

# Semiparametric Inference on the Absolute Risk Reduction and the Restricted Mean Survival Difference

Song Yang

**Abstract** For time-to-event data, when the hazards may be non-proportional, in addition to the hazard ratio, the absolute risk reduction and the restricted mean survival difference can be used to describe the time-dependent treatment effect. The absolute risk reduction measures the direct impact of the treatment on event rate or survival, and the restricted mean survival difference provides a way to evaluate the cumulative treatment effect. However, in the literature, available methods are limited for flexibly estimating these measures and making inference on them. In this article, point estimates, pointwise confidence intervals and simultaneous confidence bands of the absolute risk reduction and the restricted mean survival difference are established under a semiparametric model that can be used in a sufficiently wide range of applications. These methods are motivated by and illustrated for data from the Women's Health Initiative estrogen plus progestin clinical trial.

## Introduction

Comparison of two groups of survival data has wide applications in life testing, reliability studies, and clinical trials. Often the two sample proportional hazards model of Cox [4] is assumed and a single value of the hazard ratio is used to describe the group difference. When the hazard ratio is possibly time-dependent, a conventional approach is to give a hazard ratio estimate over each of a few time periods, by fitting a piece-wise proportional hazards model. Alternatively, a “defined” time-varying covariate can be used in a Cox regression model, resulting

---

S. Yang (✉)

Office of Biostatistics Research, National Heart, Lung, and Blood Institute,  
6701 Rockledge Dr. MSC 7913, Bethesda, MD 20892, USA  
e-mail: [yangso@nhlbi.nih.gov](mailto:yangso@nhlbi.nih.gov)

in a parametric form for the hazard ratio function (e.g. [6], Chap. 6). With these approaches, it may not be easy to pre-specify the partition of the time axis or the parametric form of the hazard ratio function.

In Yang and Prentice [22], a short-term and long-term hazards model was proposed. Assume absolutely continuous failure times and label the two groups control and treatment, with hazard functions  $\lambda_C(t)$  and  $\lambda_T(t)$ , respectively. Then the short-term and long-term hazards model postulates that

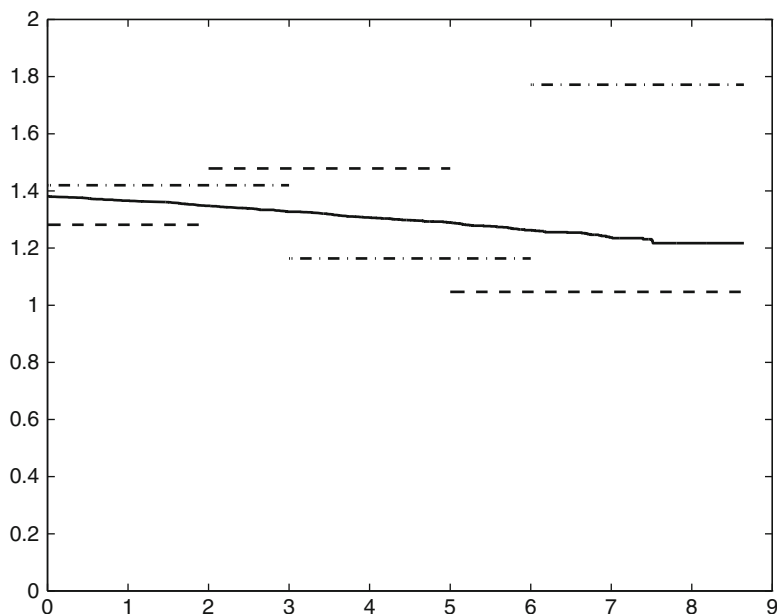
$$\lambda_T(t) = \frac{1}{e^{-\beta_2} + (e^{-\beta_1} - e^{-\beta_2})S_C(t)} \lambda_C(t), \quad t < \tau_0, \quad (1)$$

where  $\beta_1, \beta_2$  are scalar parameters,  $S_C$  is the survivor function of the control group, and

$$\tau_0 = \sup\{x : \int_0^x \lambda_C(t)dt < \infty\}. \quad (2)$$

Under this model,  $\lim_{t \downarrow 0} \lambda_T(t)/\lambda_C(t) = e^{\beta_1}$ ,  $\lim_{t \uparrow \tau_0} \lambda_T(t)/\lambda_C(t) = e^{\beta_2}$ . Thus various patterns of the hazard ratio can be realized, including proportional hazards, no initial effect, disappearing effect, and crossing hazards. In particular, model (1) includes the proportional hazards model and the proportional odds model as special cases. There is no need to specify a partition of the time axis or a parametric form of the hazard ratio function. For this model, Yang and Prentice [22] proposed a pseudo-likelihood method for estimating the parameters, and Yang and Prentice [23] studied inference procedures on the hazard ratio function. Extension of model (1) to the regression setting was also studied for current status data in Tong et al. [20].

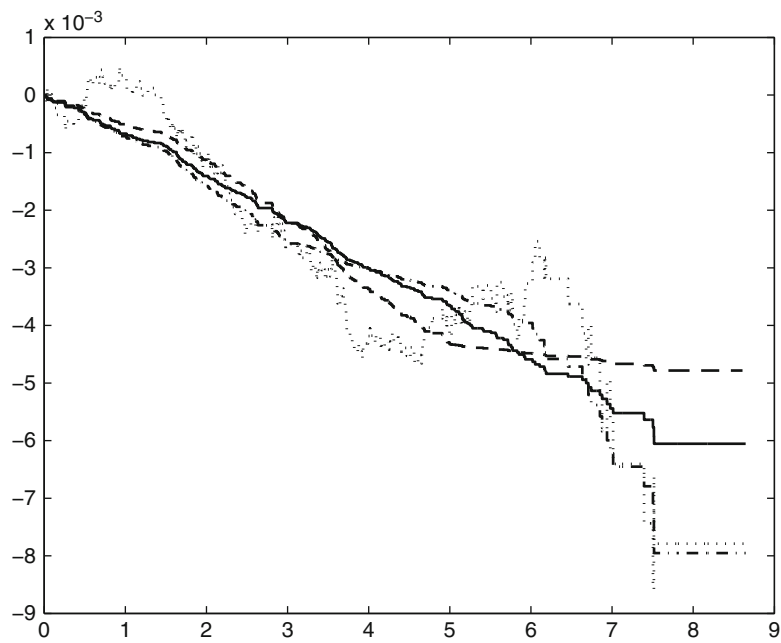
In situations with non-proportional hazards, the hazard ratio is useful for assessing temporal trend of the treatment effect, but it may not directly translate to the survival experience. For example, the hazard ratio may be less than 1 in a region where there is no improvement in the survival probability. Also, there is no simple nonparametric estimator as a reference when comparing different estimators of the hazard ratio function. In the Women's Health Initiative estrogen plus progestin clinical trial [10, 21], the hazard ratio function was decidedly non-proportional for the outcomes of coronary heart disease, venous thromboembolism, and stroke. While the estimated hazard ratios from Prentice et al. [16] and Yang and Prentice [23] are in good agreement with each other for the outcomes of coronary heart disease and venous thromboembolism, they indicate somewhat different hazard ratio shapes for stroke. Under the piece-wise Cox model with the partition of 0–2, 2–5, and 5+ years (the partition used in [16]), the hazard ratio has an upside down U-shape. On the other hand, under the piece-wise Cox model using the partition of 0–3, 3–6, and 6+ years (a plausible partition since the maximum follow-up time was almost 9 years), the hazard ratio has a U-shape. The result from Yang and Prentice [23] shows a hazard ratio that is slightly decreasing over time. Thus for stroke, the temporal trend of the hazard ratio is portrayed somewhat differently under these models. These hazard ratio estimates are displayed in Fig. 1.



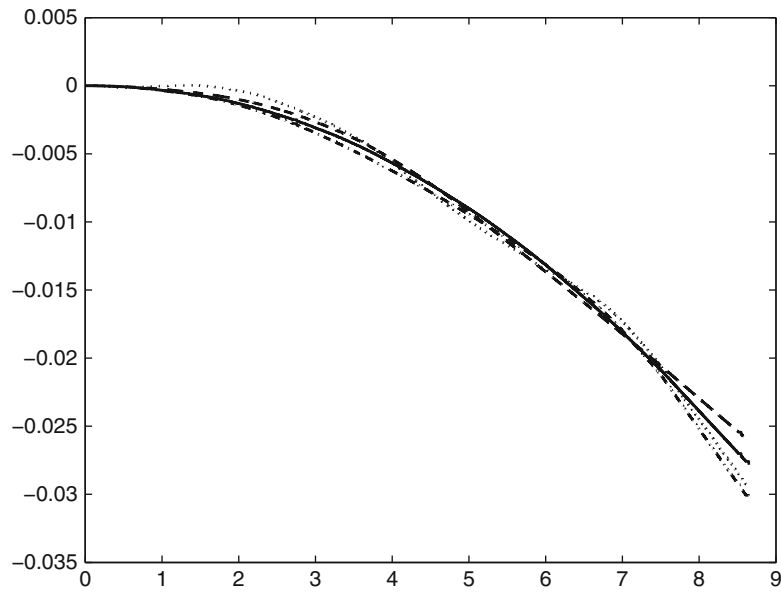
**Fig. 1** Estimated hazard ratio for the WHI clinical trial stroke data: *Solid line*—Model (1); *Dashed line*—Piece-wise Cox model with cut points at 2 and 5 years; *Dash-dotted lines*—Piece-wise Cox model with cut points at 3 and 6 years

To help compare these different results, one can consider the absolute risk reduction by the treatment. Figure 2 displays various estimators of the absolute risk reduction. From Fig. 2, several observations can be made. Between the two piece-wise Cox models with different partitions, the partition with cut points 2 and 5 years results in a better agreement with the Kaplan-Meier [7] based estimator for the early to middle portion of the data range. The other partition results in a better agreement with the Kaplan-Meier based estimator for the range beyond 6 years. The estimator based on model (1) is a good compromise between the results from the two partitions. One more comparison of these models can be made through the restricted mean survival difference, displayed in Fig. 3. The different estimators are closer to each other and are also smoother. For the piece-wise Cox models, the partition with cut points 2 and 5 years results in an estimator that is closer to the Kaplan-Meier estimator for early part of the data range, but has a more noticeable deviation near the end. Again the estimator based on model (1) results in a good compromise between the two partitions.

In this article, we consider making semiparametric inference on the absolute risk reduction and the restricted mean survival difference for two sample time-to-event data, under model (1). The absolute risk reduction is directly related to the survival experience, and is a commonly used measure in epidemiological studies. The restricted mean survival time has been used as a summary measure in various works when the hazards are non-proportional. The restricted mean survival time



**Fig. 2** Estimated absolute risk reduction for the WHI clinical trial stroke data: *Solid line*—Model (1); *Dotted line*: Kaplan-Meier; *Dashed line*—Piece-wise Cox model with cut points at 2 and 5 years; *Dash-dotted lines*—Piece-wise Cox model with cut points at 3 and 6 years



**Fig. 3** Estimated mean restricted survival difference for the WHI clinical trial stroke data: *Solid line*—Model (1); *Dotted line*: Kaplan-Meier; *Dashed line*—Piece-wise Cox model with cut points at 2 and 5 years; *Dash-dotted lines*—Piece-wise Cox model with cut points at 3 and 6 years

up to  $t$  can be thought of as the ‘ $t$ -year life expectancy’, and it approaches the unrestricted mean survival time as  $t$  approaches infinity. In clinical trials where the trial often ends after a pre-specified follow-up period, the restricted mean survival time is a more appropriate measure than the unrestricted mean survival time. In the subsequent development, the estimates, point-wise confidence intervals and simultaneous confidence bands of the absolute risk reduction and the restricted mean survival difference will be established under model (1). Such semiparametric inference procedures are sufficiently flexible for many applications, due to the various properties of model (1) mentioned before. These confidence intervals and confidence bands can be used to capture and graphically present the treatment effect. We illustrate these visual tools through applications to the clinical trial data from the Women’s Health Initiative.

There have been various works in the literature that are related to the problems considered here. Recently Schaubel and Wei [18] considered the restricted mean survival difference and other measures under dependent censoring. Royston and Parmar [17] considered inference on the restricted mean survival time by extending standard survival models to accommodate a wide range of baseline distributions. In both works, point-wise confidence intervals are constructed. In earlier works, Dabrowska et al. [5] introduced a relative change function defined in terms of cumulative hazards and found simultaneous bands for this function under the assumption of proportional hazards. Parzen et al. [13] constructed nonparametric simultaneous confidence bands for the survival probability difference. Cheng et al. [3] proposed pointwise and simultaneous confidence interval procedures for the survival probability under semiparametric transformation models. Zucker [24] and Chen and Tsiatis [2] compared the restricted mean survival time between two groups using Cox proportional hazards models. McKeague and Zhao [11] proposed simultaneous confidence bands for ratios of survival functions via the empirical likelihood method.

The article is organized as follows. In section “The Estimators and Their Asymptotic Properties” the short-term and long-term hazard ratio model and the parameter estimator are described. Pointwise confidence intervals are established for the absolute risk reduction and the restricted mean survival difference under the model. In section “Simultaneous Confidence Bands”, simultaneous confidence bands are developed for the absolute risk reduction and the restricted mean survival difference. Simulation results are presented in section “Simulation Studies”. Application to the stroke data from the Women’s Health Initiative trial is given in section “Application”. Some discussions are given in section “Discussion”.

## The Estimators and Their Asymptotic Properties

Let  $T_1, \dots, T_n$  be the pooled lifetimes of the two groups, with  $T_1, \dots, T_{n_1}$ ,  $n_1 < n$ , constituting the control group having the survivor function  $S_C$ . Let  $C_1, \dots, C_n$  be the censoring variables, and  $Z_i = I(i > n_1)$ ,  $i = 1, \dots, n$ , where  $I(\cdot)$  is the

indicator function. The available data consist of the independent triplets  $(X_i, \delta_i, Z_i)$ ,  $i = 1, \dots, n$ , where  $X_i = \min(T_i, C_i)$  and  $\delta_i = I(T_i \leq C_i)$ . We assume that  $T_i$ ,  $C_i$  are independent given  $Z_i$ . The censoring variables ( $C_i$ 's) need not be identically distributed, and in particular the two groups may have different censoring patterns. For  $t < \tau_0$  with  $\tau_0$  defined in (2), let  $R(t)$  be the odds function  $1/S_C(t) - 1$  of the control group. The model of Yang and Prentice [22] can be expressed as

$$\lambda_i(t) = \frac{1}{e^{-\beta_1 Z_i} + e^{-\beta_2 Z_i} R(t)} \frac{dR(t)}{dt}, \quad i = 1, \dots, n, t < \tau_0,$$

where  $\lambda_i(t)$  is the hazard function for  $T_i$  given  $Z_i$ .

Let  $S_T$  be the survivor function of the treatment group. Then the absolute risk reduction is

$$\Phi(t) = S_T(t) - S_C(t).$$

This function is positive if the treatment reduces the event rate and negative if the treatment increases the event rate. Under model (1),  $\Phi(t)$  depends on the parameter  $\beta = (\beta_1, \beta_2)^T$  and the baseline function  $R(t)$ , where “ $T$ ” denotes transpose. Yang and Prentice [22] studied a pseudo likelihood estimator  $\hat{\beta}$  of  $\beta$  which we describe below.

Let  $\tau < \tau_0$  be such that

$$\lim_n \sum_{i=1}^n I(X_i \geq \tau) > 0, \quad (3)$$

with probability 1. For  $t \leq \tau$ , define

$$\begin{aligned} \hat{P}(t; \mathbf{b}) &= \prod_{s \leq t} \left( 1 - \frac{\sum_{i=1}^n \delta_i e^{-b_2 Z_i} I(X_i = s)}{\sum_{i=1}^n I(X_i \geq s)} \right), \\ \hat{R}(t; \mathbf{b}) &= \frac{1}{\hat{P}(t; \mathbf{b})} \int_0^t \frac{\hat{P}_-(s; \mathbf{b})}{\sum_{i=1}^n I(X_i \geq s)} d\left(\sum_{i=1}^n \delta_i e^{-b_1 Z_i} I(X_i \leq s)\right), \end{aligned}$$

where  $\hat{P}_-(s; \mathbf{b})$  denotes the left continuous (in  $s$ ) version of  $\hat{P}(s; \mathbf{b})$ .

Let  $L(\beta, R)$  be the likelihood function of  $\beta$  under model (1) when the function  $R(t)$  is known, with the corresponding score vector  $S(\beta, R) = \partial \ln L(\beta, R) / \partial \beta$ . Define  $Q(\mathbf{b}) = S(\mathbf{b}, R)|_{R(t) = \hat{R}_n(t; \mathbf{b})}$ . Then the pseudo maximum likelihood estimator  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$  of  $\beta$  is the zero of  $Q(\mathbf{b})$ .

Once  $\hat{\beta}$  is obtained,  $R(t)$  can be estimated by  $\hat{R}(t; \hat{\beta})$ . Thus under model (1), the absolute risk reduction  $\Phi(t)$  can be estimated by

$$\hat{\Phi}(t) = \{1 + e^{-\hat{\beta}_2 + \hat{\beta}_1 \hat{R}(t; \hat{\beta})}\}^{-e^{\hat{\beta}_2}} - \frac{1}{1 + \hat{R}(t; \hat{\beta})}. \quad (4)$$

In Appendix 1, we show that  $\hat{\Phi}(t)$  is strongly consistent for  $\Phi(t)$  under model (1).

To study the distributional properties of  $\hat{\Phi}(t)$ , let

$$U_n(t) = \sqrt{n}(\hat{\Phi}(t) - \Phi(t)), \quad t \leq \tau.$$

Let  $\xi_0(t) = 1 + R(t)$ ,  $\xi(t) = e^{-\beta_1} + e^{-\beta_2}R(t)$ ,  $\hat{\xi}_0(t) = 1 + \hat{R}(t; \beta)$ ,  $\hat{\xi}(t) = e^{-\beta_1} + e^{-\beta_2}\hat{R}(t; \beta)$ , and define

$$K_1(t) = \sum_{i \leq n_1} I(X_i \geq t), \quad K_2(t) = \sum_{i > n_1} I(X_i \geq t),$$

$$A(t) = \frac{1}{\hat{\xi}(t)}(e^{-\beta_1}, e^{-\beta_2}\hat{R}(t; \beta))^T,$$

$$B(t) = \int_t^\tau \frac{A(s)K_1(s)K_2(s)}{\hat{\xi}(s)\hat{P}(s; \beta)} \left( \frac{e^{-\beta_2}}{\xi(s)} - \frac{1}{\xi_0(s)} \right) dR(s).$$

In Appendix 1, it will be shown that, with probability 1,

$$Q(\beta) = \sum_{i \leq n_1} \int_0^\tau \{\mu_1(t) + o(1)\} dM_i(t) + \sum_{i > n_1} \int_0^\tau \{\mu_2(t) + o(1)\} dM_i(t), \quad (5)$$

uniformly in  $t \leq \tau$  and  $i \leq n$ , where

$$\begin{aligned} \mu_1(t) &= -\frac{\hat{\xi}_0(t)A(t)K_2(t)}{\hat{\xi}(t)K(t)} + \frac{\hat{\xi}_0(t)\hat{P}_-(t; \beta)}{K} B(t), \\ \mu_2(t) &= A(t) \frac{K_1(t)}{K(t)} + \frac{\hat{\xi}(t)\hat{P}_-(t; \beta)}{K(t)} B(t), \end{aligned} \quad (6)$$

$$M_i(t) = \delta_i I(X_i \leq t) - \int_0^t I(X_i \geq s) \frac{dR(s)}{e^{-\beta_1}Z_i + e^{-\beta_2}Z_i R(s)}, \quad i = 1, \dots, n.$$

By Lemma A3 of Yang and Prentice [22],

$$\sqrt{n}(\hat{R}(t; \beta) - R(t)) = \frac{1}{\sqrt{n}\hat{P}(t; \beta)} \left( \sum_{i \leq n_1} \int_0^t v_1 dM_i + \sum_{i > n_1} \int_0^t v_2 dM_i \right) \quad (7)$$

where

$$v_1(t) = \frac{n\hat{\xi}_0(t)\hat{P}_-(t; \beta)}{K(t)}, \quad v_2(t) = \frac{n\hat{\xi}(t)\hat{P}_-(t; \beta)}{K(t)}.$$

Let  $\Lambda_T$  be the cumulative hazard function of the treatment group and define

$$C(t) = \frac{1}{\hat{P}(t; \beta)} \left( \frac{1}{\xi_0^2(t)} - \frac{S_T(t)}{\xi(t)} \right), \quad \Omega = \left\{ -\frac{1}{n} \frac{\partial Q(\beta)}{\partial \beta} \right\}^{-1},$$

$$D(t) = C(t) \hat{P}(t; \beta) \frac{\partial \hat{R}(t; \beta)}{\partial \beta} - S_T(t) \left( \frac{R(t)}{\xi(t)}, \Lambda_T(t) - \frac{R(t)}{\xi(t)} \right)^T.$$

For  $t \leq \tau$ , define the process

$$\begin{aligned} \tilde{U}_n(t) = & \frac{D^T(t) \Omega}{\sqrt{n}} \left( \sum_{i \leq n_1} \int_0^t \mu_1 dM_i + \sum_{i > n_1} \int_0^t \mu_2 dM_i \right) \\ & + \frac{C(t)}{\sqrt{n}} \left( \sum_{i \leq n_1} \int_0^t v_1 dM_i + \sum_{i > n_1} \int_0^t v_2 dM_i \right). \end{aligned} \quad (8)$$

With the representations for  $Q(\beta)$  and  $\sqrt{n}(\hat{R}(t; \beta) - R(t))$ , in Appendix 2 it will be shown that  $U_n$  is asymptotically equivalent to  $\tilde{U}_n$  which converges weakly to a zero-mean Gaussian process  $U^*$ . The weak convergence of  $U_n$  thus follows. The limiting covariance function  $\sigma_\Phi(s, t)$  of  $U^*$  involves the derivative vector  $\partial \hat{R}(t; \beta) / \partial \beta$  and the derivative matrix in  $\Omega$ . Although analytic forms of these derivatives are available, they are quite complicated and cumbersome. Here we approximate them by numerical derivatives. For the functions  $C(t)$ ,  $D(t)$ ,  $\mu_1(t)$ ,  $\mu_2(t)$ ,  $v_1(t)$ ,  $v_2(t)$ , define corresponding  $\hat{C}(t)$ ,  $\hat{D}(t)$ ,  $\dots$  by replacing  $\beta$  with  $\hat{\beta}$ ,  $R(t)$  with  $\hat{R}(t; \hat{\beta})$ ,  $S_T(t)$  and  $\Lambda(t)$  with model based estimators, and  $\partial \hat{R}(t; \beta) / \partial \beta$  with the numerical derivatives. Similarly, let  $\hat{\Omega}$  be the numerical approximation of  $\Omega$ . Simulation studies show that the results are fairly stable with respect to the choice of the jump size in the numerical derivatives, and that the choice of  $n^{-1/2}$  works well. With these approximations, the covariation process  $\sigma_\Phi(s, t)$ ,  $s \leq t \leq \tau$ , can be estimated by

$$\begin{aligned} \hat{\sigma}_\Phi(s, t) = & \hat{D}^T(s) \hat{\Omega} \left( \int_0^\tau \frac{\hat{\mu}_1(w) \hat{\mu}_1^T(w) K_1(w) d\hat{R}(w; \hat{\beta})}{n(1 + \hat{R}(w; \hat{\beta}))} \right. \\ & + \int_0^\tau \frac{\hat{\mu}_2(w) \hat{\mu}_2^T(w) K_2(w) d\hat{R}(w; \hat{\beta})}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \Big) \hat{\Omega}^T \hat{D}(t) \\ & + \hat{C}(s) \hat{C}(t) \left( \int_0^s \frac{\hat{v}_1^2(w) K_1(w) d\hat{R}(w; \hat{\beta})}{n(1 + \hat{R}(w; \hat{\beta}))} \right. \\ & + \int_0^s \frac{\hat{v}_2^2(w) K_2(w) d\hat{R}(w; \hat{\beta})}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \Big) \\ & + \hat{C}(t) \hat{D}^T(s) \hat{\Omega} \left( \int_0^t \frac{\hat{\mu}_1(w) \hat{v}_1(w) K_1(w) d\hat{R}(w; \hat{\beta})}{n(1 + \hat{R}(w; \hat{\beta}))} \right. \end{aligned}$$



$$\begin{aligned}
& + \int_0^t \frac{\hat{\mu}_2(w) \hat{v}_2(w) K_2(w) d\hat{R}(w; \hat{\beta})}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \\
& + \hat{C}(s) \hat{D}^T(t) \hat{\Omega} \left( \int_0^s \frac{\hat{\mu}_1(w) \hat{v}_1(w) K_1(w) d\hat{R}(w; \hat{\beta})}{n(1 + \hat{R}(w; \hat{\beta}))} \right. \\
& \left. + \int_0^s \frac{\hat{\mu}_2(w) \hat{v}_2(w) K_2(w) d\hat{R}(w; \hat{\beta})}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \right). \tag{9}
\end{aligned}$$

For a fixed  $t_0 \leq \tau$ , from the above results, an asymptotic  $100(1 - \alpha)\%$  confidence interval for  $\Phi(t_0)$  is  $\hat{\Phi}(t_0) \mp z_{\alpha/2} \sqrt{\hat{\sigma}_{\Phi}(t_0, t_0)/n}$ , where  $z_{\alpha/2}$  is the  $100(1 - \alpha/2)\%$  percentile of the standard normal distribution.

Now let us look at the restricted mean survival difference

$$\Psi(t) = \int_0^t S_T(s) ds - \int_0^t S_C(s) ds.$$

Under model (1),  $\Psi(t)$  is estimated by

$$\hat{\Psi}(t) = \int_0^t \hat{\Phi}(s) ds,$$

for  $\hat{\Phi}(t)$  in (4). In Appendix 1, it will be shown that  $\hat{\Psi}(t)$  is a strongly consistent estimator for  $\Psi(t)$ .

For  $t \leq \tau$  define

$$V_n(t) = \sqrt{n}(\hat{\Psi}(t) - \Psi(t)),$$

and

$$\tilde{V}_n(t) = \int_0^t \tilde{U}_n(s) ds,$$

for  $\tilde{U}_n$  in (8). Exchanging the order of integration yields

$$\begin{aligned}
\tilde{V}_n(t) &= \frac{\int_0^t D^T(x) dx \Omega}{\sqrt{n}} \left( \sum_{i \leq n_1} \int_0^\tau \mu_1(w) dM_i(w) + \sum_{i > n_1} \int_0^\tau \mu_2(w) dM_i(w) \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{i \leq n_1} \int_0^t v_1(w) \int_w^t C(x) dx dM_i(w) \\
&+ \frac{1}{\sqrt{n}} \sum_{i > n_1} \int_0^t v_2(w) \int_w^t C(x) dx dM_i(w). \tag{10}
\end{aligned}$$

In Appendix 2, it will be shown that the process  $V_n(t)$  is asymptotically equivalent to the process  $\tilde{V}_n(t)$  which converges weakly to the zero-mean Gaussian process

$V^*(t) = \int_0^t U^*(s)ds$ . Thus  $V_n(t)$  also converges weakly to  $V^*(t)$ . The covariation process  $\sigma_\Psi(s, t)$  of  $V^*(t)$  can be consistently estimated by

$$\begin{aligned}
\hat{\sigma}_\Psi(s, t) = & \int_0^s \hat{D}^T(x) dx \hat{\Omega} \left( \int_0^\tau \frac{\hat{\mu}_1(w) \hat{\mu}_1^T(w) K_1(w) d\hat{R}(w; \hat{\beta})}{n(1 + \hat{R}(w; \hat{\beta}))} \right. \\
& + \int_0^\tau \frac{\hat{\mu}_2(w) \hat{\mu}_2^T(w) K_2(w) d\hat{R}(w; \hat{\beta})}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \Big) \hat{\Omega}^T \int_0^t \hat{D}^T(x) dx \\
& + \int_0^s \frac{\hat{v}_1^2(w) K_1(w)}{n(1 + \hat{R}(w; \hat{\beta}))} \left( \int_w^s C(x) dx \right)^2 d\hat{R}(w; \hat{\beta}) \\
& + \int_0^s \frac{\hat{v}_2(w)^2 K_2(w)}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \left( \int_w^s C(x) dx \right)^2 d\hat{R}(w; \hat{\beta}) \\
& + \int_0^s \hat{D}^T(x) dx \hat{\Omega} \int_0^t \frac{\hat{\mu}_1(w) \hat{v}_1(w) K_1(w)}{n(1 + \hat{R}(w; \hat{\beta}))} \int_w^t C(x) dx d\hat{R}(w, \hat{\beta}) \\
& + \int_0^s \hat{D}^T(x) dx \hat{\Omega} \int_0^t \frac{\hat{\mu}_2(w) \hat{v}_2(w) K_2(w)}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \int_w^t C(x) dx d\hat{R}(w, \hat{\beta}) \\
& + \int_0^t \hat{D}^T(x) dx \hat{\Omega} \int_0^s \frac{\hat{\mu}_1(w) \hat{v}_1(w) K_1(w)}{n(1 + \hat{R}(w; \hat{\beta}))} \int_w^s C(x) dx d\hat{R}(w, \hat{\beta}) \\
& + \int_0^t \hat{D}^T(x) dx \hat{\Omega} \int_0^s \frac{\hat{\mu}_2(w) \hat{v}_2(w) K_2(w)}{n(e^{-\hat{\beta}_1} + e^{-\hat{\beta}_2} \hat{R}(w; \hat{\beta}))} \int_w^s C(x) dx d\hat{R}(w, \hat{\beta}) \quad (11)
\end{aligned}$$

From these results, an asymptotic  $100(1 - \alpha)\%$  confidence interval for  $\Psi(t_0)$  can be obtained as  $\hat{\Psi}(t_0) \mp z_{\alpha/2} \sqrt{\hat{\sigma}_\Psi(t_0, t_0)/n}$ .

## Simultaneous Confidence Bands

To make simultaneous inference on  $\Phi(t)$  over a time interval  $I = [a, b] \subset [0, \tau]$ , let  $w_n(t)$  be a data-dependent function that converges in probability to a bounded function  $w^*(t) > 0$ , uniformly in  $t$  over  $I$ . Then  $U_n/w_n$  converges weakly  $U^*/w^*$ . If  $c_\alpha$  is the upper  $\alpha$ th percentile of  $\sup_{t \in I} |U^*/w^*|$ , an asymptotic  $100(1 - \alpha)\%$  simultaneous confidence band for  $\Phi(t)$ ,  $t \in I$ , can be obtained as

$$\left( \hat{\Phi}(t) - \frac{c_\alpha w_n(t)}{\sqrt{n}}, \hat{\Phi}(t) + \frac{c_\alpha w_n(t)}{\sqrt{n}} \right).$$

It is difficult to obtain  $c_\alpha$  analytically. One obvious alternative would be the bootstrapping method. However, it is very time-consuming. More discussion on this will be given later on the application to data from the Women's Health Initiative estrogen plus progestin clinical trial. Here a normal resampling approximation will be used. Lin et al. [8] used the normal resampling approximation to simulate

the asymptotic distribution of sums of martingale residuals for checking the Cox regression model. This approach reduces computing time significantly, and has been used in many works, including Lin et al. [9], Cheng et al. [3], Tian et al. [19], and Peng and Huang [14].

For  $t \leq \tau$ , let  $N_i(t) = \delta_i I(X_i \leq t)$ ,  $i = 1, \dots, n$ , and define the process

$$\begin{aligned} \hat{U}_n(t) &= \frac{\hat{D}^T(t)\hat{\Omega}}{\sqrt{n}} \left( \sum_{i \leq n_1} \int_0^\tau \hat{\mu}_1 d(\epsilon_i N_i) + \sum_{i > n_1} \int_0^\tau \hat{\mu}_2 d(\epsilon_i N_i) \right) \\ &\quad + \frac{\hat{C}(t)}{\sqrt{n}} \left( \sum_{i \leq n_1} \int_0^t \hat{v}_1 d(\epsilon_i N_i) + \sum_{i > n_1} \int_0^t \hat{v}_2 d(\epsilon_i N_i) \right) \\ &= \frac{\hat{D}^T(t)\hat{\Omega}}{\sqrt{n}} \left( \sum_{i \leq n_1} \epsilon_i \delta_i \hat{\mu}_1(X_i) I(X_i \leq \tau) + \sum_{i > n_1} \epsilon_i \delta_i \hat{\mu}_2(X_i) I(X_i \leq \tau) \right) \\ &\quad + \frac{\hat{C}(t)}{\sqrt{n}} \left( \sum_{i \leq n_1} \epsilon_i \delta_i \hat{v}_1(X_i) I(X_i \leq t) + \sum_{i > n_1} \epsilon_i \delta_i \hat{v}_2(X_i) I(X_i \leq t) \right), \quad (12) \end{aligned}$$

where  $\epsilon_i$ ,  $i = 1, \dots, n$ , are independent standard normal variables that are also independent of the data.

Conditional on  $(X_i, \delta_i, Z_i)$ ,  $i = 1, \dots, n$ ,  $\hat{U}_n$  is a sum of  $n$  independent variables at each time point. In Appendix 2, it will be shown that  $\hat{U}_n$  given the data converges weakly to  $U^*$ . It follows that  $\sup_{t \in I} |\hat{U}_n(t)/w_n(t)|$  given the data converges in distribution to  $\sup_{t \in I} |U^*(t)/w^*(t)|$ . Therefore,  $c_\alpha$  can be estimated empirically from a large number of realizations of the conditional distribution of  $\sup_{t \in I} |\hat{U}/w|$  given the data.

Motivated from recommendations in the literature for confidence bands of the survivor function and the cumulative hazard function in the one sample case, several choices of the weight  $w_n$  can be considered. The choice  $w_n(t) = \sqrt{\hat{\sigma}_\Phi(t, t)}$  results in equal precision bands [12], which differ from pointwise confidence intervals in that  $c_\alpha$  replaces  $z_{\alpha/2}$ . The choice  $w_n(t) = 1 + \hat{\sigma}_\Phi(t, t)$  results in the Hall-Wellner type bands recommended by Bie et al. [1], which often have narrower widths in the middle of data range and wider widths near the extremes of data range [8]. One could also consider the unweighted version with  $w_n(t) \equiv 1$ . Compared with the previous two choices, this choice does not require  $\hat{\sigma}_\Phi(t, t)$ , and hence is easier to implement.

To obtain simultaneous confidence bands for  $\Psi(t)$ , again consider the weighted process  $V_n(t)/w_n(t)$  which converges weakly to the limiting process  $V^*/w^*$ . If  $\tilde{c}_\alpha$  is the upper  $\alpha$ th percentile of  $\sup_{t \in I} |V^*/w^*|$ , an asymptotic  $100(1 - \alpha)\%$  simultaneous confidence band for  $\Psi(t)$ ,  $t \in I$ , can be obtained as

$$\left( \hat{\Psi}(t) - \frac{\tilde{c}_\alpha w_n(t)}{\sqrt{n}}, \hat{\Psi}(t) + \frac{\tilde{c}_\alpha w_n(t)}{\sqrt{n}} \right).$$

To approximate the critical value  $\tilde{c}_\alpha$ , for  $t \leq \tau$ , define the process

$$\begin{aligned}
 \hat{V}_n(t) &= \frac{\int_0^t \hat{D}^T(s) ds \hat{\Omega}}{\sqrt{n}} \left( \sum_{i \leq n_1} \int_0^\tau \hat{\mu}_1 d(\epsilon_i N_i) + \sum_{i > n_1} \int_0^\tau \hat{\mu}_2 d(\epsilon_i N_i) \right) \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i \leq n_1} \int_0^t \hat{v}_1(w) \int_w^t \hat{C}(x) dx d(\epsilon_i N_i(w)) \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i > n_1} \int_0^t \hat{v}_2(w) \int_w^t \hat{C}(x) dx d(\epsilon_i N_i(w)) \\
 &= \frac{\hat{D}^T(t) \hat{\Omega}}{\sqrt{n}} \left( \sum_{i \leq n_1} \epsilon_i \delta_i \hat{\mu}_1(X_i) I(X_i \leq \tau) + \sum_{i > n_1} \epsilon_i \delta_i \hat{\mu}_2(X_i) I(X_i \leq \tau) \right) \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i \leq n_1} \epsilon_i \delta_i \hat{v}_1(X_i) I(X_i \leq t) \int_{X_i}^t \hat{C}(x) dx \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i > n_1} \epsilon_i \delta_i \hat{v}_2(X_i) I(X_i \leq t) \int_{X_i}^t \hat{C}(x) dx, \tag{13}
 \end{aligned}$$

where  $\epsilon_i$ ,  $i = 1, \dots, n$ , are independent standard normal variables that are also independent of the data. In Appendix 2, the process  $\hat{V}_n(t)$  given the data is shown to converge weakly to  $V^*(t)$ . Thus  $\tilde{c}_\alpha$  can be approximated empirically from a large number of realizations of the conditional distribution of  $\sup_{t \in [a, b]} |\hat{V}(t)/w_n|$  given the data. Similarly to the case for  $\hat{U}_n$ , the weight function  $w_n$  can be chosen to yield equal precision, Hall-Wellner type, and unweighted confidence bands.

## Simulation Studies

For stable moderate sample behavior, the range of the simultaneous confidence bands for both  $\Phi(t)$  and  $\Psi(t)$  needs to be restricted. Through a series of simulation studies, a data-dependent range was found to result in good performance for moderate samples. The range is obtained by truncating at the 25th percentile of the uncensored data at the lower end, and truncating at the 5th largest uncensored observation at the upper end. By this truncation, the confidence bands are given in a range where a reasonable amount of data are available. Also, in the estimating procedures, the function  $\hat{P}(t; \mathbf{b})$  is replaced by an asymptotically equivalent form

$$\exp\left(-\int_0^t \frac{1}{\sum_{i=1}^n I(X_i \geq s)} d\left\{\sum_{i=1}^n \delta_i e^{-b_2 Z_i} I(X_i \leq s)\right\}\right).$$

For simulation studies reported here and for the real data application in section “Application”,  $\tau$  was set to include all data in calculating  $\hat{\beta}$ . All numerical

**Table 1** Empirical coverage probabilities of the three types of simultaneous confidence bands HW, EP, and UW, for the absolute risk reduction  $\Phi$  and the restricted mean survival difference  $\Psi$ , under model (1), based on 1,000 repetitions

Hazard ratio	Censoring (%)	$n$	$\Phi$			$\Psi$		
			HW	EP	UW	HW	EP	UW
0.9 $\uparrow$ 1.2	10	100	0.968	0.963	0.974	0.964	0.974	0.955
	30		0.968	0.963	0.974	0.956	0.972	0.946
	50		0.950	0.949	0.953	0.957	0.977	0.956
	10	200	0.961	0.964	0.958	0.951	0.965	0.944
	30		0.954	0.955	0.966	0.949	0.963	0.944
	50		0.940	0.942	0.945	0.940	0.962	0.937
	10	400	0.954	0.958	0.962	0.952	0.964	0.950
	30		0.961	0.961	0.967	0.951	0.969	0.946
	50		0.949	0.945	0.954	0.949	0.962	0.945
1.2 $\downarrow$ 0.8	10	100	0.951	0.946	0.962	0.953	0.964	0.938
	30		0.951	0.947	0.969	0.959	0.979	0.956
	50		0.930	0.930	0.949	0.962	0.973	0.960
	10	200	0.956	0.956	0.955	0.952	0.969	0.947
	30		0.960	0.957	0.962	0.953	0.972	0.949
	50		0.942	0.933	0.947	0.943	0.960	0.940
	10	400	0.958	0.952	0.958	0.955	0.968	0.944
	30		0.954	0.955	0.954	0.949	0.966	0.951
	50		0.951	0.950	0.956	0.948	0.961	0.947

computations were done in *Matlab*. Some representative results are given in Table 1, where lifetime variables were generated with  $R(t)$  chosen to yield the standard exponential distribution for the control group. The values of  $\beta$  were  $(\log(0.9), \log(1.2))$  and  $(\log(1.2), \log(0.8))$ , representing 1/3 increase or decrease over time from the initial hazard ratio, respectively. The censoring variables were independent and identically distributed with the log-normal distribution, where the normal distribution had mean  $c$  and standard deviation 0.5, with  $c$  chosen to achieve various censoring rates. The data were split into the treatment and control groups by a 1:1 ratio. The empirical coverage probabilities were obtained from 1,000 repetitions, and for each repetition, the critical values  $c_\alpha$  and  $\tilde{c}_\alpha$  were calculated empirically from 1,000 realizations of relevant conditional distributions. For both  $\Phi(t)$  and  $\Psi(t)$ , the equal precision bands, Hall-Wellner type bands and unweighted bands are denoted by EP, HW and UW respectively.

Note that with 1,000 repetitions and  $1.96\sqrt{0.95 \cdot 0.05/1,000} = 0.0135$ , we expect the empirical coverage probabilities to be mostly greater than 0.9365. In Table 1, the empirical coverage probabilities are greater than 0.9365 for all but three cases. Those three cases occurred for  $\Phi(t)$ , with 50% censoring and smaller sample sizes. The phenomenon disappeared when  $n = 400$ . Various additional simulation

**Table 2** Empirical coverage probabilities of the three types of simultaneous confidence bands HW, EP, and UW, for the absolute risk reduction  $\Phi$  and the restricted mean survival difference  $\Psi$ , under a monotone hazard ratio model, based on 1,000 repetitions

Hazard ratio	Censoring (%)	$n$	$\Phi$			$\Psi$		
			HW	EP	UW	HW	EP	UW
0.9 $\uparrow$ 1.2	10	100	0.973	0.977	0.975	0.964	0.975	0.955
	30		0.983	0.983	0.987	0.971	0.984	0.963
	50		0.969	0.973	0.971	0.967	0.986	0.965
	10	200	0.967	0.951	0.961	0.955	0.967	0.937
	30		0.966	0.965	0.975	0.956	0.971	0.950
	50		0.956	0.964	0.962	0.966	0.978	0.965
	10	400	0.956	0.916	0.978	0.963	0.965	0.948
	30		0.967	0.962	0.975	0.961	0.972	0.956
	50		0.984	0.982	0.979	0.970	0.984	0.969
1.2 $\downarrow$ 0.8	10	100	0.974	0.970	0.979	0.964	0.975	0.953
	30		0.971	0.964	0.980	0.965	0.983	0.964
	50		0.966	0.971	0.978	0.976	0.989	0.974
	10	200	0.959	0.930	0.965	0.945	0.962	0.944
	30		0.971	0.972	0.971	0.947	0.967	0.937
	50		0.960	0.957	0.969	0.963	0.986	0.961
	10	400	0.935	0.872	0.975	0.960	0.968	0.953
	30		0.962	0.959	0.976	0.953	0.971	0.952
	50		0.966	0.958	0.982	0.961	0.974	0.958

studies indicated that the proposed procedures performed well for sample size close to 100 and up, with moderate censoring. Under heavy censoring, the results were still good with uncensored observations close to 50 and up in each treatment group.

To check how robust the procedures are against violation of model assumptions, various monotone hazard ratio models were also considered alternative to the model (1). The results indicated that the proposed procedures continued to perform well. For example, in Table 2, the control group lifetime variables were standard exponential. The hazard ratio was linear from 0 to the 90th percentile of the standard exponential, and continuous and constant afterwards. The initial and end hazard ratios again were (0.9, 1.2) and (1.2, 0.8) respectively, and the censoring variables were the same as before. It can be seen from Table 2 that the confidence bands performed satisfactorily.

To compare efficiency against the non-parametric alternatives based on the Kaplan-Meier estimators, for estimating  $\Phi(t)$  and  $\Psi(t)$  at various time points, the mean squared errors of the model based estimators and the Kaplan-Meier estimators were examined under model (1) in various simulation studies. Typically the model based estimators have smaller mean squared errors, more so for  $\Phi(t)$  than for  $\Psi(t)$ . Also, the efficiency is higher under heavy censoring and for time points closer to the upper tail region. This is because the Kaplan-Meier becomes increasingly unstable

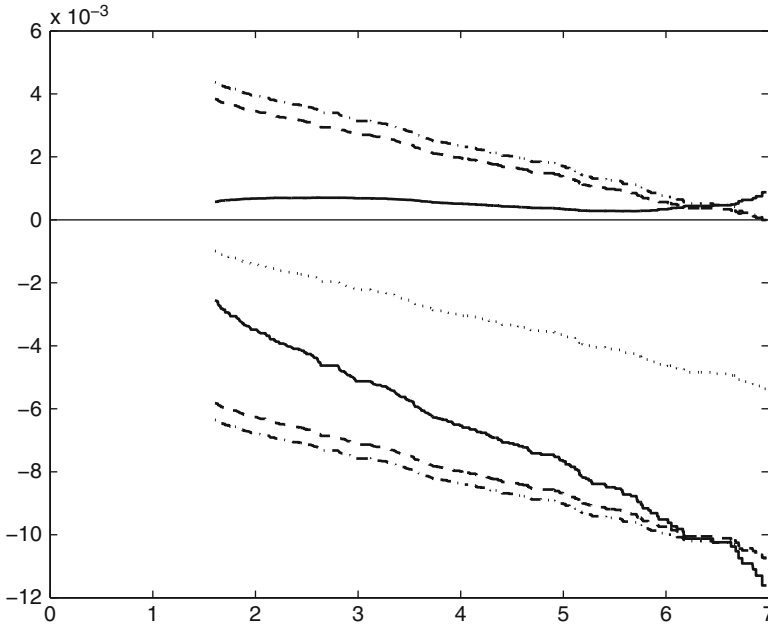
**Table 3** Ratio of mean squared errors of the model based estimators over the Kaplan-meier estimators, for  $\Phi(t)$  and  $\Psi(t)$  under model (1), at  $t = 0.5, 1, 1.5$  respectively, based on 1,000 repetitions

Hazard ratio	Censoring (%)	n	$\Phi$			$\Psi$		
			0.5	1	1.5	0.5	1	1.5
0.9 $\uparrow$ 1.2	10	100	0.6019	0.6731	0.6286	0.6341	0.7978	0.8470
	30		0.5527	0.6025	0.5248	0.5960	0.7303	0.7717
	50		0.4676	0.4008	0.2447	0.5232	0.6149	0.5872
	10	200	0.6438	0.7090	0.6865	0.6795	0.8436	0.9099
	30		0.5920	0.6436	0.5513	0.6368	0.7763	0.8407
	50		0.5150	0.4403	0.2672	0.5802	0.6744	0.6682
	10	400	0.6831	0.7191	0.6800	0.6975	0.9000	0.9530
	30		0.6321	0.6357	0.5425	0.6747	0.8271	0.8563
	50		0.5523	0.4222	0.2789	0.6255	0.7205	0.6777
1.2 $\downarrow$ 0.8	10	100	0.6251	0.6509	0.6203	0.7195	0.8275	0.8349
	30		0.5897	0.6278	0.5613	0.6850	0.7825	0.8037
	50		0.4778	0.4180	0.2631	0.5662	0.6434	0.6182
	10	200	0.6650	0.7035	0.6902	0.7324	0.8535	0.8930
	30		0.6322	0.6648	0.5816	0.6982	0.8088	0.8467
	50		0.5434	0.4837	0.2893	0.6099	0.6992	0.7015
	10	400	0.7085	0.7081	0.6992	0.7432	0.8973	0.9289
	30		0.6753	0.6742	0.6079	0.7226	0.8560	0.8794
	50		0.6015	0.4661	0.3010	0.6458	0.7595	0.7256

near upper tail region and under heavy censoring. Some representative results are given in Table 3, in terms of the ratio of the mean squared errors of the model based estimators over the Kaplan-Meier estimators, under configurations the same as those for Table 1.

Application

For the Women’s Health Initiative (WHI) randomized controlled trial of combined (estrogen plus progestin) postmenopausal hormone therapy, an elevated coronary heart disease risk was reported, with overall unfavorable health benefits versus risks over an average of 5.6 year study period [10, 21]. Few research reports have stimulated as much public response, since preceding observational research literature suggested a 40–50% reduction in coronary heart disease incidence among women taking postmenopausal hormone therapy. Analysis of the WHI observational study shows a similar discrepancy with the WHI clinical trial for coronary heart disease, stroke, and venous thromboembolism, even after adjusting for confounding factors in the observational study. Following control for time from estrogen-plus-progestin

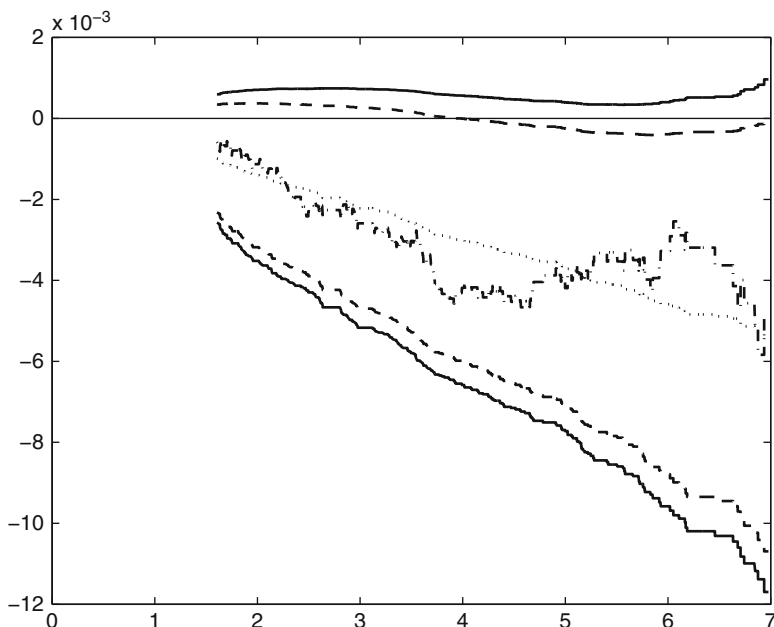


**Fig. 4** Simultaneous 95% confidence bands of the absolute risk reduction for the WHI clinical trial stroke data: *Solid line*—equal precision confidence band; *Dashed line*—Hall-Wellner type confidence band; *Dash-dotted lines*—unweighted confidence band; *Dotted line*: Estimated absolute risk reduction

initiation and confounding, hazard ratio estimates were rather similar between the clinical trial and observational study components of WHI, although there was evidence of some remaining difference for stroke [16].

In the introduction, it was mentioned that for stroke, the estimated absolute risk reduction based on model (1) provides a good compromise between the results from the two partitioning approaches under the piece-wise Cox model. Let us illustrate the methods developed in the previous sections with the stroke data from the WHI clinical trial. Among the 16,608 postmenopausal women ( $n_1 = 8,102$ ), there were 151 and 107 events observed in the treatment and control group respectively, implying about 98% censoring, primarily by the trial stopping time. Fitting model (1) to this data set, we get  $\hat{\beta} = (0.32, -1.69)^T$ . Plots of the model based survival curves and the Kaplan-Meier curves for the two groups show that the model is reasonable. The residual plot as mentioned in Yang and Prentice [23] also indicates a good model fit. These plots are not displayed here to save space. The three 95% simultaneous confidence bands for the absolute risk reduction are given in Fig. 4. From Fig. 4, it can be seen that both the Hall-Wellner type band and the unweighted band maintain a roughly constant width through the data range considered. In comparison, the equal precision band has width gradually increasing as the standard error of the estimated absolute risk reduction increases over time. Also, the width of

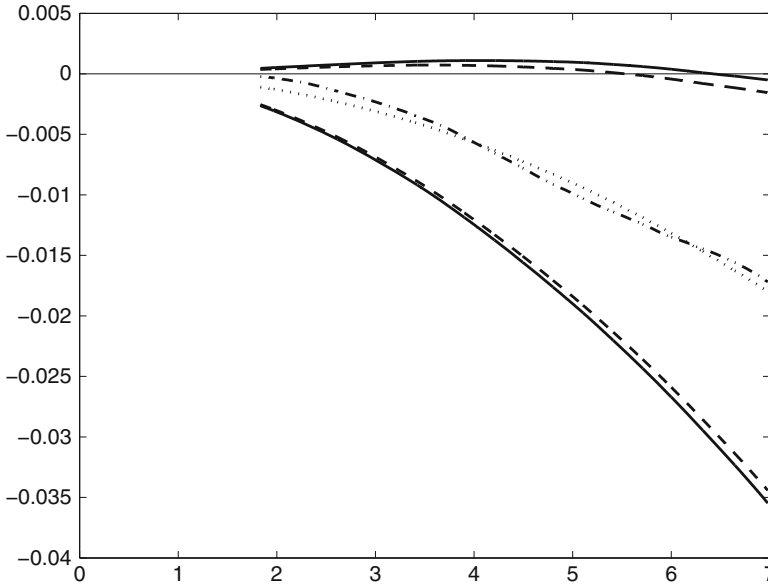




**Fig. 5** 95% equal precision confidence band and pointwise 95% confidence intervals of the absolute risk reduction for the WHI clinical trial stroke data: *Solid line*—equal precision confidence band; *Dashed line*—pointwise 95% confidence intervals; *Dotted line*: Model based estimator of the absolute risk reduction; *Dash-dotted lines*—Kaplan-Meier estimator of the absolute risk reduction

the equal precision band is narrower than those of the Hall-Wellner type band and the unweighted band through most of the range. Similar phenomena are often present in other applications not reported here. Thus it is recommended that the equal precision band be used in making inference on the absolute risk reduction under model (1). Note that the simple bootstrap method for approximating  $c_\alpha$  when  $w_n \equiv 1$  is already much more computationally intensive than the normal resampling approximation. With  $w_n(t) = \sqrt{\hat{\sigma}_\Phi(t)}$ , the bootstrap method would require one more level of bootstrapping samples, thus further increasing the computational burden. In comparison, once  $\hat{\sigma}_\Phi(t)$  is obtained with the martingale structure, the normal resampling approximation only needs a small additional computation and programming cost. Similar remarks are also applicable to the case with the restricted mean survival difference.

To compare the point-wise confidence intervals and the simultaneous confidence band, Fig. 5 displays 95% point-wise confidence intervals and the simultaneous confidence band for the stroke data. The simultaneous confidence band is slightly wider than the point-wise confidence intervals and maintains the same rate of inflation in width throughout the range. The confidence intervals and confidence band indicate some evidence that the absolute risk reduction is negative in the



**Fig. 6** 95% equal precision confidence band and pointwise 95% confidence intervals of the mean restricted survival difference for the WHI clinical trial stroke data: *Solid line*—equal precision confidence band; *Dashed line*—pointwise 95% confidence intervals; *Dotted line*: Model based estimator of the mean restricted survival difference; *Dash-dotted lines*—Kaplan-Meier estimator of the mean restricted survival difference

range of 4–7 years, but the evidence is not very strong. Figure 5 also includes the Kaplan-Meier estimator. Between the semiparametric and nonparametric estimators, The model based estimator is smoother, the Kaplan-Meier estimator is more volatile and oscillates around the model based estimator. The model based estimator captures the general decreasing trend in the absolute risk reduction, and averages out the deviations from that trend, particularly in the range of 3.5 to 7 years.

For the restricted mean survival difference, Fig. 6 displays the estimator under model (1), the 95% point-wise confidence intervals and simultaneous equal precision confidence band for the stroke data. Since the restricted mean survival difference is a summary measure, the estimators are smoother compared with those for the absolute risk reduction. Also, the semiparametric and nonparametric estimators show a better agreement compared with the case for the absolute risk reduction. Furthermore, the inflation of width by the band over the point-wise confidence intervals is smaller compared with the situation in Fig. 5. This is possibly because the restricted mean survival difference, a summary measure, may have higher correlation at different time points compared with the absolute risk reduction at those same time points. From Fig. 6, there is some evidence that the restricted mean survival difference is negative towards the end of the data range.

## Discussion

We have studied the asymptotic properties of the estimators for the absolute risk reduction and the restricted mean survival difference under the short-term and long-term hazards model. Point-wise confidence intervals and simultaneous confidence bands are developed for these measures. These procedures can have a sufficiently wide range of applications because of the flexibility of the model. In simulation studies, the confidence bands have good performance for moderate samples. Among the versions with different weights, the equal precision confidence band is recommended. It has width that is proportional to the standard error at each time point and often results in narrower width in most of the data range. It also demonstrates the inflation of the confidence interval width needed for simultaneous inference. For the restricted mean survival difference, often the measure at a fixed time, say  $t_0$  years, with  $t_0$  close to the maximum follow-up period of the clinical trial, is of interest. In those situations, the point-wise confidence intervals may suffice.

Compared with the nonparametric methods based on the Kaplan-Meier estimator, the semiparametric approach developed here produces more smooth estimators and more stable behaviors, especially near the end of the data range. Thus it provides a good alternative to the nonparametric approach should the model be appropriate. The model also permits inference on the hazard ratio function, as described in Yang and Prentice [23], where the nonparametric approach could result in wide confidence intervals at the tail regions. When the model provides good fit to the data, together the confidence intervals and bands on the hazard ratio, the absolute risk reduction and the restricted mean survival difference, present good visual tools for assessing the temporal pattern and cumulative effect of the treatment. It is also of interest to extend the results here to epidemiological studies by considering the regression setting and adjusting for covariate. These and other problems are worthy of further exploration.

**Acknowledgements** The original version of this article has previously been published in *Lifetime Data Analysis* in 2013.

## Appendix 1: Consistency

Throughout the Appendices, we assume the following regularity conditions, which is a little weaker than the conditions used in Yang and Prentice [22].

*Condition 1.*  $\lim_{n \rightarrow \infty} \frac{n_1}{n} = \rho \in (0, 1)$ .

*Condition 2.* The survivor function  $G_i$  of  $C_i$  given  $Z_i$  is continuous and satisfies

$$\frac{1}{n} \sum_{i \leq n_1} G_i(t) \rightarrow \Gamma_1, \quad \frac{1}{n} \sum_{i > n_1} G_i(t) \rightarrow \Gamma_2,$$

uniformly for  $t \leq \tau$ , for some  $\Gamma_1, \Gamma_2$ , and  $\tau < \tau_0$  such that  $\Gamma_j(\tau) > 0$ ,  $j = 1, 2$ .

*Condition 3.* The survivor functions  $S_C$  and  $S_T$  are absolutely continuous and  $S_C(\tau) > 0$ .

Under these conditions, the strong law of large numbers implies that (3) is satisfied.

For  $t \leq \tau$ , define

$$L(t) = \Gamma_1 S_C + \Gamma_2 S_T,$$

$$U_j(t; \mathbf{b}) = \int_0^t \Gamma_1 dF_C + \exp(-b_j) \int_0^t \Gamma_2 dF_T, \quad j = 1, 2,$$

$$\Lambda_j(t; \mathbf{b}) = \int_0^t \frac{dU_j(s; \mathbf{b})}{L(s)}, \quad j = 1, 2,$$

$$P(t; \mathbf{b}) = \exp\{-\Lambda_2(t; \mathbf{b})\}, \quad R(t; \mathbf{b}) = \frac{1}{P(t; \mathbf{b})} \int_0^t P(s; \mathbf{b}) d\Lambda_1(s; \mathbf{b}),$$

$$f_j^0(t; \mathbf{b}) = \frac{\exp(-b_j) R^{j-1}(t; \mathbf{b})}{\exp(-b_1) + \exp(-b_2) R(t; \mathbf{b})}, \quad j = 1, 2,$$

$$m_j(\mathbf{b}) = \left\{ \int_0^\tau f_j^0 \Gamma_2(t) dF_T(t) - \int_0^\tau \frac{f_j^0 \Gamma_2(t) S_T(t) dR(t; \mathbf{b})}{\exp(-b_1) + \exp(-b_2) R(t; \mathbf{b})} \right\}, \quad j = 1, 2,$$

and  $m(\mathbf{b}) = (m_1(\mathbf{b}), m_2(\mathbf{b}))'$ . We will also assume

*Condition 4.* The function  $m(\mathbf{b})$  is non-zero for  $b \in \mathcal{B} - \{\beta\}$ , where  $\mathcal{B}$  is a compact neighborhood of  $\beta$ .

**Theorem 1.** Suppose that Conditions 1–4 hold. Then, (i) the zero  $\hat{\beta}$  of  $Q(\mathbf{b})$  in  $\mathcal{B}$  is strongly consistent for  $\beta$ ; (ii)  $\hat{\Phi}(t)$  is strongly consistent for  $\Phi(t)$ , uniformly for  $t \in [0, \tau]$ , and  $\hat{\Psi}(t)$  is strongly consistent for  $\Psi(t)$ , uniformly on  $t \in [0, \tau]$ ; (iii)  $\hat{\Omega}$  converges almost surely to a limiting matrix  $\Omega^*$ .

*Proof.* Under Conditions 1–3, the limit of  $\sum_{i=1}^n I(X_i \geq t)/n$  is bounded away from zero on  $t \in [0, \tau]$ . Thus, with probability 1,

$$\frac{\sum_{i=1}^n \delta_i e^{-b_j Z_i} I(X_i = t)}{\sum_{i=1}^n \delta_i I(X_i \geq t)} \rightarrow 0, \quad j = 1, 2, \quad (14)$$

uniformly for  $t \in [0, \tau]$  and  $b \in \mathcal{B}$ . From this, one also has, with probability 1,

$$|\Delta \hat{P}(t; \mathbf{b})| \rightarrow 0, \quad |\Delta \hat{R}(t; \mathbf{b})| \rightarrow 0, \quad (15)$$

uniformly for  $t \in [0, \tau]$  and  $b \in \mathcal{B}$ , where  $\Delta$  indicates the jump of the function in  $t$ .

Define the martingale residuals

$$\hat{M}_i(t; \mathbf{b}) = \delta_i I(X_i \leq t) - \int_0^t I(X_i \geq s) \frac{\hat{R}(ds; \mathbf{b})}{e^{-b_1 Z_i} + e^{-b_2 Z_i} \hat{R}(s; \mathbf{b})}, \quad 1 \leq i \leq n.$$

From (12) and (13), and the fundamental theorem of calculus, it follows that, with probability 1,

$$Q(\mathbf{b}) = \sum_{i=1}^n \int_0^\tau \{f_i(t; \mathbf{b}) + o(1)\} \hat{M}_i(dt; \mathbf{b}), \quad (16)$$

uniformly in  $t \leq \tau$ ,  $b \in \mathcal{B}$  and  $i \leq n$ , where  $f_i = (f_{1i}, f_{2i})^T$ , with

$$f_{1i}(t; \mathbf{b}) = \frac{Z_i e^{-b_1 Z_i}}{e^{-b_1 Z_i} + e^{-b_2 Z_i} \hat{R}(t; \mathbf{b})}, \quad f_{2i}(t; \mathbf{b}) = \frac{Z_i e^{-b_2 Z_i} \hat{R}(t; \mathbf{b})}{e^{-b_1 Z_i} + e^{-b_2 Z_i} \hat{R}(t; \mathbf{b})}.$$

From the strong law of large numbers ([15], p. 41) and repeated use of Lemma A1 of Yang and Prentice [22], one obtain, with probability 1,

$$\hat{P}(t; \mathbf{b}) \rightarrow \hat{P}(t; \mathbf{b}), \quad \hat{R}(t; \mathbf{b}) \rightarrow R(t; \mathbf{b}), \quad Q(\mathbf{b})/n \rightarrow m(\mathbf{b}), \quad (17)$$

uniformly in  $t \leq \tau$  and  $\mathbf{b} \in \mathcal{B}$ . From these results and Condition 4, one obtains the strong consistency of  $\hat{\beta}$ ,  $\hat{\Phi}(t)$  and  $\hat{\Psi}(t)$ , and almost sure convergence of  $\hat{\Omega}$ .

## Appendix 2: Weak Convergence

For  $C(t)$ ,  $D(t)$ ,  $\mu_1(t)$ ,  $\mu_2(t)$ ,  $v_1(t)$ ,  $v_2(t)$ , let  $C^*(t)$ ,  $D^*(t)$ , etc. be their almost sure limit. In addition, let  $L_j$  be the almost sure limit of  $K_j/n$ ,  $j = 1, 2$ . For  $0 \leq s, t < \tau$ , let

$$\begin{aligned} & \sigma_\Phi(s, t) \\ &= D^{*T}(s) \Omega^* \left( \int_0^\tau \frac{\mu_1^* \mu_1^{*T}}{1+R} L_1 dR + \int_0^\tau \frac{\mu_2^* \mu_2^{*T}}{e^{-\beta_1} + e^{-\beta_2 R}} L_2 dR \right) \Omega^{*T} D^*(t) \\ &+ C^*(s) C^*(t) \left( \int_0^s \frac{v_1^{*2}}{1+R} L_1 dR + \int_0^s \frac{v_2^{*2}}{e^{-\beta_1} + e^{-\beta_2 R}} L_2 dR \right) \\ &+ C^*(t) D^{*T}(s) \Omega^* \left( \int_0^t \frac{\mu_1^* v_1^*}{1+R} L_1 dR + \int_0^t \frac{\mu_2^* v_2^*}{e^{-\beta_1} + e^{-\beta_2 R}} L_2 dR \right) \\ &+ C^*(s) D^{*T}(t) \Omega^* \left( \int_0^s \frac{\mu_1^* v_1^*}{1+R} L_1 dR + \int_0^s \frac{\mu_2^* v_2^*}{e^{-\beta_1} + e^{-\beta_2 R}} L_2 dR \right), \end{aligned} \quad (18)$$

and

$$\begin{aligned}
& \sigma_{\Psi}(s, t) \\
&= \int_0^s D^{*T}(x) dx \Omega^* \left( \int_0^{\tau} \frac{\mu_1^*(w) \mu_1^{*T}(w)}{1 + R(w)} L_1(w) dR(w) \right. \\
&\quad + \int_0^{\tau} \frac{\mu_2^*(w) \mu_2^{*T}(w)}{e^{-\beta_1} + e^{-\beta_2} R(w)} L_2(w) dR(w) \left. \right) \Omega^{*T} \int_0^t D^{*T}(x) dx \\
&\quad + \int_0^s \frac{v_1^{*2}(w)}{1 + R(w)} \left( \int_w^s C^*(x) dx \right)^2 L_1(w) dR(w) \\
&\quad + \int_0^s \frac{v_2^{*2}(w)}{e^{-\beta_1} + e^{-\beta_2} R(w)} \left( \int_w^s C^*(x) dx \right)^2 L_2(w) dR(w) \\
&\quad + \int_0^s D^{*T}(x) dx \Omega^* \int_0^t \frac{\mu_1^*(w) v_1^*(w)}{1 + R(w)} \left( \int_w^t C^*(x) dx \right) L_1(w) dR(w) \\
&\quad + \int_0^s D^{*T}(x) dx \Omega^* \int_0^t \frac{\mu_2^*(w) v_2^*(w)}{e^{-\beta_1} + e^{-\beta_2} R(w)} \left( \int_w^t C^*(x) dx \right) L_2(w) dR(w) \\
&\quad + \int_0^t D^{*T}(x) dx \Omega^* \int_0^s \frac{\mu_1^*(w) v_1^*(w)}{1 + R(w)} \left( \int_w^s C^*(x) dx \right) L_1(w) dR(w) \\
&\quad + \int_0^t D^{*T}(x) dx \Omega^* \int_0^s \frac{\mu_2^*(w) v_2^*(w)}{e^{-\beta_1} + e^{-\beta_2} R(w)} \left( \int_w^s C^*(x) dx \right) L_2(w) dR(w). \quad (19)
\end{aligned}$$

**Theorem 2.** Suppose that Conditions 1–4 hold and that the matrix  $\Omega^*$  is non-singular. Then, (i)  $U_n$  is asymptotically equivalent to the process  $\tilde{U}_n$  in (8) which converges weakly to a zero-mean Gaussian process  $U^*$  on  $[0, \tau]$ , with covariance function  $\sigma_{\Phi}(s, t)$  in (18). In addition,  $\tilde{U}_n(s)$  given the data converges weakly to the same limiting process  $U^*$ . (ii)  $V_n(t)$  is asymptotically equivalent to the process  $\tilde{V}_n$  in (11) which converges weakly to the zero-mean Gaussian process  $\int_0^t U^*(s) ds$  on  $t \in [0, \tau]$ , with covariance function  $\sigma_{\Psi}(s, t)$  in (19). The process  $\int_0^t \tilde{V}_n(s) ds$  given the data also converges weakly to the same limiting process  $\int_0^t U^*(s) ds$ .

*Proof.* (i) As in the proof for Theorem A2 (ii) in Yang and Prentice [22], from the strong embedding theorem and (16),  $Q(\beta)/\sqrt{n}$  can be shown to be asymptotically normal. Now Taylor series expansion of  $Q(\mathbf{b})$  around  $\beta$  and the non-singularity of  $\Omega^*$  imply that  $\sqrt{n}(\hat{\beta} - \beta)$  is asymptotically normal. From the  $\sqrt{n}$ -boundedness of  $\hat{\beta}$ ,

$$\sqrt{n}(\hat{R}(t; \hat{\beta}) - \hat{R}(t; \beta)) = \frac{\partial R(t; \beta)}{\partial \beta} \sqrt{n}(\hat{\beta} - \beta) + o_p(1),$$

uniformly in  $t \leq \tau$ . These results, some algebra and Taylor series expansion together show that  $U_n$  is asymptotically equivalent to  $\tilde{U}_n$ . Similarly to the proof of the

asymptotic normality of  $Q(\beta)/\sqrt{n}$ , one can show that  $\tilde{U}_n$  converges weakly to a zero-mean Gaussian process. Denote the limiting process by  $U^*$ . From the martingale integral representation of  $\tilde{U}_n$ , it follows that the covariation process of  $U^*$  is given by  $\sigma(s, t)$  in (18), which can be consistently estimated by  $\hat{\sigma}(s, t)$  in (9). By checking the tightness condition and the convergence of the finite-dimensional distributions, it can be shown that  $\tilde{U}_n(s)$  given the data also converges weakly to  $U^*$ .

(ii) From the results in (i), the assertions on  $V_n$  and  $\tilde{V}_n$  follow.

## References

1. Bie, O., Borgan, O., Liestøl, K.: Confidence intervals and confidence bands for the cumulative hazard rate function and their small-sample properties. *Scand. J. Stat.* **14**, 221–233 (1987)
2. Chen, P., Tsiatis, A.A.: Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57**, 1030–1038 (2001)
3. Cheng, S.C., Wei, L.J., Ying, Z.: Predicting survival probabilities with semiparametric transformation models. *J. Am. Stat. Assoc.* **92**, 227–235 (1997)
4. Cox, D.R.: Regression models and life-tables (with Discussion). *J. R. Stat. Soc. B* **34**, 187–220 (1972)
5. Dabrowska, D.M., Doksum, K.A., Song, J.: Graphical comparison of cumulative hazards for two populations. *Biometrika* **76**, 763–773 (1989)
6. Kalbfleisch, J.D., Prentice, R.L.: *The Statistical Analysis of Failure Time Data*, 2nd edn. Wiley, New York (2002)
7. Kaplan, E., Meier, P.: Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958)
8. Lin, D.Y., Wei, L.J., Ying, Z.: Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572 (1993)
9. Lin, D.Y., Fleming, T.R., Wei, L.J.: Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81 (1994)
10. Manson, J.E., Hsia, J., Johnson, K.C., Rossouw, J.E., Assaf, A.R., Lasser, N.L., Trevisan, M., Black, H.R., Heckbert, S.R., Detrano, R., Strickland, O.L., Wong, N.D., Crouse, J.R., Stein, E., Cushman, M., Women's Health Initiative Investigators: Estrogen plus progestin and the risk of coronary heart disease. *N. Engl. J. Med.* **349**, 523–534 (2003)
11. McKeague, I.W., Zhao, Y.: Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Stat. Probab. Lett.* **60**, 405–415 (2002)
12. Nair, V.N.: Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**, 265–275 (1984)
13. Parzen, M.I., Wei, L.J., Ying, Z.: Simultaneous confidence intervals for the difference of two survival functions. *Scand. J. Stat.* **24**, 309–314 (1997)
14. Peng, L., Huang, Y.: Survival analysis with temporal covariate effects. *Biometrika* **94**, 719–733 (2007)
15. Pollard, D.: *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics, Hayward (1990)
16. Prentice, R.L., Langer, R., Stefanick, M.L., Howard, B.V., Pettinger, M., Anderson, G., Barad, D., Curb, J.D., Kotchen, J., Kuller, L., Limacher, M., Wactawski-Wende, J., Women's Health Initiative Investigators: combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observational studies and the women's health initiative clinical trial. *Am. J. Epidemiol.* **162**, 404–414 (2005)

17. Royston, P., Parmar, M.K.: The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat. Med.* **19**, 2409–2421 (2011)
18. Schaubel, D.E., Wei, G.: Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics* **67**, 29–38 (2011)
19. Tian, L., Zucker, D., Wei, L.J.: On the Cox model with time-varying regression coefficients. *J. Am. Stat. Assoc.* **100**, 172–183 (2005)
20. Tong, X., Zhu, C., Sun, J.: Semiparametric regression analysis of two-sample current status data, with applications to tumorigenicity experiments. *Can. J. Stat.* **35**, 575–584 (2007)
21. Writing Group for the Women’s Health Initiative Investigators: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women’s health initiative randomized controlled trial. *J. Am. Med. Assoc.* **288**, 321–333 (2002)
22. Yang, S., Prentice, R.L.: Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika* **92**, 1–17 (2005)
23. Yang, S., Prentice, R.L.: Estimation of the 2-sample hazard ratio function using a semiparametric model. *Biostatistics* **12**, 354–368 (2011)
24. Zucker, D.M.: Restricted mean life with covariates: modification and extension of a useful survival analysis method. *J. Am. Stat. Assoc.* **93**, 702–709 (1998)



Risk Assessment and Evaluation of Predictions

Lee, M.-L.T.; Gail, M.; Pfeiffer, R.; Satten, G.; Cai, T.;  
Gandy, A. (Eds.)

2013, XI, 445 p., Softcover

ISBN: 978-1-4614-8980-1