ANNUAL REVIEWS

*Annual Review of Statistics and Its Application*

# Joint Modeling of Longitudinal and Survival Data

Jane-Ling Wang[1] and Qixian Zhong[2]

[1]Department of Statistics, University of California, Davis, California, USA; email: janelwang@ucdavis.edu

[2]Department of Statistics and Data Science, Xiamen University, Xiamen, China; email: qxzhong@xmu.edu.cn

Annu. Rev. Stat. Appl. 2025. 12:449–76

First published as a Review in Advance on November 14, 2024

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-112723-034334



## Keywords

censoring, time-to-event, nonparametric maximum likelihood estimator, transformation survival model, linear mixed-effects model, incomplete data, EM algorithm, expectation–maximization algorithm

## Abstract

In medical studies, time-to-event outcomes such as time to death or relapse of a disease are routinely recorded along with longitudinal data that are observed intermittently during the follow-up period. For various reasons, marginal approaches to model the event time, corresponding to separate approaches for survival data/longitudinal data, tend to induce bias and lose efficiency. Instead, a joint modeling approach that brings the two types of data together can reduce or eliminate the bias and yield a more efficient estimation procedure. A well-established avenue for joint modeling is the joint likelihood approach that often produces semiparametric efficient estimators for the finite-dimensional parameter vectors in both models. Through a transformation survival model with an unspecified baseline hazard function, this review introduces joint modeling that accommodates both baseline covariates and time-varying covariates. The focus is on the major challenges faced by joint modeling and how they can be overcome. A review of available software implementations and a brief discussion of future directions of the field are also included.

## 1. INTRODUCTION

In clinical trials, it is common to collect baseline variables from individuals at the onset of the study along with longitudinal measurements and an event time of interest. In contrast to time-varying longitudinal measurements, the baseline variables are time-independent, meaning that their values do not change during the study.

For instance, in follow-up studies of HIV clinical trials (Abrams et al. 1994, Goldman et al. 1996), researchers recorded treatment assignments and demographic information when a participant enrolled in the study, and then measured the CD4 cell counts of participants at subsequent clinic visits. The timing and the number of these clinical visits usually vary across subjects. Hence, the CD4 counts are longitudinal data, while the remaining variables are baseline covariates. The study also records the time to death for some participants, or the duration until the participant is lost to follow-up, as well as the time at which the study concludes if the participants do not experience the event during the study. For such a study, statistical analysis has two goals: (*a*) modeling the event time using the longitudinal data as time-dependent covariates, in addition to baseline covariates, and (*b*) modeling the CD4 counts as longitudinal data using baseline covariates and addressing the informative dropout of the longitudinal data caused by death, the occurrence of the event. Both of these are important problems that call for the joint modeling of the survival and longitudinal processes because marginal modeling of either process will induce bias and may be less efficient if the two components are correlated. There are already several review articles on the joint modeling of longitudinal and survival data (Tsiatis & Davidian 2004, Wu et al. 2012, Furgal et al. 2019, Papageorgiou et al. 2019). Our intention is to offer an up-to-date overview of the field, highlighting the critical issues in joint modeling and exploring future directions. We note that our reference list is not exhaustive.

Event-time data are often referred to as survival data in the biostatistics community. Hereafter, we use the terminology "survival data" liberally and define survival analysis as the statistical analysis and inference for survival data. There are two distinct features of survival data: (*a*) They are positive random variables, and (*b*) they are often incomplete data, meaning that the actual values of some of the events may not be observed, due to the design of the study and the reality that patients may drop out before the event is observed. There are many forms of incomplete data, such as right censoring, left censoring, double censoring, interval censoring, left truncation, right truncation, and a combination of several incomplete mechanisms; we refer to Klein et al. (2014) for details and focus here on the most prevalent type, random right censoring. In Section 5.2, we give a very brief discussion of the joint modeling approaches for other types of incomplete data.

The feature of positivity of survival data is easy to address, but the incomplete nature of survival data requires special attention that deviates from a modeling approach for data that can be observed completely without bias or errors. For a complete outcome variable, the standard approach is to adopt a mean regression model to describe the influence of covariates on the outcome. This approach is not ideal for incomplete survival outcomes because the upper tail probability of right-censored data is either inaccessible or difficult to estimate due to the scarcity of data in the upper (right) tail of the survival distribution. For instance, it is infeasible to estimate the upper tail if the largest observations are censored, and this happens routinely in studies that end before all events have been recorded.

Consequently, mainstream approaches in survival analysis are anchored instead on modeling the hazard function of the survival data. An added benefit of the hazard-based approach is that it automatically takes care of the nonnegative feature of survival data. This paradigm shift in survival analysis is attributed to the pioneering work of Sir D.R. Cox, who proposed the elegant proportional hazards model (Cox 1972, 1975), which then inspired alternative survival models evolving

around the hazard function. Half a century later, Cox's 1972 paper remains one of the highest cited papers in statistics, and the proportional hazards model is still the prevailing approach to model survival data due to the success of the partial likelihood approach (Cox 1975), which produces a consistent and efficient parametric estimator without involving the nonparametric baseline hazard function. To be more specific, the Cox model is a semiparametric model, where the hazard function $\lambda(t \mid \boldsymbol{W})$, for an individual with covariate vector $\boldsymbol{W} \in \mathcal{R}^d$, is specified by

$$\lambda(t \mid \boldsymbol{W}) = \lambda_0(t) \exp\{\boldsymbol{\theta}^\top \boldsymbol{W}\}. \quad 1.$$

Here, $\boldsymbol{\theta}$ is a finite-dimensional parameter and the baseline hazard function $\lambda_0$ is unspecified, hence nonparametric. It is called the proportional hazards model because the ratio of the hazard functions for any two individuals does not change over time.

When the baseline hazard function $\lambda_0$ in Equation 1 belongs to a parametric family, e.g., a Weibull distribution, this becomes a parametric model, where the joint likelihood approach is preferred (Pawitan & Self 1993) to model both components. However, the likelihood approach fails when $\lambda_0$ is unspecified, because then the likelihood is unbounded. Luckily, one can resort to the nonparametric likelihood proposed by Kiefer & Wolfowitz (1956). Later, it was discovered that Cox's partial likelihood is equivalent to the nonparametric likelihood; therefore, the maximum partial likelihood estimator of $\boldsymbol{\theta}$ is also the nonparametric maximum likelihood estimator (NPMLE). This explains the semiparametric efficiency of the maximum partial likelihood estimator in the Cox model as shown in the monograph by Tsiatis (2006, Section 5.2). For the definition and theory of semiparametric efficiency, readers are directed to Begun et al. (1983) and the monographs of Bickel et al. (1998), Van der Vaart (2000), Tsiatis (2006), and Kosorok (2008). We next explain the concept of nonparametric likelihood and NPMLE, as it is the pillar of joint modeling.

Conventional likelihood approaches are applicable to a family of distributions with a dominating measure, so the Radon–Nikodym derivative, with respect to (w.r.t.) the dominating measure, exists for each member of the family and serves as a surrogate of the likelihood. For instance, for a family of absolute continuous cumulative distribution functions (CDFs), a dominating measure is the Lebesque measure, and the Radon–Nikodym derivative of an absolute continuous CDF is its probability density function, which is used as a measure of the likelihood for such a family. Likewise, a dominating measure for a family of discrete probability measures is the counting measure, and the Radon–Nikodym derivative of a discrete CDF is the probability mass function, which is used as a measure of the likelihood for this discrete family. Therefore, the maximum likelihood approach is well suited for these two families. However, there are many other CDFs that are neither absolute continuous nor discrete and do not have a dominating measure. For such families, Kiefer & Wolfowitz (1956) proposed a pairwise comparison method for any two CDFs $F$ and $G$ using their sum $F + G$ as the dominating measure. If the Radon–Nikodym derivative of $F$ w.r.t. $F + G$ is larger than the Radon–Nikodym derivative of $G$ w.r.t. $F + G$, we declare that $F$ is more likely than $G$. With this nonparametric likelihood, the NPMLE is the CDF $F$ that is more likely than any other competitor $G$. It is not hard to see that any $F(x)$ that is continuous at $x$ will be dominated by any $G$ that is discrete at $x$, because the Radon–Nikodym derivative of $F$ w.r.t. $F + G$ is zero. By this token, the NPMLE is always a discrete probability measure that assigns nonzero mass only to the observed data. For instance, the empirical CDF is the NPMLE of all CDFs, and the Kaplan–Meier estimate (Kaplan & Meier 1958) is the NPMLE for the CDFs of all randomly right-censored data. Intriguingly, the maximum partial likelihood estimator in the Cox model is also its NPMLE.

Another key advantage of the hazard-based approach is that it leads directly to a likelihood function or a nonparametric likelihood. Both likelihood approaches produce efficient or

semiparametric efficient estimators under some regularity conditions. This is the case for the aforementioned Cox proportional hazards model as well as for many other survival models.

Despite its success, the proportional hazards assumption in the Cox model is restrictive, and hence it may not be suitable for some applications and may lead to biased estimators. Beyond the proportional hazards model, there are several alternative semiparametric models: the proportional odds model (Bennett 1983), the accelerated failure time (AFT) model (Wei 1992), the transformation survival models (Dabrowska & Doksum 1988), and the extended hazard model (Etezadi-Amoli & Ciampi 1987, Tseng & Shu 2011). Both the proportional hazards and proportional odds models belong to the broader class of transformation survival models, where the nonparametric maximum likelihood approach offers semiparametric efficiency in estimation (Zeng & Lin 2007b) in many cases. The Cox proportional hazards model, along with the AFT model, is also a special case of the extended hazard model. Generally, the NPMLE is intractable, except for the Cox model, but a pseudolikelihood approach, proposed by Zeng & Lin (2007a), can be used to obtain semiparametric efficient estimators.

All the above survival models can accommodate time-dependent covariates, in addition to baseline covariates, and the same estimation principle for baseline covariates also works for time-dependent covariates as long as the entire longitudinal trajectory can be observed without error for all subjects and variables. For instance, the Cox model with time-dependent covariates takes the form

$$\lambda(t \mid \boldsymbol{W}) = \lambda_0(t) \exp\{\boldsymbol{\theta}^T \boldsymbol{W}(t)\}, \qquad 2.$$

where $\boldsymbol{W}$ could include baseline covariates, if we regard a baseline covariate as a constant value time-dependent covariate. Equation 2 is a concurrent model as it is assumed that only the current value $\boldsymbol{W}(t)$ of the longitudinal process, and not its past values $\{\boldsymbol{W}(s) : s \leq t\}$, is predictive for the survival time. The partial likelihood principle still works if all the $\boldsymbol{W}$ trajectories can be observed fully without error. In practice, especially in longitudinal studies involving human subjects, this is typically infeasible as measurements are mostly recorded at discrete and intermittent time points. In addition, the measurements may contain random noise, aka measurement errors. Such discretely observed time-dependent covariates are referred to as longitudinal data in the literature. A naive approach to incorporate such data in the Cox model is to impute the longitudinal process using the most recent observation of a subject as the value of that subject at a current time. This is called the last-value carry forward method. Not surprisingly, this method will induce bias (Raboud et al. 1993) in the subsequent survival analysis and this bias can be substantial if the longitudinal data are measured infrequently.

A more sophisticated approach to impute the longitudinal process is to adopt a prespecified model, such as a linear mixed-effects model (Laird & Ware 1982), for the longitudinal process, and then use the subject-specific random effects to represent the longitudinal process for each subject (De Gruttola & Tu 1994, Tsiatis et al. 1995). This two-stage method can reduce the bias if the adopted model fits the longitudinal processes. However, there will still be an intrinsic bias caused by the presence of measurement errors in the observed longitudinal data. Additionally, the two-stage method does not take advantage of the information contained in the survival data, which is correlated with the longitudinal data. This leads to a loss of efficiency for the estimation of both the longitudinal and survival models. This motivated the search for alternative approaches to jointly model the two components with the aims of eliminating the bias (asymptotically) and increasing the efficiency as compared with the above two-stage method that imputes the longitudinal data first before estimating components of the survival model.

In a landmark paper, Wulfsohn & Tsiatis (1997) proposed to model the survival and longitudinal data simultaneously through their joint nonparametric likelihood. They modeled the

longitudinal data with a Gaussian linear mixed-effects model and regarded them as the covariates of the survival data in a Cox model. This approach accomplished a major goal of joint modeling, as Zeng & Cai (2005) and Dupuy et al. (2006) later showed that the estimated parameters achieve semiparametric efficiency when the baseline hazard is unspecified. It turns out that the same nonparametric likelihood approach also works for the larger transformation survival models (see Equation 5 below) and leads to semiparametric efficiency of the parametric estimators (Zeng & Lin 2007b). Due to this generality, we illustrate the joint likelihood approach through the more general transformation survival model.

We now provide an outline of the joint modeling approach. Further details are in Section 2. It is straightforward to determine the nonparametric joint likelihood function based on the joint distribution of longitudinal data and survival data (see Section 2.2). However, the likelihood in Equation 8 below involves a complicated integral due to the presence of random effects, and direct maximization of the likelihood function is challenging. Hence, the expectation–maximization (EM) algorithm (Dempster et al. 1977) has been employed, where one treats the random effects in the longitudinal covariates as missing data. The expectation step (E-step) is often implemented by Monte Carlo or other integral approximation methods, and the Newton–Raphson method is applied in the maximization step (M-step). The numerical integration in the E-step is computationally costly, especially for higher dimensional random effects. Nevertheless, the EM algorithm is computationally stable and in general produces reliable results. Section 3.1 contains further details.

Another challenge of the EM algorithm is that it is unable to provide direct estimates of the standard error (SE) of the parameters. Some solutions are discussed in Section 3.1.3, which is devoted to this topic.

To mitigate the computational complexity of the EM algorithm, an alternative method known as the conditional score method was proposed by Tsiatis & Davidian (2001) and further studied by Tsiatis & Davidian (2004). This approach can yield unbiased estimates from estimating equations and is computationally fast, albeit at the cost of lower estimation efficiency. Details are in Section 3.2. Section 3.4 briefly reviews Bayesian approaches for joint modeling.

Section 4 contains an illustration of the joint modeling approach for the aforementioned AIDS data. Section 5 showcases the joint modeling approach for other survival models and incomplete data. Available software to perform joint modeling methods is the topic of Section 6.

The aforementioned computational bottleneck has been a key obstacle for the progress and further advancement of the field of joint modeling, especially for nonparametric survival and longitudinal settings. Keep in mind that this was in the pre–artificial intelligence era, where gradient decent– and machine learning–based numerical integration methods were not widely known to statisticians, so EM was the main computational tool to solve the optimization problem in the joint likelihood. Today, we have many more tools to tackle the numerical challenges posed by the joint likelihood approach. The recent advances in deep learning, and more broadly artificial intelligence (AI), provide an opportunity to revitalize the field by deploying machine learning techniques to accelerate computation and to increase the accuracy of the estimates. Further discussion of this and other future directions are in Section 7.

## 2. JOINT MODELING

### 2.1. External and Internal Covariates

We first introduce internal and external covariates, which are two major types of covariates in the survival model. Their definitions are given below. For more detailed information, please refer to Kalbfleisch & Prentice (2002, Section 6.3.1).

**Definition 1 (External and internal covariates).** A time-dependent covariate $\boldsymbol{W}$ with history $\boldsymbol{\mathcal{W}}(t) = \{\boldsymbol{W}(s), 0 \le s < t\}$ up to $t$ is called external if, for all $u$, $v$ such that $0 < u \le v$, it satisfies the property

$$\mathrm{P}(u \le T \le u + \mathrm{d}u \mid \boldsymbol{\mathcal{W}}(u), T \ge u) = \mathrm{P}(u \le T \le u + \mathrm{d}u \mid \boldsymbol{\mathcal{W}}(v), T \ge u),$$

where $T$ is the event time. A covariate that is not external is called internal.

Whereas an external covariate $\boldsymbol{W}$ may influence the probability function over time, its future path will not influence the current value of the probability function. By this measure, all baseline covariates that do not change values over time are external covariates. Any interaction between a baseline covariate and time since entering the study is also external. More generally, any external risk factor, such as temperature and particles less than 2.5 micrometers in diameter (also called PM2.5) from air pollution, is also an external covariate. One virtue of external time-dependent covariates is that they can often be observed completely, and hence can be incorporated in a survival model directly without any model assumptions. In contrast, internal covariates are always time-dependent and are measured directly from the subject, and hence they are unlikely to be observed completely. In this review, we call such internal covariates longitudinal data (see Section 2.2) to distinguish them from time-dependent external covariates that can be observed fully. In order to incorporate longitudinal data in a survival model, the longitudinal data require specification of a dedicated model, which is estimated jointly with the survival model by maximizing their joint (nonparametric) likelihood.

Both (internal) longitudinal and (external) time-dependent covariates account for a survival model and the joint likelihood, yet their influence on the joint likelihood approach differs substantially. An external time-dependent covariate, which is fully observed, does not pose the computational challenges that an internal longitudinal covariate does. This is due to the internal longitudinal covariate requiring a model assumption, while the external time-dependent covariate does not necessitate a model assumption since its entire trajectory can be observed without errors. In other words, internal longitudinal covariates are the root cause of why a more complex joint modeling approach is necessary—specifically, to correct the bias in a marginal approach.

## 2.2. Longitudinal Data

Longitudinal studies are characterized by the repeated collection of measurements from the same individuals over time and can lead to an understanding of how responses change over time. Longitudinal data can be viewed as a series of repeated records of a random curve at different observation times, where each recording is contaminated by noise (measurement errors). In this context, the observed longitudinal data for the $i$th subject can be described by

$$V_{ij} = U_i(T_{ij}) + \epsilon_{ij},\ (i = 1, \ldots, n, j = 1, \ldots, n_i). \qquad 3.$$

Here, the pairs $(V_{ij}, T_{ij})$ represent the observed longitudinal data for the $i$th subject at time $T_{ij}$, where $U_i(\cdot)$, for $i = 1, \ldots, n$, denote independent identically distributed realizations of a random curve $U(\cdot)$, and $\epsilon_{ij}$ represents the independent measurement errors, across $i$ and $j$, with mean zero and finite but unknown variance. In the following, we denote $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{in_i})^\top$ and $\boldsymbol{V}_i = (V_{i1}, \ldots, V_{in_i})^\top$. A key feature of longitudinal data is that the $i$th subject has $n_i$ observations over time, where $n_i$ varies across subjects and may be small for some subjects.

An appealing approach to model the trend of continuous longitudinal data is through the linear mixed-effects model (Laird & Ware 1982), which assumes that the random curves follow the model

$$U_i(t) = \boldsymbol{\beta}^\top \boldsymbol{X}_i(t) + \boldsymbol{b}_i^\top \boldsymbol{Z}_i(t). \qquad 4.$$

Here $\boldsymbol{X}_i(\cdot)$ and $\boldsymbol{Z}_i(\cdot)$ are vectors of covariates for the fixed and random effects, respectively. The vectors $\boldsymbol{X}_i(\cdot)$ and $\boldsymbol{Z}_i(\cdot)$ can be user-specified covariates with overlapping components. Often, part or all of the components of the fixed effects are chosen as the random effects in the biostatistics community. In this context, the unknown regression coefficients $\boldsymbol{\beta}$ represent the fixed effects, which capture the overall mean effect or the population-level effect of the variable $\boldsymbol{X}(\cdot)$ on the outcome variable. On the other hand, the random vector $\boldsymbol{b}_i$ represents the between-subject variation around the fixed effects. Under the Gaussian assumption on $b_i$ and measurement errors $e_{ij}$, the parameters of this model, including $\boldsymbol{\beta}$ and the variance components (covariance of $\boldsymbol{b}_i$ and variance of the measurement errors) can be efficiently estimated using likelihood-based methods through the EM algorithm. In addition, the variance of the noise $\epsilon_{ij}$ can also be efficiently estimated if it has a simple parametric form, such as constant variance in a homoscedastic setting.

For discrete or categorical longitudinal data, the linear mixed effects can be replaced by a generalized mixed-effects model (Faucett et al. 1998, Yao 2008, Li et al. 2010, Pan & Yi 2011), where the joint likelihood approach can be established similarly as for the linear mixed-effects model through an EM algorithm, by adjusting the joint likelihood accordingly.

## 2.3. Survival Data

Survival data, also known as time-to-event data, are commonly encountered in fields ranging from biomedical science to sociology, industry, engineering, and economics. The primary outcome of interest is the time $T$ until an event occurs. This event could be anything from the failure of a mechanical component to the occurrence of a disease, or the death of a patient. The focus of survival analysis is to model the hazard function of an event occurring at a certain time $t$, identifying factors associated with the risk for the event of interest to happen, and making predictions about future events based on the characteristics of individuals or groups.

The unique characteristic of survival data that sets it apart from other types of outcome data is that many event times are unobserved due to censoring or other forms of incomplete data. In this article, we focus on random right censoring, which is the most common type of censoring in survival data. References for joint modeling involving other types of incomplete data are included in Section 5.2.

Right-censored event times are encountered when the event of interest has not occurred for some individuals by the end of the study period or during the follow-up period if a patient drops out of the study before the event occurs. Let $T$ and $C$ be the event and right-censoring time, respectively. We do not observe the event time for some individuals but do know that the event has not occurred at the censoring time $C$. Thus, we can only observe $Y = \min(T, C)$ and the censoring indicator $\Delta = I(T \leq C)$, where $\Delta = 1$ indicates that the observed time is the event time $T$ and $\Delta = 0$ signifies that the event time is censored by $C$.

We assume that the event time $T_i$ for the $i$th subject is associated with the latent processes $U_i(\cdot)$ plus some other time-varying processes $\boldsymbol{W}_i(\cdot)$ that can be observed completely without errors. Examples of $\boldsymbol{W}_i(\cdot)$ include baseline covariates that have constant values over time and some external time-varying covariates (see Section 2.1 for the definition of external covariates).

For an individual with covariate $(U, \boldsymbol{W})$, its survival function or hazard function can be used to characterize the event time $T$. Let $\mathcal{U}(t) = \{U(s) : 0 \leq s < t\}$ be the history of $U$ up to time $t$. Given $(U, \boldsymbol{W})$, the survival function $S(t \mid U, \boldsymbol{W})$ represents the probability that the individual survives beyond time $t$ and is defined as

$$S(t \mid U, \boldsymbol{W}) = \mathrm{P}(T > t \mid \mathcal{U}(t), \boldsymbol{W}).$$

Additionally, the hazard function for a continuous event time $T$ is defined as

$$\lambda(t \mid U, \boldsymbol{W}) = \lim_{\Delta t \to 0} \frac{\mathrm{P}(t < T \leq t + \Delta t \mid T > t, \mathcal{U}(t), \boldsymbol{W})}{\Delta t},$$

which describes the instantaneous risk of an event occurring at a given time $t$, given that the individual $(U, \boldsymbol{W})$ has survived up to that time.

We emphasize here that both the survival and hazard functions evaluated at time $t$ should only depend on the values of the longitudinal process up to time $t$ and not on the entire longitudinal trajectory. This is to avoid the influence of future values of the longitudinal process to affect the risk at a current point. However, this restriction is not necessary for an external covariate as its future has no influence on the past, according to the definition in Section 2.1.

Once we have the hazard functions, the associated probability density (or mass) function can further be represented as $f(t \mid U, \boldsymbol{W}) = \lambda(t \mid U, \boldsymbol{W})S(t \mid U, \boldsymbol{W})$, and the corresponding cumulative hazard function is $\Lambda(t \mid U, \boldsymbol{W}) = \int_0^t \lambda(s \mid U, \boldsymbol{W})\mathrm{d}s$. Any of the four functions, $\lambda(t \mid U, \boldsymbol{W})$, $\Lambda(t \mid U, \boldsymbol{W})$, $f(t \mid U, \boldsymbol{W})$, and $S(t \mid U, \boldsymbol{W})$, uniquely determines the others.

There are many different types of survival models beyond the Cox model. In this review article, we adopt transformation survival models as the platform to model survival data, because this class includes the Cox model and the joint modeling approach for such models shares many similarities with the joint modeling approach for the Cox model.

The cumulative hazard function for a transformation model is

$$\Lambda(t \mid \boldsymbol{b}_i, U, \boldsymbol{W}) = H\left(\int_0^t \lambda_0(s)\exp\{\alpha U_i(s) + \boldsymbol{\phi}^\top \boldsymbol{W}_i(s)\}\mathrm{d}s\right), \quad 5.$$

where the transformation function $H$ is a prespecified strictly increasing and continuously differentiable function, and $\lambda_0$ is an unknown baseline hazard function.

When $H(t) = t$ is the identity function, the model in Equation 5 becomes the Cox proportional hazards model (Cox 1972, 1975). When $H(t) = \log(1 + t)$, it becomes the proportional odds model (Bennett 1983). Standard choices for $H$ include Box–Cox transformations and the logarithmic transformation:

$$H_\rho(t) = \frac{\log(1 + \rho t)}{\rho}, \ \rho \geq 0. \quad 6.$$

When $\rho = 0$ and $\rho = 1$, this gives the Cox proportional hazards model and the proportional odds model, respectively. The logarithmic transformation has computational advantages as it can be converted to a Cox model with a multiplicative frailty variable (more on this will be elaborated in Section 3.1.2).

We note that the hazard function in Equation 5 is associated with the latent process $U_i$ but not with $V_i$. This is a main reason why the marginal approach to model the survival component alone will not lead to consistent estimators. However, even if we replace $U_i$ with $V_i$ as a risk factor, joint modeling would still be needed as we only have observations of $V_i$ at a few time points. For this reason, we opt for the common practice that chooses $U_i$ as the covariate rather than $V_i$.

## 2.4. Joint Modeling of Longitudinal and Survival Data

In practice, researchers often gather baseline covariates at the start of a study and subsequently track longitudinal variables up to an event time or censoring time. For example, **Figure 1** displays 20 randomly selected CD4 cell count trajectories (in square-root scale) among 467 HIV infected patients in a clinical study (Goldman et al. 1996), in which each trajectory was discretely recorded at the study's inception and during subsequent follow-ups until death, dropout from the study, or the study's conclusion. Thus, such longitudinal data are only available intermittently and may be further contaminated by measurement errors. More importantly, data missing after the event for each individual may lead to informative dropout of the longitudinal process. In addition to the CD4 counts, at the study onset each patient was randomly assigned to receive either the ddI (didanosine) or ddC (zalcitabine) treatment (`drug`), and two baseline covariates, sex (`sex`) and

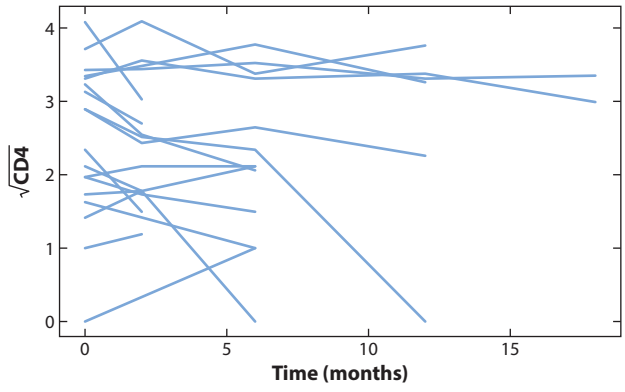


**Figure 1**

Trajectories of the square root of CD4 cell count for 20 randomly selected subjects.

previous infection at study entry (`prevDiag`), were also recorded. For such studies, we denote $\boldsymbol{T}$ as the recorded time points, $\boldsymbol{V}$ as the CD4 counts, and $\boldsymbol{W}$ as the baseline covariates `drug`, `sex`, and `prevDiag`.

Formally, the observations are $\mathcal{D}_n = \{\boldsymbol{O}_i = (\boldsymbol{T}_i, \boldsymbol{V}_i, \boldsymbol{W}_i, Y_i, \Delta_i) : i = 1, \ldots, n\}$, the definition of $(\boldsymbol{T}_i, \boldsymbol{V}_i)$ aligns with that in Section 2.2, and $\boldsymbol{W}_i$ are the completely observed covariates that may include time-independent baseline and time-dependent covariates, such as the interaction term between a baseline covariate and the time since entering the study. The observed survival data are $Y_i = \min(T_i, C_i)$ and the censoring indicator $\Delta_i = I(T_i \leq C_i)$, where $T_i$ and $C_i$ are the event and censoring time, respectively. Depending on the target of interest, there are three major approaches to jointly model longitudinal and survival data:

1. The most popular approach is to estimate the regression parameter in a time-dependent hazard model, providing a way to account for measurement error and infrequently measured values of the longitudinal variable (Tsiatis & Davidian 2004).

   When there are no longitudinal covariates $\boldsymbol{V}_i$ and the covariate $\boldsymbol{W}_i$ is time-independent, the cumulative hazard function is usually modeled as a Cox proportional hazards model (Cox 1972):

$$\Lambda(t \mid \boldsymbol{b}_i, \boldsymbol{W}_i) = \int_0^t \lambda_0(s)\mathrm{d}s \exp\{\boldsymbol{\phi}^\top \boldsymbol{W}_i\},$$

   where $\lambda_0(\cdot)$ is the baseline hazard function. This implies that the hazard ratio of two subjects remains constant over time and is determined by their respective covariate values. This elegant model has been widely adopted in practice to analyze time-to-event data using the partial likelihood approach (Cox 1972). When time-dependent covariates are available and fully observable, e.g., $\boldsymbol{W}$ contains fully observed time-dependent covariates, the partial likelihood approach still works and the Cox model becomes

$$\Lambda(t \mid \boldsymbol{b}_i, \boldsymbol{W}_i) = \int_0^t \lambda_0(s) \exp\{\boldsymbol{\phi}^\top \boldsymbol{W}_i(s)\}\mathrm{d}s.$$

   However, some time-dependent covariates may only be observed at discrete time points and may contain measurement errors (denoted as $\boldsymbol{V}$). This leads to the following joint model for longitudinal data and survival data:

$$\begin{cases} V_{ij} = U_i(T_{ij}) + \epsilon_{ij}, \\ \Lambda(t \mid \boldsymbol{b}_i, U_i, \boldsymbol{W}_i) = \int_0^t \lambda_0(s) \exp\{\alpha U_i(s) + \boldsymbol{\phi}^\top \boldsymbol{W}_i(s)\}\mathrm{d}s, \end{cases}$$

where $U_i(t) = \boldsymbol{\beta}^\top \boldsymbol{X}_i(t) + \boldsymbol{b}_i^\top \boldsymbol{Z}_i(t)$ is defined in Equation 4 of Section 2.2. This joint model, where the longitudinal data and survival data are linked through the unobserved time-varying covariate $U_i(\cdot)$, rather than the observed longitudinal data $V_{ij}$, has attracted increasing attention over the past three decades. The rationale is that $U_i(\cdot)$ is the appropriate latent biological process that determines the risk and it is only intermittently available for each subject with potential measurement errors. Moreover, if instead one would prefer to tie the observed $V_{ij}$ instead of $U_i(\cdot)$ to the risk, joint modeling cannot be avoided, because a complete trajectory $V_i(\cdot)$ that generates the observed longitudinal data $V_{ij}$ is not available. Therefore, the convention is to tie the latent process $U_i(\cdot)$ to the risk for survival, and this is the approach we adopt here.

2. When the primary focus is on the longitudinal data, the event time (or longitudinal endpoint) is used to address the sampling bias caused by the informative dropout, i.e., unavailability of further longitudinal measurements following the dropout (death, in this case). A common strategy is to model the observed time using survival models that are connected to the longitudinal endpoint through latent variables (see Little 1995, Sun & Song 2001, Roy & Lin 2002, Huang & Wang 2004, Sun et al. 2012, Han et al. 2014, Kim et al. 2017, and the references therein).
3. Another goal is to study the joint evolution of longitudinal measurements and event times, with equal interest in both components. This approach dictates that the survival and longitudinal models share some common random effects, and hence it is often referred to as the shared random effects or shared frailty model (Henderson et al. 2000). For example, the cumulative hazard function can take the form

$$\Lambda(t \mid \boldsymbol{b}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i) = \int_0^t \lambda_0(s) \exp\{\alpha \boldsymbol{b}_i^\top \boldsymbol{Z}_i(s) + \boldsymbol{\phi}^\top \boldsymbol{W}_i(s)\} \mathrm{d}s,$$

where $\boldsymbol{b}_i$ and $\boldsymbol{Z}_i(t)$ are given in Equation 4 and $\boldsymbol{W}_i$ is given in Equation 5. The work by Henderson et al. (2000) and Ratcliffe et al. (2004), and the references therein, provides further insights. An R package `JSM` (Xu et al. 2020) can be used to fit this model. The `JM` package (Rizopoulos 2010) also considered a shared random-effects approach using a different method.

Hereinafter, we primarily focus on the first approach above and consider the joint model with a transformation survival model for generality:

$$\begin{cases} V_{ij} = U_i(T_{ij}) + \epsilon_{ij}, \\ \Lambda(t \mid \boldsymbol{b}_i, U_i, \boldsymbol{W}_i) = H(\int_0^t \lambda_0(s) \exp\{\alpha U_i(s) + \boldsymbol{\phi}^\top \boldsymbol{W}_i(s)\} \mathrm{d}s), \end{cases} \quad 7.$$

where $U_i(t) = \boldsymbol{\beta}^\top \boldsymbol{X}_i(t) + \boldsymbol{b}_i^\top \boldsymbol{Z}_i(t)$ is defined in Equation 4 of Section 2.2.

## 3. ESTIMATION

The key step of a joint modeling approach is to select a meaningful joint likelihood that can correct the bias in the marginal approach. A standard approach is to adopt a parametric model for all the random quantities in the joint model. Specifically, a Gaussian model offers computational advantages, and this type of model has been prevalent in the literature. Thus, the standard assumptions are as follows:

- $\epsilon_{ij}$ and $\boldsymbol{b}_i$ are independent for all $i$ and $j$, and they follow identical normal distributions $\mathcal{N}(0, \sigma_\epsilon^2)$ and $\mathcal{N}(0, \Sigma_b)$, respectively, where $\sigma_\epsilon$ and $\Sigma_b$ are unspecified parameters.

- Conditional on $\boldsymbol{b}$ and covariate $\boldsymbol{W}$, the censoring time and longitudinal measurement times are both noninformative in the sense that they do not contribute information to the event time $T$ or the latent process $U$, once we have adjusted for the covariates effects.
- The longitudinal and survival outcomes are conditionally independent given $\boldsymbol{b}_i$.

While it is common to postulate Gaussian errors, it is not clear why the random effects should have a Gaussian distribution. There are two advantages in adopting Gaussian models: (*a*) The joint likelihood can be expressed easily and one has a close-form expression for the estimated random effect $\boldsymbol{b}_i$, and (*b*) the numerical integral involved in the joint likelihood in Equation 8 below can be evaluated through the (adaptive) Gauss–Hermite quadrature method (Pinheiro & Bates 1995) instead of other numerical integration methods, e.g., Monte Carlo integration or Laplace approximation, which are time consuming and may have less numerical precision. These are computational advantages, not a justification for the Gaussian random-effect assumption. However, a pleasant surprise is that the Gaussian assumption appears to be robust against model misspecification in numerical studies (Song et al. 2002). That is, the joint likelihood using the Gaussian assumption for the random effects often leads to similar fixed-effects estimates and survival estimates when the true random effect is non-Gaussian. An explanation was provided by Hsieh et al. (2006), namely that in the E-step the mode of the posterior distribution with Gaussian random effects approaches quickly the mode of the true posterior distribution, even when the true random effects are not Gaussian, as long as the number of repeated measurements is not too small. This discovery provides strong justification for the Gaussian assumption for the random effects in the linear mixed-effects model, as the estimates obtained through the maximum likelihood function are remarkably robust against violations of Gaussianity.

## 3.1. Estimation of Parameters

This section reviews the maximum likelihood approach for the joint model presented in Equation 7. The EM algorithm is introduced to obtain parameter estimates, followed by the calculation of SEs for these estimates to facilitate statistical inference.

**3.1.1. Joint likelihood approach.** The joint model in Equation 7 involves an unknown parameter $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}, \sigma_\epsilon, \Sigma_b)$ and an unknown function $\lambda_0$. Here, the elegant partial likelihood approach of the Cox model, which provided an efficient parametric estimate of $\beta$ without involving the nonparametric baseline hazard function $\lambda_0$, is no longer applicable because of the presence of the longitudinal components. Luckily, a nonparametric likelihood approach, as proposed by Kiefer & Wolfowitz (1956), can be employed to replace the partial likelihood, and it will lead to efficient estimation of $\boldsymbol{\theta}$. Here, the efficiency means semiparametric efficiency as defined by Bickel et al. (1998), Van der Vaart (2000), Tsiatis (2006), and Kosorok (2008). This is perhaps not surprising since the partial likelihood is also the nonparametric likelihood and the partial likelihood is known to lead to semiparametric efficient estimators in the Cox model. The more striking fact is that Zeng & Lin (2007b) showed that the semiparametric efficiency of the NPMLE extends to the transformation survival model, Equation 5, even under the joint modeling framework.

However, a main advantage of the partial likelihood is lost in the joint modeling approach. Unlike the partial likelihood approach, where the parametric components $\boldsymbol{\phi}$ and $\alpha$ in the survival component (Equation 5) can be estimated without involving the nonparametric baseline hazard function $\lambda_0$, the parametric component $\boldsymbol{\theta}$ in the joint models (Equation 7) cannot be separated from the estimation of the nonparametric components $\lambda_0$. This seems daunting at first, as $\lambda_0$ is a function and hence infinite dimensional. Luckily, the NPMLE for $\lambda_0$ corresponds to a discrete distribution. Accordingly, $\lambda_0$ can be reparametrized as a vector and easily estimated in the M-step

through the Breslow formula for the hazard function based on the current estimate of $\boldsymbol{\theta}$ (explained in Section 3.1.2).

To derive the likelihood function, we need additional assumptions: The censoring time of the event and the time schedule of the longitudinal data are noninformative, and the event time $T_i$ and the longitudinal data $V_{ij}$ are independent, conditional on the random effects $\boldsymbol{b}_i$. With these assumptions and by integrating out the unobservable random effects $\boldsymbol{b}_i$, the likelihood function of $\boldsymbol{\psi} = (\boldsymbol{\theta}, \Lambda_0)$ takes the form

$$\mathcal{L}_n(\boldsymbol{\psi} \mid \boldsymbol{O}) = \prod_{i=1}^{n} \int p_{\boldsymbol{\psi}}(Y_i, \Delta_i \mid \boldsymbol{b}_i) p_{\boldsymbol{\theta}}(\boldsymbol{U}_i \mid \boldsymbol{b}_i) p_{\boldsymbol{\theta}}(\boldsymbol{b}_i) \mathrm{d}\boldsymbol{b}_i, \tag{8.}$$

where

$$\begin{aligned} p_{\boldsymbol{\psi}}(Y_i, \Delta_i \mid \boldsymbol{b}_i) = & \left[ \lambda_0(Y_i) \exp\{\boldsymbol{\phi}^\top \boldsymbol{W}_i(Y_i) + \alpha U_i(Y_i)\} H' \left( \int_0^{Y_i} \lambda_0(s) \exp\{\boldsymbol{\phi}^\top \boldsymbol{W}_i(s) + \alpha U_i(s)\} \mathrm{d}s \right) \right]^{\Delta_i} \\ & \times \exp\left\{ -H \left( \int_0^{Y_i} \lambda_0(s) \exp\{\boldsymbol{\phi}^\top \boldsymbol{W}_i(s) + \alpha U_i(s)\} \mathrm{d}s \right) \right\}, \\ p_{\boldsymbol{\theta}}(\boldsymbol{U}_i \mid \boldsymbol{b}_i) = & \prod_{j=1}^{n_i} \frac{1}{(2\pi\sigma_\epsilon^2)^{1/2}} \exp\left\{ -\frac{(V_{ij} - \boldsymbol{\beta}^\top \boldsymbol{X}_i(T_{ij}) - \boldsymbol{b}_i^\top \boldsymbol{Z}_i(T_{ij}))^2}{2\sigma_\epsilon^2} \right\}, \\ p_{\boldsymbol{\theta}}(\boldsymbol{b}_i) = & \frac{1}{|2\pi \Sigma_b|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{b}_i^\top \Sigma_b \boldsymbol{b}_i \right\}, \end{aligned}$$

and $H'(\cdot)$ is the derivative of $H(\cdot)$ and $|\cdot|$ represents the determinant of a matrix.

The first step to numerically find the NPMLE is to determine the form of the NPMLE of $\lambda_0$. This would be intractable unless we could represent $\lambda_0$ by a vector parameter with growing dimension. This turns out to be feasible and has been explored by Cai & Cheng (2004) and Zeng & Lin (2007b). It is now well known that the NPMLE of $\lambda_0$ is a discrete hazard function whose cumulative hazard function is a step function with jumps at all uncensored observations. Thus, we can reparametrize $\lambda_0$ by a long vector whose values represent these jump sizes at the uncensored observations. The dimension of this vector would be the number of uncensored observations $m = \sum_{i=1}^{n} \Delta_i$, which is of the same order as the sample size $n$. Thus, we have converted an infinite-dimensional optimization problem to a high-dimensional nonconvex optimization problem, which is still a challenge.

**3.1.2. Expectation–maximization algorithm.** Direct optimization of the logarithm of the likelihood function in Equation 8 is challenging due to the high dimension of the functional parameters for $\lambda_0$ and the involvement of the integral of the unobserved random effects $\boldsymbol{b}_i$. Instead, the EM algorithm (Dempster et al. 1977) has been deployed to solve this optimization problem by treating the random effects $b_i$ as missing data.

The EM algorithm is known for its stability and generality, especially in finding the maximum likelihood estimators (MLEs) for statistical models that incorporate unobserved latent variables. Thus, it is widely used in the joint modeling literature. The EM algorithm for joint modeling includes recursive updates of the E-step and the M-step. In the E-step, one calculates the conditional expectation of the complete data log-likelihood function in terms of the posterior distribution of the latent random effects given the observed data and the current estimates of the parameters. In the M-step, one maximizes the function obtained in the E-step with respect to the parameters to obtain updated parameter estimates. This involves finding the parameter values that maximize the conditional expectation of the complete data log-likelihood function calculated in the E-step.

By iteratively alternating between the E-step and M-step, the EM algorithm converges to the maximum likelihood estimates of the joint model of longitudinal and survival data. However, it is computationally intensive due to the calculation of the expectation, which requires evaluating a potentially high-dimensional integral in the E-steps. In addition, there is no explicit form for the estimates in the M-step, so an optimization algorithm, such as the Newton–Raphson method, needs to be deployed.

While the E-step is computationally intensive and incurs approximation errors of the integral in the likelihood function, the M-step becomes a high-dimensional nonconvex maximization problem. To address this challenge, Xu et al. (2020) utilized Laplace transformations for a density function of the form $\exp\{-H(s)\}$, initially proposed by Tsodikov (2003) and subsequently also adopted by Zeng & Lin (2007b). This transformation helps mitigate computational complexities and is applicable to the logarithmic family in Equation 6. Xu et al. (2020) introduced a new likelihood function for the parameters $\boldsymbol{\psi} = (\boldsymbol{\theta}, \Lambda_0)$ by creating an artificial latent variable $\xi$, whose probability density function is $\pi(s)$ and satisfies $\exp\{-H(s)\} = \int_0^\infty \exp(-st)\pi(t)\mathrm{d}t$. Subsequently, they utilized the EM algorithm to optimize this new likelihood function by incorporating the new latent variables $\xi$ and $\boldsymbol{b}$. This approach leads to a closed form of the NPMLE of $\lambda_0(\cdot)$ in the M-step, resembling a Breslow-type estimate. By substituting this Breslow-type estimate back into the log-likelihood function, it becomes a function of $\boldsymbol{\theta}$, making it computationally feasible to obtain the NPMLE of $\boldsymbol{\theta}$ using the Newton–Raphson method. The trick to exploit the Laplace transformation not only simplifies the estimation of the baseline hazard function but also enhances the computational efficiency of obtaining the NPMLE of the parameters. Details of the EM algorithm for the transformation hazard model (Equation 5) are provided by Xu et al. (2020, Section 2.3). For theoretical support of this approach, readers are directed to the elegant discussion paper of Zeng & Lin (2007b), who established the semiparametric efficiency of the NPMLE of $\boldsymbol{\theta}$.

To recap, the nonparametric maximum likelihood approach produces efficient parametric estimates if the survival model is a transformation model taking the form in Equation 5. The derivation of the NPMLE is performed through the EM algorithm, which is no small feat in the joint modeling setting, as there are challenges in both the E- and M-steps of the algorithm. While one can resolve the challenges in the M-steps by reparametrization and by leveraging the Breslow formula, the E-steps are computationally intensive due to the presence of random effects in the joint model, which lead to a likelihood function that involves multivariate integrals. Our experience has been that an adaptive approach for the Gauss–Hermite quadrature rule (Pinheiro & Bates 1995) implemented in the package `JSM` (Xu & Zeger 2001) works quite well. This adaptive approach first centers and scales the quadrature points in each EM iteration according to the conditional distribution of the random effects based on the current estimate of the parameter $\boldsymbol{\theta}$. Although it seems to demand more computational effort, this adaptive rule is actually more efficient since it requires fewer quadrature points to reach the same level of precision.

**3.1.3. Standard error estimation.** To perform statistical inference for the parameter $\boldsymbol{\theta}$, it is necessary to estimate the SE of the estimates. Although the consistency and asymptotic normality of MLEs in Section 3.1.1 from the EM algorithm under the proportional hazards assumption or transformation survival model is well established by Zeng & Cai (2005) and Zeng & Lin (2007a), the asymptotic covariance matrix for the estimates of the parameter $\boldsymbol{\theta}$ has no explicit form. This leads to a challenge for the direct estimation of the SE of the parameter estimates as elucidated by Hsieh et al. (2006), who recommend using the bootstrap method to estimate the SE. The SE estimates from the bootstrap method work well in practice, even though they tend to slightly underestimate the actual SE due to the repeated use of the same observations in a bootstrap sample. The major downside of this approach is the high computational cost.

An alternative method by Louis (1982) for the SE estimation for a generic EM algorithm can be adapted to the joint model setting. However, its success hinges on a parametric model with a small number of parameters, which requires the baseline hazard function to follow a parametric structure, which substantially limits the applicability of this method.

An effective approach to mitigate the above issues emerged in the past decade. Xu et al. (2014) adopted the ideas of a forward difference method, Richardson extrapolation method, forward difference score (FDS), and Richardson extrapolation score (RES) (Jamshidian & Jennrich 2000) to estimate the derivatives of a profile Fisher score. Since the methods of Jamshidian & Jennrich (2000) cannot handle semiparametric models directly, Xu et al. (2014) first profile out the nuisance parameter, i.e., the cumulative baseline hazard function, and then use either forward difference or Richardson extrapolation to differentiate the profile Fisher score vector. The main procedures are outlined below.

Write $\boldsymbol{C}_i = (\boldsymbol{O}_i, \boldsymbol{b}_i)$ and $\mathcal{L}_n^c(\boldsymbol{\psi} \mid \boldsymbol{C}) = \prod_{i=1}^n p_{\boldsymbol{\psi}}(Y_i, \Delta_i \mid \boldsymbol{b}_i) p_{\boldsymbol{\theta}}(\boldsymbol{U}_i \mid \boldsymbol{b}_i) p_{\boldsymbol{\theta}}(\boldsymbol{b}_i)$, to describe the complete data and complete-data likelihood function. In the E-step of the EM algorithm with the current parameter $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\theta}}, \tilde{\Lambda}_0)$, define

$$Q(\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}) = \mathrm{E}[\log\{\mathcal{L}_n^c(\boldsymbol{\psi} \mid \boldsymbol{C})\} \mid O, \tilde{\boldsymbol{\psi}}] \text{ and } M(\tilde{\boldsymbol{\psi}}) = \arg\max_{\boldsymbol{\psi}} Q(\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}),$$

with $\boldsymbol{\psi} = (\boldsymbol{\theta}, \Lambda_0)$. Then calculate the profile Fisher score

$$S_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}) = \left[ \frac{\partial Q(\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}})}{\partial \boldsymbol{\theta}} \Big|_{\Lambda_0 = \hat{\Lambda}(\boldsymbol{\theta})} \right] \Big|_{\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}}$$

with $\hat{\Lambda}(\boldsymbol{\theta}) = \arg\max_{\Lambda_0} Q(\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}})$ given $\boldsymbol{\theta}$. Then the $i$th row of the information matrix $\mathcal{I}$ is

$$\text{PRES: } \mathcal{I}_{i,\cdot} = \frac{S_{\boldsymbol{\theta}}(\boldsymbol{\psi}_4^{(i)}) - 8S_{\boldsymbol{\theta}}(\boldsymbol{\psi}_2^{(i)}) + 8S_{\boldsymbol{\theta}}(\boldsymbol{\psi}_1^{(i)}) - S_{\boldsymbol{\theta}}(\boldsymbol{\psi}_3^{(i)})}{12\delta},$$

$$\text{PFDS: } \mathcal{I}_{i,\cdot} = \frac{S_{\boldsymbol{\theta}}(\boldsymbol{\psi}_1^{(i)}) - S_{\boldsymbol{\theta}}(\boldsymbol{\psi}^*)}{\delta},$$

where $\boldsymbol{\psi}_k^{(i)} = (\boldsymbol{\theta}_k^{(i)}, \Lambda^*(\boldsymbol{\theta}_k^{(i)}))$ for $k = 1, 2, 3, 4$; $\boldsymbol{\theta}_1^{(i)} = \boldsymbol{\theta}^* + \delta \boldsymbol{e}_i, \boldsymbol{\theta}_2^{(i)} = \boldsymbol{\theta}^* - \delta \boldsymbol{e}_i$; $\boldsymbol{\theta}_3^{(i)} = \boldsymbol{\theta}^* + 2\delta \boldsymbol{e}_i$; $\boldsymbol{\theta}_1^{(i)} = \boldsymbol{\theta}^* - 2\delta \boldsymbol{e}_i$; and $\boldsymbol{\psi}^* = (\boldsymbol{\theta}^*, \Lambda_0^*)$ is the NPMLE of $\boldsymbol{\psi}$, $\boldsymbol{e}_i$ is the $i$th coordinate vector, and $\Lambda^*(\boldsymbol{\theta}) := \arg\max_{\Lambda_0} \log \mathcal{L}_n(\boldsymbol{\theta}, \Lambda_0 \mid \boldsymbol{O})$ for a given $\boldsymbol{\theta}$. Finally, the asymptotic covariance matrix of $\boldsymbol{\theta}$ is $\mathcal{I}^{-1}$.

## 3.2. Conditional Score Approach

The conditional score method (Tsiatis & Davidian 2001, 2004) is an alternative approach to tackle the bias in the marginal modeling of a survival model in the presence of longitudinal data or measurement errors in covariates, as shown in Section 3.3 below. It was proposed to counter the computational complexities associated with the EM algorithm. In the joint model under the proportional hazards assumption, Tsiatis & Davidian (2001) considered the random effects $\boldsymbol{b}_i$ as nuisance parameters without imposing distributional assumptions. Tsiatis & Davidian (2001) initially obtained the ordinary least squares estimator for $U_i(t)$ using the data from the $i$th trajectory up to time $t$ and then identified a sufficient statistic for $\boldsymbol{b}_i$ by analyzing the joint conditional distribution of a counting process and the estimator for $U_i(t)$. This sufficient statistic is designed to eliminate the dependence of the conditional distribution on $\boldsymbol{b}_i$, indicating that it contains adequate information to estimate the coefficients $\alpha$ and $\boldsymbol{\phi}$, since the survival model in the joint model (Equation 7) is mainly connected to the longitudinal model through the random effects $\boldsymbol{b}_i$. This leads to the formulation of a conditional score estimating equation for $\alpha$ and $\boldsymbol{\phi}$, analogous to the derivation of classical partial likelihood score equations or generalized estimating equations. The estimates for these parameters are then derived by solving this estimating equation.

While this approach can provide computationally effective unbiased estimators for $\alpha$ and $\boldsymbol{\phi}$, and leads to consistency and asymptotically normality of the resulting estimators, it is less efficient than the joint likelihood approach. According to Tsiatis & Davidian (2001), the efficiency loss is partly attributed to the way the conditional score method handles the estimation of $U_i(t)$, which does not utilize data beyond time $t$. In contrast, the EM algorithm takes account of information contained in the entire trajectories. Nevertheless, the estimates from the conditional score method can be useful to accelerate the EM algorithm by serving as initial estimates.

### 3.3. Two-Stage Method

A simple approach is to impute (or interpolate) the longitudinal covariates and then fit a survival model based on the imputed data. There are several ways to impute the longitudinal data. A preferred approach is to do this through a longitudinal model. A general approach is given as follows:

- **Stage 1:** The longitudinal data are analyzed using a longitudinal statistical model, such as a linear mixed model or functional principal component analysis (FPCA). The random effects for each subject are then used to impute the individual longitudinal trajectories for that subject.
- **Stage 2:** The usual survival model fitting is performed, using the imputed longitudinal trajectories in stage 1 instead of the original observed longitudinal covariate.

Stage 1 aims to recover the trajectories of longitudinal covariates over time, which mitigates issues with both missing data and measurement errors to some extent. It also facilitates the survival modeling with longitudinal covariates in stage 2. While such a two-stage approach facilitates the modeling of survival outcome with longitudinal covariates, it induces bias in the final analysis in two ways: (*a*) The informative dropout caused by death in the longitudinal covariates would trigger bias in the analysis in stage 1, and (*b*) imputing longitudinal trajectories for sparsely observed longitudinal data would induce bias in the analysis of stage 2.

In contrast, a joint modeling approach does not induce any bias in the analysis and increases efficiency as it utilizes both the survival and longitudinal data simultaneously, rather than sequentially in the two-stage approach.

### 3.4. Bayesian Method

So far, we have only covered frequentist approaches for joint modeling, but various Bayesian approaches have also been proposed.

The starting point is to specify prior distributions for the parameters of interest based on existing knowledge, expert opinion, or previous studies. Let the prior probability density function of the parameters $\boldsymbol{\psi}$ be $p(\boldsymbol{\psi})$. Denote $\boldsymbol{b} = (\boldsymbol{b}_1^\top, \ldots, \boldsymbol{b}_n^\top)^\top$, $\mathcal{D}_n$ is the observed data, and the posterior distribution of $(\boldsymbol{\psi}, \boldsymbol{b})$ is calculated combining the prior distribution and the likelihood function,

$$p(\boldsymbol{\psi}, \boldsymbol{b} \mid \mathcal{D}_n) \propto \prod_{i=1}^{n} [p_{\boldsymbol{\psi}}(Y_i, \Delta_i \mid \boldsymbol{b}_i) p_{\boldsymbol{\theta}}(\boldsymbol{U}_i \mid \boldsymbol{b}_i) p_{\boldsymbol{\theta}}(\boldsymbol{b}_i)] p(\boldsymbol{\psi}).$$

Markov chain Monte Carlo methods, such as the Gibbs sampler or the Metropolis–Hastings algorithm, are then used to draw samples from the posterior distribution $p(\boldsymbol{\psi}, \boldsymbol{b} \mid \mathcal{D}_n)$. For a more comprehensive description of this approach, we refer readers to Faucett & Thomas (1996); Faucett et al. (1998), Wang & Taylor (2001), Xu & Zeger (2001), Law et al. (2002), Brown & Ibrahim (2003), Ibrahim et al. (2004), Chi & Ibrahim (2006), Rizopoulos & Ghosh (2011), and Andrinopoulou & Rizopoulos (2016), among others.

## 4. APPLICATION TO REAL DATA

Revisiting the AIDS data (Goldman et al. 1996), we illustrate joint modeling by investigating the clinical efficacy of two drugs, ddI and ddC, on the lifetime of HIV-infected patients while adjusting for the CD4 cell count. The dataset comprises 467 patients, with 188 subjects deceased during the 18-month follow-up period, resulting in a 59.7% censoring rate. At the study onset, each patient was randomly assigned to receive either the ddI or ddC treatment (ddI = 1, ddC = 0), and two baseline covariates, `sex` (male=1, female=0) and previous infection at study entry (`prevDiag`) (AIDS diagnosis = 1, no AIDS diagnosis = 0), were also recorded. The CD4 counts were assessed at the study entry and subsequently at 2, 6, 12, and 18 months until death, loss to follow-up, or study completion.

For this analysis, we employed the model

$$\begin{cases} V_{ij} = U_i(T_{ij}) + \epsilon_{ij} = \beta_0 + \beta_1 \texttt{drug}_i + \beta_2 \texttt{sex}_i + \beta_3 \texttt{prevDiag}_i + \beta_4 T_{ij} + \beta_5 T_{ij}^2 \\ \qquad\quad + b_{0i} + b_{1i} T_{ij} + b_{2i} T_{ij}^2 + \epsilon_{ij}, \\ \Lambda(t \mid \boldsymbol{b}_i, U_i, \boldsymbol{W}_i) = H_\rho \left( \int_0^t \lambda_0(s) \exp\{\alpha U_i(s) + \phi_1 \texttt{drug}_i + \phi_2 \texttt{sex}_i + \phi_3 \texttt{prevDiag}_i\} \mathrm{d}s \right), \end{cases}$$

where $V_{ij}$ is the observed CD4 counts, $\boldsymbol{W}_i = (\texttt{drug}_i, \texttt{sex}_i, \texttt{prevDiag}_i)^\top$, and $H_\rho(t) = \log(1 + \rho t)/\rho$ represents the logarithmic transformation, with $\rho$ being a tuning parameter. Using the R package `JSM` (Xu et al. 2020), we fitted several models for various values of $\rho$ in a candidate grid {0, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 2, 3, 4, 5} and chose the optimal one using the Akaike information criterion (AIC). **Figure 2** displays the AICs from different $\rho$ values and suggests that the model with $\rho = 0$ (i.e., the Cox model) provides the smallest AIC. The resulting coefficient estimates under the Cox model are shown in **Table 1**.

The results in **Table 1** show that there is a significant linear time trend for the longitudinal CD4 counts and prior diagnosis with AIDS is a significant covariate. For the risk of death due to AIDS, ddC is a more effective treatment than ddI, subjects with prior diagnosis with AIDS have significantly higher risk, and lower CD4 counts are significantly associated with higher risk.

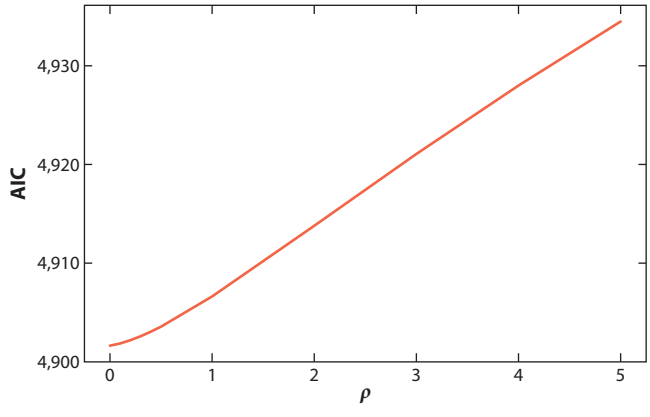


**Figure 2**

AIC curve for model fitting as a function of the logarithmic transformation parameter $\rho$ for the AIDS data. Abbreviation: AIC, Akaike information criterion.

**Table 1 Coefficient estimates for the AIDS data**

| Coefficient | Estimate | Standard error | *p*-Value |
|---|---|---|---|
| $\beta_0$ | 3.098 | 0.134 | $<2.2 \times 10^{-16}$ |
| $\beta_1$ | 0.070 | 0.074 | 0.342 |
| $\beta_2$ | −0.012 | 0.127 | 0.925 |
| $\beta_3$ | −0.916 | 0.078 | $<2.2 \times 10^{-16}$ |
| $\beta_4$ | −0.045 | 0.007 | $5.6 \times 10^{-10}$ |
| $\beta_5$ | 0.0006 | 0.0005 | 0.265 |
| $\phi_1$ | 0.333 | 0.151 | 0.026 |
| $\phi_2$ | −0.298 | 0.253 | 0.241 |
| $\phi_3$ | 0.774 | 0.219 | $4.2 \times 10^{-4}$ |
| $\alpha$ | −0.934 | 0.125 | $6.7 \times 10^{-14}$ |

## 5. EXTENSIONS TO OTHER MODELS AND DATA TYPES

### 5.1. Extensions to Other Survival Models

In the previous sections, our emphasis is primarily on the transformation survival model within the joint model (Equation 7). Alternative survival models, such as the AFT (Wei 1992) and the extended hazard (Ciampi & Etezadi-Amoli 1985) models could also be used and have their own appealing properties.

Given the latent process $U$ and covariates $\boldsymbol{W}$ as outlined in Section 2.4, the hazard function of an AFT model is

$$\lambda(t \mid \boldsymbol{b}, U, \boldsymbol{W}) = \lambda_0 \left( \int_0^t \exp\{\alpha U(s) + \boldsymbol{\phi}^\top \boldsymbol{W}(s)\} \mathrm{d}s \right) \exp\{\alpha U(t) + \boldsymbol{\phi}^\top \boldsymbol{W}(t)\}, \qquad 9.$$

where $\lambda_0(\cdot)$ represents an unspecified baseline hazard function. In contrast with the Cox model, where the baseline hazard function is solely a function of time, the AFT model incorporates covariate effects on the baseline hazard function as well. This provides the interpretation of the accelerated failure schedule, hence the name AFT. The AFT model in Equation 9 is equivalent to $e^\epsilon = \int_0^T \exp\{\alpha U(s) + \boldsymbol{\phi}^\top \boldsymbol{W}(s)\} \mathrm{d}s$, where $\lambda_0$ is the hazard function of $e^\epsilon$. Moreover, it is quite easy to derive the equivalence of the AFT model to the following log-linear model, when there are only baseline covariates involved:

$$\log T = -\boldsymbol{\phi}^T \boldsymbol{W} + \epsilon,$$

where the hazard function of $e^\epsilon$ is $\lambda_0$. This intriguing alternative expression of the AFT model sets it apart from other survival models, as it resembles a conventional regression model and can predict the event time directly without going through the hazard function. This is an attractive feature of the AFT model, but it poses computational challenges as $T$ is not observed for a portion of the data, so the estimation of $\boldsymbol{\phi}$ still needs to be routed through the hazard function in Equation 9. Moreover, the fact that $\boldsymbol{\phi}$ depicts both the risk factor and an accelerated failure schedule within the baseline component complicates the estimation of $\boldsymbol{\phi}$, as the NPMLE is intractable.

To tackle this challenge, Tseng et al. (2005) proposed a pseudolikelihood approach by approximating the baseline hazard function through a step function that remains constant between two consecutive observed failure times. This reparameterization scheme substantially simplifies the EM algorithm and produces reliable estimates in numerical studies. Later, Zeng & Lin (2007a) proposed another type of pseudolikelihood that produces a smooth baseline hazard estimate and semiparametric efficient estimators for the regression parameter in an AFT model with only baseline covariates. Their approach was generalized by Tseng et al. (2015) to the joint modeling setting

with longitudinal covariates and a scenario where the survival data belongs to an extended hazard model.

The hazard function for the extended hazard model takes the following form:

$$\lambda(t \mid \boldsymbol{b}, U, \boldsymbol{W}) = \lambda_0 \left( \int_0^t \exp\{\alpha_1 U(s) + \boldsymbol{\phi}_1^\top \boldsymbol{W}(s)\} \mathrm{d}s \right) \exp\{\alpha_2 U(t) + \boldsymbol{\phi}_2^\top \boldsymbol{W}(s)\}, \qquad 10.$$

where $\alpha_1$, $\alpha_2$, $\boldsymbol{\phi}_1$, and $\boldsymbol{\phi}_2$ are unknown parameters, and $\lambda_0(\cdot)$ represents an unspecified baseline hazard function. This extended hazard model offers a more flexible and broader range of models compared with the Cox and AFT models, and it includes both of these models as a special case. Specifically, the extended hazard model reduces to the Cox model when $\alpha_1 = 0$ and $\boldsymbol{\phi}_1 = \mathbf{0}$, and it becomes the AFT model when $\alpha_1 = \alpha_2$ and $\boldsymbol{\phi}_1 = \boldsymbol{\phi}_2$. Inspired by the work of Zeng & Lin (2007a) on the AFT survival model, Tseng et al. (2015) systematically examined the joint modeling of the extended hazard model and its longitudinal covariates in terms of theory and implementation, establishing asymptotic normality and semiparametric efficiency of parametric estimators.

## 5.2. Other Types of Incomplete Survival Data

In addition to right-censored event times, there is a diverse array of incomplete survival data, including left-censored, double-censored, interval-censored, left-truncated, and right-truncated data, as well as cure data and data resulting from competing risks. The extension to other forms of incomplete data follows the same path: establishing the joint likelihood of the survival and longitudinal data, then finding a suitable likelihood approach that can facilitate the corresponding EM algorithm. Below, we briefly summarize these types of survival data along with the respective joint modeling approaches.

- Double-censored data arise when individuals experience left censoring or right censoring; the former occurs when the event of interest for an individual has already occurred before the study begins, leading to incomplete information on the exact event time. With $L$ and $C$ being the left- and right-censored times, respectively, the observed data include the observed time $Y = \max(L, \min(T, C))$ and the right- and left-censoring indicators $\Delta = I(T \leq C)$ and $\eta = I(T \geq L)$, respectively. The PhD thesis of Xu (2014) contains a thorough analysis of the joint likelihood approach for doubly censored survival data and associated longitudinal covariates, based on a general class of transformation survival models. Additional work for doubly censored data in joint modeling settings includes that of Su & Wang (2016) and Li et al. (2020).
- Interval-censored data in survival analysis are present in a situation where the exact event time is not known, but it is known to have occurred within a specific time interval (see, e.g., Sun 2006, Sun & Chen 2022, Yi et al. 2022).
- In addition to the usual right censoring, left truncation is common in studies with delayed entry. For instance, individuals whose events occurred before the study entry point may have been excluded from the analysis, resulting in a biased sample. In this regard, left truncation is very different from left censoring, which produces incomplete, but not biased, data (see, e.g., Su & Wang 2012).
- Cure survival data occur in a situation where a subset of individuals in a study population is considered cured or immune to the event of interest, meaning that they will never experience the event. Thus, their event time takes the value $\infty$. This often occurs in studies where a portion of individuals either are not susceptible to the risk under investigation or are cured of the disease after some time. As a result, the population is a mixture of individuals who are susceptible to the event and those who are cured or immune to it (for details, see Law et al.

2002; Brown & Ibrahim 2003; Chen et al. 2004; Yu et al. 2004, 2008; Kim et al. 2013; Amico & Van Keilegom 2018).

- Competing risks or multiple failure types in survival analysis are encountered when individuals in a study may be at risk for more than one type of event, but the occurrence of one event, such as death due to this event, precludes the occurrence of all other events. In other words, all the other potential event times serve as right-censoring times, because individuals who experience a competing event are no longer at risk of the primary event and are therefore censored at the time of the competing event. The complication here is that the usual independence assumption between the event time and censoring time no longer holds as these competing risks are inherently correlated within a subject. Consequently, the analysis of survival data subject to competing risks requires careful handling of the censoring scheme. Ignoring competing risks or treating them as independent censored observations will lead to biased estimates of the event of interest (for further details, see Elashoff et al. 2007, 2008, 2016; Hu et al. 2009; Li et al. 2010, 2012; Xu 2014).

## 5.3. Multivariate Longitudinal Processes

So far, for simplicity of presentation, we have implicitly assumed that there is only one latent longitudinal covariate $U$. In many clinical studies, it is common to repeatedly record multiple longitudinal outcomes along with an event time for the same individual. Suppose that there are $p$ longitudinal covariates. For individual $i$, denote as $\boldsymbol{V}_{ik}$ a column vector of length $n_{ik}$ of repeated observations for the $k$th component of the latent process $U_{ik}(\cdot)$. That is, $\boldsymbol{V}_{ik} = (V_{i1k}, \ldots, V_{in_{ik}k})^\top$, with

$$V_{ijk} = U_{ik}(T_{ijk}) + \epsilon_{ijk}, i = 1, \ldots, n; j = 1, \ldots, n_{ik}; k = 1, \ldots, p,$$

where the $\epsilon_{ijk}$ are measurement errors. Under the transformation survival model with user-specified covariate $(\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t))$ and external covariate $\boldsymbol{W}_i$, the joint model of multivariate longitudinal and survival data is

$$\begin{cases} U_{ik}(t) = \boldsymbol{\beta}_k^\top \boldsymbol{X}_i(t) + \boldsymbol{b}_{ik}^\top \boldsymbol{Z}_i(t), k = 1, \ldots, p, \\ \Lambda(t \mid \boldsymbol{\mathcal{B}}_i, \boldsymbol{U}_i, \boldsymbol{W}_i) = H(\int_0^t \lambda_0(s) \exp\{\sum_{k=1}^p \alpha_k U_{ik}(s) + \boldsymbol{\phi}^\top \boldsymbol{W}_i(s)\} \mathrm{d}s), \end{cases}$$

where $\boldsymbol{\beta}_k$ and $\boldsymbol{b}_{ik}$ represent fixed and random effects, $\boldsymbol{\mathcal{B}}_i = (\boldsymbol{b}_{i1}^\top, \ldots, \boldsymbol{b}_{ip}^\top)^\top$, and $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{ip})^\top$. The random effects $\boldsymbol{\mathcal{B}}_i$ capture the correlations among different longitudinal trajectories and are often assumed to follow a multivariate normal distribution. However, increasing the number of random effects poses a significant computational burden, primarily as one needs to compute multidimensional integrals over the random effects. For more details, readers are directed to Elashoff et al. (2016, chapter 6) and the review article by Hickey et al. (2016).

## 5.4. Nonparametric Approaches for Longitudinal or Survival Data

Although the predominant approach for the linear mixed-effects models in Equation 4 has been parametric, it can be extended to a nonparametric model via a basis expansion approach for the time effects. For instance, if we choose $\boldsymbol{X}(t)$ and $\boldsymbol{Z}(t)$ by B-splines (De Boor 1978), a nonparametric model emerges when the number of spline basis functions grows with the sample size, where both the fixed and random effects can be modeled nonparametrically. Specifically,

$$U_i(t) = \sum_{k=1}^{q_1} \beta_k B_k(t) + \sum_{k=1}^{q_2} b_k \tilde{B}_k(t),$$

where $\{B_k(\cdot)\}_{k=1}^{q_1}$ and $\{\tilde{B}_k(\cdot)\}_{k=1}^{q_2}$ are B-spline functions, and $\beta_k$ and $b_k$ are fixed and random coefficients, respectively (for further details, see Rice & Wu 2001).

**5.4.1. Functional principal component analysis approaches.** One drawback with the B-splines is the involvement of a large number of random effects and the associated difficulty of the required computation of a high-dimensional integral in the E-step of the EM algorithm. This can be rectified with an alternative nonparametric approach developed independently in another community, where one treats longitudinal data as sparsely observed functional data. The latent processes $U_i$ in Equation 7 can be regarded as functional data, which in this case are random curves defined on an interval. A brief overview of functional data analysis (FDA) is provided in the review article by Wang et al. (2016). It is well known in the FDA community that the FPCA approach is the most parsimonious way to accommodate the random effects in functional data. The FPCA approach initially did not gain traction in the longitudinal community in its original version as it could not handle sparsely measured functional data, which correspond to longitudinal data. A turning point was when Yao et al. (2005) proposed a nonparametric smoothing approach to estimate the mean and covariance functions for sparsely measured functional data and further leveraged the Karhunen–Loéve expansion to produce random effects and implement the PACE (Principal Analysis by Conditional Estimation) method (Zhou et al. 2024) to estimate these random effects.

Under the Cox proportional hazards assumption, Yao (2007) applied the FPCA approach to model the random trajectory jointly with survival data. The latent process is modeled as

$$U_i(t) = \mu(t) + \boldsymbol{\beta}^\top \boldsymbol{X}_i(t) + \sum_{k=1}^{\infty} \xi_k \varphi_k,$$

where $\mu(t) = \mathrm{E}[U_i(t)]$, $\boldsymbol{X}_i(\cdot)$ is a vector of covariates, and $(\xi_k, \varphi_k)$ are paired scores and eigenfunctions. This formulation assumes that $\mathrm{E}[U_i(t) - \mu(t) \mid \boldsymbol{X}_i(t)]$ is a linear predictor $\boldsymbol{\beta}^\top \boldsymbol{X}_i(t)$ of $\boldsymbol{X}_i(t)$. Yao (2007) employed the EM algorithm to estimate $\mu(t)$, represented by B-splines, and treated the number of eigenfunctions in the model as a tuning parameter chosen by the AIC. Specifically, $\mu(t) = \sum_{k=1}^{q_1} \beta_k B_k(t)$, where $\{B_k(\cdot)\}$ are B-spline functions, and $\beta_k$ are fixed coefficients that expand the mean function $\mu(t)$.

While the FPCA approach can be viewed as a linear mixed-effects model, it is intrinsically a nonparametric approach as both the fixed and random effects structures are nonparametric. Such nonparametric models offer greater flexibility to encompass a wider range of function families and are typically data adaptive, automatically capturing important features of an individual curve.

Aside from this difference, there are fundamental and philosophical distinctions between the longitudinal and FDA communities. This was elucidated in an overview article (Rice 2004) in a special 2004 issue of *Statistica Sinica* titled "Emerging Issues in Longitudinal and Functional Data Analysis" (Davidian et al. 2004).

**5.4.2. Varying coefficient models.** Another nonparametric approach involves extending the varying coefficient model to the joint model, either for the longitudinal data or the survival model, or for both. This extension can be achieved by modeling the latent process $U_i(\cdot)$ in Equation 7 as

$$U_i(t) = \boldsymbol{\beta}^\top(t) \boldsymbol{X}_i(t) + \boldsymbol{b}_i^\top \boldsymbol{Z}_i(t),$$

or by modeling the cumulative hazard function as

$$\Lambda(t \mid \boldsymbol{b}_i, U_i, \boldsymbol{W}_i) = H\left( \int_0^t \lambda_0(s) \exp\left\{ \alpha(s) U_i(s) + \boldsymbol{\phi}^\top(s) \boldsymbol{W}_i(s) \right\} \mathrm{d}s \right),$$

where $\boldsymbol{\beta}(\cdot)$, $\alpha(\cdot)$, and $\boldsymbol{\phi}(\cdot)$ are unknown coefficient functions (for references on varying-coefficient longitudinal models, see Brumback & Rice 1998; Hoover et al. 1998; Guo 2002; Huang et al. 2002, 2004; Fan et al. 2003; Morris & Carroll 2006; Müller & Zhang 2005, and for survival models, see Murphy 1993, Marzec & Marzec 1997, Cai & Sun 2003, Tian et al. 2005).

**5.4.3. Machine learning approaches.** With modern advances in technology and AI, new tools from the machine learning community can now be deployed to ease the computational burden and orient the field of joint modeling to move toward nonparametric approaches. For instance, recurrent neural networks, a type of neural network designed for sequence data, can be adopted to modeling longitudinal data to provide more accurate risk prediction (Lee et al. 2019, Nagpal et al. 2021, Wiegrebe et al. 2024). In addition, deep neural networks (deep learning) can be deployed to nonparametrically model the survival models. Zhong et al. (2021) demonstrate this for the extended hazard model in Equation 10, but their approach has yet to be extended to the joint modeling setting.

## 6. SOFTWARE

Several software packages are publicly available for the joint analysis of longitudinal and survival data. For example, SAS (SAS Inst. Inc. 2013) and WinBUGS (Spiegelhalter et al. 2003) assume that the baseline hazard function follows a Weibull distribution, which may be quite restrictive in practice. The SAS macro `JMFit` (Zhang et al. 2016), and the R packages `JM` (Rizopoulos 2010), `JMbayes` (Rizopoulos 2016), `JMbayes2` (Rizopoulos et al. 2024), `lcmm` (Proust-Lima et al. 2023), `frailtypack` (Rondeau et al. 2024), and `INLAjoint` (Rustand et al. 2024), as well as the Stata module `stjm` (Crowther et al. 2013), provide more flexible models by allowing the baseline hazard function to be represented by piecewise constants or spline functions. However, in these packages the tuning parameters for the survival component are prefixed and not data adaptive.

Moreover, the R packages `joineR` (Philipson et al. 2018) and `joineRML` (Hickey et al. 2018) model the baseline hazard completely unspecified and obtain SE estimates of the regression parameters using the bootstrap method (Efron 1994).

A recent R package, `JSM` (Xu et al. 2020), employs B-spline basis functions to model the latent process and provides computationally efficient SE estimates for a broad class of semiparametric models, such as the Cox proportional hazards model and the logarithmic class of transformation models. It also contains an implementation of the nonparametric multiplicative random effects model of Ding & Wang (2008) and the shared random-effects model of Henderson et al. (2000).

## 7. CONCLUSIONS AND FUTURE DIRECTIONS

It has been nearly 30 years since the pioneering work of Pawitan & Self (1993), De Gruttola & Tu (1994), and Wulfsohn & Tsiatis (1997) that advocated the importance of modeling the survival and longitudinal components through a joint likelihood approach. Because of the need for a likelihood function, the approaches to model the effects of covariates have been predominantly parametric, except for the baseline hazard function, which can be estimated nonparametrically thanks to the ingenious idea of nonparametric likelihood from Kiefer & Wolfowitz (1956). Arguably, this and other nonparametric likelihood approaches and the EM algorithm form the bedrock of the field. Together, they have facilitated the implementation of the joint likelihood approach and carried the field into the new millennium. Below we present a few future directions that reflect our subjective view.

- More complex survival and longitudinal models: A major challenge of the joint modeling approach is its extension to more flexible models, including a fully nonparametric survival or longitudinal model. While B-splines could be deployed to fit either the survival or longitudinal component as described in Section 5.4, the EM algorithm may not be stable in the face of a high-dimensional optimization problem, especially when the baseline function is estimated nonparametrically. The Breslow formula for estimating the baseline cumulative

hazard function provides much relief and stability in the M-step of the EM algorithm, but the joint likelihood is still intrinsically a high-dimensional nonconvex optimization problem. Therefore, we need more efficient and stable optimization methods that are scalable for big data.

- Scalable nonparametric approaches: Another challenge is the need for numerical integration in the joint likelihood function, which limits the number of random effects that can be realistically employed in the E-step of the EM algorithm. The challenges in the E- and M-steps can be alleviated by deploying modern machine learning methods, such as deep neural networks.

  In this era of AI and big data, there is a pressing need for scalable nonparametric approaches in the field of joint modeling. For instance, survival distributions in the presence of a high-dimensional covariate may exhibit a complex yet low-dimensional structure (Bauer & Kohler 2019, Schmidt-Hieber 2020, Jiao et al. 2023), which cannot be detected by traditional nonparametric approaches, such as kernel and B-spline smoothing methods. Therefore, these classical methods are subject to the curse of dimensionality. In this regard, neural networks can be a promising tool to overcome this curse, as they are capable of automatically detecting the low-dimensional structure of a function embedded in a high-dimensional space. Recent work on deep learning for survival data (Kvamme et al. 2019; Katzman et al. 2018; Zhong et al. 2021, 2022) underscores the effectiveness of deep learning for survival data. However, little is known about how this pans out for joint modeling. While some empirical approaches have emerged (Lee et al. 2019, Nagpal et al. 2021, Lin & Luo 2022, Zeng et al. 2024), theoretical support is still lacking. This presents an opportunity for the field to progress and evolve in response to the fast development of new AI tools.
- Beyond the concurrent model: So far, the approaches are concurrent—only the current longitudinal values at time $t$ affect the risk at time $t$. However, there may be a time lag for the effect of the longitudinal data to kick in. Moreover, maybe the entire history of the longitudinal data, or part of the history, could be useful to model the survival data. For instance, we could use the average longitudinal values up to the current time, the entire history of the longitudinal process, or the values in the most recent period to model and predict the event time. We could even use the slope of the longitudinal process for survival modeling. Many of these possibilities require the development of new survival models that allow an entire stochastic process, e.g., the history process, to link to the risk of a subject. While this will trigger further challenges in an already challenging framework, addressing these presents an opportunity to advance the field.
- More complex longitudinal data: Although this review primarily focuses on univariate/multivariate longitudinal data in the Euclidean space, new data types have emerged lately. These include high-dimensional longitudinal data that allow the dimension to grow with sample size, such as gene expression profiles (Molyneaux et al. 2017, Sun et al. 2019) and the human microbiome, where compositional data are repeatedly recorded using high-throughput sequencing technologies (Livanos et al. 2016, Hu et al. 2022). Lastly, there is also a rising interest in longitudinal imaging data (Kang & Song 2023, Zeng et al. 2024, Zou et al. 2023, Zhou & Song 2024). These new data types pose challenges that necessitate new statistical models, methodologies, and computational methods for joint modeling.
- Hypothesis testing: So far, we have only touched upon one branch of statistical inference, estimation, in the joint modeling framework. This reflects the evolution of the field as results on hypothesis testing are scarce. Clearly, there is much room to explore hypothesis testing. Tests that only involve testing a finite-dimensional parameter might be relatively

straightforward, but testing whether a proposed survival or longitudinal model is suitable for the data in hand could be complex. More research in this direction would be very welcome.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We thank an anonymous reviewer for many helpful suggestions that improved the quality of the article. J.-L.W.'s research has been supported by the US National Science Foundation grant DMS22-10891, and Q.Z. was supported by the National Natural Science Foundation of China (12201527 and 12271286).

## LITERATURE CITED

Abrams D, Goldman A, Launer C, Korvick J, Neaton J, et al. 1994. Comparative trial of didanosine and zalcitabine in patients with human immunodeficiency virus infection who are intolerant of or have failed zidovudine therapy. *N. Engl. J. Med.* 330(10):657–62

Amico M, Van Keilegom I. 2018. Cure models in survival analysis. *Annu. Rev. Stat. Appl.* 5:311–42

Andrinopoulou ER, Rizopoulos D. 2016. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Stat. Med.* 35(26):4813–23

Bauer B, Kohler M. 2019. On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Stat.* 47(4):2261–85

Begun JM, Hall WJ, Huang WM, Wellner JA. 1983. Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Stat.* 11(2):432–52

Bennett S. 1983. Analysis of survival data by the proportional odds model. *Stat. Med.* 2(2):273–77

Bickel PJ, Klaassen CA, Bickel PJ, Ritov Y, Klaassen J, et al. 1998. *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins Univ. Press

Brown ER, Ibrahim JG. 2003. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics* 59(3):686–93

Brumback BA, Rice JA. 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Stat. Assoc.* 93(443):961–76

Cai T, Cheng S. 2004. Semiparametric regression analysis for doubly censored data. *Biometrika* 91(2):277–90

Cai Z, Sun Y. 2003. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand. J. Stat.* 30(1):93–111

Chen MH, Ibrahim JG, Sinha D. 2004. A new joint model for longitudinal and survival data with a cure fraction. *J. Multivariate Anal.* 91(1):18–34

Chi YY, Ibrahim JG. 2006. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* 62(2):432–45

Ciampi A, Etezadi-Amoli J. 1985. A general model for testing the proportional hazards and the accelerated failure time hypotheses in the analysis of censored survival data with covariates. *Commun. Stat. Theory Methods* 14(3):651–67

Cox DR. 1972. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* 34(2):187–202

Cox DR. 1975. Partial likelihood. *Biometrika* 62(2):269–76

Crowther MJ, Abrams KR, Lambert PC. 2013. Joint modeling of longitudinal and survival data. *Stata J.* 13(1):165–84

Dabrowska DM, Doksum KA. 1988. Partial likelihood in transformation models with censored data. *Scand. J. Stat.* 15(1):1–23

Davidian M, Lin X, Wang JL, eds. Emerging issues in longitudinal and functional data analysis. *Stat. Sin.* 14(3)

De Boor C. 1978. *A Practical Guide to Splines*. New York: Springer-Verlag

De Gruttola V, Tu XM. 1994. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* 50(4):1003–14

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39(1):1–22

Ding J, Wang JL. 2008. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* 64(2):546–56

Dupuy JF, Grama I, Mesbah M. 2006. Asymptotic theory for the Cox model with missing time-dependent covariate. *Ann. Stat.* 34(2):903–24

Efron B. 1994. Missing data, imputation, and the bootstrap. *J. Am. Stat. Assoc.* 89(426):463–75

Elashoff RM, Li G, Li N. 2007. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat. Med.* 26(14):2813–35

Elashoff RM, Li G, Li N. 2008. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* 64(3):762–71

Elashoff RM, Li N, Li G. 2016. *Joint Modeling of Longitudinal and Time-to-Event Data*. Boca Raton, FL: Chapman and Hall/CRC

Etezadi-Amoli J, Ciampi A. 1987. Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics* 43(1):181–92

Fan J, Yao Q, Cai Z. 2003. Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B* 65(1):57–80

Faucett CL, Schenker N, Elashoff RM. 1998. Analysis of censored survival data with intermittently observed time-dependent binary covariates. *J. Am. Stat. Assoc.* 93(442):427–37

Faucett CL, Thomas DC. 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Stat. Med.* 15(15):1663–85

Furgal AK, Sen A, Taylor JM. 2019. Review and comparison of computational approaches for joint longitudinal and time-to-event models. *Int. Stat. Rev.* 87(2):393–418

Goldman AI, Carlin BP, Crane LR, Launer C, Korvick JA, et al. 1996. Response of CD4 lymphocytes and clinical consequences of treatment using ddI or ddC in patients with advanced HIV infection. *JAIDS* 11(2):161–69

Guo W. 2002. Functional mixed effects models. *Biometrics* 58(1):121–28

Han M, Song X, Sun L, Liu L. 2014. Joint modeling of longitudinal data with informative observation times and dropouts. *Stat. Sin.* 24(4):1487–504

Henderson R, Diggle P, Dobson A. 2000. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1(4):465–80

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. 2016. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med. Res. Methodol.* 16:117

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. 2018. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Med. Res. Methodol.* 18:50

Hoover DR, Rice JA, Wu CO, Yang LP. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85(4):809–22

Hsieh F, Tseng YK, Wang JL. 2006. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* 62(4):1037–43

Hu J, Wang C, Blaser MJ, Li H. 2022. Joint modeling of zero-inflated longitudinal proportions and time-to-event data with application to a gut microbiome study. *Biometrics* 78(4):1686–98

Hu W, Li G, Li N. 2009. A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat. Med.* 28(11):1601–19

Huang CY, Wang MC. 2004. Joint modeling and estimation for recurrent event processes and failure time data. *J. Am. Stat. Assoc.* 99(468):1153–65

Huang JZ, Wu CO, Zhou L. 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 89(1):111–28

Huang JZ, Wu CO, Zhou L. 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Stat. Sin.* 14(3):763–88

Ibrahim JG, Chen MH, Sinha D. 2004. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Stat. Sin.* 14(3):863–83

Jamshidian M, Jennrich RI. 2000. Standard errors for EM estimation. *J. R. Stat. Soc. Ser. B* 62(2):257–70
Jiao Y, Shen G, Lin Y, Huang J. 2023. Deep nonparametric regression on approximate manifolds: nonasymptotic error bounds with polynomial prefactors. *Ann. Stat.* 51(2):691–716
Kalbfleisch J, Prentice R. 2002. *The Statistical Analysis of Failure Time Data*. New York: Wiley. 2nd ed.
Kang K, Song XY. 2023. Joint modeling of longitudinal imaging and survival data. *J. Comput. Graph. Stat.* 32(2):402–12
Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53(282):457–81
Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. 2018. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* 18:24
Kiefer J, Wolfowitz J. 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* 27(4):887–906
Kim S, Zeng D, Li Y, Spiegelman D. 2013. Joint modeling of longitudinal and cure-survival data. *J. Stat. Theory Pract.* 7:324–44
Kim S, Zeng D, Taylor JM. 2017. Joint partially linear model for longitudinal data with informative drop-outs. *Biometrics* 73(1):72–82
Klein JP, Van Houwelingen HC, Ibrahim JG, Scheike TH. 2014. *Handbook of Survival Analysis*. Raton, FL: CRC Press Boca
Kosorok MR. 2008. *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer
Kvamme H, Borgan Ø, Scheel I. 2019. Time-to-event prediction with neural networks and Cox regression. *J. Mach. Learn. Res.* 20(129):1–30
Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. *Biometrics* 38:963–74
Law NJ, Taylor JM, Sandler H. 2002. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* 3(4):547–63
Lee C, Yoon J, Van Der Schaar M. 2019. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Trans. Biomed. Eng.* 67(1):122–33
Li N, Elashoff RM, Li G, Saver J. 2010. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Stat. Med.* 29(5):546–57
Li N, Elashoff RM, Li G, Tseng CH. 2012. Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Stat. Med.* 31(16):1707–21
Li S, Sun J, Tian T, Cui X. 2020. Semiparametric regression analysis of doubly censored failure time data from cohort studies. *Lifetime Data Anal.* 26:315–38
Lin J, Luo S. 2022. Deep learning for the dynamic prediction of multivariate longitudinal and survival data. *Stat. Med.* 41(15):2894–907
Little RJ. 1995. Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Stat. Assoc.* 90(431):1112–21
Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, et al. 2016. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nat. Microbiol.* 1:16140
Louis TA. 1982. Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc. Ser. B* 44(2):226–33
Marzec L, Marzec P. 1997. On fitting Cox's regression model with time-dependent coefficients. *Biometrika* 84(4):901–8
Molyneaux PL, Willis-Owen SA, Cox MJ, James P, Cowman S, et al. 2017. Host–microbial interactions in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 195(12):1640–50
Morris JS, Carroll RJ. 2006. Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B* 68(2):179–99
Müller HG, Zhang Y. 2005. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* 61(4):1064–75
Murphy SA. 1993. Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Stat.* 21(1):35–50
Nagpal C, Jeanselme V, Dubrawski A. 2021. Deep parametric time-to-event regression with time-varying covariates. *Proc. Mach. Learn. Res.* 146:184–93

Pan Q, Yi GY. 2011. An estimation method of marginal treatment effects on correlated longitudinal and survival outcomes. *Stat. Interface* 4(4):499–509

Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. 2019. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annu. Rev. Stat. Appl.* 6:223–40

Pawitan Y, Self S. 1993. Modeling disease marker processes in aids. *J. Am. Stat. Assoc.* 88(423):719–26

Philipson P, Sousa I, Diggle P, Williamson P, Kolamunnage-Dona R, Henderson R. 2018. joineR: joint modelling of repeated measurements and time-to-event data. *R Package*, version 1.2.4. **https://CRAN.R-project.org/package=joineR**

Pinheiro JC, Bates DM. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. Comput. Graph. Stat.* 4(1):12–35

Proust-Lima C, Philipps V, Diakite A, Liquet B. 2023. `lcmm`: Extended mixed models using latent classes and latent processes. *R Package*, version 2.1.0. **https://cran.r-project.org/package=lcmm**

Raboud J, Reid N, Coates R, Farewell V. 1993. Estimating risks of progressing to aids when covariates are measured with error. *J. R. Stat. Soc. Ser. A* 156(3):393–406

Ratcliffe SJ, Guo W, Ten Have TR. 2004. Joint modeling of longitudinal and survival data via a common frailty. *Biometrics* 60(4):892–99

Rice JA. 2004. Functional and longitudinal data analysis: perspectives on smoothing. *Stat. Sin.* 14:631–47

Rice JA, Wu CO. 2001. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 57(1):253–59

Rizopoulos D. 2010. JM: an R package for the joint modelling of longitudinal and time-to-event data. *J. Stat. Softw.* 35:1–33

Rizopoulos D. 2016. The R package `JMbayes` for fitting joint models for longitudinal and time-to-event data using MCMC. *J. Stat. Softw.* 72(7):1–46

Rizopoulos D, Ghosh P. 2011. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat. Med.* 30(12):1366–80

Rizopoulos D, Papageorgiou G, Afonso PM. 2024. `JMbayes2`: extended joint models for longitudinal and time-to-event data. *R Package*, version 0.5-0. **https://CRAN.R-project.org/package=JMbayes2**

Rondeau V, Gonzalez JR, Mazroui Y, Mauguen A, Diakite A, et al. 2024. `frailtypack`: General frailty models: shared, joint and nested frailty models with prediction. *R Package*, version 3.6.3. **https://CRAN.R-project.org/package=frailtypack**

Roy J, Lin X. 2002. Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. *J. Am. Stat. Assoc.* 97(457):40–52

Rustand D, van Niekerk J, Krainski ET, Rue H. 2024. Joint modeling of multivariate longitudinal and survival outcomes with the R package INLAjoint. arXiv:2402.08335 [stat.ME]

SAS Inst. Inc. 2013. *SAS Software*, version 9.4. SAS Inst. Inc., Cary, NC

Schmidt-Hieber J. 2020. Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Stat.* 48(4):1875–97

Song X, Davidian M, Tsiatis AA. 2002. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 58(4):742–53

Spiegelhalter D, Thomas A, Best N. 2003. *WinBUGS user manual, version 1.4*. Manual, MRC Biostat. Unit, Cambridge, UK

Su YR, Wang JL. 2012. Modeling left-truncated and right-censored survival data with longitudinal covariates. *Ann. Stat.* 40(3):1465

Su YR, Wang JL. 2016. Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data. *Ann. Stat.* 44(3):1298

Sun J. 2006. *The Statistical Analysis of Interval-Censored Failure Time Data*. New York: Springer

Sun J, Chen DG. 2022. *Emerging Topics in Modeling Interval-Censored Survival Data*. New York: Springer

Sun J, Herazo-Maya JD, Molyneaux PL, Maher TM, Kaminski N, Zhao H. 2019. Regularized latent class model for joint analysis of high-dimensional longitudinal biomarkers and a time-to-event outcome. *Biometrics* 75(1):69–77

Sun J, Song PXK. 2001. Statistical analysis of repeated measurements with informative censoring times. *Stat. Med.* 20(1):63–73

Sun L, Song X, Zhou J, Liu L. 2012. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *J. Am. Stat. Assoc.* 107(498):688–700
Tian L, Zucker D, Wei L. 2005. On the Cox model with time-varying regression coefficients. *J. Am. Stat. Assoc.* 100(469):172–83
Tseng YK, Hsieh F, Wang JL. 2005. Joint modelling of accelerated failure time and longitudinal data. *Biometrika* 92(3):587–603
Tseng YK, Shu KN. 2011. Efficient estimation for a semiparametric extended hazards model. *Commun. Stat. Simul. Comput.* 40(2):258–73
Tseng YK, Su Y, Mao M, Wang J. 2015. An extended hazard model with longitudinal covariates. *Biometrika* 102(1):135–50
Tsiatis AA. 2006. *Semiparametric Theory and Missing Data*. New York: Springer
Tsiatis AA, Davidian M. 2001. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 88(2):447–58
Tsiatis AA, Davidian M. 2004. Joint modeling of longitudinal and time-to-event data: an overview. *Stat. Sin.* 14:809–34
Tsiatis AA, Degruttola V, Wulfsohn MS. 1995. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and CD4 counts in patients with AIDS. *J. Am. Stat. Assoc.* 90(429):27–37
Tsodikov A. 2003. Semiparametric models: a generalized self-consistency approach. *J. R. Stat. Soc. Ser. B* 65(3):759–74
Van der Vaart AW. 2000. *Asymptotic Statistics*. Cambridge, UK: Cambridge Univ. Press
Wang JL, Chiou JM, Müller HG. 2016. Functional data analysis. *Annu. Rev. Stat. Appl.* 3:257–95
Wang Y, Taylor JM. 2001. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Am. Stat. Assoc.* 96(455):895–905
Wei LJ. 1992. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat. Med.* 11(14–15):1871–79
Wiegrebe S, Kopper P, Sonabend R, Bischl B, Bender A. 2024. Deep learning for survival analysis: a review. *Artif. Intel. Rev.* 57(3):65
Wu L, Liu W, Yi GY, Huang Y. 2012. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *J. Probability Stat.* 2012:640153
Wulfsohn MS, Tsiatis AA. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics* 53(1):330–39
Xu C. 2014. *Semiparametric analysis of incomplete survival data*. PhD Thesis, Univ. Calif., Davis
Xu C, Baines PD, Wang JL. 2014. Standard error estimation using the EM algorithm for the joint modeling of survival and longitudinal data. *Biostatistics* 15(4):731–44
Xu C, Hadjipantelis PZ, Wang JL. 2020. Semi-parametric joint modeling of survival and longitudinal data: the R package JSM. *J. Stat. Softw.* 93(2):1–29
Xu J, Zeger SL. 2001. Joint analysis of longitudinal data comprising repeated measures and times to events. *J. R. Stat. Soc. Ser. C* 50(3):375–87
Yao F. 2007. Functional principal component analysis for longitudinal and survival data. *Stat. Sin.* 17(3):965–83
Yao F. 2008. Functional approach of flexibly modelling generalized longitudinal data and survival time. *J. Stat. Plan. Inference* 138(4):995–1009
Yao F, Müller HG, Wang JL. 2005. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* 100(470):577–90
Yi F, Tang N, Sun J. 2022. Simultaneous variable selection and estimation for joint models of longitudinal and failure time data with interval censoring. *Biometrics* 78(1):151–64
Yu M, Law NJ, Taylor JM, Sandler HM. 2004. Joint longitudinal-survival-cure models and their application to prostate cancer. *Stat. Sin.* 14(3):835–62
Yu M, Taylor JM, Sandler HM. 2008. Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *J. Am. Stat. Assoc.* 103(481):178–87
Zeng D, Cai J. 2005. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann. Stat.* 35(5):2132–63

Zeng D, Lin D. 2007a. Efficient estimation for the accelerated failure time model. *J. Am. Stat. Assoc.* 102(480):1387–96

Zeng D, Lin D. 2007b. Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Stat. Soc. Ser. B* 69(4):507–64

Zeng L, Zhang J, Chen W, Ding Y. 2024. tdCoxSNN: Time-dependent Cox survival neural network for continuous-time dynamic prediction. arXiv:2307.05881 [stat.ML]

Zhang D, Chen MH, Ibrahim JG, Boye ME, Shen W. 2016. JMFit: a SAS macro for joint models of longitudinal and survival data. *J. Stat. Softw.* 71(3):1–24

Zhong Q, Mueller J, Wang JL. 2022. Deep learning for the partially linear Cox model. *Ann. Stat.* 50(3):1348–75

Zhong Q, Mueller JW, Wang JL. 2021. Deep extended hazard models for survival analysis. In *NIPS'21: Proceedings of the 35th International Conference on Neural Information Processing Systems*, ed. M Ranzato, A Beygelzimer, Y Dauphin, PS Liang, J Wortman Vaughan, pp. 15111–24. Red Hook, NY: Curran

Zhou X, Song X. 2024. Joint analysis of multivariate longitudinal, imaging, and time-to-event data. *J. R. Stat. Soc. Ser. C* 73(4):921–34

Zhou Y, Chen H, Iao SI, Kundu P, Zhou H, et al. 2024. `fdapace`: Functional data analysis and empirical dynamics. *R Package*, version 0.6.0. **https://CRAN.R-project.org/package=fdapace**

Zou H, Xiao L, Zeng D, Luo S. 2023. Multivariate functional mixed model with MRI data: an application to Alzheimer's disease. *Stat. Med.* 42(10):1492–511