ARTICLE TEMPLATE

Comparing crossing hazard rate functions by joint modeling of survival and longitudinal data

Kayoung Park[a] and Peihua Qiu[b]

[a]Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, U.S.A.; [b]Department of Biostatistics, University of Florida, Gainesville, FL 32611, U.S.A.

ARTICLE HISTORY
Compiled August 13, 2019

ABSTRACT
It is one of the important issues in survival analysis to compare two hazard rate functions to evaluate treatment effect. It is quite common that the two hazard rate functions cross each other at one or more unknown time points, representing temporal changes of the treatment effect. In certain applications, besides survival data, we also have related longitudinal data available regarding some time-dependent covariates. In such cases, a joint model that accommodates both types of data can allow us to infer the association between the survival and longitudinal data and to assess the treatment effect better. In this paper, we propose a modeling approach for comparing two crossing hazard rate functions by joint modeling survival and longitudinal data. Maximum likelihood estimation is used in estimating the parameters of the proposed joint model using the EM algorithm. Asymptotic properties of the maximum likelihood estimators are studied. To illustrate the virtues of the proposed method, we compare the performance of the proposed method with several existing methods in a simulation study. Our proposed method is also demonstrated using a real dataset obtained from an HIV clinical trial.

KEYWORDS
censoring data; crossing hazard rates; longitudinal data; proportional hazards regression; survival analysis

1. Introduction

In survival data analysis, we are often interested in comparing two hazard rate functions for evaluating a treatment effect [1–3]. In various applications, the two hazard rate functions would cross each other, reflecting temporal changes of the treatment effect [4–8]. This paper focuses on the efficient comparison of two hazard rate functions in such cases.

In practice, the phenomenon of crossing hazard rates is common. For example, radiation and chemotherapy can usually improve patients' prospects for short-term survival; but they have little or no long-term medical benefits. Surgery may cause high mortality in a short period; but, in the long run, it will often improve patients' long-term health. In these cases, the two related hazard rate functions would often cross each other, reflecting the different treatment effects in different time periods. In the literature, some procedures have been proposed for comparing two hazard rate functions.

CONTACT Kayoung Park. Email: kypark@odu.edu

Early methods, including the logrank, Gehan-Wilcoxon, and Peto-Peto tests, among several others (cf., [3], Chapter 7) do not take into account the crossing phenomenon. It has been well demonstrated that these methods are ineffective in comparing two crossing hazard rates because early differences between the two hazard rates would be canceled out by later differences of opposite signs in their test statistics [7,9]. To overcome this limitation, several authors, including [6] and [10], define their test statistics using absolute or squared differences between the two hazard rates. Another group of methods handles the crossing hazard rates problem by choosing special weights in the weighted logrank test, which change signs before and after a potential crossing point. See, e.g., [8,11–13], and [14] for different weighting schemes. Some other methods employ the modeling approach, by explicitly including the crossing structure of the hazard rates in a model [7,15–17]. For recent development on nonparametric estimation of crossing hazard rates, see [18] and [19], and the references cited therein. Some additional models have been developed for covariate-dependent heteroscedasticity by [20] and [21]. They proposed generalized proportional hazards model for analysis of survival regression data with cross-effects of survival functions. While models in [20] and [21] handle cross-effect caused by observable covariates, models in other papers described above, including the one discussed in the current paper, deal with cross-effect that cannot be explained by any observable covariates.

Although some modeling approaches [7] can accommodate some time-independent baseline information (e.g., a patient's age, gender, etc., at the time when he/she was first included in the study), most existing methods for handling the crossing hazard rates problem, including the ones mentioned above, use survival data only. However, in medical research, it is a common practice to collect both the time-dependent and time-independent data, besides patients' survival times. There are some existing methods in the literature for jointly analyzing the longitudinal and survival data. See, for instance, [22–24], and [25], and the references cited there. Almost all these papers do not specifically handle the case when two hazard rate functions cross.

One motivation example of our current research on comparing two crossing hazard rate functions by joint modeling longitudinal and survival data is the ddI/ddC data that are obtained from an HIV clinical trial [26]. The major goal of this study is to compare the efficacy of two drugs ddI and ddC for HIV-infected patients. See Section 4 for more detailed description. In the data, besides patients' survival times (in months), their longitudinal observations of the number of CD4 cells at different time points are also available. Intuitively, the longitudinal CD4 cell counts are related to patients' survival times; a low number of CD4 cells is prognostic of poor survival outcome. Therefore, such longitudinal information should be accommodated when comparing the hazard rate functions of the two treatment groups of patients. The life-table estimates of the two hazard rate functions are shown in Figure 1, from which it can be seen that they cross each other around the 2-month and 15-month time points. However, the statistical analysis of this data in [27] did not consider the crossing hazard rates phenomenon, although they jointly modeled the longitudinal and survival data. Because the positive differences of the two hazard rates at certain time points and their negative differences at certain other time points could be canceled out, they concluded that the two drugs did not have a significant difference in their efficacy, which may not be appropriate here.

Model selection and diagnostics for the proposed joint model has been discussed in [28]. In this paper, we rigorously justify the proposed model by investigating its identifiability, consistency, and asymptotic normality. We also perform extensive numerical studies to investigate its numerical performance. The rest part of this article is orga-

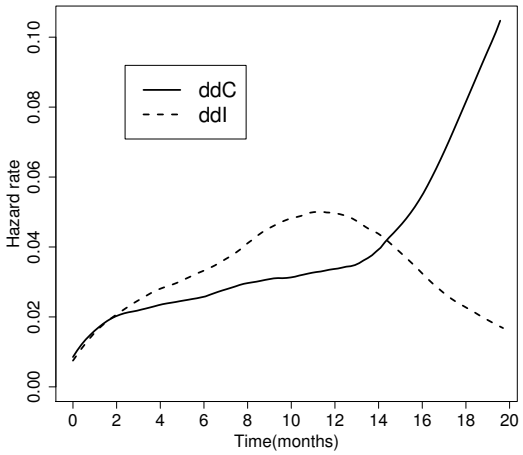

Figure 1. Life-table estimates of the two hazard rate functions of the ddI and ddC patient groups in the HIV clinical trial example.

nized as follows. Section 2 describes the proposed joint model, the model estimation procedure using the EM algorithm, and some statistical properties of the maximum likelihood estimators of the model coefficients. Section 3 presents a simulation study to compare the numerical performance of the proposed method and some existing methods. Section 4 demonstrates the proposed method using the ddI/ddC data mentioned above. Section 5 gives some concluding remarks. Proofs of some theoretical results are given in a supplementary file.

2. Proposed Method

In this section, we describe our proposed method in three parts. Subsection 2.1 describes the proposed joint model, Subsection 2.2 describes the model estimation, and Subsection 2.3 discusses some asymptotic properties of the estimated model.

2.1. Proposed joint modeling procedure

Assume that there are n subjects involved in a study in question. For each individual, we observe both survival and longitudinal data. Let T_i denote the true survival time and C_i denote the censoring time of the ith subject, for $i = 1, \cdots, n$. Then, for the ith subject, we actually observe (O_i, Δ_i), where $O_i = \min(T_i, C_i)$, and $\Delta_i = I\{T_i \leq C_i\}$ which equals 1 if $T_i \leq C_i$ and 0 otherwise. For the longitudinal part of the data, let $Y(t)$ be the longitudinal process. For the ith subject, it is measured at discrete time points $\{t_{ij}, t_{ij} \leq O_i, j = 1, \ldots, n_i\}$. Then, for the ith subject, the observed survival and longitudinal data are $(O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t_i})$, where $\boldsymbol{Y_i} = (Y(t_{i1}), \ldots, Y(t_{in_i}))'$ and $\boldsymbol{t_i} = (t_{i1}, \ldots, t_{in_i})'$. We further assume that the longitudinal process $Y(t)$ follows the following linear mixed effects model:

$$\begin{aligned} Y(t_{ij}) &= M(t_{ij}) + \varepsilon_{ij} \\ &= \boldsymbol{X}(t_{ij})'\boldsymbol{\beta} + \boldsymbol{Z}(t_{ij})'\boldsymbol{b}_i + \varepsilon_{ij}, \end{aligned} \tag{1}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the covariates of the fixed effects and the random effects, respectively, $\boldsymbol{\beta}$ is the fixed effects coefficients, and $\boldsymbol{b}_i$ is the random effects coefficients with $\boldsymbol{b}_i \overset{iid}{\sim} N(0, \Sigma_b)$. We further assume that the random errors $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$ are independent of the random effects coefficients $\boldsymbol{b}_i$.

Note that longitudinal outcome is usually measured only at discrete time points, and thus no longitudinal measurements exist when an event occurs between scheduled follow-up visits. In the literature, various approaches are available to handle missing longitudinal data and missing covariates in survival analysis. See, for instance, [29], [30], [31], and [32]. Moreover, [33] discussed the imputation of longitudinal data in informative dropout setting, and [34] recently proposed a multiple imputation approach to impute missing data in both longitudinal and survival data.

For the survival data, in cases when the two hazard rate functions of the case and control groups cross only once, we assume that the hazard model for the ith subject is defined by

$$\lambda\left(t_{ij}|M(t_{ij}), g_i\right) = \lambda_0(t_{ij}) \exp\left\{\psi M(t_{ij}) + \phi\left(t_{ij} - \gamma\right) g_i\right\}, \tag{2}$$

where $\lambda_0(t_{ij})$ is a baseline hazard, $M(t_{ij})$ is the conditional mean component of the longitudinal response $Y(t_{ij})$ conditional on the random effects coefficients $\boldsymbol{b}_i$ of the ith subject (cf., model (1)), g_i is the group indicator that equals 1 if the ith subject is in the treatment group and 0 otherwise, ψ and ϕ are unknown coefficients, and γ is the unknown crossing point. In model (2), the coefficient ψ measures the association between the survival and longitudinal data. In cases when $\psi = 0$, the information in the longitudinal data would not affect the survival of each patient. In other words, the survival and longitudinal information is independent of each other in such cases. If the ith subject is in the treatment group (i.e., $g_i = 1$), then, from model (2), his/her hazard function is

$$\lambda\left(t_{ij}|M(t_{ij}), g_i = 1\right) = \lambda_0(t_{ij}) \exp\left\{\psi M(t_{ij}) + \phi\left(t_{ij} - \gamma\right)\right\}.$$

Similarly, in cases when the ith subject is in the control group (i.e., $g_i = 0$), the hazard function becomes

$$\lambda\left(t_{ij}|M(t_{ij}), g_i = 0\right) = \lambda_0(t_{ij}) \exp\left\{\psi M(t_{ij})\right\}.$$

Therefore, the ratio of the two hazard functions is

$$\frac{\lambda\left(t_{ij}|M(t_{ij}), g_i = 1\right)}{\lambda\left(t_{ij}|M(t_{ij}), g_i = 0\right)} = \exp\left\{\phi\left(t_{ij} - \gamma\right)\right\}. \tag{3}$$

From (3), it can be seen that the hazard ratio depends on ϕ and γ. In cases when $\phi = 0$, the hazard ratio equals 1 all the time. In such cases, the two hazard functions are identical. In cases when $\phi > 0$, the hazard ratio is less than 1 when $t_{ij} < \gamma$, equal to 1 when $t_{ij} = \gamma$, and greater than 1 when $t_{ij} > \gamma$. In such cases, the two hazard functions cross at γ, the hazard function of the case group is below the hazard function of the control group before the crossing point, and their positions are switched after the crossing point. The magnitude of ϕ controls how different the two hazard functions are. In cases when $\phi < 0$, the two hazard functions also cross at γ, but the hazard function of the case group is above the hazard function of the control group before

the crossing point in such cases. From this description, we can see that the crossing pattern of the two hazard functions depends on both ϕ and γ.

Model (2) can be generalized in several different ways. For instance, in certain applications, besides the longitudinal data described in model (1) and the group indicator g_i, there might be other covariates. Let $\boldsymbol{W_i}$ be the vector of these covariates evaluated for the ith patient. Then, we can consider the following generalized model:

$$\lambda\left(t_{ij}|M(t_{ij}), g_i, \boldsymbol{W_i}\right) = \lambda_0(t_{ij}) \exp\{\psi M(t_{ij}) + \phi\left(t_{ij} - \gamma\right) g_i + \boldsymbol{W_i}'\boldsymbol{\eta}\},$$

where $\boldsymbol{\eta}$ is a vector of coefficients. Model (2) can also be generalized to include two or more crossing points. In cases when two crossing points are present and the covariate vector $\boldsymbol{W_i}$ is also considered, the model becomes

$$\begin{aligned}\lambda\left(t_{ij}|M(t_{ij}), g_i, \boldsymbol{W_i}\right) &= \lambda_0(t_{ij}) \exp\left\{\psi M(t_{ij}) + \phi_1\left(t_{ij} - \gamma_1\right) I\{t_{ij} \le \kappa\} g_i + \right. \\ &\quad \left. \phi_2\left(t_{ij} - \gamma_2\right) I\{t_{ij} > \kappa\} g_i + \boldsymbol{W_i}'\boldsymbol{\eta}\right\},\end{aligned}$$

where γ_1 and γ_2 are the two crossing points, κ is the change point, and ϕ_1 and ϕ_2 are coefficients. It should be pointed out that, theoretically speaking, our model can accommodate as many crossing points as needed. However, in practice, there are only one or two crossing points in most applications. Also, survival data are often censored. In cases when there are multiple crossing points, we usually only consider the first one or two crossing points [18], because the observed survival data usually do not provide enough information for us to estimate the remaining crossing points with a reasonable accuracy.

To take into account the longitudinal data when modeling the survival data, besides the joint modeling approach described above (cf., (1) and (2)), there are some alternative approaches. For instance, in the proposed joint modeling approach, only the component $M(t_{ij})$ of the longitudinal response $Y(t_{ij})$ (cf., expression (1)) is included in the Cox proportional hazards model (2). A simpler approach is to include $Y(t_{ij})$ as a time-dependent covariate in the Cox proportional hazards model (2) directly. Namely, we concentrate on the Cox proportional hazards model alone by considering the model

$$\lambda\left(t_{ij}|Y(t_{ij}), g_i\right) = \lambda_0(t_{ij}) \exp\left\{\psi Y(t_{ij}) + \phi\left(t_{ij} - \gamma\right) g_i\right\}. \tag{4}$$

This approach is called the naive approach in this paper. The frailty modeling approach described below is an extension of the above naive modeling approach. It aims to account for heterogeneity caused by unobserved covariates. It can be considered as a random effect model for describing survival data where the random effect has a multiplicative effect on the baseline hazard function. More specifically, the frailty model corresponding to the naive model (4) can be formulated as

$$\lambda\left(t_{ij}|\nu_i, Y(t_{ij}), g_i\right) = \nu_i \lambda_0(t_{ij}) \exp\left\{\psi Y(t_{ij}) + \phi\left(t_{ij} - \gamma\right) g_i\right\}, \tag{5}$$

where the longitudinal response $Y(t_{ij})$ is included in the survival model directly as a time-dependent covariate, and ν_i is the random effects term (also called the frailty). In the literature, $\{\nu_i, i = 1, \ldots, n\}$ is often assumed to be an i.i.d. sequence of random variables following a given distribution. Among many possible distributions, the most common frailty distribution is the gamma distribution because of its relatively simple computation [35]. Estimation of the frailty model (5) can be accomplished by the R-package frailtypack [36]. By comparing the naive model (4), the frailty model (5), and

the joint models (1) and (2), we can have the following several observations. (i) By using $M(t_{ij})$ in (2), we try to remove the measurement error from the longitudinal response $Y(t_{ij})$ so that the resulting survival model (2) could be more efficient. But the underlying model assumptions in (1) could compromise this benefit when the assumptions are invalid. (ii) Because the frailty model (5) has the frailty ν_i included, which can accommodate certain variability in the survival data, it should be more robust to the model specification, compared to the joint models (1) and (2) and the naive model (4).

2.2. Model estimation

The joint model defined by (1) and (2) can be estimated by maximizing the log likelihood of the observed survival and longitudinal data. Remember that, for the ith patient, the observed data are $(O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t}_i)$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_y, \boldsymbol{\theta}_b, \boldsymbol{\theta}_t)$ be a vector of all parameters in models (1) and (2), where $\boldsymbol{\theta}_y = \left(\boldsymbol{\beta}, \sigma_\varepsilon^2\right)$, $\boldsymbol{\theta}_b$ is a vector of all parameters in Σ_b of the random effects term in (1), and $\boldsymbol{\theta}_t = (\psi, \phi, \gamma, \Lambda_0)$ includes all parameters in model (2) in which $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$ is the cumulative baseline hazard. Then, the likelihood of the observed data can be written as

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} L_i\left(\boldsymbol{\theta} | O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t_i}\right) \\ &= \prod_{i=1}^{n} \left[\int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n_i} f_1\left(Y(t_{ij}) | \boldsymbol{b}_i, \boldsymbol{\theta}_y\right) \right\} f_2\left(\boldsymbol{b}_i | \boldsymbol{\theta}_b\right) f_3\left(O_i, \Delta_i | \boldsymbol{b}_i, \boldsymbol{\theta}_t\right) d\boldsymbol{b}_i \right], \end{aligned} \tag{6}$$

where

$$f_1\left(Y(t_{ij}) | \boldsymbol{b}_i, \boldsymbol{\theta}_y\right) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{ -\frac{\left(Y(t_{ij}) - M(t_{ij})\right)^2}{2\sigma_\epsilon^2} \right\},$$

$$f_2\left(\boldsymbol{b}_i | \boldsymbol{\theta}_b\right) = \frac{1}{(2\pi)^{q/2} \left|\Sigma_b\right|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{b}_i' \Sigma_b^{-1} \boldsymbol{b}_i \right\},$$

and

$$\begin{aligned} f_3\left(O_i, \Delta_i | \boldsymbol{b}_i, \boldsymbol{\theta}_t\right) &= \left[\lambda_0(O_i) \exp\left\{\psi M(O_i) + \phi\left(O_i - \gamma\right) g_i\right\}\right]^{\Delta_i} \\ &\times \exp\left[-\int_0^{O_i} \lambda_0(u) \exp\left\{\psi M(u) + \phi\left(u - \gamma\right) g_i\right\} du \right]. \end{aligned}$$

The maximum of this likelihood over the infinite parameter space $\{\lambda_0(t) \geq 0, \text{ for all } t \in [0, \infty)\}$ does not exist. Thus, the maximum likelihood principle is not applicable here. Nevertheless, we can use the nonparametric likelihood after modifying the likelihood by discretizing the cumulative baseline hazard $\Lambda_0(t)$ as a step function with jumps at the observed follow-up times. More specifically, let us re-order the survival times such that $u_1 < \cdots < u_p$. Then, the baseline hazard is evaluated at each of the ordered survival times, and denoted by $\lambda_0(u_k)$, where $k = 1, \cdots, p$. A

nonparametric likelihood is obtained by changing $f_3\left(O_i, \Delta_i | \boldsymbol{b}_i, \boldsymbol{\theta}_t\right)$ in (6) to

$$\begin{aligned} f_3^{'}\left(O_i, \Delta_i | \boldsymbol{b}_i, \boldsymbol{\theta}_t\right) &= \left[\prod_{k=1}^{p} \lambda_0\left(u_k\right)^{\Delta_i I\{O_i = u_k\}}\right] \times \left[\exp\left\{\psi M(O_i) + \phi\left(O_i - \gamma\right) g_i\right\}\right]^{\Delta_i} \\ &\times \exp\left[-\sum_{k=1}^{p} \lambda_0\left(u_k\right) \exp\left\{\psi M(u_k) + \phi\left(u_k - \gamma\right) g_i\right\} I\{O_i \geq u_k\}\right]. \end{aligned}$$

This nonparametric likelihood approach is well discussed in survival analysis, and similar approaches have been proposed for joint modeling of survival and longitudinal data by several authors, including [22] and [37].

For maximizing the likelihood of a joint model, the expectation-maximization (EM) algorithm is a standard method [22], which has two alternating steps. The E-step computes the expectation of the log-likelihood of the uncensored observations, conditional on all observed data and on the current estimates of the parameters, and the expectation can be evaluated using the numerical integration technique of Gauss-Hermite quadrature [24,38]. The M-step updates the parameter estimates by maximizing the expectation obtained from the E-step. More specifically, the conditional expectation of the log-likelihood of the complete data can be factorized into three parts

$$\begin{aligned} & E\left[\log \prod_{i=1}^{n} \prod_{j=1}^{n_i} f_1\left(Y(t_{ij}) | \boldsymbol{b}_i, \boldsymbol{\theta}_y\right) \Big| O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t}_i, \hat{\boldsymbol{\theta}}\right] + \\ & E\left[\log \prod_{i=1}^{n} f_2\left(\boldsymbol{b}_i | \boldsymbol{\theta}_b\right) \Big| O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t}_i, \hat{\boldsymbol{\theta}}\right] + \\ & E\left[\log \prod_{i=1}^{n} f_3\left(O_i, \Delta_i | \boldsymbol{b}_i, \boldsymbol{\theta}_t\right) \Big| O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t}_i, \hat{\boldsymbol{\theta}}\right]. \end{aligned} \tag{7}$$

Each of the above three parts has the form $E\{h(\boldsymbol{b}_i) | O_i, \Delta_i, \boldsymbol{Y_i}, \boldsymbol{t}_i, \hat{\boldsymbol{\theta}}\}$, where $h(\cdot)$ is a specific function. For simplicity, we denote it as $E_i\{h(\boldsymbol{b}_i)\}$. Then, the equation (7) becomes

$$\begin{aligned} & -\frac{1}{2} \sum_{i=1}^{n} n_i \log\left(2\pi\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} E_i\left\{Y(t_{ij}) - M(t_{ij})\right\}^2 \\ & -\frac{n}{2} \log\left|\left(2\pi\right)^q \Sigma_b\right| - \frac{1}{2} \sum_{i=1}^{n} E_i\left\{\boldsymbol{b}_i' \Sigma_b^{-1} \boldsymbol{b}_i\right\} \\ & + \sum_{i=1}^{n} \sum_{k=1}^{p} \Delta_i I\{O_i = u_k\} \log \lambda_0(u_k) + \sum_{i=1}^{n} \Delta_i \left\{\psi M(O_i) + \phi\left(O_i - \gamma\right) g_i\right\} \\ & - \sum_{i=1}^{n} \sum_{k=1}^{p} \lambda_0\left(u_k\right) E_i\left[\exp\left\{\psi M(u_k) + \phi\left(u_k - \gamma\right) g_i\right\}\right] I\{O_i \geq u_k\}. \end{aligned} \tag{8}$$

Then, the closed-form maximum likelihood estimators of σ_ε^2, Σ_b, and $\lambda_0(u)$ can be

obtained as follows.

$$\begin{aligned}
\hat{\sigma}_{\varepsilon}^{2} &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} E_i\left\{Y(t_{ij}) - M(t_{ij})\right\}^2}{\sum_{i=1}^{n} n_i} \\
\hat{\Sigma}_{b} &= \frac{1}{n}\sum_{i=1}^{n} E_i\left\{\boldsymbol{b}_i\boldsymbol{b}_i'\right\} \\
\hat{\lambda}_0(u) &= \frac{\sum_{i=1}^{n}\Delta_i I\left(O_i = u\right)}{\sum_{i=1}^{n} E_i\left[\exp\left\{\psi M(u) + \phi\left(u-\gamma\right) g_i\right\}\right] I\{O_i \geq u\}}.
\end{aligned}$$

No closed-form solutions exist for the remaining parameters, including $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}, \psi, \phi, \gamma)$, which can be obtained by using the Newton-Raphson algorithm. In the algorithm, the observed score and information evaluated at the current values of the parameters should be computed beforehand in order to update the parameter estimates. If $\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}$ is the current estimate of $\tilde{\boldsymbol{\theta}}$, then an updated estimate would be

$$\hat{\tilde{\boldsymbol{\theta}}}_{(l)} = \hat{\tilde{\boldsymbol{\theta}}}_{(l-1)} + \mathcal{I}^{-1}_{\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}} S_{\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}},$$

where $S_{\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}}$ is the observed score evaluated at $\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}$, which can be obtained by the first derivative of the conditional expectation of the log-likelihood of the complete data (8), and $\mathcal{I}_{\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}}$ is the observed information evaluated at $\hat{\tilde{\boldsymbol{\theta}}}_{(l-1)}$, which is equal to the negative derivative of the above quantity. More details can be found in [22] and [39]. In this iterative algorithm, the initial values of $\boldsymbol{\beta}$ can be set to be their estimates from the linear mixed effects model (1) when the model is estimated alone. Similarly, the initial values of the remaining parameters ψ, ϕ, and γ can be set to be their estimates in the Cox proportional hazards model when the model is estimated in the way that the longitudinal data are incorporated as a time-dependent covariate in the model, which is called the naive model in Section 3.

The inverse observed information matrix can be used as an approximate variance-covariance matrix for the estimates [40], which is used in our paper to estimate approximate standard errors of the estimates. Several other approaches, such as the bootstrap approach [41] and the profile likelihood approach [42] have been proposed in the joint modeling literature. For recent development on standard error estimation for the joint modeling of survival and longitudinal data, see [43], and the references cited therein.

2.3. Theoretical properties

Let $\mathbb{E}_n$ be the expectation with respect to the empirical distribution of the data, and E_0 be the expectation with respect to the underlying true distribution of the data. Further, let $\boldsymbol{\zeta}$ denote $(\boldsymbol{\beta}, \sigma_{\varepsilon}^2, \boldsymbol{\theta}_b, \psi, \phi, \gamma)$. Then, the parameter set $\boldsymbol{\theta}$ can be expressed as $\boldsymbol{\theta} = (\boldsymbol{\zeta}, \Lambda_0)$. Some theoretical results in this section require the following conditions:

(C.1) There are some individuals still at risk at the end of the study, namely, $P\left(O \geq \tau\right) > 0$, where τ denotes the time when the study ends.

(C.2) The parameter set $\boldsymbol{\zeta}$ is an interior point of some known compact set $Z \subset \mathbb{R}^r$, where r is the dimension of $\boldsymbol{\zeta}$. The cumulative hazard rate function Λ_0 is an absolutely continuous function with respect to the Lebesque measure on $[0, \infty)$,

and it is nondecreasing with $\Lambda_0(0) = 0$ and $\Lambda_0(\tau) < \infty$. The set containing all such functions is denoted as M.

(C.3) Let Θ denote the parameter space $Z \times M$. Then, it is assumed that the true values of all parameters in $\boldsymbol{\theta}$ are contained in Θ.

(C.4) In model (1), it is assumed that the matrices with $\{\boldsymbol{X}(t_{ij})', j = 1, \ldots, n_i, i = 1, \ldots, n\}$ and $\{\boldsymbol{Z}(t_{ij})', j = 1, \ldots, n_i, i = 1, \ldots, n\}$ as their rows have full column ranks. Moreover, $\boldsymbol{X}(t)$ and $\boldsymbol{Z}(t)$ are continuously differentiable in $[0, \tau]$. We denote the derivatives of $\boldsymbol{X}(t)$ and $\boldsymbol{Z}(t)$ with respect to t as $\nabla_t \boldsymbol{X}(t)$ and $\nabla_t \boldsymbol{Z}(t)$, respectively. Then, $\max_{[0,\tau]} \|\nabla_t \boldsymbol{X}(t)\|$ and $\max_{[0,\tau]} \|\nabla_t \boldsymbol{Z}(t)\|$ are finite, where $\|\cdot\|$ is the Euclidean norm.

(C.5) From model (2), $\lambda(t|M(t), g)$ can be re-expressed as

$$\begin{aligned} & \lambda_0(t) \exp\{\psi \left(\boldsymbol{X}(t)'\boldsymbol{\beta} + \boldsymbol{Z}(t)'\boldsymbol{b}\right) + \phi\,(t - \gamma)\, g\} \\ = \; & \lambda_0(t) \exp\{\psi \boldsymbol{Z}(t)'\boldsymbol{b} + \boldsymbol{X}(t)'\psi\boldsymbol{\beta} + \phi t g - \phi\gamma g\} \\ = \; & \lambda_0(t) \exp\{\psi \boldsymbol{Z}(t)'\boldsymbol{b} + \boldsymbol{W}(t)'\boldsymbol{\eta}\}, \end{aligned}$$

where $\boldsymbol{W}(t)' = [\boldsymbol{X}(t)', tg, g]$, $\boldsymbol{\eta}' = (\boldsymbol{\eta}_1', \eta_2, \eta_3)$, $\boldsymbol{\eta}_1 = \psi\boldsymbol{\beta}, \eta_2 = \phi$, and $\eta_3 = -\phi\gamma$. We assume that if there exist constant vectors $\boldsymbol{V_1}$ and $\boldsymbol{V_2}$ such that $\boldsymbol{Z}(t)'\boldsymbol{V_1} = 0$ and $\boldsymbol{W}(t)'\boldsymbol{V_2} = m(t)$ for all $t \in [0, \tau]$ where $m(t)$ is a non-random function of t, then $\boldsymbol{V_1}, \boldsymbol{V_2}$ and $m(t)$ are all 0. Furthermore, $\boldsymbol{W}(t)$ is continuously differentiable in $[0, \tau]$, and $\max_{[0,\tau]} \|\nabla_t \boldsymbol{W}(t)\| < \infty$.

For the maximum likelihood estimation problem discussed in the previous subsection, we have the following results.

Lemma 2.1 (Existence)**.** *Under the assumptions (C.1)–(C.5), the maximum likelihood estimators $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\zeta}}, \hat{\Lambda}_0)$ for maximizing the likelihood function defined in* (6) *exist, and they satisfy the equation*

$$\hat{\Lambda}_0(t) = \int_0^t \frac{dH_n(u)}{W_n(u; \hat{\boldsymbol{\theta}})},$$

where $H_n(u) = \frac{1}{n}\sum_{i=1}^n \Delta_i I\,(O_i \le u)$ and $W_n(u; \hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^n E_i\big[\exp\{\hat{\psi}M(u) + \hat{\phi}(u - \hat{\gamma})g_i\}\big] I\{O_i \ge u\}$.

Lemma 2.2 (Identifiability)**.** *Under the assumptions (C.1)–(C.5), all parameters in the maximum likelihood estimation problem discussed in Subsection 2.2 are identifiable. That is, if there are $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ in Θ such that $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}^*)$ almost surely, then we have $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.*

Lemma 1 says that the maximum likelihood estimators of $\boldsymbol{\theta}$ exist, and Lemma 2 confirms that they are unique under some regularity conditions. Therefore, the maximum likelihood estimators $\hat{\boldsymbol{\theta}}$ are well defined. The next theorem builds their strong consistency.

Theorem 2.3 (Consistency)**.** *Under the assumptions (C.1) - (C.5), the maximum likelihood estimators $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\zeta}}, \hat{\Lambda}_0)$ of $\boldsymbol{\theta} = (\boldsymbol{\zeta}, \Lambda_0)$ have the properties that $\|\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\|$ and $\|\hat{\Lambda}_0 - \Lambda_0\|_\infty$ both converge almost surely to zero as $n \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_\infty$ denotes the supremum norm on $[0, \tau]$.*

Next, we provide a result about the asymptotic normality of $\boldsymbol{\theta}$. To this end, some

notation is required. Let $H = \{h = (\boldsymbol{h}_1, h_2) \,|\|\boldsymbol{h}_1\| \leq 1, \|h_2\|_v \leq 1\}$, where $\|\cdot\|$ is the Euclidean norm and $\|h_2\|_v$ is defined to be the absolute value of $h_2(0)$ plus the total variation of h_2 on the $[0, \tau]$. We define $\boldsymbol{\theta}(h) = \boldsymbol{\zeta}'\boldsymbol{h}_1 + \int_0^\tau h_2(u)d\Lambda(u)$. Then, the parameter vector $\boldsymbol{\theta}$ can be considered as a functional on H, and the parameter space Θ can be considered as a subset of $l^\infty(H)$, which is the space of bounded real-valued functions on H. We further define the supremum norm as $\|U\|_\infty = \sup_{h\in H} |U(h)|$ on $l^\infty(H)$. Then, we have the following result.

Theorem 2.4 (Asymptotic Normality)**.** *Under the assumptions (C.1) - (C.5),* $\sqrt{n}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}, \hat{\Lambda}_0 - \Lambda_0)$ *weakly converges to a Gaussian process G in* $l^\infty(H)$ *with mean zero and covariance process*

$$cov\left[G(\boldsymbol{g}), G(\boldsymbol{g}^*)\right] = \omega_1^{-1}(\boldsymbol{g}^*)'\boldsymbol{g}_1 + \int_0^\tau \omega_2^{-1}(\boldsymbol{g}^*)g_2(u)d\Lambda_0(u),$$

where $\boldsymbol{g} = (\boldsymbol{g}_1, g_2)$ *and* $\boldsymbol{g}^*$ *are two elements of* H*, and* $\omega_1(g^*)$ *and* $\omega_2(g^*)$ *are two continuously invertible linear operator on* H *that are defined in the proof of this theorem.*

3. Simulation Study

In this section, we present some simulation results to compare the performance of the proposed joint modeling approach with some alternative approaches. The alternative approaches considered here include the joint modeling method without considering any crossing points [23,27], the modeling approach by [7] for comparing two crossing hazard rate functions, the frailty modeling approach, and the naive modeling approach described in Section 2.1. In model (2) of our proposed joint modeling approach, only one crossing point is assumed present. In this section, we also consider the case when two crossing points are assumed present. In such a case, model (2) becomes

$$\begin{aligned}\lambda\left(t_{ij}|M(t_{ij}), g_i\right) &= \lambda_0\left(t_{ij}\right)\exp\left\{\psi M(t_{ij}) + \phi_1\left(t_{ij} - \gamma_1\right)I\{t_{ij} \leq \kappa\}g_i + \right. \\ &\quad \left.\phi_2\left(t_{ij} - \gamma_2\right)I\{t_{ij} > \kappa\}g_i\right\},\end{aligned}$$

where γ_1 and γ_2 are the two crossing time points, κ is the change point, and ψ, ϕ_1, and ϕ_2 are parameters. To distinguish the two cases, our proposed joint modeling approach with one crossing point assumed present is denoted as NEW1, and the one with two crossing points assumed present is denoted as NEW2. Similarly, in both the frailty modeling approach and the naive approach, we also consider cases with either one or two crossing points present, and the corresponding models are denoted as FRAILTY1, FRAILTY2, NAIVE1 and NAIVE2, respectively. In the frailty model (5), we assume that $\nu_i \sim \Gamma\left(\frac{1}{\rho}, \frac{1}{\rho}\right)$, $E(\nu_i) = 1$, $Var(\nu_i) = \rho$, and the value of ρ is estimated by MLE. In the conventional joint modeling approach without considering any crossing points, which is noted as OLD in this section, the longitudinal data are still modeled by (1), but the survival data are modeled by

$$\lambda\left(t_{ij}|M(t_{ij}), g_i\right) = \lambda_0\left(t_{ij}\right)\exp\left\{\psi M(t_{ij}) + \phi t_{ij}g_i\right\}.$$

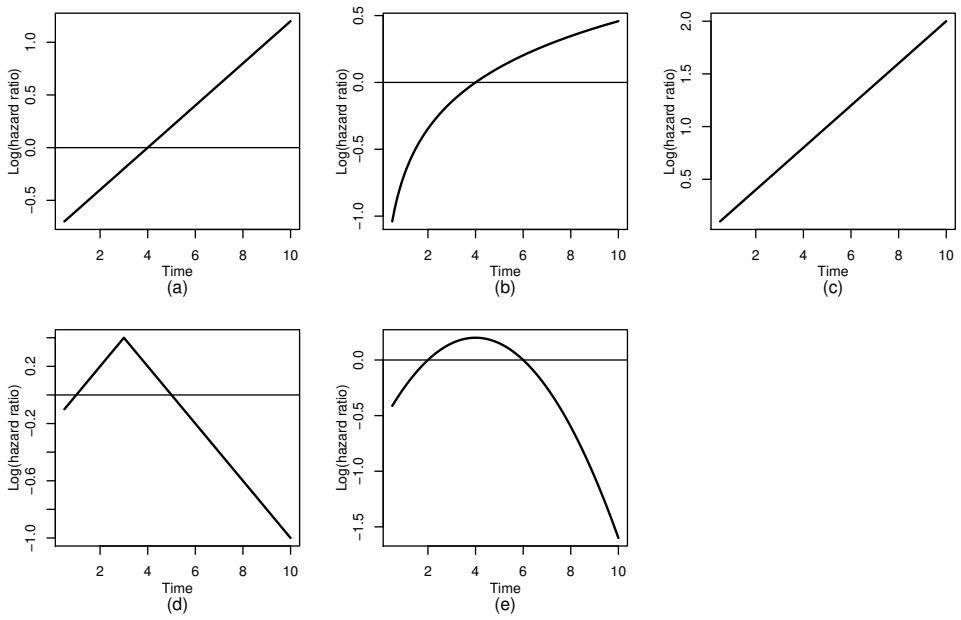

Figure 2. Log ratios of two hazard rate functions in five different cases considered in the simulation study.

In the modeling approach by [7], denoted as LQS here, the survival data are modeled by

$$\lambda\left(t_{ij}|g_i\right) = \lambda_0\left(t_{ij}\right)\exp\left\{\phi\left(BC_\alpha(t_{ij}) - BC_\alpha(\gamma)\right)g_i\right\},$$

where

$$BC_\alpha(t_{ij}) = \begin{cases} t_{ij}^\alpha, & \text{if } \alpha \neq 0 \\ \log(t_{ij}), & \text{if } \alpha = 0, \end{cases}$$

ϕ and α are two parameters, and γ is the crossing point.

In the simulation study, we consider five different patterns of the hazard rate functions of the treatment and control groups. The log ratios of the hazard rate functions in the five different cases are shown in Figure 2 (a)-(e), respectively. In case (a), the log ratio of the hazard rate functions is linear over time, and the two hazard rate functions have one crossing point at time=4. In case (b), the two hazard rate functions still have one crossing point at time=4, but their log ratio is non-linear. In case (c), the log ratio of the hazard rate functions is linear, but the hazard rate functions do not have any crossing point. Case (d) considers two crossing points at time=1 and time=5, and the log ratio of the hazard rate functions is piecewise linear. Case (e) also considers two crossing points at time=2 and time=6, but the log ratio of the hazard rate functions is not piecewise linear. More specifically, the hazard rate functions in these five cases

are given by the following five expressions, respectively:

$$\begin{aligned}
\lambda(t) &= \lambda_0(t)\exp\{0.5M(t)+0.2(t-4)g\},\\
\lambda(t) &= \lambda_0(t)\exp\{0.5M(t)+0.5(\log(t)-\log(4))g\},\\
\lambda(t) &= \lambda_0(t)\exp\{\psi M(t)+0.2tg\},\\
\lambda(t) &= \lambda_0(t)\exp\{0.5M(t)+0.2(t-1)1\{t\le 3\}g-0.2(t-5)1\{t>3\}g\},\\
\lambda(t) &= \lambda_0(t)\exp\left\{0.5M(t)+\left[-0.05(t-4)^2+0.2\right]g\right\},
\end{aligned}$$

where $g = 0$ or 1 is a group indicator, and $M(t)$ is the conditional mean component of the longitudinal response $Y(t)$ (cf., model (1)). In each case, the two hazard rate functions can be specified from the related expression of $\lambda(t)$ by replacing g by 0 and 1, respectively. In all cases considered, a constant baseline hazard function is used for λ_0. The true longitudinal covariate is modeled by

$$M(t) = \beta_0 + (\beta_1 + b_i)\,t,$$

where $\beta_0 = 1$, $\beta_1 = 0.5$, and b_i is the random effect coefficient generated from a zero-mean normal distribution with variance $\sigma_b^2 = 0.1$. The observed longitudinal data (i.e., observations of $Y(t)$) are generated from $N(M(t), \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = 0.5$. In each setting, both sample sizes of the treatment and control groups are chosen to be $n_1 = 100$ or 200. The censoring time for each subject is generated from the uniform distribution on the interval $[1, T]$, where T is adjusted to reach a pre-specified censoring rate. We choose the left end of the interval to be 1 to allow a minimum follow-up time of one time unit. In the simulation, we consider the two censoring rates of 20% and 50%.

First, we compare the performance of NEW1, NAIVE1, and FRAILTY1 in the first three cases when there is only one crossing point or when there is no crossing point. The MSE values of their estimated parameters based on 200 replicated simulations are presented in Table 1. In case (b), because the true log hazard ratio is not linear over t and thus the true value of ϕ is unavailable, the MSE values of ϕ cannot be computed. Similarly, in case (c) when there is no crossing point in the survival model, the MSE values of γ cannot be computed either. From the table, it can be seen that, in case (a) when the linear Cox proportional hazards model (2) with one crossing point is valid, the method NEW1 is more efficient for estimating the coefficients in that model, compared to the method NAIVE1. This is especially true for estimating ψ which is the coefficient for connecting the longitudinal data with the survival model. Compared to FRAILTY1, NEW1 is still much better in estimating ψ, and marginally better in estimating ϕ (except in the case when $n_1 = 100$ and $CR = 50\%$). But, FRAILTY1 is better in estimating the crossing point γ. In case (b) when the linear log-hazard-ratio assumption is invalid in all three models, we can see that FRAILTY1 performs the best, which confirms our observation made at the end of Section 2.1 that it is quite robust to its model specification compared to the other two models. In case (c) when the linear Cox proportional hazards model is valid but it contains no crossing points (i.e., $\gamma \le 0$ in model (2)), the method NEW1 performs the best uniformly in all cases considered.

Next, we compare the performance of NEW2, NAIVE2, and FRAILTY2 in cases (d) and (e) when there are two crossing points in the true survival model. The MSE values of their estimated parameters based on 200 replicated simulations are presented in Table 2. In case (e), because the true values of ϕ_1 and ϕ_2 are unavailable, their MSE

Table 1. MSE values of the estimated parameters of the methods NEW1, NAIVE1, and FRAILTY1 in the first three cases when there is only one crossing point or when there is no crossing point. In the table, n_1 denotes the number of subjects in the case (or control) group, and CR denotes the censoring rate.

Case	n_1	CR	Method	ψ	ϕ	γ	β_0	β_1	σ_ε^2	σ_b^2
(a)	100	20%	NEW1	0.0750	0.0086	1.7837	0.0005	0.0003	0.0001	0.0001
			NAIVE1	0.1643	0.0099	1.8693				
			FRAILTY1	0.1567	0.0095	1.0318				
		50%	NEW1	0.1471	0.0335	2.1464	0.0005	0.0003	0.0001	0.0002
			NAIVE1	0.1803	0.0353	2.1392				
			FRAILTY1	0.1852	0.0239	1.6920				
	200	20%	NEW1	0.0538	0.0037	0.5528	0.0003	0.0001	0.0001	0.0000
			NAIVE1	0.1332	0.0041	0.7360				
			FRAILTY1	0.1375	0.0041	0.6485				
		50%	NEW1	0.1198	0.0114	1.6661	0.0003	0.0002	0.0001	0.0001
			NAIVE1	0.1903	0.0133	2.0055				
			FRAILTY1	0.1897	0.0123	1.1175				
(b)	100	20%	NEW1	0.2946		3.2598	0.0004	0.0001	0.0001	0.0000
			NAIVE1	0.2845		2.8488				
			FRAILTY1	0.1534		1.5055				
		50%	NEW1	0.3161		3.3015	0.0004	0.0001	0.0001	0.0001
			NAIVE1	0.3002		3.1890				
			FRAILTY1	0.1563		1.9263				
	200	20%	NEW1	0.2779		1.6740	0.0002	0.0000	0.0000	0.0000
			NAIVE1	0.2619		1.3726				
			FRAILTY1	0.1483		1.0089				
		50%	NEW1	0.2789		2.0686	0.0002	0.0001	0.0000	0.0000
			NAIVE1	0.2700		2.0542				
			FRAILTY1	0.1600		1.0717				
(c)	100	20%	NEW1	0.0859	0.0151		0.0007	0.0005	0.0001	0.0002
			NAIVE1	0.1837	0.0228					
			FRAILTY1	0.1903	0.0221					
		50%	NEW1	0.1831	0.0634		0.0008	0.0005	0.0002	0.0004
			NAIVE1	0.2347	0.0837					
			FRAILTY1	0.2050	0.0736					
	200	20%	NEW1	0.0758	0.0087		0.0003	0.0002	0.0001	0.0001
			NAIVE1	0.1758	0.0111					
			FRAILTY1	0.1738	0.0100					
		50%	NEW1	0.1596	0.0271		0.0004	0.0003	0.0001	0.0002
			NAIVE1	0.2513	0.0377					
			FRAILTY1	0.2224	0.0357					

Table 2. MSE values of the estimated parameters of the methods NEW2, NAIVE2, and FRAILTY2 in the cases (d) and (e) when there are two crossing points in the survival model. In the table, n_1 denotes the number of subjects in the case (or control) group, and CR denotes the censoring rate.

Case	n_1	CR	Method	ψ	κ	ϕ_1	γ_1	ϕ_2	γ_2	β_0	β_1	σ_ε^2	σ_b^2
(d)	100	20%	NEW2	0.0967	0.0356	0.0401	0.7007	0.0380	1.9027	0.0006	0.0003	0.0001	0.0001
			NAIVE2	0.1554	0.0335	0.0336	0.5963	0.0468	2.4047				
			FRAILTY2	0.1529	0.0319	0.0313	0.9275	0.0325	1.9834				
		50%	NEW2	0.1501	0.0354	0.0419	0.7622	0.0920	2.4763	0.0006	0.0004	0.0001	0.0002
			NAIVE2	0.1980	0.0332	0.0454	0.4867	0.1066	3.0328				
			FRAILTY2	0.2128	0.0314	0.0445	0.6774	0.0855	4.3216				
	200	20%	NEW2	0.0708	0.0357	0.0196	0.3542	0.0167	1.2533	0.0003	0.0002	0.0001	0.0001
			NAIVE2	0.1680	0.0325	0.0161	0.3429	0.0205	1.1788				
			FRAILTY2	0.1684	0.0320	0.0143	0.6292	0.0110	1.7654				
		50%	NEW2	0.1118	0.0345	0.0225	0.5336	0.0604	2.5771	0.0003	0.0002	0.0001	0.0001
			NAIVE2	0.1843	0.0326	0.0261	0.4659	0.0619	2.9668				
			FRAILTY2	0.2027	0.0317	0.0261	0.5538	0.0762	3.3739				
(e)	100	20%	NEW2	0.2966	0.1376		1.1704		1.6083	0.0004	0.0001	0.0001	0.0000
			NAIVE2	0.2757	0.1324		1.0274		1.6632				
			FRAILTY2	0.1255	0.1297		1.9155		3.1192				
		50%	NEW2	0.3287	0.1413		0.8106		1.7911	0.0003	0.0001	0.0001	0.0001
			NAIVE2	0.2924	0.1307		0.7115		2.1246				
			FRAILTY2	0.1954	0.1333		1.0489		2.3273				
	200	20%	NEW2	0.2812	0.1452		0.7451		1.4151	0.0002	0.0000	0.000	0.0000
			NAIVE2	0.2624	0.1364		0.6981		1.4724				
			FRAILTY2	0.1130	0.1353		1.5640		2.5908				
		50%	NEW2	0.2908	0.1358		0.6854		1.7467	0.0002	0.0001	0.0001	0.0000
			NAIVE2	0.2719	0.1353		0.7235		1.8478				
			FRAILTY2	0.1843	0.1215		0.6384		2.5220				

values cannot be computed. From the table, it can be seen that, in case (d) when the linear Cox proportional hazards model with two crossing points is valid, NEW2 is much better than the other two methods for estimating ψ, it is close to the best in estimating ϕ_2 and γ_2, and the three methods are comparable for estimating κ, ϕ_1, and γ_1. In case (e) when the linear Cox proportional hazards model is invalid, FRAILTY2 is much more efficient in estimating ψ compared to the other two methods, which is consistent with the results found in Table 1. Its performance is close to the best in estimating κ. But, it does not do a good job in estimating the two crossing points γ_1 and γ_2.

Finally, for each of the eight methods NEW1, NAIVE1, FRAILTY1, LQS, NEW2, NAIVE2, FRAILTY2, and OLD, we compute the mean squared residuals of the estimated log hazard ratio in each of the five cases considered. Then, the mean squared residuals is averaged over 200 replicated simulations. The averaged value of the mean squared residuals is presented in Table 3. Because some methods, including NAIVE1, FRAILTY1, LQS, NAIVE2 and FRAILTY2 do not specifically model the longitudinal data, the corresponding results for the longitudinal data are not presented. From the table, it can be seen that the proposed method NEW1 outperforms all other methods in case (a) except the method FRAILTY1 which is slightly better in the case when $n_1 = 100$ and $CR = 50\%$. In case (b) when the linear Cox proportional hazards model is invalid, FRAILTY1 performs the best among all methods, NEW1 and NAIVE1 perform similarly well, but the methods LQS, NEW2, NAIVE2, FRAILTY2, and OLD perform much worse. In case (c) when there is no crossing point, the conventional joint modeling approach OLD performs the best, as expected, and our proposed method NEW1 performs the second best. In case (d) when the linear Cox proportional haz-

Table 3. Averaged mean squared residuals of the estimated log hazard ratio by the methods NEW1, NAIVE1, FRAILTY1, LQS, NEW2, NAIVE2, FRAILTY2 and OLD in all five cases considered, based on 200 replicated simulations.

Case	n_1	CR	Method							
			NEW1	NAIVE1	FRAILTY1	LQS	NEW2	NAIVE2	FRAILTY2	OLD
(a)	100	20%	0.0557	0.0652	0.0632	0.1217	0.1995	0.2130	0.2186	0.2064
		50%	0.1023	0.1228	0.0911	0.1244	0.3092	0.3550	0.1516	0.2724
	200	20%	0.0258	0.0299	0.0295	0.0657	0.0960	0.0404	0.0409	0.1874
		50%	0.0446	0.0561	0.0534	0.0943	0.1769	0.0930	0.0739	0.2560
(b)	100	20%	0.0934	0.0952	0.0887	0.2328	0.3340	0.2662	2.7131	0.2019
		50%	0.1484	0.1542	0.1312	0.3087	0.6543	0.6246	5.2855	0.3136
	200	20%	0.0708	0.0745	0.0673	0.2044	0.1735	0.1664	0.1450	0.1827
		50%	0.0955	0.1000	0.0939	0.2276	0.1838	0.4687	1.8622	0.2731
(c)	100	20%	0.0767	0.0969	0.0936	0.1413	0.2434	0.2095	0.1015	0.0374
		50%	0.1369	0.1867	0.1509	0.6488	0.3416	0.2450	0.1677	0.0535
	200	20%	0.0385	0.0487	0.0433	0.0499	0.1321	0.1364	0.0764	0.0274
		50%	0.0660	0.0882	0.0754	0.2939	0.2133	0.1684	0.0985	0.0286
(d)	100	20%	0.1078	0.1118	0.1199	0.4999	0.0787	0.0837	0.1035	0.0844
		50%	0.1784	0.1928	0.1056	1.0871	0.0986	0.1121	0.2174	0.1394
	200	20%	0.0766	0.0778	0.0807	0.3685	0.0408	0.0432	0.0428	0.0678
		50%	0.1026	0.1148	0.0563	0.5004	0.0625	0.0676	0.0605	0.0950
(e)	100	20%	0.2068	0.1722	0.2330	0.2652	0.1238	0.1169	0.2221	0.1788
		50%	0.5015	0.3895	0.1497	0.5838	0.2244	0.2602	0.1486	0.2882
	200	20%	0.1837	0.1540	0.2042	0.2359	0.0835	0.0789	0.1755	0.1633
		50%	0.3631	0.1739	0.1006	0.4196	0.1580	0.1758	0.0822	0.2291

ards model with two crossing points is invalid, the method NEW2 performs the best among all methods except FRAILTY1 and FRAILTY2 which perform slightly better than NEW2 when $n_1 = 200$ and $CR = 50\%$. In case (e) when the linear Cox proportional model is invalid and there are two crossing points in the model, NAIVE2 performs the best when $CR = 20\%$, NEW2 performs the second best in such a case, and all other methods do not perform well. In case (e) when $CR = 50\%$, FRAILTY2 performs the best, FRAILTY1 is slightly worse, and NEW2 performs the third best. From this example, we can conclude that (i) when the linear Cox proportional hazards model is valid, our proposed method NEW performs well, (ii) when the linear Cox proportional hazards model is invalid, the frailty method and sometimes the naive method could be more reliable, and (iii) in the latter case, the performance of NEW is still reasonably good.

In the case when $n_1 = 100$ and the censoring rate is 20%, the estimated log hazard ratios by the eight methods in the five cases considered are shown in Figure 3. From the plots in the figure, it can be seen that, in cases (a) and (b), the methods NEW1, NAIVE1, FRAILTY1, NEW2, NAIVE2, and FRAILTY2 all estimate the true log hazard ratio reasonably well, the method LQS can catch the overall trend but its estimate is quite far away from the true hazard ratio, and the method OLD cannot provide a reasonable estimate at all. In case (c) when there is no crossing point and the linear Cox proportional hazards model is valid, the conventional method OLD performs the best, the estimates by the method LQS are quite far away from the true hazard ratios, and the other six methods all perform reasonable well. In case (d), the methods NEW2, NAIVE2, and FRAILTY2 perform well, and all the other five methods do not perform well. In case (e), the methods NEW2 and NAIVE2 perform well, while the method FRAILTY2 catches the trend but its estimates do not look

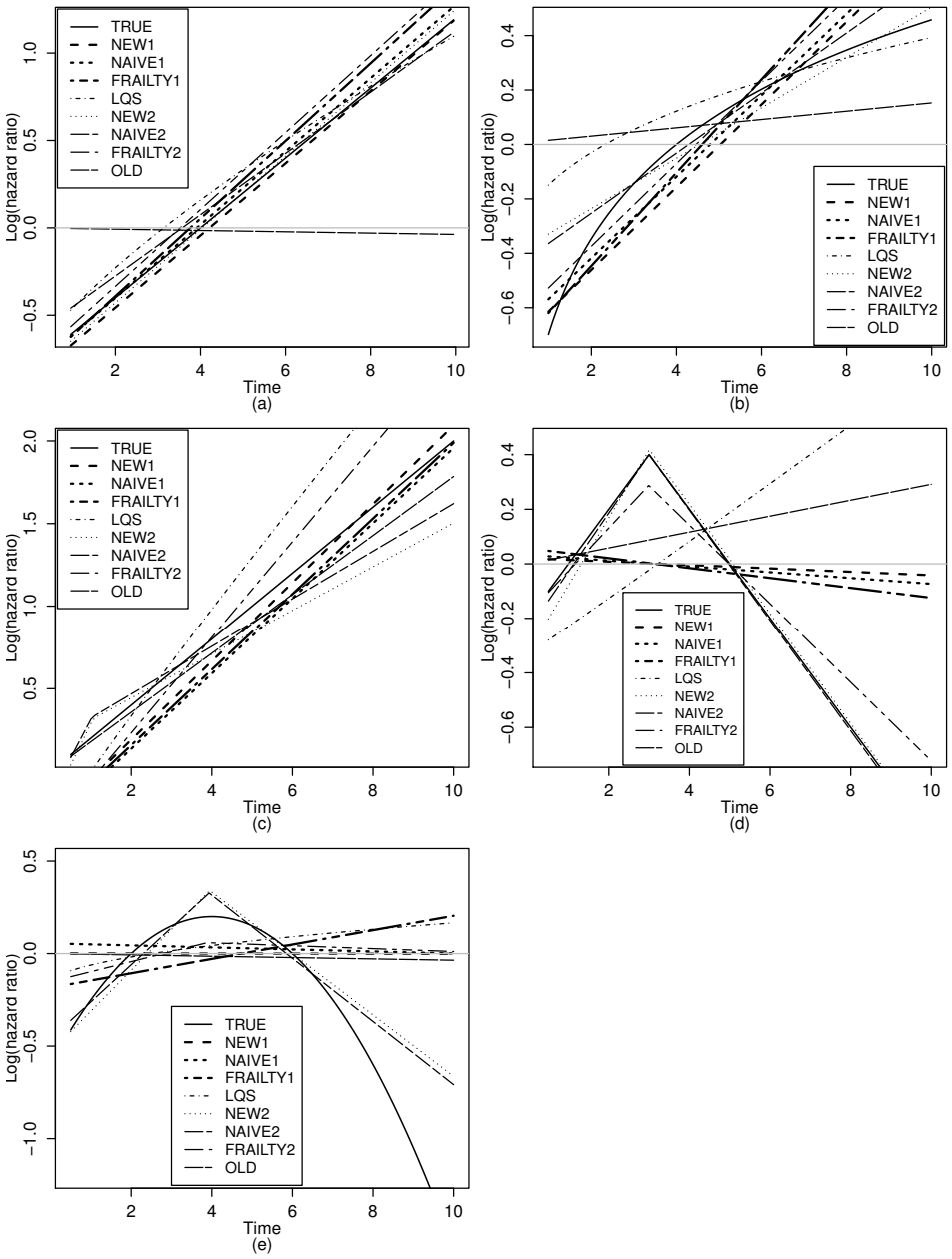

Figure 3. Estimates of the log hazard ratios in the five different cases considered in the simulation study when $n_1 = 100$ and the censoring rate is 20%.

reasonable. As we can expect, all the other five methods do not provide reasonable estimates in this case.

4. A Real-Data Example

In this section, we apply our proposed joint modeling approach to the HIV clinical trial example mentioned in Section 1. The goal of the study was to compare the efficacy and safety of two drug treatments named didanosine (ddI) and zalcitabine (ddC) for HIV-infected patients who had failed or were intolerant of the zidovudine (AZT) therapy. More details about how the study was conducted can be found in [26] and [44]. The ddI/ddC data in this example were previously analyzed by [27] and [45], using joint modeling approaches. In the data, there were totally 467 eligible HIV-infected patients included, 230 of them were randomly assigned to receive the ddI treatment, and the remaining 237 patients were assigned to receive the ddC treatment. For each patient, his/her survival time was defined to be the period from the start of the study to the time when the patient developed AIDS symptoms or when he/she was dead. During the course of the study, CD4 cell counts were recorded for each patient at the 0 (i.e., the time when the study started), 2, 6, 12, and 18 months. For the survival data, the number of censored observations is 130 in the ddI group and 149 in the ddC group. Besides the survival times and the longitudinal CD4 cell counts, there are also four time-independent covariates, described below using the same notations and assignments of numerical values as those in [27].

- Drug=1 if the patient was in the ddI group, and Drug=0 otherwise.
- Gender=1 if the patient was male, and Gender=-1 otherwise.
- PrevOI=1 if the patient received AIDS diagnosis before the study, and PrevOI=0 otherwise.
- Stratum=1 if the patient failed the AZT therapy before the study, and Stratum=-1 if the patient was found intolerant of the AZT therapy before the study.

[27] has justified that the following mixed-effects model is appropriate for describing the longitudinal data:

$$\begin{aligned} Y\left(t_{ij}\right) &= \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}\mathrm{Drug}_i + \beta_3 \mathrm{Gender}_i \\ &+ \beta_4 \mathrm{PrevOI}_i + \beta_5 \mathrm{Stratum}_i + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}, \end{aligned} \tag{9}$$

where Y is the squared root of the observed CD4 cell count, $\mathbf{b}_i = (b_{0i}, b_{1i})'$ are the random effects coefficients with $\mathbf{b}_i \overset{iid}{\sim} N(0, \Sigma_b)$, $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ are coefficients, and ε_{ij} are i.i.d. random errors.

From Figure 1, the two estimated hazard rate functions of the two treatment groups have two possible crossing points. So, a natural survival model is the following Cox proportional hazards model with two crossing points, denoted as NEW2:

$$\begin{aligned} \lambda(t_{ij}) &= \lambda_0(t_{ij}) \exp\left\{\psi M(t_{ij}) + \phi_1 \left(t_{ij} - \gamma_1\right) I\{t_{ij} \le \kappa\} \mathrm{Drug}_i\right\} \\ &\times \exp\left\{\phi_2 \left(t_{ij} - \gamma_2\right) I\{t_{ij} > \kappa\} \mathrm{Drug}_i\right\} \\ &\times \exp\left\{\eta_1 \mathrm{Gender}_i + \eta_2 \mathrm{PrevOI}_i + \eta_3 \mathrm{Stratum}_i\right\}, \end{aligned} \tag{10}$$

where $M(t_{ij})$ is the right-hand side of (9) without the random error term, γ_1 and γ_2 are

Table 4. Point estimates and the 95% CIs of the parameters in models (9) and (10) by our proposed joint modeling approach NEW2 , the naive approach NAIVE2, and the frailty modeling approach FRAILTY2, respectively.

		NEW2		NAIVE2		FRAILTY2	
	Parameter	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Model (10)	ψ	-0.235	$(-0.289, -0.182)$	-0.165	$(-0.215, -0.115)$	-0.162	$(-0.212, -0.112)$
	ϕ_1	0.076	$(-0.016, 0.168)$	-0.130	$(-0.593, 0.334)$	-0.125	$(-0.366, 0.116)$
	γ_1	1.230	$(-5.153, 7.613)$	7.524	$(3.608, 11.441)$	7.778	$(4.313, 11.243)$
	ϕ_2	-0.217	$(-0.383, -0.051)$	-1.569	$(-4.199, 1.061)$	-2.097	$(-3.155, -1.039)$
	γ_2	15.079	$(12.535, 17.624)$	11.172	$(10.186, 12.157)$	11.345	$(10.978, 11.711)$
	η_1	-0.200	$(-0.417, 0.016)$	-0.161	$(-0.401, 0.078)$	-0.161	$(-0.403, 0.081)$
	η_2	0.300	$(0.098, 0.502)$	0.377	$(0.148, 0.605)$	0.389	$(0.160, 0.619)$
	η_3	0.062	$(-0.100, 0.225)$	0.057	$(-0.101, 0.216)$	0.058	$(-0.101, 0.216)$
Model (9)	β_0	8.775	$(8.415, 9.134)$				
	β_1	-0.184	$(-0.225, -0.143)$				
	β_2	0.028	$(-0.028, 0.085)$				
	β_3	-0.142	$(-0.484, 0.201)$				
	β_4	-2.565	$(-2.868, -2.262)$				
	β_5	0.590	$(0.264, 0.916)$				

the two crossing points, and $(\psi, \phi_1, \phi_2, \kappa, \eta_1, \eta_2, \eta_3)$ are all coefficients. In model (10), if $M(t_{ij})$ is replaced by $Y(t_{ij})$ and the resulting model is estimated alone, then the corresponding approach is the naive approach denoted as NAIVE2 here. The frailty model with two crossing points, denoted as FRAILTY2, is also considered, which is obtained by replacing $M(t_{ij})$ with $Y(t_{ij})$ in (10) and then multiplying by the frailty ν_i on the right-hand-side. As in the simulation examples in Section 3, we assume that $\nu_i \sim \Gamma\left(\frac{1}{\rho}, \frac{1}{\rho}\right)$, $E(\nu_i) = 1$, $Var(\nu_i) = \rho$, and the value of ρ is estimated by MLE.

The estimated parameters and the corresponding 95% confidence intervals (CIs) of the methods NEW2, NAIVE2, and FRAILTY2 are presented in Table 4. From the table, it can be seen that, for the method NEW2, the parameters $\beta_0, \beta_1, \beta_4$, and β_5 in the longitudinal model (9) are all significantly different from 0 at the significance level of 0.05, and the parameters β_2 and β_3 are not significantly different from 0. These results along with the values of the parameter estimates show that 1) the CD4 cell counts decreases over time, 2) the decreasing rates of the two treatment groups are not significantly different, 3) gender does not have a significant impact on the CD4 cell counts, 4) patients with previous AIDS diagnostic experience before the study have significantly lower CD4 cell counts, and 5) those who had failed the AZT therapy before the study have significantly higher levels of CD4 cell counts than those who were found intolerant of the AZT therapy before the study. For the survival part of the results, the estimate of ψ by NEW2 is -0.235 and its 95% CI is $(-0.289, -0.182)$, which provides a strong evidence of a negative association between the hazard of death and the CD4 cell counts. The estimates $\hat{\phi}_1 = 0.076$ and $\hat{\gamma}_1 = 1.230$ suggest that the ddI group has a little better survival rate than the ddC group before the first crossing point at $\hat{\gamma}_1 = 1.230$ months, but the ddC group has a slightly better survival rate than the ddI group after the first crossing point and before the estimated change point at $\hat{\kappa} = 11.5$ months. However, because the CIs of ϕ_1 and γ_1 both contain 0, the first crossing point is not significant. On the other hand, the CI of neither ϕ_2 nor γ_2 contains 0, suggesting that the second crossing point is significant at the significance level of 0.05. The estimates $\hat{\phi}_2 = -0.217$ and $\hat{\gamma}_2 = 15.079$ suggest that the ddI group has a little worse survival rate than the ddC group after the estimated changing point at $\hat{\kappa} = 11.5$ months and before the estimated second crossing point $\hat{\gamma}_2 = 15.079$ months. However, after the estimated second crossing point, the ddI group has a better survival rate than the ddC group. Further, for the three time-independent covariates Gender,

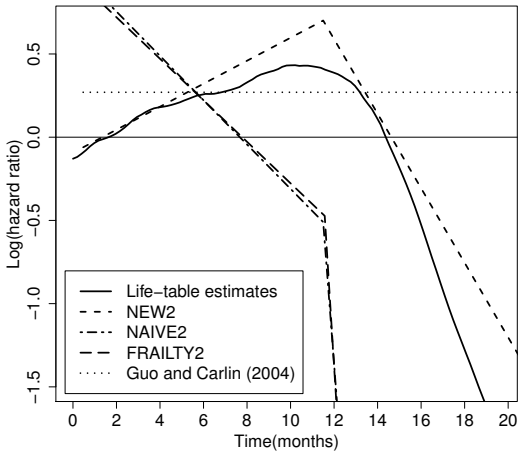

Figure 4. Estimated log hazard ratios by our proposed method NEW2 (dashed line), the naive method NAIVE2 (dot-dashed line), the frailty method FRAILTY2 (long-dashed line), and the method of [27] (dotted line). The bold solid curve denotes the life-table estimator of the log hazard ratio computed directly from the observed survival data.

PrevOI, and Stratum, it seems that only PrevOI has a significantly positive impact on patients' hazard rates. Regarding the results of NAIVE2 and FRAILTY2, we can see from Table 4 that their estimates of the first crossing point γ_1 are 7.524 and 7.778, respectively, and their estimates of the second crossing point γ_2 are 11.172 and 11.345, respectively. None of the CIs for γ_1 and γ_2 contain 0. Compared to the life-table estimates of the two hazard rate functions shown in Figure 1, these results are obviously unreliable. Therefore, the estimates of other parameters by them would not be reliable either.

To further illustrate the results in Table 4, the estimated hazard ratio of the ddI and ddC groups by the method NEW2 when Gender=1, PrevOI=1, and Stratum=1 is shown in Figure 4 by the short-dashed line. As a comparison, the estimated hazard ratios by NAIVE2, FRAILTY2, the method of [27], and the life-table estimation are shown in the same plot by the dot-dashed, long-dashed, dotted, and bold solid lines, respectively. Because the estimated log hazard ratio by the life-table estimation is computed from the observed survival data directly without using any models, and the other four estimated hazard ratios are based on four different models, this plot can be used as a model checking plot [46], for evaluating the efficacy of the four modeling approaches. From this plot, it seems that our proposed approach NEW2 performs reasonably well, the approach by [27] cannot describe the increasing-and-then-decreasing trend of the hazard ratio well, and both the NAIVE2 and FRAILTY2 approaches cannot estimate the two crossing points properly.

5. Discussion

In this article, we have proposed a modeling approach for comparing two crossing hazard rate functions by jointly modeling survival and longitudinal data. This approach has the flexibility to accommodate more than one crossing point. Numerical results show that it works reasonably well in various cases. Theoretically, we have proved that the maximum likelihood estimates of our proposed models exist, are well-defined, and

have the strong consistency and the asymptotic normality properties.

Many research questions are still open in the proposed method. For instance, although the proposed method is already much more general than some existing methods, it is restricted to cases with piecewisely linear log hazard ratios. While it is efficient when the piecewise linearity is valid, it may not be appropriate for applications in which log hazard ratios are not piecewisely linear. In the literature, there has been some discussion about cases when the log hazard ratios are nonlinear [7,18]. However, almost all such research focuses on analyzing survival data only. It, therefore, requires much future research to generalize our proposed survival model properly for comparison of two crossing hazard rate functions when longitudinal data are available. From the numerical results in Section 3, we can see that the frailty modeling approach seems more robust to the piecewise linearity assumption than our proposed joint modeling approach. Systematic theoretical and numerical comparison between these two types of approaches is needed in our future research. Further, in our proposed joint models, only numerical continuous longitudinal data are considered. In practice, many longitudinal data might be discrete or even categorical. It is therefore important to generalize our method to such cases as well.

Acknowledgments

The authors thank the editor and two referees for their valuable comments which greatly improved the quality of this paper.

Supplementary Materials

Proofs of some theoretical results from Section 2.3 are given in Supplementary Materials.

References

[1] Lawless JF. Statistical models and methods for lifetime data. Vol. 362. John Wiley & Sons; 2011.
[2] Bain L, Englehardt M. Statistical analysis of reliability and life-testing models: theory and methods. Vol. 115. CRC Press; 1991.
[3] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media; 2005.
[4] O'Quigley J, Pessione F. Score tests for homogeneity of regression effect in the proportional hazards model. Biometrics. 1989;:135–144.
[5] O'Quigley J, Pessione F. The problem of a covariate-time qualitative interaction in a survival study. Biometrics. 1991;:101–115.
[6] Lin X, Wang H. A new testing approach for comparing the overall homogeneity of survival curves. Biometrical Journal. 2004;46(5):489–496.
[7] Liu K, Qiu P, Sheng J. Comparing two crossing hazard rates by cox proportional hazards modelling. Statistics in medicine. 2007;26(2):375–391.
[8] Qiu P, Sheng J. A two-stage procedure for comparing hazard rate functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2008;70(1):191–208.
[9] O'Quigley J. On a two-sided test for crossing hazards. The Statistician. 1994;:563–569.
[10] Fleming TR, O'Fallon JR, O'Brien PC, et al. Modified kolmogorov-smirnov test procedures with application to arbitrarily right-censored data. Biometrics. 1980;:607–625.

[11] Chen Z, Huang H, Qiu P. Comparison of multiple hazard rate functions. Biometrics. 2016; 72:39–45.
[12] Chen Z, Huang H, Qiu P. An improved two-stage procedure to compare hazard curves. Journal of Statistical Computation and Simulation. 2017;87:1877–1886.
[13] Mantel N, Stablein DM. The crossing hazard function problem. The Statistician. 1988; :59–64.
[14] Moreau T, Maccario J, Lellouch J, et al. Weighted log rank statistics for comparing two distributions. Biometrika. 1992;79(1):195–198.
[15] Anderson J, Senthilselvan A. A two-step regression model for hazard functions. Applied Statistics. 1982;:44–51.
[16] Breslow NE, Edler L, Berger J. A two-sample censored-data rank test for acceleration. Biometrics. 1984;:1049–1062.
[17] Zhang J, Peng Y. Crossing hazard functions in common survival models. Statistics & probability letters. 2009;79(20):2124–2130.
[18] Cheng MY, Qiu P, Tan X, et al. Confidence intervals for the first crossing point of two hazard functions. Lifetime data analysis. 2009;15(4):441–454.
[19] Muggeo VM, Tagliavia M. A flexible approach to the crossing hazards problem. Statistics in medicine. 2010;29(18):1947–1957.
[20] Hsieh F. On heteroscedastic hazards regression models: theory and application. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001;63(1):63–79.
[21] Bagdonavičius V, Hafdi MA, Nikulin M. Analysis of survival data with cross-effects of survival functions. Biostatistics. 2004;5(3):415–425.
[22] Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997;:330–339.
[23] Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica. 2004;:809–834.
[24] Zeng D, Cai J. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. The Annals of Statistics. 2005;33(5):2132–2163.
[25] Dupuy JF, Grama I, Mesbah M, et al. Asymptotic theory for the cox model with missing time-dependent covariate. The Annals of Statistics. 2006;34(2):903–924.
[26] Abrams DI, Goldman AI, Launer C, et al. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. New England Journal of Medicine. 1994;330(10):657–662.
[27] Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. The american statistician. 2004;58(1):16–24.
[28] Park K, Qiu P. Model selection and diagnostics for joint modeling of survival and longitudinal data with crossing hazard rate functions. Statistics in medicine. 2014;33(26):4532–4546.
[29] Little RJ, Rubin DB. Statistical analysis with missing data. Vol. 793. John Wiley & Sons; 2019.
[30] Yang X, Li J, Shoptaw S. Imputation-based strategies for clinical trial longitudinal data with nonignorable missing values. Statistics in medicine. 2008;27(15):2826–2849.
[31] Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in medicine. 1999;18(6):681–694.
[32] White IR, Royston P. Imputing missing covariate values for the cox model. Statistics in medicine. 2009;28(15):1982–1998.
[33] Ali MW, Siddiqui O. Multiple imputation compared with some informative dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. Journal of biopharmaceutical statistics. 2000;10(2):165–181.
[34] Hu B, Li L, Greene T. Joint multiple imputation for longitudinal outcomes and clinical events that truncate longitudinal follow-up. Statistics in medicine. 2016;35(17):2991–3006.
[35] Hougaard P. Analysis of multivariate survival data. Springer Science & Business Media; 2012.
[36] Rondeau V, Mazroui Y, Gonzalez JR. frailtypack: an r package for the analysis of corre-

lated survival data with frailty models using penalized likelihood estimation or parametrical estimation. J Stat Softw. 2012;47(4):1–28.
[37] Dupuy JF, Mesbah M. Joint modeling of event time and nonignorable missing longitudinal data. Lifetime Data Analysis. 2002;8(2):99–115.
[38] Evans M, Swartz T. Approximating integrals via monte carlo and deterministic methods. Vol. 20. OUP Oxford; 2000.
[39] Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. Statistics in Medicine. 2007;26(14):2813–2835.
[40] Yu M, Law NJ, Taylor JM, et al. Joint longitudinal-survival-cure models and their application to prostate cancer. Statistica Sinica. 2004;:835–862.
[41] Efron B. Missing data, imputation, and the bootstrap. Journal of the American Statistical Association. 1994;89(426):463–475.
[42] Murphy SA, Van der Vaart AW. On profile likelihood. Journal of the American Statistical Association. 2000;95(450):449–465.
[43] Xu C, Baines PD, Wang JL. Standard error estimation using the em algorithm for the joint modeling of survival and longitudinal data. Biostatistics. 2014;15(4):731–744.
[44] Goldman AI, Carlin BP, Crane LR, et al. Response of cd4 lymphocytes and clinical consequences of treatment using ddi or ddc in patients with advanced hiv infection. JAIDS Journal of Acquired Immune Deficiency Syndromes. 1996;11(2):161–169.
[45] Rizopoulos D. Jm: An r package for the joint modelling of longitudinal and time-to-event data. Journal of Statistical Software (Online). 2010;35(9):1–33.
[46] Cook RD, Weisberg S. Applied regression including computing and graphics. Vol. 488. John Wiley & Sons; 2009.