Noname manuscript No.
(will be inserted by the editor)

Evaluation of the Treatment Time-Lag Effect for Survival Data

Received: date / Accepted: date

Address(es) of author(s) should be given

Abstract Medical treatments often take a period of time to reveal their impact on subjects, which is the so-called time-lag effect in the literature. In the survival data analysis literature, most existing methods compare two treatments in the entire study period. In cases when there is a substantial time-lag effect, these methods would not be effective in detecting the difference between the two treatments, because the similarity between the treatments during the time-lag period would diminish their effectiveness. In this paper, we develop a novel modeling approach for estimating the time-lag period and for comparing the two treatments properly after the time-lag effect is accommodated. Theoretical arguments and numerical examples show that it is effective in practice.

Keywords Cox proportional hazards model · Crossing hazard rates · Lag effect · Survival analysis · Treatment comparison

1 Introduction

Survival analysis is a powerful tool for analyzing time to event data. One important task in survival analysis is to evaluate treatment effect by comparing two hazard rate functions (Klein and Moeschberger, 2005). To this end, many testing procedures (e.g., Log-rank, Gehan-Wilcoxon, and Peto-Peto tests) have been developed in the literature. Most of them are under the assumption that the Cox proportional hazards model is appropriate (Bain and Englehardt, 1991). There are also some discussions about cases when the two hazard rate functions cross each other (Chen et al, 2016; Cheng et al, 2009; Lin and Wang, 2004; Liu et al, 2007; O'Quigley, 1994; Park and Qiu, 2014; Qiu and Sheng, 2008). This paper focuses on cases when the two hazard rate functions are similar in the first period of time and quite different afterwards. This scenario occurs when treatments have time-lag effects, and is popular in practice.

Figure 1 presents the life-table estimates of the hazard rate functions of the rats data from Mantel et al (1977). The data contain a drug group of 50 rats and a placebo group of 100 rats. Rats in the first group were treated with a drug and those in the second group were given a placebo. The event of interest in this example is the formation of tumor. From Figure 1, it can be seen that the two hazard curves are almost the same until about 80 days, and the one of the drug group tends to be above that of the placebo group afterwards. A more detailed description of the rats data and its statistical analysis can be found in Section 4.

The above example shows the treatment time-lag effect, and such examples are common in practice. In the survival analysis literature, however, most existing methods compare the two hazard rate functions (or survival functions) in the entire study period. If these traditional methods are used in cases with treatment time-lag effect, then they would not be effective in detecting the difference between the two treatment groups, because the similarity of the two treatment groups during the time-lag period could attenuate their overall difference. In applications with treatment time-lag effect, it is important to estimate how long the time-lag effect lasts, and whether the two treatment groups are significantly different afterwards. In the literature, there is a limited existing discussion about these issues. For instance, Dinse et al (1993) proposed an estimate of the time-lag period based on Kaplan-Meier estimates of the

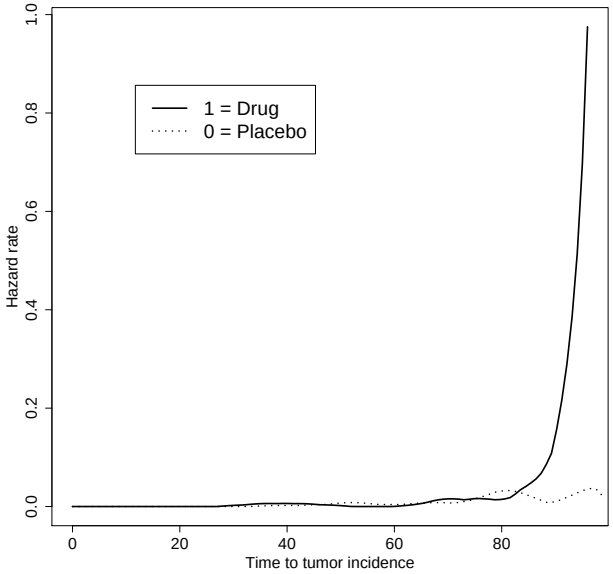

Fig. 1 Estimates of the hazard rate functions for the rats data.

survival functions. Because this estimate was defined as the largest time point τ^* such that a two-sided Z-test for comparing the two survival functions in $[0, \tau^*]$ gave an insignificant result, it would over-estimate the true time-lag period constantly, as demonstrated in Section 3. Also, that paper did not discuss how to compare the two hazard rate functions after the time-lag point was estimated. Zucker and Lakatos (1990) proposed two weighted log-rank type statistics for comparing two survival functions that were robust to treatment time-lag effects. However, to use these methods, we need to specify a potential time-lag point, which is often difficult in practice. In both papers, covariates effect cannot be accommodated.

This paper tries to address estimation of the time-lag point and proper comparison of the two hazard rate functions simultaneously by proposing a modeling approach. Our model is a generalization of the conventional Cox proportional hazards model, and it can accommodate both treatment time-lag effect and covariates effect. To estimate its parameters, the maximum partial likelihood estimation and a bootstrap resampling procedure are used. Certain asymptotic properties of the estimators, including consistency and asymptotic distributions, are provided. Numerical examples show that this method works well in practice.

The remaining part of the paper is organized as follows. In Section 2, our proposed model that addresses the treatment time-lag effect is described. Model estimation and some theoretical properties of the estimators are also discussed. Simulation studies about the proposed method are presented in Section 3. Applications to two real-data examples are discussed in Section 4. Several remarks conclude the paper in

Section 5. Some technical details, including theorem proofs and several lemmas, are given in a supplementary file.

2 Proposed Method

2.1 Proposed model

Assume that there are n subjects. For the ith subject, T_i^o denotes the true survival time, and C_i is the censoring time, where $i = 1, \ldots, n$. Then, we observe (T_i, δ_i), where $T_i = \min(T_i^o, C_i)$ and $\delta_i = I(T_i^o \leq C_i)$. Let $\mathbf{X}(t)$ be the p-dimensional covariate vector that may depend on the time t, and g be the group indicator (i.e., $g_i = 1$ if the ith subject is in the treatment group, and 0 otherwise). For the n subjects, the observed data are then $\{(T_i, \delta_i, \mathbf{X}_i(t), g_i)$, for $i = 1, \ldots, n\}$. Among the subjects, we assume that their observations are independent. Also, we assume that the true survival time T_i^o and the censoring time C_i are independent, given $\mathbf{X}_i(t)$, for $i = 1, \ldots, n$. Then, our proposed model is

$$\lambda(t|\mathbf{X}(t), g) = \lambda_0(t) \exp\{\alpha' \mathbf{X}(t) + \beta' \mathbf{X}(t)(t-\tau) I(t > \tau) g\}, \tag{1}$$

where $\lambda_0(t)$ is a baseline hazard rate function, α and β are p-dimensional regression coefficient vectors, and τ denotes the lag time. From model (1), when $t \leq \tau$, the hazard ratio is

$$\frac{\lambda(t|\mathbf{X}(t), g = 1)}{\lambda(t|\mathbf{X}(t), g = 0)} = 1,$$

and when $t > \tau$, it becomes

$$\frac{\lambda(t|\mathbf{X}(t), g = 1)}{\lambda(t|\mathbf{X}(t), g = 0)} = \exp\{\beta' \mathbf{X}(t)(t-\tau)\}.$$

From the above two expressions, we can see that in the case when $\beta = \mathbf{0}$, the two hazard rate functions are identical all the time. In the case when $\beta \neq \mathbf{0}$, the two hazard rate functions are the same when $t \leq \tau$, different when $t > \tau$, and the magnitude of difference depends on the value of β. From this description of model (1), we can see that the time-lag period is described by τ, and the difference between the two hazard rate functions after τ is described by β.

Note that model (1) guarantees the two hazard rate functions to be continuous when accommodating the time-lag effect. We can consider several variants. For instance, if the hazard ratio does not need to be a continuous function of t, then the following simpler model can be considered:

$$\lambda(t|\mathbf{X}(t), g) = \lambda_0(t) \exp\{\alpha' \mathbf{X}(t) + \beta' \mathbf{X}(t) I(t > \tau) g\}.$$

The log hazard ratio changes from 0 to $\beta' \mathbf{X}(t)$ before and after τ in such cases. Model (1) also can be generalized to a more flexible one by including more terms, such as the following

$$\lambda(t|\mathbf{X}(t), g) = \lambda_0(t) \exp\{\alpha' \mathbf{X}(t) + \beta' \mathbf{X}(t)(t-\tau) I(t > \tau) g + \gamma' \mathbf{X}(t)(t-\tau)^2 I(t > \tau) g\},$$

where γ is a regression coefficient vector.

It is worth mentioning that in practice, the time-lag treatment effect and covariate effect may not be necessarily related. So, we can also consider a model such as the following:

$$\lambda(t|\mathbf{X}(t),g) = \lambda_0(t)\exp\{\alpha'\mathbf{X}(t)+\beta(t-\tau)I(t>\tau)g\}.$$

Estimation of this model is similar to that of (1).

When there are many covariates in a model, we can choose the best sub-model using a model selection criterion, such as the AIC, AICc, BIC, and BICc (cf., Park and Qiu, 2014). The formulas of these model selection criteria are

$$-2\log(L_{max})+\psi(n,k,r),$$

where L_{max} is the maximum value of the likelihood function of a sub-model under consideration, k is the number of parameters in the model, r is the number of uncensored survival times, and $\psi(n,k,r) = 2k, 2k+2k(k+1)/(n-k-1), k\log(n)$, and $k\log(r)$, respectively, for the criteria AIC, AICc, BIC, and BICc.

2.2 Model estimation

Estimation of model (1) can be accomplished by maximizing the Cox partial likelihood function (Cox, 1972) of the parameter vector $\theta = (\alpha', \beta', \tau)'$:

$$L_n(\theta) = \prod_{i=1}^{n}\left\{\frac{\exp\{\alpha'\mathbf{X}_i(T_i)+\beta'\mathbf{X}_i(T_i)(T_i-\tau)I(T_i>\tau)g_i\}}{\sum_{j=1}^{n}Y_j(T_i)\exp\{\alpha'\mathbf{X}_i(T_i)+\beta'\mathbf{X}_i(T_i)(T_i-\tau)I(T_i>\tau)g_j\}}\right\}^{\delta_i}, \quad (2)$$

where $Y_j(t) = I(T_j \geq t)$ indicates whether the jth subject is under risk at time t. Then, the log partial likelihood function $l_n(\theta) = \log L_n(\theta)$ is given by

$$\begin{aligned} l_n(\theta) = &\sum_{i=1}^{n}\delta_i\left\{\alpha'\mathbf{X}_i(T_i)+\beta'\mathbf{X}_i(T_i)(T_i-\tau)I(T_i>\tau)g_i\right\} \\ &-\delta_i\log\left\{\sum_{j=1}^{n}Y_j(T_i)\exp\{\alpha'\mathbf{X}_i(T_i)+\beta'\mathbf{X}_i(T_i)(T_i-\tau)I(T_i>\tau)g_j\}\right\}. \end{aligned} \quad (3)$$

Let $\psi = (\alpha', \beta')'$. Then, for each fixed τ, we can obtain the maximum partial likelihood estimators $\widehat{\psi}(\tau) = (\widehat{\alpha}(\tau)', \widehat{\beta}(\tau)')'$ such that

$$\widehat{\psi}(\tau) = \arg\max_{\psi} l_n(\psi,\tau).$$

From the equation (2), the score function $U(\psi)$ is given by the first derivative of $l_n(\psi,\tau)$ with respect to ψ, and the information matrix $I(\psi)$ is given by the negative of the second derivative of $l_n(\psi,\tau)$ with respect to ψ. Then, the maximum partial likelihood estimate of ψ for each fixed τ can be obtained by solving the equation $U(\psi) = 0$. Numerically, this can be achieved by using the Newton-Raphson iterative

algorithm. More specifically, let $\psi^{(k)}$ be the estimate of ψ in the kth iteration. Then, the updated estimate in the $(k+1)$th iteration would be

$$\psi^{(k+1)} = \psi^{(k)} + I(\psi^{(k)})^{-1} U(\psi^{(k)}),$$

where $U(\psi^{(k)})$ and $I(\psi^{(k)})$ are respectively the score function and information matrix evaluated at $\psi^{(k)}$.

The inverse of the information matrix, evaluated at $\widehat{\psi}(\tau)$, can be used as an estimate of the variance-covariance matrix of $\widehat{\psi}(\tau)$, from which we can obtain standard errors of the estimated regression coefficients. Then, we can construct confidence intervals for ψ. More specifically, let $z_{1-\alpha/2}$ be the $(1-\alpha/2)$th percentile of a standard normal distribution. Then, an asymptotic $100(1-\alpha)\%$ confidence interval for ψ is given by

$$\widehat{\psi}(\tau) \pm z_{1-\alpha/2} I(\widehat{\psi}(\tau))^{-1/2}.$$

For τ, we can obtain its estimator by maximizing $l_n(\widehat{\psi}(\tau), \tau)$. To this end, we can plot $l_n(\widehat{\psi}(\tau), \tau)$ against τ to identify the maximizer. Or, the grid search algorithm can be employed here for finding the grid point of τ resulting in the maximum likelihood. However, the maximum likelihood estimate of τ may not be unique. Therefore, our final estimator of τ is defined by

$$\widehat{\tau} = \arg\min\{\tau^* : \tau^* = \arg\max_{\tau} l_n(\widehat{\psi}(\tau), \tau)\}.$$

As a side note, we have assumed that there are no ties among uncensored event times in the above model estimation procedure for simplicity. Otherwise, the Breslow's or Efron's likelihoods should be used in (2) (cf., Chapter 8, Klein and Moeschberger, 2005).

For obtaining a confidence interval for τ, we consider a bootstrap method described below. Remember that the observed data are $\mathscr{D} = \{(T_i, \delta_i, \mathbf{X}_i(T_i), g_i), i = 1, \ldots, n\}$. To keep the sample sizes of the treatment and control groups, we resample the two groups of data separately. Assume that there are n_1 subjects in the control group, and n_2 subjects in the treatment group. Then, we repeatedly draw n_1 observations from the control group, and n_2 observations from the treatment group. The two parts are then combined as a single bootstrap sample. From each bootstrap sample, we can compute a point estimate of τ by using the proposed method described above. This process is repeated B times to obtain B bootstrap estimates of τ, from which the distribution of $\widehat{\tau}$ can be estimated empirically. The 95% confidence interval for τ is then constructed using the 2.5th and 97.5th percentiles of the B bootstrap estimates of τ.

2.3 Theoretical properties

We discuss some statistical properties of the parameter estimates described in the previous part. First, introduce some notation. Let $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_n(t))'$, $\mathbf{Z}_i(t) =$

$(\mathbf{Z}_{1i}(t)', \mathbf{Z}_{2i}(t))'$, $\mathbf{Z}_{1i}(t) = \mathbf{X}_i(t)$, $\mathbf{Z}_{2i}(t) = \mathbf{X}_i(t)(t-\tau)I(t>\tau)g_i$, and

$$
\begin{aligned}
S^{(0)}(t;\theta) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\exp\{\psi'\mathbf{Z}_i(t)\} = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\exp\{\alpha'\mathbf{Z}_{1i}(t)+\beta'\mathbf{Z}_{2i}(t)\},\\
\mathbf{S}^{(1)}(t;\theta) &= \frac{\partial S^{(0)}(t;\theta)}{\partial\psi} = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\mathbf{Z}_i(t)\exp\{\psi'\mathbf{Z}_i(t)\},\\
\mathbf{S}^{(2)}(t;\theta) &= \frac{\partial \mathbf{S}^{(1)}(t;\theta)}{\partial\psi} = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\mathbf{Z}_i(t)^{\otimes 2}\exp\{\psi'\mathbf{Z}_i(t)\},\\
\mathbf{E}(t;\theta) &= \frac{\mathbf{S}^{(1)}(t;\theta)}{S^{(0)}(t;\theta)},\\
\mathbf{V}(t;\theta) &= \frac{\mathbf{S}^{(2)}(t;\theta)}{S^{(0)}(t;\theta)} - \mathbf{E}(t;\theta)^{\otimes 2},
\end{aligned}
\tag{4}
$$

$\mathbf{x}^{\otimes 0} = 1$, $\mathbf{x}^{\otimes 1} = \mathbf{x}$, and $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}'$. Note that $S^{(0)}(t;\theta)$ is a scalar, $\mathbf{S}^{(1)}(t;\theta)$ and $\mathbf{E}(t;\theta)$ are $2p$-dimensional vectors, and $\mathbf{S}^{(2)}(t;\theta)$ and $\mathbf{V}(t;\theta)$ are $2p\times 2p$ matrices. In the case when $t<\tau$, $S^{(k)}(t;\alpha) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\mathbf{Z}_{1i}(t)^{\otimes k}\exp\{\alpha'\mathbf{Z}_{1i}(t)\}$, for $k=0,1,2$. To establish the asymptotic properties of the estimators, the following assumptions are needed and some of them are adapted from Andersen and Gill (1982).

(A.1) The basline hazard rate function $\lambda_0(t)$ is continuous in a neighborhood of τ, $\inf_{t\in[0,t^*]}\lambda_0(t) > 0$, and $\int_0^{t^*}\lambda_0(t)dt < \infty$, where $[0,t^*]$ denotes the study period.

(A.2) For $k=0,1,2$,

$$
E\left[\sup_{t\in[0,t^*],\theta}\left\{\|\mathbf{Z}(t)\|^k\exp\left(\psi'\mathbf{Z}(t)\right)\right\}^2\right] < \infty,
$$

where $\|\mathbf{x}(t)\|$ is the supremum norm across all elements of the vector $\mathbf{x}$.

(A.3) For $k=0,1,2$,

$$
\sup_{t\in[0,t^*],\theta}\left\|S^{(k)}(t;\theta) - s^{(k)}(t;\theta)\right\| \xrightarrow{P} 0,
$$

where $s^{(k)}(t;\theta) = E[S^{(k)}(t;\theta)]$.

(A.4) There exists a constant $\varepsilon > 0$ such that

$$
n^{-\frac{1}{2}}\sup_{i,t}\sqrt{\mathbf{Z}_i(t)'\mathbf{Z}_i(t)}Y_i(t)I\left(\psi'\mathbf{Z}_i(t) > -\varepsilon\sqrt{\mathbf{Z}_i(t)'\mathbf{Z}_i(t)}\right) \xrightarrow{P} 0.
$$

(A.5) Let $\mathbf{e}(t;\theta) = \mathbf{s}^{(1)}(t;\theta)/s^{(0)}(t;\theta)$ and $\mathbf{v}(t;\theta) = \mathbf{s}^{(2)}(t;\theta)/s^{(0)}(t;\theta) - \mathbf{e}(t;\theta)^{\otimes 2}$. Assume that $s^{(0)}(t;\theta)$, $\mathbf{s}^{(1)}(t;\theta)$, and $\mathbf{s}^{(2)}(t;\theta)$ are continuous bounded functions of θ and t, and uniformly continuous in $t\in[0,t^*]$, and the matrix

$$
\int_0^t \mathbf{v}(\tilde{t};\theta)s^{(0)}(\tilde{t};\theta)\lambda_0(\tilde{t})d\tilde{t}
$$

is positive definite.

Theorem 1 *Under the assumptions (A.1) - (A.5), (i) the estimators $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\tau}$ are all consistent in probability, (ii) the convergence rate of each component in $\widehat{\alpha}$ and $\widehat{\beta}$ to the corresponding component in α and β is $O(n^{-1/2})$, and (iii) the convergence rate of $\widehat{\tau}$ to τ is $O(n^{-1})$.*

Let $v^-(t)$ and $v^+(t)$ be two independent jump processes such that $v^-(t)$ is a Poisson random variable at each t with rate parameter $-p_0^-(\tau;\mathbf{Z}_2)t$ on $\mathbb{R}_-$ and $v^+(t)$ is a Poisson variable at each t with rate parameter $p_0^+(\tau^+;\mathbf{Z}_2)s$ on $\mathbb{R}_+$, where $\tau^+ = \max(\tau,0)$,

$$
\begin{aligned}
p_0^-(t;\mathbf{Z}_2) &= E\left[\lambda_0(t)Y(t)\exp(\alpha'\mathbf{Z}_1(t))|\mathbf{Z}_2(t)\right], \\
p_0^+(t;\mathbf{Z}_2) &= E\left[\lambda_0(t)Y(t)\exp(\alpha'\mathbf{Z}_1(t)+\beta'\mathbf{Z}_2(t))|\mathbf{Z}_2(t)\right].
\end{aligned}
$$

For $k \geq 1$, let V_k^- and V_k^+ be independent sequences of i.i.d. random variables with respective conditional characteristic functions

$$
\begin{aligned}
\varphi^-(t;\mathbf{Z}_2) &= E\left[\exp(itV_k^-)|\mathbf{Z}_2(\tau)\right], \\
\varphi^+(t;\mathbf{Z}_2) &= E\left[\exp(itV_k^+)|\mathbf{Z}_2(\tau^+)\right].
\end{aligned}
$$

Conditional on $\mathbf{Z}_2(\tau)$, both V_k^- and V_k^+ are assumed to be independent of v^- and v^+. Let $Q(t) = Q^-(t)I(t<0) - Q^+(t)I(t>0)$ be the right-continuous jump process with

$$
\begin{aligned}
Q^-(t) &= \sum_{k=0}^{v^-(t)} V_k^- + v^-(t)\log\left\{\frac{s^{(0)}(\tau;\alpha)}{s^{(0)}(\tau;\psi)}\right\} = \sum_{k=0}^{v^-(t)} V_k^- + v^-(t)d_0(\tau), \\
Q^+(t) &= \sum_{k=0}^{v^+(t)} V_k^+ + v^+(t)\log\left\{\frac{s^{(0)}(\tau;\alpha)}{s^{(0)}(\tau;\psi)}\right\} = \sum_{k=0}^{v^+(t)} V_k^+ + v^+(t)d_0(\tau),
\end{aligned}
$$

where $d_0(t) = \log\left\{s^{(0)}(t;\alpha)/s^{(0)}(t;\psi)\right\}$, and $s^{(0)}(t;\alpha)$ is $s^{(0)}(t;\theta)$ when $\beta = \mathbf{0}$.

Theorem 2 *Under the assumptions (A.1) - (A.5),*

(i) $\sqrt{n}(\widehat{\psi}-\psi)$ *has an asymptotically normal distribution* $N(0,\Sigma(\theta)^{-1})$*, where*

$$
\Sigma(\theta) = \begin{pmatrix} \Sigma_{11}(\theta) & \Sigma_{12}(\theta) \\ \Sigma_{21}(\theta) & \Sigma_{22}(\theta) \end{pmatrix},
$$

$\Sigma_{11}(\theta)$ is the upper-left $p\times p$ submatrix of $\int_0^t v(\tilde{t};\theta)s^{(0)}(\tilde{t};\theta)\lambda_0(\tilde{t})d\tilde{t}$, $\Sigma_{12}(\theta)$ is the upper-right $p\times p$ submatrix of $\int_\tau^t v(\tilde{t};\theta)s^{(0)}(\tilde{t};\theta)\lambda_0(\tilde{t})d\tilde{t}$, $\Sigma_{21}(\theta) = \Sigma_{12}(\theta)'$, and $\Sigma_{22}(\theta)$ is the lower-right $p\times p$ submatrix of $\int_\tau^t v(\tilde{t};\theta)s^{(0)}(\tilde{t};\theta)\lambda_0(\tilde{t})d\tilde{t}$.

(ii) $n(\widehat{\tau}-\tau)$ *coverges in probability to* $M_Q = \inf\{m : m = \arg\max_t Q(t)\}$.

(iii) $\widehat{\psi}$ *and* $\widehat{\tau}$ *are asymptotically independent.*

2.4 Testing of β

From model (1), comparison of the two hazard rate functions of the treatment and control groups can be accomplished by testing the hypotheses

$$H_0 : \beta = \mathbf{0} \qquad \text{versus} \qquad H_a : \beta \neq \mathbf{0}. \tag{5}$$

Based on the asymptotic normality of $\widehat{\beta}$ in Theorem 2, we can conduct the Wald test using the test statistic

$$\widehat{\beta}' \left[\text{Var}(\widehat{\beta}) \right]^{-1} \widehat{\beta}, \tag{6}$$

and the asymptotic null distribution of this statistic is χ^2_p. For testing the hypotheses in (5), an alternative approach is to use the likelihood ratio test (LRT) with the test statistic

$$-2 \left(\log L_R - \log L_F \right), \tag{7}$$

where L_F and L_R denote the likelihoods of the full model (i.e., model (1)) and the reduced model (i.e., model (1) with $\beta = \mathbf{0}$), respectively. The asymptotic null distribution of this statistic is also χ^2_p.

3 Simulation Study

In this section, we present some simulation results to evaluate the numerical performance of the proposed method described in Section 2. First, we consider the following simple model without any covariates:

$$\lambda(t|\mathbf{X}(t), g) = \lambda_0(t) \exp\{\beta(t-\tau) I(t > \tau) g\}, \tag{8}$$

in which $\beta = 1$, $\tau = 1$, and the baseline hazard rate is assumed to be $\lambda_0(t) = 0.5$. The numbers of subjects in the treatment and control groups are assumed to be the same (i.e., $n_1 = n_2$). They are either 150 or 200. The censoring time is generated from the uniform distribution on the interval $[a, b]$, where a and b are adjusted to reach a pre-specified censoring rate. In the simulation, we consider the two censoring rates of 20% and 40%.

For comparison purposes, besides the proposed method, we also consider 6 commonly-used existing methods: Log-rank, Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, a modified Peto-Peto, and Fleming-Harrington (cf., Chapter 7, Klein and Moeschberger, 2005). Their test statistics all take the following form:

$$\frac{\sum_{i=1}^{k} w(t_i) \left(d_{i1} - y_{i1} \frac{d_i}{y_i} \right)}{\sqrt{\sum_{i=1}^{k} w(t_i)^2 \frac{y_{i1}}{y_i} \left(1 - \frac{y_{i1}}{y_i} \right) \left(\frac{y_i - d_i}{y_i - 1} \right) d_i}}, \tag{9}$$

where $t_1 < \cdots < t_k$ are the distinct observed event times of both the control and treatment groups, $\{w(t_i)\}$ are weights, d_{ij} is the ith event in the jth group, $d_i = d_{i1} + d_{i2}$, y_{ij} is the number of subjects in group j who are at risk at time t_i, and $y_i = y_{i1} + y_{i2}$, for $j =$

1,2. The weight $w(t_i)$ is $1, y_i, \sqrt{y_i}, \widetilde{S}(t_i), \widetilde{S}(t_i)\frac{y_i}{y_i+1}$, and $\left\{\widehat{S}(t_{i-1})\right\}^{r_1}\left\{1-\widehat{S}(t_{i-1})\right\}^{r_2}$, respectively, for the methods Log-rank, Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, modified Peto-Peto, and Fleming-Harrington, where $\widehat{S}(t)=\prod_{t_i\le t}\left(1-\frac{d_i}{y_i}\right)$ is the Kaplan-Meier estimator of the survival function $S(t)$, $\widetilde{S}(t)=\prod_{t_i\le t}\left(1-\frac{d_i}{y_i+1}\right)$ is its modification, and $r_1, r_2 \ge 0$ are two parameters. Furthermore, the estimate of the time-lag point by Dinse et al (1993) with $\alpha = 0.05$ and the robust survival function comparison method by Zucker and Lakatos (1990) are also considered here (see Section 1 for a brief description). In the method by Zucker and Lakatos (1990), there are two different versions of the test statistic. We use the one based on the maximum efficiency, as recommended by the authors of that paper. To use this method, we need to specify the time-lag point. To this end, three cases when the specified time-lag point is equal to, smaller than or larger than the true time-lag point are considered.

First, we evaluate the accuracy of the parameter estimates. Based on 1,000 replicated simulations, the averaged point estimates along with the bias and mean square error values of the estimated parameters are summarized in Table 1. Because the six methods based on (9) and the one by Zucker and Lakatos (1990) do not provide estimates of the time-lag point, they are not included. From the table, we can see that (i) the point estimates of our proposed method are closer to the true parameter values when the sample sizes are larger, (ii) both the bias and mean square error values of its estimated parameters are smaller when the sample sizes increase, (iii) the mean square error values of $\widehat{\beta}$ are larger when the censoring rates increase, and (iv) the estimate of the time-lag point by the method in Dinse et al (1993) is much larger than the true parameter value. The results (i) and (ii) are consistent with the consistency results in Theorem 1, and the result (iv) is consistent with our intuition as mentioned in Section 1.

Table 1 Averaged point estimates and bias and mean square error (MSE) values of the estimated parameters, based on 1000 replicated simulations for model (8). In the table, n_1 and n_2 denote the number of subjects in each group, and CR denotes the censoring rate.

CR	Method	True Parameters	$n_1 = n_2 = 150$			$n_1 = n_2 = 200$		
			Point Estimates	Bias	MSE	Point Estimates	Bias	MSE
20%	Proposed	$\tau = 1$	1.233	0.233	0.119	1.248	0.248	0.118
		$\beta = 1$	1.031	0.031	0.158	1.006	0.006	0.119
	Dinse et al.	$\tau = 1$	3.213	2.213	5.162	3.155	2.155	4.988
40%	Proposed	$\tau = 1$	1.095	0.095	0.110	1.101	0.101	0.104
		$\beta = 1$	1.211	0.211	0.413	1.158	0.158	0.317
	Dinse et al.	$\tau = 1$	2.659	1.659	2.920	2.595	1.595	2.768

To compare the hazard rate functions of the two groups, we conduct the hypothesis testing of (5) using both the Wald test (6), the likelihood ratio test (7), the 6 existing methods mentioned above related to (9), and the method by Zucker and Lakatos (1990). For the 6 existing methods considered, besides the regular comparison of the hazard rate functions in the entire study period, we also use them to compare the hazard rate functions in the time period after $\widehat{\tau}$. In the method by Zucker and Lakatos, we

Table 2 The proportions of rejecting the null hypothesis of various different methods for testing the difference between the two hazard rate functions related to model (8) based on a significance level 0.05. Some entries have two values, with the first one being proportion of rejections computed in the entire study period and the second one computed in the time period after $\widehat{\tau}$. The results are based on 1,000 replicated simulations. In the table, n_1 and n_2 denote the number of subjects in each group, and CR denotes the censoring rate.

CR	Testing Method	$n_1 = n_2 = 150$		$n_1 = n_2 = 200$	
		Entire	After $\widehat{\tau}$	Entire	After $\widehat{\tau}$
20%	Wald test		0.966		0.995
	Likelihood ratio test		0.967		0.995
	Log-rank	0.051	0.850	0.062	0.950
	Gehan-Wilcoxon	0.071	0.120	0.086	0.232
	Tarone-Ware	0.026	0.494	0.024	0.680
	Peto-Peto	0.050	0.300	0.050	0.513
	Modified Peto-Peto	0.050	0.280	0.054	0.495
	Fleming-Harrington ($r_1 = 0, r_2 = 1$)	0.589	0.965	0.732	0.996
	Fleming-Harrington ($r_1 = 1, r_2 = 0$)	0.050	0.305	0.050	0.515
	Fleming-Harrington ($r_1 = 1, r_2 = 1$)	0.038	0.919	0.046	0.971
	Zucker and Lakatos ($\tau^* = 0.5$)	0.105		0.130	
	Zucker and Lakatos ($\tau^* = 1.0$)	0.299		0.410	
	Zucker and Lakatos ($\tau^* = 2.0$)	0.762		0.878	
40%	Wald test		0.915		0.961
	Likelihood ratio test		0.920		0.965
	Log-rank	0.111	0.751	0.155	
	Gehan-Wilcoxon	0.022	0.170	0.014	
	Tarone-Ware	0.021	0.426	0.018	
	Peto-Peto	0.021	0.372	0.015	
	Modified Peto-Peto	0.021	0.352	0.015	
	Fleming-Harrington ($r_1 = 0, r_2 = 1$)	0.631	0.921	0.782	
	Fleming-Harrington ($r_1 = 1, r_2 = 0$)	0.022	0.382	0.015	
	Fleming-Harrington ($r_1 = 1, r_2 = 1$)	0.155	0.890	0.216	
	Zucker and Lakatos ($\tau^* = 0.5$)	0.217		0.315	
	Zucker and Lakatos ($\tau^* = 1.0$)	0.435		0.603	
	Zucker and Lakatos ($\tau^* = 2.0$)	0.745		0.872	

need to specify the time-lag point. In this study, we specify it to be $\tau^* = 0.5$, 1.0, or 2.0. With a significance level 0.05, the proportions of rejecting the null hypothesis by these methods based on 1,000 replicated simulations are presented in Table 2. From the table, we can see that both the Wald test and the likelihood ratio test give very significant results, while none of the 6 alternative methods based on (9) could detect the difference between two hazard rate functions in the entire study period and in the time period after $\widehat{\tau}$, except the Fleming-Harrington test when $r_2 = 1$ and the time period of comparison is after $\widehat{\tau}$. It is well known that the Fleming-Harrington test gives more weights to departures occurring late in time when $r_1 = 0$ and $r_2 > 0$ and middle in time when $r_1 = r_2$, and this explains why it performs well in this case. Regarding the method by Zucker and Lakatos, it is quite significant only when the specified time-lag point is much larger than the true time-lag point (i.e., when $\tau^* = 2.0$). Even in this case, it has much less power compared to the proposed Wald test and the likelihood ratio test.

The empirical coverage probabilities and empirical mean length for the 95% and 99% confidence intervals for the parameters related to model (8) based on 100 repli-

cated simulations are presented in Table 3. The empirical coverage probabilities are reported as the proportions of the confidence intervals that cover the true parameters, and the empirical mean lengths of the confidence intervals are reported as the averaged lengths of the confidence intervals. From Table 3, it can be seen that (i) the empirical coverage probabilities appear to be a little smaller than the confidence levels, (ii) the empirical mean length decreases as the sample size increases, and (iii) for a given sample size, an increase in confidence level yields greater empirical coverage probabilities and wider empirical mean lengths.

Table 3 The empirical coverage probabilities and empirical mean lengths of the 95% and 99% confidence intervals for the parameters in model (8). The results are based on 100 replicated simulations. In the table, n_1 and n_2 denote the number of subjects in each group, and CR denotes the censoring rate.

CR	Parameters	Confidence levels (%)	Empirical coverage probabilities		Empirical mean lengths	
			$n_1 = n_2 = 150$	$n_1 = n_2 = 200$	$n_1 = n_2 = 150$	$n_1 = n_2 = 200$
20%	τ	95	0.87	0.87	1.377	1.224
		99	0.98	0.97	1.692	1.543
	β	95	0.91	0.88	0.992	0.860
		99	0.95	0.92	1.304	1.131
40%	τ	95	0.96	0.93	1.640	1.501
		99	0.99	0.99	1.853	1.771
	β	95	0.93	0.88	1.393	1.167
		99	0.99	0.97	1.830	1.534

Next, we consider the following more general model containing a time-independent covariate and a time-dependent covariate:

$$\lambda(t|\mathbf{X}(t),g) = \lambda_0(t)\exp\{\alpha_1 x_1 + \alpha_2 x_1 t + (\beta_0 + \beta_1 x_1)(t-\tau)I(t>\tau)g\}, \qquad (10)$$

where the baseline hazard rate is assumed to be $\lambda_0(t) = 0.4$, and the model parameter values are set to be $\alpha_1 = -1$, $\alpha_2 = 0.5$, $\beta_0 = 0.5$, $\beta_1 = -0.7$, and $\tau = 2$. The covariate x_1 is assumed to follow the Bernoulli distribution with the probability of success 0.5. The sample sizes are set to be $n_1 = n_2 = 150$ or $n_1 = n_2 = 300$. The censoring time is generated from the uniform distribution on $[1,8]$, and the resulting censoring rate is about 20%. Table 4 presents the results about the point estimates of the parameters, based on 1,000 replicated simulations. Similar conclusions can be made from this table to those from Table 1. The results of the hypothesis testing based on $H_0 : \beta_0 = \beta_1 = 0$ of various different methods are shown in Table 5. Again, only the Wald test and the likelihood ratio test can detect the difference between the two hazard rate functions in this example. Table 6 presents the empirical coverage probabilities and empirical mean lengths of the 95% and 99% confidence intervals for the parameters in model (10), based on 100 replicated simulations. Similar conclusions can be made from this table to those from Table 3.

4 Applications

In this section, we illustrate our proposed method using two real-data examples. The first one is about tumorigenesis of a drug described in Section 1 (cf., Figure 1). Rats

Table 4 Averaged point estimates and bias and mean square error (MSE) values of the estimated parameters, based on 1,000 replicated simulations for model (10). In the table, n_1 and n_2 denote the number of subjects in each group.

Method	True Parameters	$n_1 = n_2 = 150$ Point Estimates	Bias	MSE	$n_1 = n_2 = 300$ Point Estimates	Bias	MSE
Proposed	$\tau = 2$	2.097	0.097	0.459	2.046	0.046	0.324
	$\alpha_1 = -1$	-1.003	-0.003	0.062	-1.000	0.000	0.030
	$\alpha_2 = 0.5$	0.693	0.193	0.263	0.587	0.087	0.084
	$\beta_0 = 0.5$	0.508	0.008	0.020	0.503	0.003	0.009
	$\beta_1 = -0.7$	-0.943	-0.243	0.366	-0.817	-0.117	0.122
Dinse et al.	$\tau = 2$	6.268	4.268	18.554	6.626	4.626	21.736

Table 5 The proportions of rejections of various different methods for testing the difference between the two hazard rate functions related to model (10) based on a significance level 0.05. Some entries have two values, with the first one being proportion of rejections computed in the entire study period and the second one computed in the time period after $\widehat{\tau}$. The results are based on 1,000 replicated simulations. In the table, n_1 and n_2 denote the number of subjects in each group.

Testing Method	$n_1 = n_2 = 150$ Entire	After $\widehat{\tau}$	$n_1 = n_2 = 200$ Entire	After $\widehat{\tau}$
Wald test		0.951		1.000
Likelihood ratio test		0.726		0.955
Log-rank	0.028	0.078	0.034	0.134
Gehan-Wilcoxon	0.022	0.031	0.029	0.046
Tarone-Ware	0.022	0.042	0.023	0.067
Peto-Peto	0.022	0.034	0.025	0.056
Modified Peto-Peto	0.022	0.033	0.025	0.057
Fleming-Harrington ($r_1 = 0, r_2 = 1$)	0.057	0.133	0.098	0.241
Fleming-Harrington ($r_1 = 1, r_2 = 0$)	0.022	0.034	0.026	0.056
Fleming-Harrington ($r_1 = 1, r_2 = 1$)	0.025	0.099	0.028	0.121
Zucker and Lakatos ($\tau^* = 0.5$)	0.032		0.043	
Zucker and Lakatos ($\tau^* = 1.0$)	0.047		0.074	
Zucker and Lakatos ($\tau^* = 2.0$)	0.139		0.229	

Table 6 Empirical coverage probabilities and empirical mean lengths of the 95% and 99% confidence intervals for the parameters in model (10). The results are based on 100 replicated simulations. In the table, n_1 and n_2 denote the number of subjects in each group.

Parameters	Confidence levels (%)	Empirical coverage probabilities $n_1 = n_2 = 150$	$n_1 = n_2 = 200$	Empirical mean length $n_1 = n_2 = 150$	$n_1 = n_2 = 200$
τ	95	0.98	0.95	1.885	1.715
	99	1.00	0.98	1.980	1.882
α_1	95	0.94	0.92	0.971	0.684
	99	1.00	0.99	1.276	0.899
α_2	95	0.79	0.77	1.091	0.657
	99	0.87	0.91	1.433	0.864
β_0	95	0.94	0.95	0.514	0.356
	99	0.97	0.99	0.676	0.468
β_1	95	0.80	0.78	1.352	0.831
	99	0.91	0.91	1.777	1.093

were taken from 50 distinct female litters. From each litter, one rat was randomly selected and applied the drug, and another two rats were randomly selected and applied

the placebo. The event of interest was the formation of tumor, and the study lasts for 100 days. So, there were a total of 50 rats in the drug group (i.e., $g = 1$), and a total of 100 rats in the placebo group ($g = 0$). The censoring rate for the drug group was about 48%, and the censoring rate for the placebo group was about 66%. This dataset was previously analyzed by several authors, including Mantel et al (1977), and it is available in the R package survival. Because there are no covariates in this data, the following model is considered:

$$\lambda(t|g) = \lambda_0(t)\exp\{\beta(t-\tau)I(t>\tau)g\}. \tag{11}$$

By the model estimation procedure described in Subsection 2.2, the point estimate of τ is 81, its 95% confidence interval is $(72.0, 83.3)$, the point estimate of β is 0.259, and its 95% confidence interval is $(0.059, 0.458)$. Because the 95% confidence interval for β is above 0, we can conclude that the hazard rate of the drug group is significantly higher than the hazard rate of the placebo group after the estimated time-lag point of 81 days. By the way, the confidence interval for τ is obtained by the bootstrap procedure with the bootstrap sample size $B = 1,000$. The point estimate of τ by the method in Dinse et al (1993) is 92, which seems too large by checking with Figure 1. The hypothesis testing results by various different methods are presented in Table 7. For the 6 existing methods considered, besides the regular comparison of the hazard rate functions in the entire study period, we also use them to compare the hazard rate functions in the time period after $\widehat{\tau}$. In the method by Zucker and Lakatos (1990), we specify the time-lag point to be 80, 81 (the point estimate by the proposed method), or 85. From the table, it can be seen that (i) both the Wald test and likelihood ratio test provide significant results, (ii) none of the Log-rank, Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, and modified Peto-Peto could detect the difference between the two hazard rate functions in the entire study period and in the time period after $\widehat{\tau}$, (iii) the Fleming-Harrington test can detect the difference between the two hazard rate functions only in the time period after $\widehat{\tau}$ when $r_2 = 1$ or in the entire study period when $r_1 = 0$ and $r_2 = 1$, and (iv) the method by Zucker and Lakatos is significant only when the specified time-lag point is quite large.

The second example is about the sequential primary biliary cirrhosis (PBC) data collected in Mayo Clinic between 1974 and 1984. This dataset contains sequential laboratory measurements on the 312 patients of the study. A detailed description about the study and the dataset can be found in Fleming and Harrington (1991) and in the R package survival. In this data, besides the survival data, there are a number of covariates to consider, which are listed below.

- trt: 1 for the treatment D-penicillmain and 0 for placebo,
- sex: 1 for female and 0 for male,
- age: in years,
- ascites: presence of ascites,
- hepato: presence of hepatomegaly or enlarged liver,
- spiders: blood vessel malformations on the skin,
- ast: aspartate aminotransferase (U/ml),
- edema: 0 for no edema, 0.5 for moderate, and 1 for severe edema,
- bili : serum bilirunbin (mg/dI),

Table 7 Computed p-values of various different methods for testing the difference between the two hazard rate functions related to models (11) and (12) for analysing the rats data and the primary biliary cirrhosis data. Some entries have two values, with the first one being p-value computed in the entire study period and the second one computed in the time period after $\widehat{\tau}$.

Testing Method	Rats Data		PBC Data	
Wald test		0.011		0.021
Likelihood ratio test		0.003		0.017
Log-rank	0.231	0.114	0.210	0.089
Gehan-Wilcoxon	0.686	0.750	0.579	0.082
Tarone-Ware	0.512	0.379	0.370	0.078
Peto-Peto	0.454	0.293	0.486	0.073
Modified Peto-Peto	0.478	0.327	0.492	0.072
Fleming-Harrington ($r_1 = 0, r_2 = 1$)	0.038	0.001	0.046	0.290
Fleming-Harrington ($r_1 = 1, r_2 = 0$)	0.421	0.262	0.486	0.074
Fleming-Harrington ($r_1 = 1, r_2 = 1$)	0.112	0.001	0.054	0.163
Zucker and Lakatos ($\tau^* = 80$ [Rats Data], 3.0 [PBC Data])	0.114		0.102	
Zucker and Lakatos ($\tau^* = 81$ [Rats Data], 3.31 [PBC Data])	0.050		0.057	
Zucker and Lakatos ($\tau^* = 85$ [Rats Data], 4.0 [PBC Data])	0.012		0.056	

- albumin: serum albumin (mg/dl),
- alk.phos: alkaline phosphotase (U/liter),
- platelet: platelet count,
- protime: standardized blood clotting time.

In this study, we are mainly concerned about the comparison between male and female patients regarding their survival times, since previous analyses confirmed that the treatment D-penicillmain did not have significant effect (cf., Fleming and Harrington, 1991; Therneau and Grambsch, 2000). Also, the previous analyses indicated that bili, albumin, alk.phos, platelet and protime should be considered in log scale. In the data, the female group has 101 patients with about 34% censoring rate, while the male group has 21 patients with about 19% censoring rate. The two groups are highly unbalanced because it is known that primary biliary cirrhosis is more likely for women. The estimated hazard rate functions of male and female patients are shown in Figure 2. From the plot, it can be seen that the two functions are not quite different before 3.3 years and very different afterwards.

Since there are some covariates involved in the data, we first try to exclude some less important variables from our model fitting. By using the model selection procedure discussed in Park and Qiu (2014), the selected final model is

$$\begin{aligned}\lambda(t|\mathbf{X}(t),g) = \lambda_0(t)\exp\big[&\alpha_1 \text{age} + \alpha_2 I(\text{edema} = 0.5) + \alpha_3 I(\text{edema} = 1) + \alpha_4 \log(\text{bili}) \\ &+ \alpha_5 \log(\text{albumin}) + \alpha_6 \log(\text{protime}) + \beta_0 (t-\tau) I(t > \tau) g\big]\,.\end{aligned} \tag{12}$$

The estimated parameter values and the corresponding 95% confidence intervals (CIs) for model (12) are presented in Table 8. From the table, it can be seen that all parameters except α_2 are significantly different from 0 at the significance level of 0.05. The most interesting parameters in this analysis are τ and β_0. The estimates $\widehat{\tau} = 3.31$ and $\widehat{\beta}_0 = -1.613$ suggest that the risk of death for female patients and male patients is similar until $\widehat{\tau} = 3.31$ years, and the female patients have significantly less

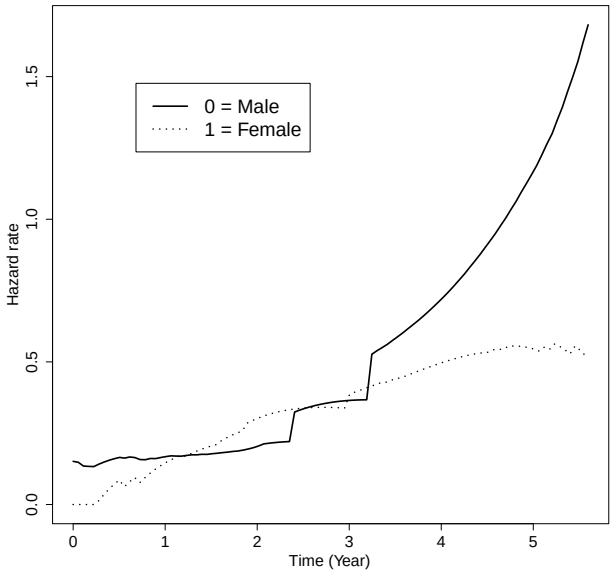

Fig. 2 Estimated hazard rate functions for the male and female patients in the primary biliary cirrhosis data.

Table 8 Point estimates and 95% confidence intervals for the parameters in model (12) in the primary biliary cirrhosis data example.

Parameters	Point estimates	95% confidence intervals
τ	3.31	(1.03 , 3.32)
α_1	0.077	(0.039, 0.114)
α_2	-0.104	(-1.199, 0.991)
α_3	1.979	(0.984, 2.973)
α_4	0.629	(0.178, 1.079)
α_5	3.983	(0.848, 7.118)
β_0	-1.613	(-2.977, -0.249)

risk afterwards. These results match those from Figure 2 well. The point estimate of τ by the method in Dinse et al (1993) is 5.596, which is again too large.

For testing the difference between the two hazard rate functions, the p-values of the Wald test, the likelihood ratio test, and the 7 alternative tests are presented in the last two columns of Table 7. From the table, it can be seen that the Wald test and likelihood ratio test based on model (12) are both significant in this example, all but one alternative methods are not, and the alternative Fleming-Harrington test with $r_1 = 0$ and $r_2 = 1$ is marginally significant.

5 Concluding Remarks

In this paper, we have presented a semiparametric modeling approach for comparing two hazard rate functions with a possible treatment time-lag effect. Model estimation and theoretical properties of the estimated model are also discussed. From the numerical examples and theoretical results, it seems that our proposed method is effective in handling cases with treatment time-lag effect. However, there are still many issues that need to be addressed in our future research. For instance, more general versions of model (1) can be considered, as discussed at the end of Subsection 2.1, and model selection and model diagnosis can be considered as well.

Acknowledgments

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

Supplementary material

Some technical details, including proofs of theoretical results from Subsection 2.2 and several lemmas, are given in a supplementary material.

References

Andersen PK, Gill RD (1982) Cox's regression model for counting processes: a large sample study. The annals of statistics 10:1100–1120

Bain L, Englehardt M (1991) Statistical analysis of reliability and life-testing models: theory and methods, vol 115. CRC Press

Chen Z, Huang H, Qiu P (2016) Comparison of multiple hazard rate functions. Biometrics 72:39–45

Cheng MY, Qiu P, Tan X, Tu D (2009) Confidence intervals for the first crossing point of two hazard functions. Lifetime data analysis 15(4):441–454

Cox DR (1972) Regression models and life-tables. Journal of the Royal Statistical Society Series B (Methodological) 34(2):187–220

Dinse GE, Piegorsch WW, Boos DD (1993) Confidence statements about the time range over which survival curves differ. Applied statistics 42:21–30

Fleming TR, Harrington DP (1991) Counting processes and survival analysis. Wiley, New York

Klein J, Moeschberger M (2005) Survival Analysis: Techniques for Censored and Truncated Data (Second Edition). Springer-Verlag, New York

Lin X, Wang H (2004) A new testing approach for comparing the overall homogeneity of survival curves. Biometrical Journal 46(5):489–496

Liu K, Qiu P, Sheng J (2007) Comparing two crossing hazard rates by cox proportional hazards modelling. Statistics in medicine 26(2):375–391

Mantel N, Bohidar NR, Ciminera JL (1977) Mantel-haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. Cancer Research 37(11):3863–3868

O'Quigley J (1994) On a two-sided test for crossing hazards. The Statistician 43:563–569

Park K, Qiu P (2014) Model selection and diagnostics for joint modeling of survival and longitudinal data with crossing hazard rate functions. Statistics in Medicine 33:4532–4546

Qiu P, Sheng J (2008) A two-stage procedure for comparing hazard rate functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70(1):191–208

Therneau TM, Grambsch PM (2000) Modeling Survival Data: Extending the Cox Model. Springer, New York

Zucker DM, Lakatos E (1990) Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. Biometrika 77(4):853–864