) and the AFT model (6.1) with $h_0(\cdot)$ being the Weibull baseline risk function. On the other hand, only the Cox model (2.1) is employed for the two-stage method for comparison. These analyses may be implemented by using the functions $coxph()$, $lme()$, and $jointModel()$ in R software.

Table 1 presents the resulting estimates of main parameters of interest and their standard errors. We can see from Table 1 that under either the Cox or the AFT survival models, both the two-stage and the joint model methods produce similar estimates for the covariate (CD4) longitudinal model. However, the two-stage method may underestimate the standard errors of the parameter estimates since it does not incorporate the survival information in the estimation procedure. Note that the parameters in the Cox model and the AFT model have different interpretations, due to different model formulations, so they are not directly comparable.

The parameter β_1 in the survival models measures the effect of the true time-dependent covariate CD4 values on event times, so its estimate and the associated P value can be used to check if the true CD4 values are predictive for the times to viral suppression. Since the covariate CD4 is measured with errors, addressing the measurement error is the main focus of joint modelling in this application. Thus, the estimation of β_1 is of primary interest. For the joint model method, under either the Cox or the AFT models, there is some evidence that covariate CD4 is associated with the time to viral suppression, after measurement error has been addressed. It is seen that evidence of significance of the covariate effect is the strongest under the Cox model. On the other hand, the two-stage method may severely underestimate the covariate CD4 effect (the small value of $\widehat{\beta}_1$). Moreover, the naive two-stage method underestimates the standard error of $\widehat{\beta}_1$, due to failing to incorporate the estimating

Table 1: Analyses of the AIDS data under different models.

Model	Method		β_0	β_1	α_0	α_1	σ
Cox model	Two-stage	Estimate	—	0.315	−0.233	1.345	0.605
		SE	—	0.208	0.113	0.154	—
		P value	—	0.129	0.040	<0.001	—
		BSE	—	0.237	—	—	—
Cox model	Joint model	Estimate	—	0.648	−0.201	1.342	0.603
		SE	—	0.234	0.135	0.162	—
		P value	—	0.006	0.137	<0.001	—
AFT model	Joint model	Estimate	0.168	−0.487	−0.237	1.341	0.604
		SE	0.260	0.289	0.125	0.156	—
		P value	0.517	0.091	0.059	<0.001	—

SE: standard error; BSE: bootstrap standard error. The *P* values are from the Wald tests for testing if the corresponding parameters are zero or not.

uncertainty from the first step. This underestimation of standard error is somewhat corrected by the bootstrap method.

7.2. A Simulation Study

In this section, we conduct a simulation study to compare the joint model method and the two-stage method with/out bootstrap standard error correction. We generate 500 datasets from the time-dependent covariate CD4 process (7.1) in the example of the previous section and the Cox model (2.1) with constant baseline hazard function $\lambda_0(t) \equiv 1$ with emphasis on the effect of the time-dependent covariate. The measurement time points used in the simulation are the same as those in the example of the previous section. The true values of model parameters, given in Table 2, are similar to those in the example, and the variance-covariance matrix A of the random effect $\mathbf{a}_i$ is set to be diagonal. Again, we take $B = 500$ bootstrap samples for the bootstrap standard error in the two-stage method.

In Table 2, we report the simulation results of averages parameter estimates (Est), empirical standard errors (ESE), averages of asymptotic standard errors (ASE), average of bootstrap standard errors (BSE), empirical biases (Bias), mean square errors (MSE), and coverage rates (CR) for the 95% confidence intervals. We can see from Table 2 that the two-stage and the joint model methods produce similar parameter estimates (close to true parameters) except the one for the covariate effect β_1. In particular, the estimate $\widehat{\beta}_1$ based on the joint model method is very close to its true value, while the estimate based on the two-stage method is about one-third of its true value, which indicates that the two-stage method may underestimate the time-dependent covariate effect severely. The joint model method provides smaller mean square errors and more reasonable coverage rates for the 95% confidence intervals than the two-stage method. Moreover, the two-stage method may underestimate the standard deviation of $\widehat{\beta}_1$ and the bootstrap correction on this standard error seems plausible.

From the above results, we see that the joint likelihood method produces less biased estimates and more reliable standard errors than the two-stage method. These results have important implications. For example, if one uses Wald-type tests for model selection, the likelihood method would give more reliable results. However, two-stage methods are generally simpler and computationally quicker to output estimates than likelihood methods.

Table 2: Comparison of the two-stage Method and the joint likelihood method via a simulation study.

Method	Parameter	β_1	α_0	α_1	σ	A_{11}	A_{22}
	True value	0.6	−0.2	1.3	0.6	0.5	0.3
Two-stage	Est	0.183	−0.183	1.303	0.598	0.501	0.353
	ESE	0.216	0.111	0.164	0.025	0.114	0.209
	ASE	0.201	0.111	0.164	—	—	—
	BSE	0.250	—	—	—	—	—
	Bias	−0.417	0.017	0.003	−0.002	−0.001	0.053
	MSE	0.221	0.013	0.027	0.0007	0.013	0.046
	CR	42.8	95.6	94.4	—	—	—
Joint model	Est	0.6004	−0.175	1.296	0.598	0.492	0.321
	ESE	0.256	0.103	0.161	0.020	0.092	0.156
	ASE	0.249	0.099	0.163	—	—	—
	Bias	0.0004	0.025	−0.004	−0.002	−0.008	0.021
	MSE	0.066	0.011	0.026	0.0004	0.008	0.025
	CR	95.6	95.8	95.2	—	—	—

We can also compare the two methods with Bayesian methods. Note that, however, Bayesian methods are equivalent to the likelihood method when noninformative priors are used. We expect that Bayesian methods have similar performance to likelihood methods.

8. Discussion

We have provided a brief review of common joint models and methods for inference. In practice, when we need to consider a longitudinal process and an event process and suspect that the two processes may be associated, such as survival models with time-dependent covariates or longitudinal models with informative dropouts, it is important to use joint model methods for inference in order to avoid biased results. The literature on model selection for joint models is quite limited. In practice, the best longitudinal model can be selected based on the observed longitudinal data, and the best survival model can be selected based on the survival data, using standard model selection procedures for these models. Then, we specify reasonable link between the two models, such as shared random effects. To choose methods for inference, the joint likelihood method generally produces most reliable results *if* the assumed models and distributions are correct. On the other hand, the two-stage methods may be computationally simpler, and many existing models and methods for longitudinal data and survival data can be easily adapted. However, two-stage methods may not completely eliminate the biases in parameter estimates in some cases.

When the longitudinal covariate process terminates at event times, that is, when the longitudinal values are unavailable at and after the event times such as deaths or dropouts, the covariates are sometimes called *internal time-dependent covariates*. Sometimes, however, longitudinal covariate information is available at and after the event times. For example, CD4 measurements may be still available after patients have been diagnosed with AIDS. Such covariates are sometimes called *external covariates*. In joint models, it is important to distinguish internal and external time-dependent covariates. In particular, for internal time-dependent covariates, joint models are more desirable since separate analysis in this case may lead to more severe bias.

Survival models with measurement errors in time-dependent covariates have received much attention in the joint models literature. Another common situation is longitudinal models with informative dropouts, in which survival models can be used to model the dropout process. Both situations focus on characterizing the association between the longitudinal and survival processes. Some authors have also considered joint models in which the focus is on more efficient inference of the survival model, using longitudinal data as auxiliary information [15, 33, 34] or assume that the longitudinal process and the survival process are governed by a common latent process [4]. Nathoo and Dean [7] considered an interesting joint model in which an NLME model is used to model tree growth, with spatial correlation incorporated.

Joint models can also be extended to *multivariate cases*, in which more than one longitudinal processes and more than one event processes can be modelled simultaneously. Extensions are often conceptually straightforward, but compuation and implementation can be more tedious than univariate cases. See Henderson et al. [4], Xu and Zeger [35], and Song et al. [26].

Zeng and Cai [36] derived some asymptotic results for maximum likelihood estimators in joint analysis of longitudinal and survival data. They showed the consistency of the maximum likelihood estimators, derived their asymptotic distributions, and showed that the maximum likelihood estimators in joint analysis are semiparametrically efficient.

Although there has been extensive research in joint models in the last two decades and the importance of joint models has been increasingly recognized, joint models are still not widely used in practice. A main reason is perhaps lack of software. Recently, Dimitris Rizopoulos has developed an R package called *JM* that can be used to fit joint models with normal longitudinal responses and event times under a maximum likelihood approach. Various options for the survival model and optimization/integration algorithms are provided, such as Cox models and AFT models for survival data and the Gauss-Hermite integration methods and Laplace approximations.

Acknowledgments

References

[1] A. A. Tsiatis and M. Davidian, "An overview of joint modeling of longitudinal and time-to-event data," *Statistica Sinica*, vol. 14, pp. 793–818, 2004.

[2] V. De Gruttola and X. M. Tu, "Modelling progression of CD4-lymphocyte count and its relationship to survival time," *Biometrics*, vol. 50, no. 4, pp. 1003–1014, 1994.

[3] M. S. Wulfsohn and A. A. Tsiatis, "A joint model for survival and longitudinal data measured with error," *Biometrics*, vol. 53, no. 1, pp. 330–339, 1997.

[4] R. Henderson, P. Diggle, and A. Dobson, "Joint modelling of longitudinal measurements and event time data," *Biostatistics*, vol. 1, pp. 465–480, 2000.

[5] Y. Wang and J. M. G. Taylor, "Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 895–905, 2001.

[6] J. Ding and J.-L. Wang, "Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data," *Biometrics*, vol. 64, no. 2, pp. 546–556, 2008.

[7] F. S. Nathoo and C. B. Dean, "Spatial multistate transitional models for longitudinal event data," *Biometrics*, vol. 64, no. 1, p. 271–279, 326, 2008.

[8] W. Ye, X. Lin, and J. M. G. Taylor, "Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach," *Biometrics*, vol. 64, no. 4, pp. 1238–1246, 2008.

[9] P. S. Albert and J. H. Shih, "On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure," *Biometrics*, vol. 66, no. 3, pp. 983–987, 2010.

[10] H. Jacqmin-Gadda, C. Proust-Lima, J. M. G. Taylor, and D. Commenges, "Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model," *Biometrics*, vol. 66, no. 1, pp. 11–19, 2010.

[11] D. Rizopoulos, G. Verbeke, and E. Lesaffre, "Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data," *Journal of the Royal Statistical Society Series B*, vol. 71, no. 3, pp. 637–654, 2009.

[12] L. Wu, W. Liu, and X. J. Hu, "Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors," *Biometrics*, vol. 66, no. 2, pp. 327–335, 2010.

[13] Y. Huang, G. Dagne, and L. Wu, "Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors," *Statistics in Medicine*, vol. 30, no. 24, pp. 2930–2946, 2011.

[14] Q. Pan and G. Y. Yi, "An estimation method of marginal treatment effects on correlated longitudinal and survival outcomes," *Statistics and Its Inference*. In press.

[15] J. Xu and S. L. Zeger, "Joint analysis of longitudinal data comprising repeated measures and times to events," *Journal of the Royal Statistical Society Series C*, vol. 50, no. 3, pp. 375–387, 2001.

[16] J. G. Ibrahim, M. Chen, and D. Sinha, "Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials," *Statistica Sinica*, vol. 14, no. 3, pp. 863–883, 2004.

[17] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York, NY, USA, 2nd edition, 2002.

[18] S. Self and Y. Pawitan, "Modeling a marker of disease progression and onset of disease," in *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell, Eds., Birkhäuser, Boston, Mass, USA, 1992.

[19] A. A. Tsiatis, V. DeGruttola, and M. S. Wulfsohn, "Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS," *Journal of the American Statistical Association*, vol. 90, pp. 27–37, 1995.

[20] P. Bycott and J. Taylor, "A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model," *Statistics in Medicine*, vol. 17, no. 18, pp. 2061–2077, 1998.

[21] U. G. Dafni and A. A. Tsiatis, "Evaluating surrogate markers of clinical outcome when measured with error," *Biometrics*, vol. 54, no. 4, pp. 1445–1462, 1998.

[22] R. L. Prentice, "Covariate measurement errors and parameter estimation in a failure time regression model," *Biometrika*, vol. 69, no. 2, pp. 331–342, 1982.

[23] Y. K. Tseng, F. Hsieh, and J. L. Wang, "Joint modelling of accelerated failure time and longitudinal data," *Biometrika*, vol. 92, no. 3, pp. 587–603, 2005.

[24] F. Hsieh, Y.-K. Tseng, and J.-L. Wang, "Joint modeling of survival and longitudinal data: likelihood approach revisited," *Biometrics*, vol. 62, no. 4, pp. 1037–1043, 2006.

[25] A. A. Tsiatis and M. Davidian, "A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error," *Biometrika*, vol. 88, no. 2, pp. 447–458, 2001.

[26] X. Song, M. Davidian, and A. A. Tsiatis, "A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data," *Biometrics*, vol. 58, no. 4, pp. 742–753, 2002.

[27] X. Song and C. Y. Wang, "Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients," *Biometrics*, vol. 64, no. 2, pp. 557–566, 2008.

[28] Y. Lee, J. A. Nelder, and Y. Pawitan, *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*, vol. 106 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, London, UK, 2006.

[29] C. L. Faucett and D. C. Thomas, "Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach," *Statistics in Medicine*, vol. 15, no. 15, pp. 1663–1685, 1996.

[30] N. J. Law, J. M. G. Taylor, and H. M. Sandler, "The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure," *Biostatistics*, vol. 3, pp. 547–563, 2002.

[31] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, vol. 10, no. 4, pp. 325–337, 2000.

[32] L. Wu, X. J. Hu, and H. Wu, "Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data," *Biostatistics*, vol. 9, no. 2, pp. 308–320, 2008.

[33] C. L. Faucett, N. Schenker, and J. M. G. Taylor, "Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data," *Biometrics*, vol. 58, no. 1, pp. 37–47, 2002.

[34] J. W. Hogan and N. M. Laird, "Model-based approaches to analysing incomplete longitudinal and failure time data," *Statistics in Medicine*, vol. 16, no. 1–3, pp. 259–272, 1997.

[35] J. Xu and S. L. Zeger, "The evaluation of multiple surrogate endpoints," *Biometrics*, vol. 57, no. 1, pp. 81–87, 2001.

[36] D. Zeng and J. Cai, "Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time," *Annals of Statistics*, vol. 33, no. 5, pp. 2132–2163, 2005.