Biometrics 64, 724–732
September 2008
DOI: 10.1111/j.1541-0420.2007.00947.x

Estimating Cumulative Treatment Effects in the Presence of Nonproportional Hazards

Guanghui Wei* **and Douglas E. Schaubel**

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.
**email:* ghwei@umich.edu

Summary. Often in medical studies of time to an event, the treatment effect is not constant over time. In the context of Cox regression modeling, the most frequent solution is to apply a model that assumes the treatment effect is either piecewise constant or varies smoothly over time, i.e., the Cox nonproportional hazards model. This approach has at least two major limitations. First, it is generally difficult to assess whether the parametric form chosen for the treatment effect is correct. Second, in the presence of nonproportional hazards, investigators are usually more interested in the cumulative than the instantaneous treatment effect (e.g., determining if and when the survival functions cross). Therefore, we propose an estimator for the aggregate treatment effect in the presence of nonproportional hazards. Our estimator is based on the treatment-specific baseline cumulative hazards estimated under a stratified Cox model. No functional form for the nonproportionality need be assumed. Asymptotic properties of the proposed estimators are derived, and the finite-sample properties are assessed in simulation studies. Pointwise and simultaneous confidence bands of the estimator can be computed. The proposed method is applied to data from a national organ failure registry.

Key words: Confidence bands; Cumulative hazards; Observational studies; Stratification; Survival analysis; Time-dependent effect.

1. Introduction

In medical studies featuring survival time data, nonproportional hazards are very common. In Cox (1972) regression modeling, the most frequent solution is to apply a model that assumes that the treatment effect is either piecewise constant or varies smoothly over time. However, it is generally difficult to assess whether the parametric form chosen for the treatment effect is correct. Even if the correct form is chosen, investigators are usually more interested in the cumulative than the instantaneous treatment effect. This is particularly true in settings where the hazard ratio changes direction over time, in which case researchers are often interested in if and when the two survival curves cross. Therefore, we propose an estimator of the cumulative treatment effect under nonproportional hazards. Under our proposed method, the treatment effect is viewed as a process that unfolds over time and is measured by the ratio of cumulative hazards; no functional form need be assumed for the nonproportionality.

The analysis that motivated our research aims to compare survival of end-stage renal disease patients on two dialysis methods: hemodialysis (HD) and peritoneal dialysis (PD). Peritoneal dialysis is less expensive than HD, but newer and hence less established; PD has long been suspected of providing reduced survival relative to HD. The debate over PD versus HD is one of the most contentious issues in medicine and, helping to fuel the debate, previous studies have produced conflicting results (Bloembergen et al., 1995; Fenton et al., 1997). Fenton et al. (1997) compared PD to HD using nonproportional hazards models assuming a piecewise constant hazard ratio. The authors found that the hazard ratio (PD versus HD) is significantly decreased early in the follow-up period, but that the effect changed direction later on. Because the cumulative effect was not evaluated, one cannot tell which therapy is better in terms of survival based on their results. Applying our method to national registry data, we compare PD and HD covariate-adjusted survival, without assuming proportional hazards. We can estimate the time-dependent cumulative effect of PD relative to HD on mortality without assuming any functional form for that effect. The treatment effect is viewed as a process over time, which is reflected by our inference procedures.

Several methods have been proposed for the comparison of survival or cumulative hazard functions in nonparametric settings. Dabrowska, Doksum, and Song (1989) introduced a relative change function involving the survival functions for two populations and constructed pointwise confidence intervals. Simultaneous confidence bands for this function were constructed under a proportional hazards assumption. Parzen, Wei, and Ying (1997) constructed simulation-based confidence bands for the difference of survival functions. McKeague and Zhao (2002) derived simultaneous confidence bands for ratios of survival functions based on empirical likelihood. Kalbfleisch and Prentice (1981) estimated an average hazard ratio using a weight function. Because each of the above methods was designed for nonparametric settings, they would be suitable for randomized clinical trials but would generally

not apply to observational data where covariate adjustment is required. In the context of covariate adjustment, Schemper (1992) suggested the estimation of average hazard ratio of the two populations through a weighted Cox model. Xu and O'Quigley (2000) estimated the average regression effect through weighted score equation, under a nonproportional hazards model with time-varying regression coefficients.

In this article, we propose an estimator based on the treatment-specific baseline cumulative hazards estimated under a stratified Cox model. The treatment effect is viewed as a process that unfolds over time, and can be related directly to the treatment-specific survival functions. Pointwise confidence intervals and simultaneous confidence bands of our measure are constructed.

The remainder of this article is organized as follows. In the next section, the proposed measure and its estimator are described. We develop the asymptotic properties of the proposed estimator in Section 3. Section 4 evaluates the applicability of the derived asymptotic results to finite samples through simulation. In Section 5, we apply our proposed method to compare survival on HD and PD using data from a national organ failure registry. We provide some discussion of the proposed and related methods in Section 6.

2. Methods

We first set up the notation used throughout the article. Let $m + 1$ be the number of treatment groups (numbered $j = 0, 1, \ldots, m$), where the first group ($j = 0$) represents a reference category to which the remaining treatment groups are compared. The total number of subjects is denoted by n. Let T_i be the survival time for subject i. The survival time of a subject is potentially right censored, with censoring time given by C_i. The observation time and observed event indicator are given by $X_i = T_i \wedge C_i$ and $\Delta_i = I(T_i \leq C_i)$, respectively, where $a \wedge b = \min\{a, b\}$ and $I(A)$ is an indicator function taking the value 1 when condition A holds and 0 otherwise. The event counting processes are defined as $N_i(t) = \Delta_i I(X_i \leq t)$. The risk indicators are denoted by $Y_i(t) = I(X_i \geq t)$. Let G_i denote the treatment group for subject i and $G_{ij} = I(G_i = j)$. Correspondingly, we set $Y_{ij}(t) = Y_i(t)G_{ij}$ and $dN_{ij}(t) = dN_i(t)G_{ij}$. The observed data consist of n independent vectors, $(X_i,\ \Delta_i,\ G_i,\ \mathbf{Z}_i)$, where $\mathbf{Z}_i$ is a vector of adjustment covariates.

We assume that T_i follows a stratified Cox model, with hazard function

$$\lambda_{ij}(t) = \lambda_i(t \,|\, G_i = j) = \lambda_{0j}(t) \exp\left\{\boldsymbol{\beta}_0^T \mathbf{Z}_i\right\}, \tag{1}$$

where $\lambda_{0j}(t)$ is an unspecified treatment-specific baseline hazard function, and $\boldsymbol{\beta}_0$ is an unknown parameter vector. Under (1), proportionality of the hazard functions is not assumed to hold across treatment groups, but is assumed with respect to the adjustment covariates. Note that in the set-up we consider, the adjustment covariate vector is treated as time-constant. We revisit the issue of time-dependent covariates in Section 6.

The partial likelihood (Cox, 1975) estimator of $\boldsymbol{\beta}_0$ is denoted by $\hat{\boldsymbol{\beta}}$, and is given by the solution to $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ where $\mathbf{0}$ is a vector of zeros and

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{j=0}^{m} \int_0^{\tau} \{\mathbf{Z}_i - \bar{\mathbf{Z}}_j(t, \boldsymbol{\beta})\} dN_{ij}(t),$$

$$\bar{\mathbf{Z}}_j(t, \boldsymbol{\beta}) = \mathbf{S}_j^{(1)}(t, \boldsymbol{\beta}) \big/ S_j^{(0)}(t, \boldsymbol{\beta}),$$

with $\mathbf{S}_j^{(d)}(t, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} Y_{ij}(t) \mathbf{Z}_i^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i\}$ for $d = 0$, 1, 2, where $\mathbf{a}^{\otimes 0} = 1, \mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for a vector $\mathbf{a}$. The quantity τ satisfies $P(X_i > \tau) > 0$ and would ordinarily be set to the maximum observation time such that all observed events are included in the analysis.

To compare each treatment group to the reference group, we propose the following measure,

$$\theta_j(t) = \frac{\Lambda_{0j}(t)}{\Lambda_{00}(t)}, \quad \text{for } j = 1, \ldots, m, \tag{2}$$

where $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s)\, ds$ is the cumulative baseline hazard for treatment group j. Under (1), $\theta_j(t)$ can be used as a measure of the aggregate treatment effect across the $(0, t]$ interval.

Let $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)\, ds$. Note that, under model (1),

$$\frac{\Lambda_{ij}(t \,|\, \mathbf{Z}_i = \mathbf{z})}{\Lambda_{i0}(t \,|\, \mathbf{Z}_i = \mathbf{z})} = \theta_j(t).$$

That is, contrasting patients who have the same covariate pattern but receive different treatments, the ratio of cumulative hazards and ratio of baseline cumulative hazards are equal. Note also that the proposed cumulative hazard ratio reduces to the hazard ratio if proportionality holds. That is, if proportionality holds across the treatment groups, such that the model $\lambda_{ij}(t) = \lambda_0(t)\exp\{\rho_j + \boldsymbol{\beta}_0^T \mathbf{Z}_i\}$ applies, then

$$\frac{\Lambda_{ij}(t \,|\, \mathbf{Z}_i = \mathbf{z})}{\Lambda_{i0}(t \,|\, \mathbf{Z}_i = \mathbf{z})} = \exp\{\rho_j\}.$$

In this light, one could view the proposed ratio of cumulative hazards as a generalization of the familiar hazard ratio.

The proposed cumulative effect measure, $\theta_j(t)$, can be estimated by

$$\hat{\theta}_j(t) = \frac{\hat{\Lambda}_{0j}(t, \hat{\boldsymbol{\beta}})}{\hat{\Lambda}_{00}(t, \hat{\boldsymbol{\beta}})}, \quad \text{for } j = 1, \ldots, m, t \in [t_L, t_U], \tag{3}$$

where t_L is chosen sufficiently large to avoid the situation where $\hat{\Lambda}_{00}(t_L, \hat{\boldsymbol{\beta}}) = 0$, while t_U is chosen to avoid well-known instability that exists in the tail of the observation time distribution. The cumulative baseline hazards can be estimated through the Breslow (1972) estimator, $\hat{\Lambda}_{0j}(t, \hat{\boldsymbol{\beta}})$, where

$$\hat{\Lambda}_{0j}(t, \hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dN_{ij}(s)}{S_j^{(0)}(s, \hat{\boldsymbol{\beta}})}.$$

In the next section, we derive the asymptotic properties of the proposed estimator.

3. Asymptotic Properties

To derive the large-sample properties of $\hat{\theta}_j(t)$, we assume the following regularity conditions for $i = 1, \ldots, n$ and $j = 0, \ldots, m$.

(a) $(X_i,\ \Delta_i,\ G_i,\ \mathbf{Z}_i)$ are independent and identically distributed random vectors.

(b) Z_{ik} have bounded total variation, i.e., $|Z_{ik}| < \kappa$ for all $i = 1, \ldots, n$ and $k = 1, \ldots, p$, where κ is a constant and Z_{ik} is the kth component of Z_i.
(c) $\int_0^\tau \lambda_0(t)\, dt < \infty$ where τ is a prespecified time point.
(d) Continuity of the following functions:

$$\mathbf{s}_j^{(1)}(t,\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{s}_j^{(0)}(t,\boldsymbol{\beta}),\ \mathbf{s}_j^{(2)}(t,\boldsymbol{\beta}) = \frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} \mathbf{s}_j^{(0)}(t,\boldsymbol{\beta}),$$

where $\mathbf{s}_j^{(d)}(t,\ \boldsymbol{\beta})$ is the limiting value of $\mathbf{S}_j^{(d)}(t,\ \boldsymbol{\beta})$ for $d = 0,\ 1,\ 2$, with $\mathbf{s}_j^{(1)}(t,\ \boldsymbol{\beta})$ and $\mathbf{s}_j^{(2)}(t,\ \boldsymbol{\beta})$ bounded and $s_j^{(0)}(t,\ \boldsymbol{\beta})$ bounded away from 0 for $t \in [0,\ \tau]$ and $\boldsymbol{\beta}$ in an open set.
(e) Positive-definiteness of the matrix $\boldsymbol{\Omega}(\boldsymbol{\beta})$ where

$$\begin{aligned} \boldsymbol{\Omega}(\boldsymbol{\beta}) &= \sum_{j=0}^{m} \int_0^\tau \mathbf{v}_j(t,\boldsymbol{\beta}) s_j^{(0)}(t,\boldsymbol{\beta}) \lambda_{0j}(t)\, dt, \\ \mathbf{v}_j(t,\boldsymbol{\beta}) &= \mathbf{s}_j^{(2)}(t,\boldsymbol{\beta}) \big/ s_j^{(0)}(t,\boldsymbol{\beta}) - \bar{\mathbf{z}}_j(t,\boldsymbol{\beta})^{\otimes 2}, \end{aligned} \tag{4}$$

and $\bar{\mathbf{z}}_j(t,\boldsymbol{\beta}) = \mathbf{s}_j^{(1)}(t,\boldsymbol{\beta}) / s_j^{(0)}(t,\boldsymbol{\beta})$ is the limiting value of $\bar{\mathbf{Z}}_j(t,\boldsymbol{\beta})$.
(f) $P(G_{ij} = 1) > 0$.

The asymptotic behavior of our estimator is summarized by the following two theorems.

THEOREM 1: *Under conditions (a) to (f), $\hat{\theta}_j(t)$ converges to $\theta_j(t)$ almost surely and uniformly for $t \in [\tau_L,\ \tau_U]$.*

The consistency of $\hat{\theta}_j(t)$ follows from the uniform consistency of $\hat{\Lambda}_{0j}(t,\hat{\boldsymbol{\beta}}), \hat{\Lambda}_{00}(t,\hat{\boldsymbol{\beta}})$, and $\hat{\boldsymbol{\beta}}$ as well as the functional delta method (Pollard, 1990) and various results from empirical processes theory (Bilias, Gu, and Ying, 1997).

THEOREM 2: *Under conditions (a) to (f), $n^{1/2}[\hat{\theta}_j(t) - \theta_j(t)]$ converges asymptotically to a zero-mean Gaussian process with covariance function $\sigma_j(s,\ t) = E[\xi_{ij}(s,\ \boldsymbol{\beta}_0)\ \xi_{ij}(t,\ \boldsymbol{\beta}_0)]$, where:*

$$\xi_{ij}(t,\boldsymbol{\beta}) = \frac{1}{\Lambda_{00}(t)} \Phi_{ij}(t,\boldsymbol{\beta}) - \frac{\Lambda_{0j}(t)}{\Lambda_{00}(t)^2} \Phi_{i0}(t,\boldsymbol{\beta}), \tag{5}$$

$$\begin{aligned} \Phi_{ij}(t,\boldsymbol{\beta}) &= \mathbf{h}_j^T(t,\boldsymbol{\beta}) \boldsymbol{\Omega}(\boldsymbol{\beta})^{-1} \Psi_i(\boldsymbol{\beta}) \\ &\quad + \int_0^t s_j^{(0)}(s,\boldsymbol{\beta})^{-1} dM_{ij}(s,\boldsymbol{\beta}), \end{aligned} \tag{6}$$

$$\mathbf{h}_j(t,\boldsymbol{\beta}) = -\int_0^t \bar{\mathbf{z}}_j(s,\boldsymbol{\beta})\, d\Lambda_{0j}(s), \tag{7}$$

$$\Psi_i(\boldsymbol{\beta}) = \sum_{j=0}^{m} \int_0^\tau \{\mathbf{Z}_i - \bar{\mathbf{z}}_j(t,\boldsymbol{\beta})\}\, dM_{ij}(t,\boldsymbol{\beta}), \tag{8}$$

$$dM_{ij}(t,\boldsymbol{\beta}) = dN_{ij}(t) - Y_{ij}(t) \exp\left\{\boldsymbol{\beta}^T \mathbf{Z}_i\right\} d\Lambda_{0j}(t). \tag{9}$$

The covariance function can be consistently estimated by $\hat{\sigma}_j(s,t,\hat{\boldsymbol{\beta}})$ where:

$$\hat{\sigma}_j(s,t,\hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_{ij}(s,\hat{\boldsymbol{\beta}}) \hat{\xi}_{ij}(t,\hat{\boldsymbol{\beta}}), \tag{10}$$

with $\hat{\xi}_{ij}(t,\hat{\boldsymbol{\beta}})$ obtained by replacing all limiting values in $\xi_{ij}(t,\ \boldsymbol{\beta}_0)$ with their empirical counterparts.

The asymptotic normality of $n^{1/2}[\hat{\theta}_j(t) - \theta_j(t)]$ can be proved by first writing $\{\hat{\theta}_j(t) - \theta_j(t)\}$ as

$$\frac{1}{\Lambda_{00}(t)} \{\hat{\Lambda}_{0j}(t,\hat{\boldsymbol{\beta}}) - \Lambda_{0j}(t)\} + \hat{\Lambda}_{0j}(t,\hat{\boldsymbol{\beta}}) \left\{ \frac{1}{\hat{\Lambda}_{00}(t,\hat{\boldsymbol{\beta}})} - \frac{1}{\Lambda_{00}(t)} \right\}.$$

The quantity $\{\hat{\Lambda}_{00}(t,\hat{\boldsymbol{\beta}})^{-1} - \Lambda_{00}(t)^{-1}\}$ can be written as a function of $\{\hat{\Lambda}_{00}(t,\hat{\boldsymbol{\beta}}) - \Lambda_{00}(t)\}$ by using the functional delta method. The proof involves decomposing $\{\hat{\Lambda}_{0j}(t,\hat{\boldsymbol{\beta}}) - \Lambda_{0j}(t)\}$ into $\{\hat{\Lambda}_{0j}(t,\hat{\boldsymbol{\beta}}) - \hat{\Lambda}_{0j}(t,\boldsymbol{\beta}_0)\} + \{\hat{\Lambda}_{0j}(t,\boldsymbol{\beta}_0) - \Lambda_{0j}(t)\}$. The central limit theorem and various results from the theory of empirical processes are applied in the proof, which is outlined in the Web Appendix A.

Some comments on model misspecification are in order. If model (1) is misspecified, Lin and Wei (1989) demonstrated that $\hat{\boldsymbol{\beta}}$ converges to a vector $\boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0$. Further, if the true model is $\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\boldsymbol{\beta}_0^T f(\mathbf{Z}_i)\}$, while the assumed model is $\lambda_{ij}(t) = \lambda_{0j}^*(t)\exp\{\boldsymbol{\beta}^T g(\mathbf{Z}_i)\}$, where $f(\mathbf{Z}_i)$ and $g(\mathbf{Z}_i)$ are functions of covariates $\mathbf{Z}_i$, under a misspecified model, $\hat{\Lambda}_{0j}(t)$(Gerds and Schumacher, 2001) converges to $\Lambda_{0j}^*(t) \neq \Lambda_{0j}(t)$. We examine this issue numerically in Section 4.

In certain situations, investigators will want to estimate $\theta_j(t)$ at a prespecified value, $t = t_0$ (e.g., 1 year, 5 years, etc.). In these cases, inference could be based on a Wald-type test because $n^{1/2}[\hat{\theta}_j(t_0) - \theta_j(t_0)]\sigma_j(t_0)^{-1}$ will asymptotically follow a standard normal distribution, with $\sigma_j^2(t) \equiv \sigma_j(t,\ t)$. However, in many practical applications, it makes more sense to view $\theta_j(t)$ as a process over time, and this view should be captured by the corresponding inference procedures. For instance, in our motivating example, based on analyses reported in the literature, we anticipate that the effect of PD (versus HD) will depend on time and there is no single specific time point at which we wish to conduct our inference. Lin, Fleming, and Wei (1994) proposed a method to construct simultaneous confidence bands for survival curve under the Cox model. We extend this to our estimator. The idea is to approximate the normalized distribution of $\hat{Q}(t) = n^{1/2}[\hat{\theta}_j(t) - \theta_j(t)]$ for $t \in [t_L,\ t_U]$ by a zero-mean Gaussian process $\widetilde{Q}(t) = n^{-1/2}\sum_{i=1}^n \hat{\xi}_{ij}(t,\hat{\boldsymbol{\beta}})R_i$, where R_i is a standard normal random variable. The distribution of $\hat{Q}(t)$ is generated through simulation by repeatedly generating independent standard normal random samples $R_i (i = 1, \ldots, n)$. To avoid the resulting lower bound of the band being negative, we consider a log-transformed process, $n^{1/2}[\log\{\hat{\theta}_j(t)\} - \log\{\theta_j(t)\}]$, whose distribution can be approximated by $\hat{Q}(t)/\hat{\theta}_j(t)$ after applying the functional delta method. In addition, a weight function, $w(t)$, is chosen to adjust the width of the band at different time points. By using the weight function, $w(t) = \hat{\theta}_j(t)/\hat{\sigma}(t)$, suggested by Nair (1984) and the previously described simulation method, we may obtain an approximate $100(1 - \alpha)\%$ empirical quantile, $\hat{q}_\alpha$, satisfying

$$\Pr\left\{ \sup_{t\in[t_L,t_U]} \left| n^{-1/2} w(t) \hat{\theta}_j(t)^{-1} \sum_{i=1}^{n} \hat{\xi}_{ij}(t,\hat{\boldsymbol{\beta}}) R_i \right| > \hat{q}_\alpha \right\} = \alpha.$$

With the log transformation, a $100(1 - \alpha)\%$ simultaneous confidence band for $\theta_j(t)$ over $[t_L,\ t_U]$ is given by $\hat{\theta}_j(t)\exp\{\pm n^{-1/2}\hat{q}_\alpha / w(t)\}$.

4. Simulation Study

The finite sample properties of the proposed estimator were evaluated through a series of simulation studies. For convenience, we consider two treatment groups. Death times were generated as

$$T_i = \left\{-\log(U_i)/\left[\alpha_j \exp\left\{\boldsymbol{\beta}_0^T \mathbf{Z}_i\right\}\right]\right\}^{1/\gamma_j},$$

for $i = 1, \ldots, n$ and $j = 0, 1$, where U_i is a Uniform(0,1) random variable, $\boldsymbol{\beta}_0 = 0.5$, and $\mathbf{Z}_i$ is a bivariate vector with each element following a Bernoulli (0.5) distribution. This set-up implies that T_i follows a Weibull model with hazard function

$$\lambda_{ij}(t) = \lambda_i(t \mid G_i = j) = \alpha_j \gamma_j t^{\gamma_j - 1} \exp\left\{\beta_0^T Z_i\right\}.$$

Nonproportionality of the hazard functions for groups 0 and 1 is induced when $\gamma_1 \neq \gamma_0$. Various values of γ_j were used to make the hazard ratio constant, decrease, and increase through time. Censoring times were generated from a Uniform($\tau/2$, τ) distribution with $\tau = 5$. Different values of α_j were used to vary the percent of censoring (denoted by C%). For each data configuration, the no-censoring setting was also examined. We varied the sample size as $n =$ 50, 100, 200, 500, and each data configuration was replicated 1000 times. We compared the ratio of cumulative hazard to its true value at the 75th percentile of the observation time distribution, which we denote by $t_{0.75}$. Results are shown in Tables 1 and 2.

The proposed estimator generally performs well in finite samples, $n = 100$, 200, 500 (Table 1). Even in the presence of a very high proportion of censoring, the empirical mean of $\hat{\theta}_1(t)$ is approximately unbiased for sample sizes of $n = 500$ and $n = 200$, and almost all simulations with size of $n = 100$. In general, the bias is reduced as the number of subjects in each treatment group increases. The average asymptotic standard error (ASE) is generally close to the empirical standard deviation (ESD), while the coverage probabilities (CP) are consistent with the nominal value of 0.95.

For smaller sample sizes (e.g., $n = 50$), the bias of $\hat{\theta}_1$ is relatively large and the coverage probabilities are notably lower than the nominal value of 0.95 (Table 2). However, if

Table 1
Simulation results for the proposed estimator

n	γ_0	γ_1	α_0	α_1	C%	$\theta_1(t_{0.75})$	BIAS	ASE	ESD	CP
500	1.4	1.2	0.4	0.35	0%	0.739	0.002	0.084	0.087	0.94
200							0.008	0.132	0.139	0.94
100							0.020	0.188	0.206	0.92
500	1.4	1.2	0.4	0.35	10%	0.741	0.006	0.084	0.084	0.95
200							0.014	0.134	0.132	0.96
100							0.030	0.191	0.199	0.94
500	1.4	1.2	0.1	0.07	0%	0.468	0.005	0.055	0.058	0.93
200							−0.001	0.085	0.088	0.93
100							0.011	0.121	0.125	0.93
500	1.4	1.2	0.1	0.07	54%	0.536	0.001	0.081	0.077	0.95
200							0.014	0.130	0.133	0.94
100							0.017	0.184	0.188	0.93
500	1	1.5	0.5	0.3	0%	0.926	0.004	0.104	0.109	0.94
200							0.012	0.165	0.170	0.93
100							0.029	0.235	0.238	0.94
500	1	1.5	0.5	0.3	10%	0.912	0.007	0.103	0.103	0.95
200							0.019	0.165	0.170	0.94
100							0.015	0.230	0.238	0.92
500	1	1.5	0.2	0.1	0%	1.123	0.014	0.128	0.131	0.95
200							0.023	0.203	0.214	0.94
100							0.042	0.289	0.308	0.94
500	1	1.5	0.2	0.1	40%	0.933	0.006	0.121	0.118	0.95
200							0.014	0.192	0.197	0.94
100							0.016	0.272	0.287	0.93
500	1.5	1.5	0.4	0.2	0%	0.500	0.003	0.058	0.058	0.95
200							0.004	0.091	0.093	0.95
100							0.018	0.131	0.138	0.94
500	1.5	1.5	0.4	0.2	10%	0.500	0.003	0.058	0.056	0.96
200							0.002	0.091	0.091	0.94
100							0.018	0.131	0.138	0.94
500	1.5	1.5	0.1	0.05	0%	0.500	0.002	0.058	0.057	0.95
200							0.010	0.092	0.092	0.94
100							0.016	0.130	0.134	0.93
500	1.5	1.5	0.1	0.05	52%	0.500	0.009	0.075	0.076	0.94
200							0.014	0.120	0.121	0.94
100							0.014	0.169	0.177	0.93

Table 2
Simulation results for the proposed estimator ($n = 50$)

	γ_0	γ_1	α_0	α_1	C%	$\theta_1(t_{0.75})$	BIAS	ASE	ESD	CP
	1.4	1.2	0.4	0.35	0%	0.739	0.042	0.266	0.294	0.91
					10%	0.741	0.034	0.265	0.292	0.92
	1.4	1.2	0.1	0.07	0%	0.468	0.023	0.169	0.182	0.91
					54%	0.536	0.048	0.269	0.300	0.90
	1.0	1.5	0.5	0.3	0%	0.926	0.048	0.331	0.347	0.92
					10%	0.912	0.068	0.334	0.381	0.91
$\hat{\theta}_1$	1.0	1.5	0.2	0.1	0%	1.123	0.086	0.415	0.480	0.92
					40%	0.933	0.083	0.403	0.461	0.90
	1.5	1.5	0.4	0.2	0%	0.500	0.028	0.182	0.205	0.92
					10%	0.500	0.031	0.183	0.202	0.92
	1.5	1.5	0.1	0.05	0%	0.500	0.027	0.182	0.199	0.91
					52%	0.500	0.038	0.244	0.271	0.90
	γ_0	γ_1	α_0	α_1	C%	$\log\theta_1(t_{0.75})$	BIAS	ASE	ESD	CP
	1.4	1.2	0.4	0.35	0%	−0.303	−0.010	0.342	0.364	0.94
					10%	−0.300	−0.020	0.342	0.365	0.94
	1.4	1.2	0.1	0.07	0%	−0.759	−0.021	0.347	0.376	0.93
					54%	−0.624	−0.034	0.470	0.502	0.95
	1.0	1.5	0.5	0.3	0%	−0.077	−0.009	0.339	0.352	0.94
					10%	−0.092	0.001	0.341	0.381	0.93
$\log(\hat{\theta}_1)$	1.0	1.5	0.2	0.1	0%	0.117	0.003	0.342	0.373	0.92
					40%	−0.069	−0.008	0.398	0.436	0.94
	1.5	1.5	0.4	0.2	0%	−0.693	−0.018	0.348	0.387	0.93
					10%	−0.693	−0.009	0.347	0.379	0.93
	1.5	1.5	0.1	0.05	0%	−0.693	−0.016	0.347	0.377	0.92
					52%	−0.693	−0.042	0.461	0.488	0.94

$\log\hat{\theta}_1(t_{0.75})$ is considered, the bias reduces dramatically and the coverage probability is quite good. Therefore, when the sample size is very small (e.g., $n = 50$), inference should be based on $\log\theta_1(t)$.

We also looked at the performance of our estimator at various percentiles of the observation time distribution. The scenario where the hazard ratio increases with time and sample size $n = 500$ are considered. We find that our estimator is approximately unbiased even at the 10th and 90th percentiles of the observation distribution and that the coverage probabilities are close to the nominal value of 0.95 (Web Table 1).

We explored the performance of our estimator and its variance under models with functional misspecification and incorrect covariate adjustment (Web Table 2). We find that under a misspecified model, the proposed estimator is biased, although the bias is relatively small; as well, the ASE is close to the ESD.

5. Analysis of the Dialysis Data

We applied our proposed methods to compare patient survival on HD and PD. Hemodialysis served as the reference category ($j = 0$), while PD was labeled as $j = 1$. Hence, the parameter of interest is $\theta_1(t)$, which contrasts the cumulative hazard for PD relative to HD. Data were obtained from the Canadian Organ Replacement Register (CORR), a nation-wide and population-based organ failure registry that is maintained by the Canadian Institute for Health Information. The mortality hazard on dialysis was investigated for end stage renal disease patients who were either on HD or PD at the time of renal replacement therapy initiation. The dialysis method is inherently time dependent because a patient may switch therapies. We carried out two separate analyses. The first analysis, in the spirit of an intent-to-treat (ITT) analysis, classified patients based on first method of dialysis; that is, the type of dialysis received at the initiation of renal replacement therapy. The ITT analysis compares the risk of death between patients initially placed on PD (versus HD) knowing that patients may switch therapies. The second analysis censored the follow-up time at the first dialysis therapy switch (CAFS). The CAFS analysis compares the risk of death for patients who stay on PD to patients who remain on HD.

The study population included $n = 23{,}254$ registered patients aged 18 years and above who initiated dialysis between 1990 and 1998. Patients began follow-up at the date of dialysis initiation and were followed until the earliest of death, loss to follow-up, kidney transplantation, or the end of the observation period (December 31, 1998). For the ITT analysis, approximately 38% of HD patients ($n_0 = 17{,}766$) were observed to die, while 36% of patients on PD ($n_1 = 5488$) died. For the CAFS analysis, the proportion observed to die for patients on HD was approximately 30% and 25% for patients on PD. Approximately 17% of patients initially placed on HD and 27% of patients initially placed on PD switched therapy at least once.

Cox regression was employed, stratified by dialysis modality, and adjusting for age, gender, race, underlying renal diagnosis, region, and various comorbid illnesses (cerebrovascular accident, cardiovascular disease [CVD], chronic

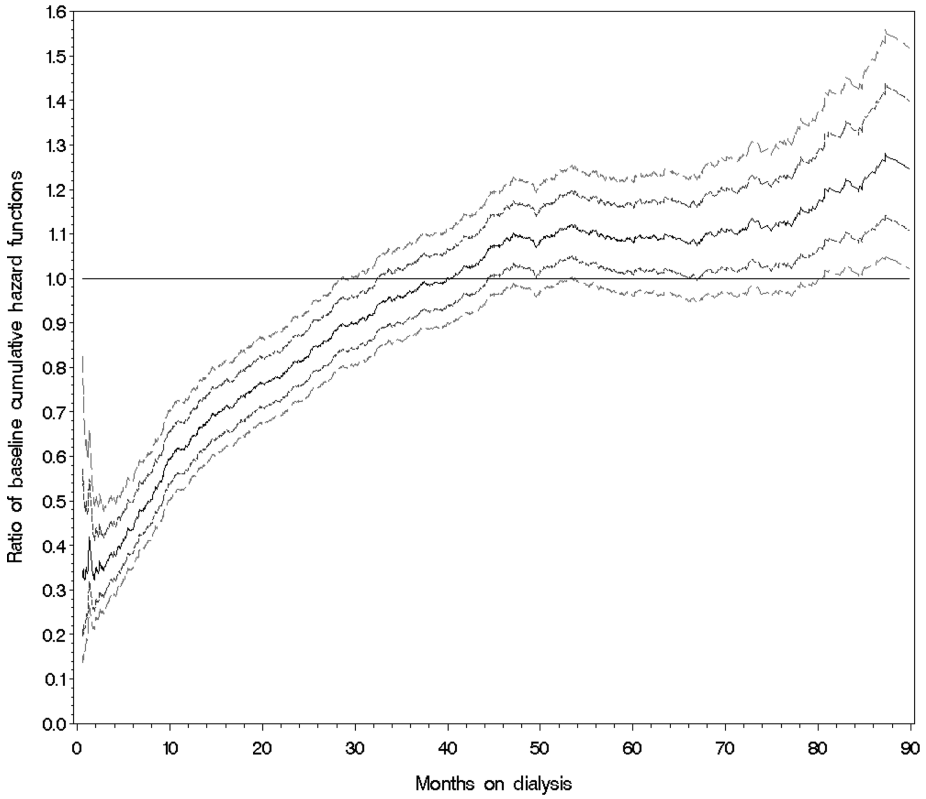

Figure 1. Estimator and 95% pointwise confidence intervals and simultaneous confidence bands for the ratio of cumulative hazard functions (PD/HD), $\theta_1(t)$, for the ITT analysis. Middle solid line: point estimate; inner lines: pointwise confidence intervals; outside lines: simultaneous confidence bands.

obstructive pulmonary disease, malignancy, peripheral vascular disease, other illnesses). Through stratification, a distinct baseline mortality hazard is allowed for HD and PD, which allows the effect of dialysis method to be nonproportional and assumes no specific functional form for the nonproportionality. The resulting 95% pointwise and simultaneous confidence bands of $\theta_1(t)$ in time interval [0.5, 90] months are given in Figure 1 for the ITT analysis and Figure 2 for the CAFS analysis. This time interval is chosen to avoid imprecision at the beginning of follow-up due to too few deaths occurring in the HD (reference) group, and instability at the tail of the observation time distribution. Based on the ITT analysis (treating dialysis method as fixed at $t = 0$), relative to HD, patients initially placed on PD had significantly increased covariate-adjusted survival probability over the [0.5, 28] months interval with $\hat{\theta}_1(t)$ ranging from a low of $\hat{\theta}_1(t) = 0.33$ at $t = 0.5$ month, to a high of $\hat{\theta}_1(t) = 0.90$ at $t = 28$ months. Survival was not significantly different for patients on PD relative to HD during the (28, 80] months interval. Long-term survival was significantly reduced for patients on PD after approximately 80 months with $\hat{\theta}_1(t) \geq 1.17$. For the CAFS analysis (censored at first therapy switch), survival probability is higher for patients on PD than HD for approximately the first 31 months, while the survival was not significantly different after that point (Figure 2).

Comparing the ITT and CAFS analyses, as is evident from Figures 1 and 2, the ITT analysis is more precise because deaths following therapy switches are not censored. In the short term, PD patients have significantly better survival under either analysis. In the long run, PD survival is not significantly different from that of HD under the CAFS analysis, but significantly lower under the ITT analysis. Supplementary analysis revealed that both $\hat{\Lambda}_0(t)$ and $\hat{\Lambda}_1(t)$ were greater for the ITT than the CAFS analysis (Web Figures 1 and 2), implying that switching therapies (in either direction) is associated with increased mortality hazard. Because PD patients were more likely than HD patients to switch, it would make sense that PD would be viewed more favorably under a CAFS (relative to ITT) approach.

6. Discussion

In the survival analysis of biomedical studies, nonproportional hazards are frequently encountered. In this manuscript, we introduce a measure of the cumulative treatment effects when the proportional hazards assumption does not hold across the treatment groups. No functional form for the nonproportionality need be assumed for our proposed estimator. In cases where hazards are in fact proportional, the proposed measure reduces to the well-known hazard ratio. Simulation studies provide evidence that the proposed estimator is approximately unbiased, while the estimated standard errors are quite accurate. Applying our method to CORR dialysis data, we found that long-term survival (after approximately 80 months) is significantly reduced for patients initially placed

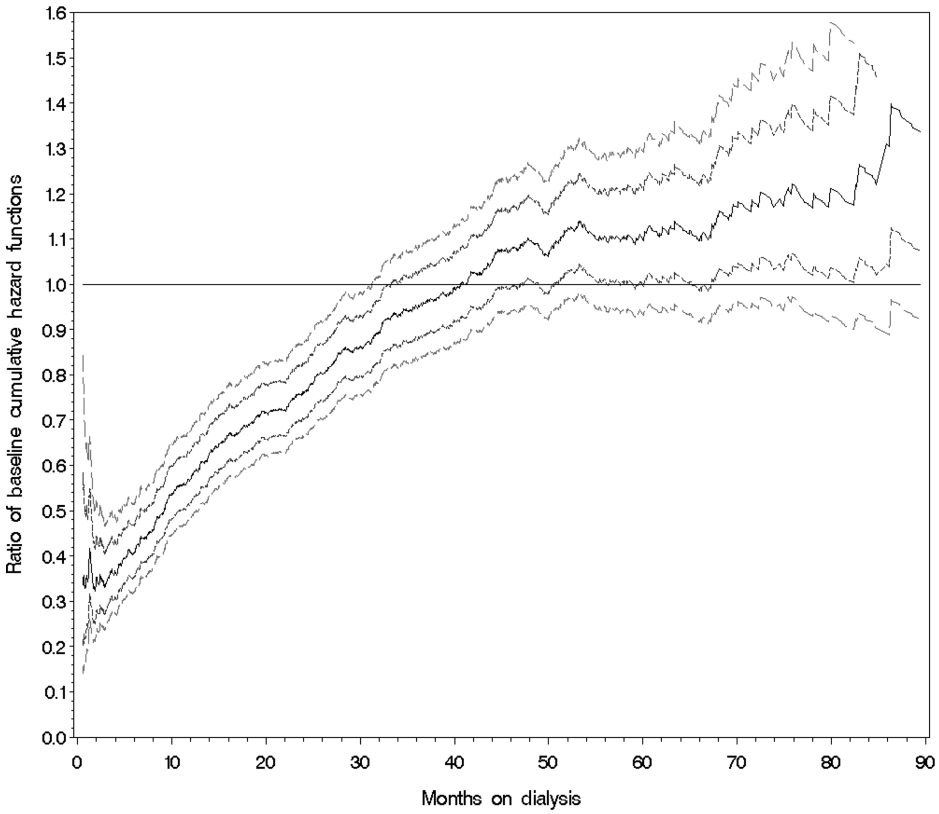

Figure 2. Estimator and 95% pointwise confidence intervals and simultaneous confidence bands for the ratio of cumulative hazard functions (PD/HD), $\theta_1(t)$, for the CAFS analysis. Middle solid line: point estimate; inner lines: pointwise confidence intervals; outside lines: simultaneous confidence bands.

on PD relative to HD (ITT analysis). The difference in long-term survival is nonsignificant after approximately the first 31 months based on the analysis with censoring at first therapy switch.

Because dialysis modality was not randomized, our results must be interpreted with caution. We did find that patients who were initially put on PD are healthier than those were put on HD in terms of comorbidity profile. This does imply that selection bias due to unmeasured covariates may be an issue.

Various methods previously proposed to account for nonproportional hazards in a Cox regression model have featured a time-varying regression coefficient, $\boldsymbol{\beta}(t)$ (e.g., Sleeper and Harrington, 1990; Zucker and Karr, 1990; Murphy and Sen, 1991; Sargent, 1997; Gustafson, 1998; Xu and O'Quigley, 2000; Martinussen, Scheike, and Skovgaard, 2002; Scheike and Martinussen, 2004). A limitation of these and related approaches is that the estimator represents an instantaneous metric and, in the presence of nonproportional hazards, investigators are usually more interested in the cumulative than the instantaneous effect. The quantity $\int_0^t \boldsymbol{\beta}(s)\,ds$ is often proposed to estimate the cumulative effect. Despite its utility, a drawback of this approach is that the integral cannot generally be connected back to the treatment-specific cumulative hazard and hence survival functions. For example, in comparing treatment ($G_i = 1$) and placebo ($G_i = 0$), $\int_0^t \boldsymbol{\beta}(s)\,ds = 0$ generally will not imply $S_0(t) = S_1(t)$ and usually it would not be straightforward without further assumptions to determine b_0 such that $\int_0^t \boldsymbol{\beta}(s)\,ds = b_0$ implies equal survival. Our proposed approach does not consider the estimation of the instantaneous treatment effect, but proposes a direct measure for the cumulative effect. In terms of the survival function, equal survival at time t among the treatments being compared is implied by $\theta_j(t) = 1$. In the situation where researchers are interested in whether and when two survival curves cross, our method is preferable. In addition, an advantage of the method proposed in this manuscript is that it is computationally straightforward.

We derived the variance for the proposed estimator using the modern theory of empirical processes, instead of the martingale central limit theorem (Fleming and Harrington, 1991). Although the asymptotic results are easier to derive using Martingale theory, the sandwich-type asymptotic variance derived through empirical processes should be more robust to model misspecification, such as missing covariate information, covariate measurement error, and mismodeling of adjustment covariates. In addition, the proposed variance could be easily extended to recurrent event setting, wherein the event of interest can be experienced more than once per subject. When proportionality does not hold across the treatment groups, we could fit a stratified version of the proportional means model (Lin et al., 2000), $E[N_{ij}(t)] \equiv \mu_{ij}(t) = \mu_{0j}(t)\exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i\}$, for $i = 1, \ldots, n$, where $\mu_{0j}(t)$ is unspecified baseline mean function for the jth treatment group. Among the methods

available for recurrent event data (e.g., see Cai and Schaubel, 2004), the marginal means approach of Lin et al. (2000) would be considered a suitable method for comparing treatments. To compare treatment group $j(>0)$ to the reference group ($j = 0$), one could use the ratio of the mean numbers of events, $\theta_j^*(t) = \mu_{0j}(t)/\mu_{00}(t)$ as a metric for the cumulative treatment effect. The estimate for $\theta_j^*(t)$ has the same expression as in the univariate survival case, but with $N_{ij}(t)$ representing the number of events in $(0, t]$ instead of a time-dependent observed death indicator. The asymptotic results would be essentially the same after adding the condition that $N_{ij}(t) < \eta < \infty$. The asymptotic variance of $\hat{\theta}_j^*(t)$ could be consistently estimated by that based on Theorem 2 of the current report, upon replacing $\Lambda_{0j}(t)$ with $\mu_{0j}(t)$.

In this article, our focus has been on the treatment effect. When the proportional hazard assumption does not hold for an adjustment covariate, traditional methods can be applied to remedy the nonproportionality, e.g., interactions with t.

Note that our proposed estimation procedure considers the case where the adjustment covariate vector is assumed to be time independent. This is not a limitation for at least two important reasons. First, the assumption of time-independent adjustment covariates matches the reality in most cases, such as the application in Section 5. Second, in settings where $\mathbf{Z}_i(t) \neq \mathbf{Z}_i$, it would be preferable to use $\mathbf{Z}_i = \mathbf{Z}_i(0)$ (i.e., the baseline covariate value) anyway, due to interpretation issues. For ease of illustration, suppose treatment is fixed at $t = 0$ but that the adjustment covariate, $Z_i(t)$, varies over time; $Z_i(0)$, as opposed to $Z_i(t)$, is included as an adjustment covariate. Consider two cases: (i) $Z_i(t)$ is uncorrelated with treatment (ii) $Z_i(t)$ is correlated with treatment. In case (i), $\hat{\theta}_1(t)$ would be estimating the same quantity whether or not the adjustment covariate was coded as time dependent, rendering the use of $Z_i(t)$ (in place of Z_i) unnecessary. In case (ii), $\hat{\theta}_1(t)$ could be substantially biased towards 1 if the adjustment covariate was coded as time dependent in the model. If $Z_i(t)$ is correlated with treatment after adjusting for $Z_i(0)$, it is much more likely that treatment is at least in part causing the variation in $Z_i(t)$, directly or indirectly, than the other way around, i.e., considering the temporality. Take the dialysis data in Section 5 as an example. We adjust for comorbid conditions, which are coded at time $t = 0$. In the CORR database, serial comorbidity data are not available. But, even if they were, we would prefer to compare PD and HD only adjusting for time 0 comorbidity. It is quite plausible that, in addition to affecting the mortality hazard, dialysis method has other intermediate consequences relating to (for example) hospitalizations and the incidence of comorbid conditions. Suppose that PD (relative to HD) reduces mortality and decreases the incidence of CVD, and that CVD onset increases mortality risk. If we adjust for time-dependent CVD, then we end up, essentially, comparing PD and HD patients of similar prognosis, therefore underestimating the magnitude of the difference between therapies with respect to mortality. In understanding this phenomenon, it helps to think of time-dependent covariates as intermediate end-points. It is well known in survival analysis that adjusting for components of the causal pathway is inappropriate; as is made clear in survival-related causal inference approaches, e.g., Robins and Greenland (1994), Hernan, Brumback, and Robins (2001), who proposed marginal structural models for use when adjustment covariates are time dependent. If time-dependent comorbidity data were available, they could perhaps be incorporated by a marginal structural-type extension of the methods proposed in this article.

7. Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 4 and 5 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

The authors thank the CORR of the Canadian Institute for Health Information for providing the access to their end stage renal disease data. They also thank the editor, associate editor, and reviewers for their constructive comments and suggestions, which led to substantial improvement of the manuscript. This work was supported by National Institutes of Health grant R01 DK-70869.

References

Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a general asymptotic theory for the Cox model with staggered entry. *The Annals of Statistics* **25,** 662–682.

Bloembergen, W. E., Port, F. K., Mauger, E. A., and Wolfe, R. A. (1995). A comparison of mortality between patients treated with hemodialysis and peritoneal dialysis. *Journal of American Society of Nephrology* **6,** 177–183.

Breslow, N. (1972). Contribution to the discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Cai, J. and Schaubel, D. E. (2004). Analysis of recurrent event data. *Handbook of Statistics* **23,** 603–623.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34,** 187–202.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Dabrowska, D. M., Doksum, K. A., and Song, J. (1989). Graphical comparison of cumulative hazards for two populations. *Biometrika* **76,** 763–773.

Fenton, S. S., Schaubel, D. E., Desmeules, M., Morrison, H. I., Mao, Y., Copleston, P., Jeffery, J. R., and Kjellstrand, C. M. (1997). Hemodialysis versus peritoneal dialysis: A comparison of adjusted mortality rates. *American Journal of Kidney Diseases* **30,** 334–342.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Gerds, T. A. and Schumacher, M. (2001). On functional misspecification of covariates in the Cox regression model. *Biometrika* **88,** 572–580.

Gustafson, P. (1998). Flexible Bayesian modeling for survival data. *Lifetime Data Analysis* **4,** 281–299.

Hernan, M. A., Brumback, B., and Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* **96,** 440–448.

Kalbfleisch, J. D. and Prentice, R. L. (1981). Estimation of the average hazard ratio. *Biometrika* **68,** 105–112.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84,** 1074–1078.

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81,** 73–81.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62,** 711–730.

Martinussen, T., Scheike, T. H., and Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* **29,** 57–74.

McKeague, I. W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics & Probability Letters* **60,** 405–415.

Murphy, S. A. and Sen, P. K. (1991). Time dependent coefficients in a Cox-type regression model. *Stochastic Processes and Their Applications* **39,** 153–180.

Nair, V. N. (1984). Confidence bands for survival functions with censored data: A comparative study. *Technometrics* **26,** 265–275.

Parzen, M. I., Wei, L. J., and Ying, Z. (1997). Simultaneous confidence intervals for the difference of the two survival functions. *Scandinavian Journal of Statistics* **24,** 309–314.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Robins, J. M. and Greenland, S. (1994). Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* **89,** 737–749.

Sargent, D. J. (1997). A flexible approach to time-varying coefficients in the Cox regression setting. *Lifetime Data Analysis* **3,** 13–25.

Scheike, T. H. and Martinussen, T. (2004). On estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics* **31,** 51–62.

Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *The Statistician* **41,** 455–465.

Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association* **85,** 941–949.

Xu, R. and O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1,** 423–439.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *The Annals of Statistics* **18,** 329–353.

Received March 2007. *Revised September* 2007.
Accepted September 2007.