Published in final edited form as:

Stat Biosci. 2015 May ; 7(1): 1–18. doi:10.1007/s12561-013-9089-6.

Cumulative Hazard Ratio Estimation for Treatment Regimes in Sequentially Randomized Clinical Trials

Xinyu Tang and

Tang Biostatistics Program, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR 72202, USA

Abdus S. Wahed

Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, USA

Abstract

The proportional hazards model is widely used in survival analysis to allow adjustment for baseline covariates. The proportional hazard assumption may not be valid for treatment regimes that depend on intermediate responses to prior treatments received, and it is not clear how such a model can be adapted to clinical trials employing more than one randomization. Besides, since treatment is modified post-baseline, the hazards are unlikely to be proportional across treatment regimes. Although Lokhnygina and Helterbrand (*Biometrics* 63: 422–428, 2007) introduced the Cox regression method for two-stage randomization designs, their method can only be applied to test the equality of two treatment regimes that share the same maintenance therapy. Moreover, their method does not allow auxiliary variables to be included in the model nor does it account for treatment effects that are not constant over time. In this article, we propose a model that assumes proportionality across covariates within each treatment regime but not across treatment regimes. Comparisons among treatment regimes are performed by testing the log ratio of the estimated cumulative hazards. The ratio of the cumulative hazard across treatment regimes is estimated using a weighted Breslow-type statistic. A simulation study was conducted to evaluate the performance of the estimators and proposed tests.

Keywords

Cumulative treatment effect; Non-proportional hazards; Sequentially randomized clinical trial; Stratified proportional hazards model

1 Introduction

It is common practice in treatments of chronic diseases for patients to start with an initial therapy based on specific diagnoses. Patients stay on this initial therapy until a predetermined milestone, commonly referred to as response, is achieved, or physicians determine that the treatment is not providing the intended benefit, commonly referred to as non-response. In either case, further treatment could be suggested, which could be a rescue

therapy if patients show primary resistance to the initial therapy, or a maintenance therapy if a response is observed. Thus, treatments applied at later stages depend on the responses to the initial therapies, and a comparison across treatments at various stages would be misleading. Instead, dynamic treatment sequences are usually formed based on stage-specific treatments and responses to these treatments during the course of therapy. In order to study the effect of a treatment sequence on the survival outcome, sequentially randomized designs are used in clinical trials. In a sequentially randomized clinical trial, eligible patients are first randomized to receive one of the initial therapies. Patients reaching the second stage will continue to participate in a second randomization to either rescue or maintenance therapies. The procedure continues. This design results in different treatment regimes consisting of an initial therapy, the intermediate response and a second-stage therapy. The analysis and comparisons of such treatment regimes can be done based on classic survival methods, weighted using inverse probabilities of treatment allocations.

The idea of inverse-probability weighting (IPW) to estimate treatment regime effect in two-stage randomization designs was first introduced by Lunceford, Davidian and Tsiatis [2]. Since then the strategy of IPW has been incorporated into many other classical statistical methods to make them applicable to two-stage randomization designs. For example, Guo and Tsiatis [3] extended the Aalen–Nelson estimator to a Weighted Risk Set Estimator (WRSE) using IPW. Hernan, Lanoy, Costagliola and Robins [4] applied IPW in comparing treatment strategies from observational studies using a Cox proportional hazards model under artificial censoring. Lokhnygina and Helterbrand [1] incorporated IPW into the Cox regression method for two-stage randomization designs. Feng and Wahed [5] presented a modified weighted log-rank test for comparing different treatment strategies using IPW. Miyahara and Wahed [6] introduced IPW to Kaplan–Meier estimators with both fixed and time-dependent weights. More recently, Goldberg and Kosorok [7] used IPW methods in Q-learning to estimate treatment effect for censored data from multistage designs. However, except for the work of Hernan et al. [4], all the aforementioned models failed to adjust for baseline covariates.

Proportional hazards models are widely used in analyzing data from clinical trials with time-to-event endpoints. The advantage of the proportional hazards model over nonparametric methods (e.g. a log-rank test) is that auxiliary covariates can be included in the model to explain the variability of the dependent variable further. Even in randomized clinical trials, baseline characteristics are often imbalanced between treatment groups and hence adjustment becomes a consideration. This imbalance is even more problematic for two stages of randomization, since patients proceeding to the second stage are randomized differently based on their intermediate responses. Lokhnygina and Helterbrand [1] proposed to use a Cox regression method to test the equality of two treatment regimes that share the same maintenance therapy in a two-stage randomization setting. However, their method does not allow auxiliary variables to be included in the model nor does it account for treatment effects that are not constant over time, an issue in many medical studies. Therefore, comparisons among treatment regimes in the presence of non-proportional hazards is of importance to health research communities. Wei and Schaubel [8] proposed cumulative treatment effect estimation based on treatment-specific cumulative baseline hazards using a stratified proportional hazards model. Their model assumed proportional

hazards with respect to baseline covariates within each treatment group and non-proportional hazards across treatment groups. In this article, we take a similar approach to estimate the cumulative treatment effect for treatment regimes from sequentially randomized clinical trials. Comparisons among treatment regimes are performed by testing the ratio of the estimated cumulative hazards. A simulation study is conducted to evaluate the performance of the estimators and proposed tests. The estimators and proposed tests are also applied to the neuroblastoma data to compare different treatment regimes in a neuroblastoma study with respect to the overall survival.

2 Design Setting and Statistical Model

2.1 Notation

We consider a two-stage randomization design, in which n eligible patients are first randomized to one of the J initial therapies ($A_1, \dots, A_J$) and patients achieving a clinical response are then randomized to one of the K second-stage therapies ($B_1, \dots, B_K$). By design, non-responders in the first stage are not eligible to receive any treatment in the second stage, similarly to the neuroblastoma study (Sect. 5) that motivated this work. Thus, there are a total of $J \times K$ treatment regimes based on this design, namely, $A_1B_1, \dots, A_JB_K$. The treatment regime $A_jB_k, j = 1, \dots, J; k = 1, \dots, K$, is defined as “treat with A_j as initial therapy, then B_k as second-stage therapy if responds to A_j ” [2]. This treatment regime consists of an initial treatment (A_j), the intermediate response status, and a second-stage treatment (B_k) in the event the patient responds to the initial treatment. Thus, estimation of the effect of treatment regime A_jB_k will not only include patients who were randomized to receive initial therapy A_j , responded, and then randomized to B_k as second-stage therapy, but also patients who were randomized to A_j and did not respond.

Let X_{ji} be the indicator for initial therapy A_j such that $X_{ji} = 1$ if the i th patient receives initial treatment $A_j \left(\sum_{j=1}^J X_{ji} = 1 \right)$. The response status is denoted by R_i , with $R_i = 1$ for responders and $R_i = 0$ for non-responders. For responders ($R_i = 1$), let Z_{ki} be the indicator for second-stage therapy B_k , i.e., $Z_{ki} = 1$ if the i th patient responds and receives second-stage treatment $B_k \left(R_i \sum_{k=1}^K Z_{ki} = R_i \right)$. Let T_i and C_i be the survival and censoring times from the time of first randomization for the i th patient, respectively. We assume that the survival time is independently right censored. Then the observed time and event indicator would be defined as $U_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. If we define V_i as the vector of baseline covariates, the observed data for the i th individual can be described as the set of random vectors $\{X_{ji}, R_i, R_i Z_{ki}, U_i, \delta_i, V_i, j = 1, \dots, J; k = 1, \dots, K, i = 1, 2, \dots, n\}$. Note that in this consideration we assume that R_i is always observed, which may not be true for patients who are censored or died prior to response evaluation. Customarily such patients are treated as non-responders and R_i is set to zero for them [2]. Whenever there is no ambiguity, we will drop the subscript i to represent a generic observation from the population.

2.2 Model

Our proposed model assumes proportional hazards within each treatment regime with respect to the baseline covariates V_i . However, the hazard functions for $J \times K$ treatment regimes are left unspecified, and could be proportional or non-proportional across regimes. In other words, a stratified proportional hazards model is used to account for the non-proportionality across treatment regimes. More specifically, let us denote the hazard and the cumulative hazard functions for treatment regime $A_j B_k$ as $\lambda_{jk}(t)$ and $\Lambda_{jk}(t) = \int_0^t \lambda_{jk}(s) ds$, respectively. Based on stratified proportional hazards with treatment regimes as strata, the hazard function for treatment regime $A_j B_k$, $j = 1, \dots, J$ and $k = 1, \dots, K$, could be written as

$$\lambda_{jk}(t) = \lambda_{jk0}(t) \exp\{\beta^T V\}, \quad j=1, \dots, J \text{ and } k=1, \dots, K, \quad (1)$$

where $\lambda_{jk0}(t)$ is the baseline hazard function for treatment regime $A_j B_k$, and β is a vector of coefficients corresponding to baseline covariates V . Note that in model (1), the parameter vector β is assumed to be constant across regimes, which implies that no interaction between treatment regimes and baseline covariates is assumed. For a general treatment regime, please see Sect. 2.1. Moreover, no functional form is specified for the regime-specific baseline hazard functions. For example, under a two-stage randomization design with $J = K = 2$, the hazard functions for treatment regimes $A_1 B_1$, $A_1 B_2$, $A_2 B_1$ and $A_2 B_2$ can be written as

$$\lambda_{11}(t) = \lambda_{110}(t) \exp\{\beta^T V\},$$

and similarly, $\lambda_{12}(t) = \lambda_{120}(t) \exp\{\beta^T V\}$, $\lambda_{21}(t) = \lambda_{210}(t) \exp\{\beta^T V\}$, and $\lambda_{22}(t) = \lambda_{220}(t) \exp\{\beta^T V\}$, respectively. Thus, irrespective of the treatment regimes, the effect of baseline covariate V on the hazard can be quantified by the log hazard ratio parameter β . The forms of the baseline hazard functions $\lambda_{110}(t)$, $\lambda_{120}(t)$, $\lambda_{210}(t)$, and $\lambda_{220}(t)$ are left unspecified, and could be non-proportional. However, we assume proportionality within each regime with respect to the baseline covariates, and the effects of the baseline covariates do not vary across regimes.

3 Inference

Our objective is to draw inference about treatment regimes $A_j B_k$, $j = 1, \dots, J$; $k = 1, \dots, K$. Note that the focus of the inference in model (1) is not the parameter vector β , rather it is to compare the hazards across treatment regimes. Based on the analytical framework of IPW (see for details in [9]) we define the weight function for treatment regime $A_j B_k$, $j = 1, \dots, J$ and $k = 1, \dots, K$, as

$$W_{jki} = X_{ji} \{(1 - R_i) + R_i Z_{ki} / \pi_{jk}\} / \pi_j,$$

where $\pi_j = P(X_{ji} = 1)$, and $\pi_{jk} = P(Z_{ki} = 1 | X_{ji} = 1, R_i = 1)$. Thus, both responders ($R_i = 1$) to A_j and non-responders ($R_i = 0$) to A_j are weighted with the inverse of the probability of randomization when evaluating the effect of the treatment regime $A_j B_k$. The probabilities of being assigned to B_k , $k = 1, \dots, K$, could be different for initial treatments A_j , $j = 1, \dots, J$.

For example, in some two-stage randomization studies, if the patients receive one initial treatment, particular choices in the second stage might be more toxic than others, and hence these patients are randomized with less probability into such choices. Based on the counting process notation described in Fleming and Harrington [10], the event and risk indicators for the i th patient are defined as $N_i(t) = I(U_i \leq t)$, and $Y_i(t) = I(U_i > t)$, respectively. We define the weighted event and risk indicators for treatment regime $A_j B_k$ as $N_{jki} = W_{jki} I(U_i \leq t)$ and $Y_{jki}(t) = W_{jki} I(U_i > t)$.

The partial likelihood estimate of β can be obtained by solving the pseudo-score equation [1]:

$$U(\beta) = \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \bar{V}_{jk}(t, \beta)\} dN_{jki}(t) = 0,$$

where

$$\bar{V}_{jk}(t, \beta) = \frac{\sum_{p=1}^n V_p Y_{jkp}(t) \exp\{\beta^T V_p\}}{\sum_{p=1}^n Y_{jkp}(t) \exp\{\beta^T V_p\}}.$$

Note that in defining $U(\cdot)$, we have utilized the weighted event and risk processes. The estimated vector of coefficients is denoted by $\hat{\beta}$. Then the Breslow estimator [12] of the cumulative baseline hazard for treatment regime $A_j B_k$ can be obtained as

$$\hat{\Lambda}_{jk0}(t, \hat{\beta}) = \sum_{i=1}^n \int_0^t \frac{dN_{jki}(s)}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}}.$$

A comparison of different treatment regimes can then be carried out in terms of the ratio of the cumulative baseline hazards. The ratio of the cumulative baseline hazards for comparing treatment regimes $A_j B_k$ and $A_{j'} B_{k'}$ is defined as

$$\theta_{jkj'k'}(t) = \frac{\Lambda_{jk0}(t)}{\Lambda_{j'k'0}(t)}. \quad (2)$$

This ratio of the cumulative baseline hazards equals the ratio of the cumulative hazards given the same values for covariates, because

$$\theta_{jkj'k'}(t) = \frac{\Lambda_{jk0}(t)}{\Lambda_{j'k'0}(t)} = \frac{\Lambda_{jk0}(t) \exp\{\beta^T V_i\}}{\Lambda_{j'k'0}(t) \exp\{\beta^T V_i\}} = \frac{\Lambda_{jk}(t)}{\Lambda_{j'k'}(t)}.$$

The ratio of the cumulative baseline hazards can be estimated by replacing the estimated cumulative baseline hazards in (2). Let us denote the corresponding estimator by $\hat{\theta}_{jkj'k'}(t)$.

For unweighted group comparisons, Wei and Schaubel [8] showed that such a ratio converges asymptotically to a Gaussian process. A similar argument outlined in Appendix A can be used to show that $\hat{\theta}_{jkj'k'}(t)$ follows a Gaussian process with mean $\theta_{jkj'k'}(t)$ and variance function $\sigma_{jkj'k'}^2(t)$, where

$$\begin{aligned}\sigma_{jkj'k'}^2(t) &= E \left\{ \xi_{jkj'k'}^2(t, \beta) \right\}, \\ \xi_{jkj'k'}(t, \beta) &= \frac{\Phi_{jki}(t, \beta)}{\Lambda_{j'k'0}(t)} - \frac{\Lambda_{jk0}(t) \Phi_{j'k'i}(t, \beta)}{\Lambda_{j'k'0}^2(t)}, \\ \Phi_{jki}(t, \beta) &= h_{jk}^T(t, \beta) \Omega^{-1}(\beta) \Psi_i(\beta) + \int_0^t y_{jk}^{(0)}(s, \beta)^{-1} dM_{jki}(s, \beta), \\ h_{jk}(t, \beta) &= - \int_0^t \bar{v}_{jk}(s, \beta) d\Lambda_{jk0}(s), \\ \Omega(\beta) &= \sum_{j=1}^J \sum_{k=1}^K \int_0^L \tau_{jk}(t, \beta) y_{jk}^{(0)}(t, \beta) d\Lambda_{jk0}(t), \\ \Psi_i(\beta) &= \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \bar{v}_{jk}(t, \beta)\} dM_{jki}(t, \beta), \\ dM_{jki}(t, \beta) &= dN_{jki}(t) - Y_{jki}(t) \exp \left\{ \beta^T V_i \right\} d\Lambda_{jk0}(t).\end{aligned}\quad (3)$$

Besides, $y_{jk}^{(0)}(t, \beta)$, $\bar{v}_{jk}(t, \beta)$ and $\tau_{jk}(t, \beta)$ are the limiting values of

$$Y_{jk}^{(0)}(t, \beta) = \frac{1}{n} \sum_{p=1}^n Y_{jkp}(t) \exp \left\{ \beta^T V_p \right\}, \quad \bar{V}_{jk}(t, \beta), \text{ and}$$

$$\frac{\sum_{p=1}^n V_p V_p^T Y_{jkp}(t) \exp \left\{ \beta^T V_p \right\}}{\sum_{p=1}^n Y_{jkp}(t) \exp \left\{ \beta^T V_p \right\}} - \bar{V}_{jk}(t, \beta) \bar{V}_{jk}^T(t, \beta), \text{ respectively. The variance function}$$

can be estimated by $\hat{\sigma}_{jkj'k'}^2(t) = \frac{1}{n} \sum_{i=1}^n \hat{\xi}_{jkj'k'}^2(t, \hat{\beta})$. The detailed steps for computing $\hat{\xi}_{jkj'k'}(t, \hat{\beta})$ are described in the Appendix B. One might expect that the asymptotic distribution of $\ln \left\{ \hat{\theta}_{jkj'k'}(t) \right\}$ would be closer to a Gaussian process than $\hat{\theta}_{jkj'k'}(t)$. Based on the delta method, the variance of $\ln \left\{ \hat{\theta}_{jkj'k'}(t) \right\}$ can be estimated by

$$\hat{\sigma}_{jkj'k'}'^2(t) = \frac{\hat{\sigma}_{jkj'k'}^2(t)}{\hat{\theta}_{jkj'k'}(t)}.$$

We will use simulation to investigate the properties of $\hat{\theta}_{jkj'k'}(t)$ and $\ln \left\{ \hat{\theta}_{jkj'k'}(t) \right\}$ in a sequentially randomized setting in Sect. 4.

Based on the log ratio estimator of the cumulative baseline hazards, a Wald-type test can be used for comparing different treatment regimes. For example, for comparing treatment regimes A_{jB_k} and $A_{j'B_k}$ at a specific time point t_0 , the null hypothesis can be described as $H_{jkj'k'}: \ln \{ \theta_{jkj'k'}(t_0) \} = 0$, and the test statistic can be written as

$$D_{jkj'k'} = \frac{\ln \left\{ \hat{\theta}_{jkj'k'}(t_0) \right\}}{\hat{\sigma}_{jkj'k'}'(t_0)}.$$

This test statistic is then compared to a standard normal distribution. Because the cumulative hazard can be viewed as an “accumulation” of the hazard over time, the choice of t_0 will

depend on the disease-specific survival patterns as well as the treatments. If there are more than two treatment regimes involved, multiple treatment regimes can be compared by performing an overall test of difference among all regimes using the Wald Chi-square test. More specifically, in our case, let $\theta(t_0) = [\theta_{jk11}(t_0), j = 1, \dots, J; k = 1, \dots, K; jk \neq 11]^T$ be the ratios $(JK - 1)$ -dimensional vector of cumulative hazard ratios at t_0 , and

$$\Sigma(t_0) = \begin{pmatrix} \sigma_{1111}^2(t_0) & \sigma_{1111 \cdot 1211}(t_0) & \cdots & \sigma_{1111 \cdot JK11}(t_0) \\ \sigma_{1111 \cdot 1211}(t_0) & \sigma_{1211}^2(t_0) & \cdots & \sigma_{1211 \cdot JK11}(t_0) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1111 \cdot JK11} & \sigma_{1211 \cdot JK11} & \cdots & \sigma_{JK11}^2(t_0) \end{pmatrix}$$

be the corresponding variance-covariance matrix. The diagonal elements of $\sigma_{jk11}^2(t_0)$ are the variances of $\hat{\theta}_{jk11}(t)$, $j = 1, \dots, J; k = 1, \dots, K$. The off-diagonal elements are the covariance between $\hat{\theta}_{jk11}(t)$ and $\hat{\theta}_{j'k'11}(t)$, $j = 1, \dots, J; k = 1, \dots, K; jk \neq j'k'$. Then for testing $H_0: A_{11}(t_0) = \cdots = A_{JK}(t_0)$ or equivalently $H_0: \ln\{\theta(t_0)\} = 0$, the test statistic can be written as

$$\chi_{JK-1}^2 = \left[\ln\{\hat{\theta}(t_0)\} \right]^T \hat{\Sigma}^{-1}(t_0) \left[\ln\{\hat{\theta}(t_0)\} \right],$$

where

$$\hat{\Sigma}'(t_0) = \left(\frac{\partial \ln\{\hat{\theta}(t_0)\}}{\partial \hat{\theta}(t_0)} \right)^T \hat{\Sigma}(t_0) \left(\frac{\partial \ln\{\hat{\theta}(t_0)\}}{\partial \hat{\theta}(t_0)} \right).$$

This test statistic is then compared to a Chi-square distribution with $(JK - 1)$ degrees of freedom.

4 Simulation Study

For simplicity, a simulation study was carried out under a two-stage randomization design with only two treatment options for each stage ($J = K = 2$). Under this design, the performance of the ratio estimator, log ratio estimator and proposed tests were assessed under the following scenarios:

- Scenario I: The survival distributions for treatment regimes A_1B_1 and A_1B_2 are the same (note that the treatment regimes A_1B_1 and A_1B_2 share the same induction therapy A_1);
- Scenario II: The survival distributions for treatment regimes A_1B_1 and A_2B_1 are the same (note that the treatment regimes A_1B_1 and A_2B_1 share the same second-stage therapy B_1);
- Scenario III: The survival distributions for treatment regimes A_1B_1 and A_2B_2 are the same (note that the treatment regimes A_1B_1 and A_2B_2 have different induction and second-stage therapies);

– Scenario IV: No pattern is specified in the population.

An indicator (X_{1i}) for initial treatment A_1 was generated following the Bernoulli(0.5) distribution. An indicator for response (R_i) was generated from the Bernoulli distribution with 60 % response rate. An indicator (Z_{1i}) for second-stage treatment B_1 was drawn from the Bernoulli(0.5) distribution where $R_i = 1$. Based on the stratified proportional hazards model (1), the death time T_{jki} for the i th patient in the jk th regime is calculated based on the Weibull distribution using

$$T_{jki} = \left\{ -\log(\phi_i) / \left[\alpha_{jk} \exp \left\{ \beta^T V_i \right\} \right] \right\}^{1/\gamma_{jk}},$$

for $j = 1, 2$, and $k = 0, 1, 2$, where ϕ_i was generated from the Uniform(0,1) distribution. $V_i = [V_{1i}, V_{2i}]^T$ and both V_{1i} and V_{2i} were generated from the Bernoulli(0.5) distribution. The parameter vector β was set to $[0.5, 0.5]^T$. A variety of values were used for α_{jk} and γ_{jk} to assess the properties of the ratio estimator, log ratio estimator and proposed tests under different scenarios (Table 1). Null hypotheses $H_{1211} : \ln\{\theta_{1211}(t)\} = 0$, $H_{2111} : \ln\{\theta_{2111}(t)\} = 0$, and $H_{2211} : \ln\{\theta_{2211}(t)\} = 0$ were true in the population under scenarios I, II and III, respectively. For scenario IV, parameter values were chosen with no pattern. The true cumulative hazard ratios for comparing A_1B_2 , A_1B_2 , and A_1B_2 to A_1B_2 under different scenarios are plotted in Fig. 1. Censoring time was generated according to the Uniform($\tau/2$, τ) distribution, where the value of τ was chosen to be 5, resulting in censoring percentage ranging from 10 to 30 %. The sample size was varied from 400 to 1200, and 1000 replications were performed under each scenario. Time t_0 was chosen to be the 75th percentile of the observed time under each scenario. For each sample, the ratios and the log ratios of the cumulative baseline hazards for comparing A_1B_2 , A_2B_1 , and A_2B_2 to A_1B_1 were calculated, and the Wald-type tests were performed based on the log ratio estimates.

Table 2 shows the ratio and log ratio estimates (EST), absolute bias (BIAS), asymptotic standard error (ASE), empirical standard deviation (ESD) and 95 % coverage probability (CP) under each scenario. For example, under scenario I and at a sample size of 800, the estimate of $\theta_{1211}(t_0)$ was 1.01 with an absolute bias of 0.01. The asymptotic standard error was 0.119, close to the empirical standard deviation of 0.121. The corresponding coverage probability was 93 % based on the estimate and asymptotic standard error. Similarly, the estimate for the $\ln\{\theta_{1211}(t_0)\}$ was 0.00 with an absolute bias of 0.01. The asymptotic standard error (0.118) and empirical standard deviation (0.120) were close. The coverage probability was 94 %, better than that for $\hat{\theta}_{1211}(t_0)$. In most of the cases, the estimates were approximately unbiased; with the absolute biases differed from 0.00 to 0.02. The asymptotic standard errors were close to the empirical standard deviations, demonstrating that the estimated standard errors were consistent. Most coverage probabilities were close to 95 %, attaining the nominal level. As sample size increased, the estimated standard error decreased and the coverage probability increased. Some of the log ratio estimates were less biased than the corresponding ratio estimates, and the overall coverage probabilities were closer to 95 % based on the log ratio estimates than those based on the ratio estimates.

Table 3 presents the rejection rates for testing null hypotheses $H_{1211} : \ln\{\theta_{1211}(t)\} = 0$, $H_{2111} : \ln\{\theta_{2111}(t)\} = 0$, and $H_{2211} : \ln\{\theta_{2211}(t)\} = 0$, separately under each scenario. Because the null hypotheses $H_{1211} : \ln\{\theta_{1211}(t)\} = 0$, $H_{2111} : \ln\{\theta_{2111}(t)\} = 0$, and $H_{2211} : \ln\{\theta_{2211}(t)\} = 0$, were true under scenarios I, II and III, respectively, the rejection rates for H_{1211} , H_{2111} , and H_{2211} were close to the nominal level of 0.05 under scenarios I, II and III, respectively, suggesting that the tests were approximately unbiased. The rejection rates were relatively larger under the smaller sample size of 400. As sample size increased, the rejection rate approached 0.05.

5 Analysis of Neuroblastoma Data

From 1991 to 1996, the Children's Cancer Group conducted a randomized clinical trial to study the effect of a combination of myeloablative chemotherapy, total-body irradiation and transplantation of autologous bone marrow purged of cancer cells (ABMT) to a standard chemotherapy in treating children with high risk neuroblastoma [13]. Since relapse is common after completion of induction therapies among children with high risk neuroblastoma, patients without progressive disease (PD) or histologically confirmed disease (HCD) after induction therapies were then randomly assigned to receive either 13-cis-retinoic acid (cis-RA) or no further therapy. Therefore, the high risk neuroblastoma clinical trial followed a two-stage randomization design. By the end of the study, a total of 379 children participated in the first randomization, with 190 children assigned to ABMT, and 189 children assigned to chemotherapy. After the completion of the first-stage therapy, 203 children had no PD or HCD, and thus participated in the second randomization. During the second randomization, 102 children were assigned to cis-RA, and 101 patients were assigned to no further therapy. Matthay et al. [13] reported 55 more children who participated in the second randomization, resulting in 130 children assigned to cis-RA and 128 children assigned to no further therapy. These children were not included in our setting because they did not participate in the first randomization. This two-stage randomization design resulted in four treatment regimes: (i) treat with ABMT, followed by cis-RA if no PD or HCD (AC); (ii) treat with ABMT, followed by no further therapy if no PD or HCD (AN); (iii) treat with chemotherapy, followed by cis-RA if no PD or HCD (CC); (iv) treat with chemotherapy, followed by no further therapy if no PD or HCD (CN).

A stratified proportional hazards model based on (1) was applied to the neuroblastoma data, using treatment regimes as strata. The only baseline characteristics found imbalanced among different treatment groups was "Evan's stage" in Matthay et al. [13], and thus was included in the model as one of the covariates. Additionally, "age" was also included in the model as a continuous covariate. The time to second randomization did not differ significantly across the first stage treatments ($p = 0.26$). The resulting log ratio estimates of the cumulative baseline hazards and their corresponding 95 % confidence intervals within the time interval of [0, 2000] days are shown in Fig. 2. From the plot of the log ratio estimate of the cumulative baseline hazards comparing treatment regime AN to AC, we observed a notable difference around 450 days. However, the horizontal "zero" line was within the confidence interval most of the time, suggesting that there was no significant difference between treatment regimes AN and AC. In the plot for comparing treatment regime CC to AC the confidence band moved further away from "zero" as time increased. Thus, we would suspect

that children following the treatment regime AC had better overall survival compared to those following the regime CC. Besides, the confidence band also became narrower as time increased since there was higher variability with fewer events at the beginning of the study. Similar results were observed from the plot for comparing treatment regime CN to AC. The hazard ratios (95 % confidence interval) for comparing AN, CC, and CN to AC at year 3 were estimated to be 1.10 (0.78, 1.43), 1.13 (0.74, 1.52), and 1.02 (0.67, 1.38), respectively. The hazard ratios (95 % confidence interval) for comparing AN, CC and CN to AC at year 5 were estimated to be 1.22 (0.86, 1.57), 1.38 (0.92, 1.85), and 1.43 (0.95, 1.91), respectively. We also used the Wald Chi-square test to evaluate the overall difference among four treatment regimes at year 5. It resulted in a p-value of 0.17, showing that there was no overall significant difference in the cumulative hazard among four treatment regimes at 5 years. Therefore, we would suspect that children with high risk neuroblastoma would have a similar overall 5-year survival irrespective of the induction therapy they received. After the completion of induction therapies, whether they would be subsequently treated with cis-RA or not, did not significantly improve the overall 5-year survival either.

6 Discussion

In this article, we proposed a stratified proportional hazards model to estimate the cumulative treatment effect for treatment regimes from sequentially randomized clinical trials. This approach is similar to that advocated by Wei and Schaubel [8] for a single-stage randomization. Comparisons among treatment regimes were performed by testing the log ratio of the estimated cumulative hazards. Simulation results showed that the ratio estimator was approximately unbiased, and the coverage probabilities were close to the nominal level. However, the log ratio estimator performed better, with smaller absolute biases and better coverage probabilities. The comparative hypothesis testing can also be performed maintaining adequate type I errors from the proposed model under moderate sample size based on the Wald-type tests. In this paper we assumed that the randomization probabilities are known (by design). Estimation of these probabilities from the observed data might result in more efficient estimators.

Wald-type tests were used to assess the survival difference between treatment regimes at a specific time point. Although such point-wise comparisons are of interest in many diseases-specific areas, comparisons of treatment regimes based on overall hazard curves may be of importance alongside the construction of simultaneous confidence bands. This issue is beyond the scope of this manuscript and is being considered as a separate publication.

Acknowledgements

We would like to thank COG Neuroblastoma Disease Committee for kind permission to use the neuroblastoma data set, especially Dr. Wendy London for her help during the application process. We thank Dr. Susan Ellenberg from University of Pennsylvania for her insightful comments. We also thank the referees of this article for their helpful comments. Dr. Wahed's research was in part supported by a National Institute of Mental Health Grant P30 MH090333.

Appendix A: Outline of the asymptotic normality of $\hat{\theta}_{jkj'k'}(t)$

The following are the equivalent assumptions in sequentially randomized clinical trials to those outlined in Sect. 3 of Wei and Schaubel [8].

- a. $(X_{ji}, R_i, R_i Z_{ki}, U_i, V_i)$ are i.i.d. random vectors for $i = 1, \dots, n$.
- b. Elements of V_i have bounded total variation for $i = 1, \dots, n$.
- c. The cumulative hazard is finite over a pre-specified interval $[0, L]$ such that $P(U_i > L) > 0$.
- d. (d)

$$y_{jk}^{(1)}(t, \beta) = \frac{\partial}{\partial \beta} y_{jk}^{(0)}(t, \beta) \quad \text{and} \quad y_{jk}^{(2)}(t, \beta) = \frac{\partial^2}{\partial \beta \partial \beta^T} y_{jk}^{(0)}(t, \beta),$$

where

$$y_{jk}^{(d)}(t, \beta) = \lim_{n \rightarrow \infty} Y_{jk}^{(d)}(t, \beta) \quad \text{for } d=0, 1, 2,$$

where

$$Y_{jk}^{(d)}(t, \beta) = n^{-1} \sum_{i=1}^n Y_{jki}(t) V_i^{\otimes d} \exp\{\beta^T V_i\} \quad \text{for } d=0, 1, 2,$$

where $V_i^{\otimes 0} = 1, V_i^{\otimes 1} = V_i, V_i^{\otimes 2} = V_i V_i^T,$

with $y_{jk}^{(1)}(t, \beta)$ and $y_{jk}^{(2)}(t, \beta)$ bounded away from 0 for $t \in [0, L]$ and β in an open set.

- e. Positive-definiteness of the matrix

$$\Omega(\beta) = \sum_{j=1}^J \sum_{k=1}^K \int_0^L \tau_{jk}(t, \beta) y_{jk}^{(0)}(t, \beta) d\Lambda_{jk0}(t, \beta),$$

where

$$\tau_{jk}(t, \beta) = y_{jk}^{(2)}(t, \beta) / y_{jk}^{(0)}(t, \beta) - \bar{v}_{jk}(t, \beta)^{\otimes 2},$$

Consistency and asymptotic normality of $\hat{\theta}_{jkj'k'}(t)$ Consistency and asymptotic normality of $\hat{\theta}_{jkj'k'}(t)$ can be established in a similar manner as in the proof of Theorems 1 and 2 given in the Web Appendix of Wei and Schaubel [8] as long as the following results hold:

1. $\hat{\beta} \xrightarrow{a.s.} \beta_0$.
2. $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically normal

3. $\hat{\Lambda}_{jk}(t, \hat{\beta})$ is a uniformly consistent estimator of $\Lambda_{jk}(t, \beta)$.

1. Consistency of $\hat{\beta}$

Recall that β is a solution to the equation

$$U_n(\beta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \bar{V}_{jk}(t, \beta)\} dN_{jki}(t) = 0.$$

First note that the processes $Y_{jk}(t)$ and $N_{jk}(t)$ are both cadlag processes and hence they are Donsker. Since the classes $\{\beta \in B\}$ and $\{V\}$ are trivially Donsker, the functions $Y_{jk}(t) \exp\{\beta^T V\}$, $VY_{jk}(t) \exp\{\beta^T V\}$, $VV^T Y_{jk}(t) \exp\{\beta^T V\}$ are all Donsker for $t \in [0, L]$, $\beta \in B$. The derivative of $U_n(\beta)$ with respect to β is $\Omega_n(\beta)$, where

$$\Omega_n(\beta) = \sum_{j=1}^J \sum_{k=1}^K \int_0^L \left[\frac{\sum_{p=1}^n V_p V_p^T Y_{jkp}(t) \exp(\beta^T V_p)}{\sum_{p=1}^n Y_{jkp}(t) \exp(\beta^T V_p)} - \left\{ \frac{\sum_{p=1}^n V_p Y_{jkp}(t) \exp(\beta^T V_p)}{\sum_{p=1}^n Y_{jkp}(t) \exp(\beta^T V_p)} \right\}^{\otimes 2} \right] \frac{1}{n} \sum_{i=1}^n dN_{jki}(s),$$

where $\otimes 2$ denotes the outer product. All functions in the above expressions are Glivenko–

Cantelli and the limiting value $y_{jk}^{(0)}(t, \beta)$ of $\frac{1}{n} \sum_{p=1}^n Y_{jkp}(t) \exp(\beta^T V_p)$ is bounded away from zero. Therefore,

$$\sup_{\beta \in B} |\Omega_n(\beta) - \Omega(\beta)| \xrightarrow{\text{a.s.}} 0,$$

where $\Omega(\beta)$ is defined in (3). Since $\Omega(\beta)$ is positive semidefinite, $U_n(\beta)$ is almost surely convex for large n . Therefore, $\hat{\beta} \xrightarrow{\text{a.s.}} \beta_0$.

2. Asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta_0)$

We write $\Psi_n(\beta)$ as

$$\Psi_n(\beta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \bar{v}_{jk}(t, \beta)\} dM_{jki}(t) = 0,$$

where

$$M_{jki}(t, \beta) = N_{jki}(t) - \int_0^L Y_{jki}(t) \exp(\beta^T V_i) d\Lambda_{jki}(t).$$

Let

$$\Psi(\beta) = E \left[\sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V - \bar{v}_{jk}(t, \beta)\} dM_{jk}(t) \right],$$

where

$$\bar{v}_{jk}(t, \beta) = \frac{E \left[V Y_{jk}(t) \exp(\beta^T V) \right]}{E \left[Y_{jk}(t) \exp(\beta^T V) \right]}.$$

Then the arguments on pages 56 and 57 in Kosorok (2008, Chap. 4, [11]) can be applied to show that $\sqrt{n}(\hat{\beta} - \beta_0)$ weakly converges to a mean zero normal random vector with covariance matrix $\Omega^{-1}(\beta_0)$.

3. Almost sure convergence of $\hat{\Lambda}_{jk}(t, \hat{\beta})$

Note that

$$\hat{\Lambda}_{jk}(t, \hat{\beta}) = \int_0^t \frac{P_n dN_{jk}(s)}{P_n Y_{jk}(s) \exp(\hat{\beta}^T V)},$$

where P_n is the empirical measure, namely, $P_n f = \frac{1}{n} \sum_{i=1}^n f(x_i)$ for real-valued functions $f: \mathcal{X} \rightarrow \mathcal{R}$, $\mathcal{X}$ being the sample space. This is basically the same estimator defined in the first display on page 57 of Kosorok (2008, Chap. 4, [11]). Hence, a straightforward application of the argument therein leads to

$$\sup_{t \in [0, L]} \left| \hat{\Lambda}_{jk}(t, \hat{\beta}) - \Lambda_{jk0}(t) \right| \xrightarrow{\text{a.s.}} 0.$$

With the above results (1–3), it is now straightforward to apply the arguments in the Web Appendix A of Wei and Schaubel [8] to establish that $\hat{\theta}_{jkj'k'}(t)$ is uniformly consistent and asymptotically normal.

Appendix B: Estimation of $\xi^{jkj'k'}(t, \beta)$

The detailed steps for computing $\hat{\xi}_{jkj'k'}(t, \hat{\beta})$ are as follows:

Step 1: Calculate
$$\hat{V}_{jk}(s, \hat{\beta}) = \frac{\sum_{p=1}^n V_p Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}}.$$

Step 2: Calculate

$$\begin{aligned}\hat{h}_{jk}(t, \hat{\beta}) &= -\int_0^t \hat{V}_{jk}(s, \hat{\beta}) d\hat{\Lambda}_{jk0}(s, \hat{\beta}) \\ &= -\int_0^t \hat{V}_{jk}(s, \hat{\beta}) \frac{\sum_{i=1}^n \frac{dN_{jki}(s)}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}}}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} \\ &= -\sum_{i=1}^n \int_0^t \frac{\hat{V}_{jk}(s, \hat{\beta}) dN_{jki}(s)}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}}.\end{aligned}$$

Step 3: Define

$$\hat{\tau}_{jk}(t, \hat{\beta}) = \frac{\sum_{p=1}^n V_p V_p^T Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} - \hat{V}_{jk}(t, \hat{\beta}) \hat{V}_{jk}^T(t, \hat{\beta}).$$

Step 4: Calculate

$$\begin{aligned}\hat{\Omega}(\hat{\beta}) &= \frac{1}{n} \sum_{j=1}^J \sum_{k=1}^K \int_0^L \hat{\tau}_{jk}(t, \hat{\beta}) \frac{\sum_{p=1}^n Y_{jkp}(t) \exp\{\hat{\beta}^T V_p\} d\hat{\Lambda}_{jk0}(t, \hat{\beta})}{\sum_{i=1}^n \frac{dN_{jki}(t)}{\sum_{p=1}^n Y_{jkp}(t) \exp\{\hat{\beta}^T V_p\}}} \\ &= \frac{1}{n} \sum_{j=1}^J \sum_{k=1}^K \int_0^L \hat{\tau}_{jk}(t, \hat{\beta}) \frac{\sum_{p=1}^n Y_{jkp}(t) \exp\{\hat{\beta}^T V_p\}}{\sum_{i=1}^n \frac{dN_{jki}(t)}{\sum_{p=1}^n Y_{jkp}(t) \exp\{\hat{\beta}^T V_p\}}} \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^K \int_0^L \hat{\tau}_{jk}(t, \hat{\beta}) dN_{jki}(t).\end{aligned}$$

Step 5: Calculate

$$\begin{aligned}\hat{\Psi}_i(\hat{\beta}) &= \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \hat{V}_{jk}(t, \hat{\beta})\} d\hat{M}_{jki}(t, \hat{\beta}) \\ &= \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \hat{V}_{jk}(t, \hat{\beta})\} [dN_{jki}(t) - Y_{jki} \exp\{\hat{\beta}^T V_i\} d\hat{\Lambda}_{jk0}(t, \hat{\beta})] \\ &= \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \hat{V}_{jk}(t, \hat{\beta})\} dN_{jki}(t) \\ &\quad - \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \hat{V}_{jk}(t, \hat{\beta})\} Y_{jki} \exp\{\hat{\beta}^T V_i\} \\ &\quad \times \sum_{i=1}^n \frac{dN_{jki}(t)}{\sum_{p=1}^n Y_{jkp}(t) \exp\{\hat{\beta}^T V_p\}} \\ &= \sum_{j=1}^J \sum_{k=1}^K \int_0^L \{V_i - \hat{V}_{jk}(t, \hat{\beta})\} dN_{jki}(t) \\ &\quad - \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^K \int_0^L \frac{\{V_i - \hat{V}_{jk}(t, \hat{\beta})\} Y_{jki} \exp\{\hat{\beta}^T V_i\}}{\sum_{p=1}^n Y_{jkp}(t) \exp\{\hat{\beta}^T V_p\}} dN_{jki}(t).\end{aligned}$$

Step 6: Define

$$\begin{aligned}
\hat{\Phi}_{jki}^L(t, \hat{\beta}) &= \int_0^t \frac{n}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} d\hat{M}_{jki}(s, \hat{\beta}) \\
&= \int_0^t \frac{n}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} \\
&\quad \times \left[dN_{jki}(s) - Y_{jki} \exp\{\hat{\beta}^T V_i\} d\hat{\Lambda}_{jk0}(s, \hat{\beta}) \right] \\
&= \int_0^t \frac{n dN_{jki}(s)}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} \\
&\quad - \int_0^t \frac{n Y_{jki} \exp\{\hat{\beta}^T V_i\}}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} \frac{\sum_{i=1}^n dN_{jki}(s)}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} \\
&= \int_0^t \frac{n dN_{jki}(s)}{\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\}} \\
&\quad - \sum_{i=1}^n \int_0^t \frac{n Y_{jki} \exp\{\hat{\beta}^T V_i\} dN_{jki}(s)}{\left[\sum_{p=1}^n Y_{jkp}(s) \exp\{\hat{\beta}^T V_p\} \right]^2}.
\end{aligned}$$

Step 7: Calculate $\hat{\Phi}_{jki}(t, \hat{\beta}) = \hat{h}_{jk}^T(t, \hat{\beta}) \hat{\Omega}^{-1}(\hat{\beta}) \hat{\Psi}_i(\hat{\beta}) + \hat{\Phi}_{jki}^L(t, \hat{\beta})$.

Step 8: Calculate $\hat{\xi}_{jkj'k'i}(t, \hat{\beta}) = \frac{\hat{\Phi}_{jki}(t, \hat{\beta})}{\hat{\Lambda}_{j'k'0}(t, \hat{\beta})} - \frac{\hat{\Lambda}_{jk0}(t, \hat{\beta}) \hat{\Phi}_{j'k'i}(t, \hat{\beta})}{\hat{\Lambda}_{j'k'0}^2(t, \hat{\beta})}$.

References

1. Lokhnygina Y, Helderbrand JD. Cox regression methods for two-stage randomization designs. *Biometrics*. 2007; 63:422–428. [PubMed: 17425633]
2. Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2002; 58:48–57. [PubMed: 11890326]
3. Guo X, Tsiatis AA. A weighted risk estimator for survival distributions in two-stage randomization designs with censored survival data. *Int J Biostat*. 2005; 1:1–15.
4. Hernan MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006; 98:237–242. [PubMed: 16611197]
5. Feng W, Wahed AS. Supremum weighted log-rank test and sample size for comparing two-stage adaptive treatment strategies. *Biometrika*. 2008; 95:695–707.
6. Miyahara S, Wahed AS. Weighted Kaplan–Meier estimators for two-stage treatment regimes. *Stat Med*. 2010; 29:2581–2591. [PubMed: 20799259]
7. Goldberg Y, Kosorok MR. Q-learning and censored data. *Ann Stat*. 2012 doi:10.1214/12-AOS968.
8. Wei G, Schauble DE. Estimating cumulative treatment effects in the presence of nonproportional hazards. *Biometrics*. 2008; 64:724–732. [PubMed: 18162108]
9. Wahed AS, Tsiatis AA. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2004; 60:124–133. [PubMed: 15032782]
10. Fleming, TR.; Harrington, DP. Counting processes and survival analysis. Wiley; New York: 1991.
11. Kosorok, MR. Introduction to Empirical Processes and Semiparametric Inference. Springer; New York: 2008.

12. Breslow NE. Contribution to the discussion of the paper by D.R. Cox. J R Stat Soc, Ser B. 1972; 34:187–220.
13. Matthay KK, Reynolds P, Seeger RC, Shimada H, Adkins ES, Haas-Kogan A, et al. Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cis-retinoic acid: A Children's Oncology Group study. J Clin Oncol. 2009; 27:1007–1013. [PubMed: 19171716]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1. The true cumulative hazard ratios for comparing A_1B_2 , A_2B_1 , and A_2B_2 to A_1B_1 under different scenarios. *Solid line* A_1B_2 vs. A_1B_1 , *dashed line* A_2B_1 vs. A_1B_1 , *dotted line* A_2B_2 vs. A_1B_1

Fig. 2.

Estimated log ratios of the cumulative baseline hazards and the corresponding 95 % pointwise confidence intervals from the neuroblastoma study. AC: “treat with ABMT, and then followed by cis-RA if no PD or HCD,” AN: “treat with ABMT, and then followed by no further therapy if no PD or HCD,” CC: “treat with chemotherapy, and then followed by cis-RA if no PD or HCD,” CN: “treat with chemotherapy, and then followed by no further therapy if no PD or HCD”

Table 1

Different values for α_{jk} and γ_{jk} , $j = 1, 2$, and $k = 0, 1, 2$, under different scenarios (for scenario details see Sect. 4)

Scenario	γ_{10}	γ_{11}	γ_{12}	γ_{20}	γ_{21}	γ_{22}	α_{10}	α_{11}	α_{12}	α_{20}	α_{21}	α_{22}	Censoring (%)
I	1.4	1.4	1.4	1.2	1.2	1.2	0.7	0.4	0.4	0.3	0.2	0.25	10
II	1.0	1.5	1.2	1.0	1.5	1.3	0.5	0.2	0.1	0.5	0.2	0.05	23
III	1.2	1.0	1.2	1.2	1.5	1.0	0.1	0.4	0.7	0.1	0.04	0.4	30
IV	1.0	1.5	1.2	1.5	1.0	1.4	0.7	0.4	0.35	0.3	0.2	0.25	10

Table 2

Simulation results under each scenario (for scenario details see Sect. 4). EST: ratio or log ratio estimate; BIAS: absolute bias; ASE: asymptotic standard error; ESD: empirical standard deviation; and CP: 95 % coverage probability

Scenario	n	$\theta_{211}(t_0)$					$\ln(\theta_{211}(t_0))$				
		EST	BIAS	ASE	ESD	CP	EST	BIAS	ASE	ESD	CP
I	400	1.01	0.00	0.166	0.175	0.93	-0.01	0.02	0.165	0.171	0.94
	800	1.01	0.01	0.119	0.121	0.93	0.00	0.01	0.118	0.120	0.94
	1200	1.00	0.01	0.097	0.102	0.93	0.00	0.01	0.097	0.101	0.94
II	400	0.51	0.01	0.084	0.083	0.95	-0.69	0.00	0.167	0.166	0.95
	800	0.51	0.01	0.060	0.061	0.96	-0.69	0.01	0.119	0.121	0.95
	1200	0.51	0.00	0.049	0.051	0.95	-0.69	0.00	0.098	0.100	0.95
III	400	1.14	0.01	0.113	0.105	0.97	0.13	0.00	0.099	0.092	0.97
	800	1.14	0.01	0.082	0.073	0.97	0.13	0.00	0.072	0.064	0.97
	1200	1.14	0.00	0.067	0.058	0.98	0.13	0.00	0.059	0.05	0.98
IV	400	0.79	0.01	0.121	0.126	0.93	-0.24	0.02	0.153	0.157	0.95
	800	0.79	0.01	0.087	0.089	0.92	-0.24	0.02	0.109	0.113	0.93
	1200	0.79	0.01	0.071	0.072	0.94	-0.24	0.02	0.089	0.090	0.94

Scenario	n	$\theta_{2111}(t_0)$					$\ln(\theta_{2111}(t_0))$				
		EST	BIAS	ASE	ESD	CP	EST	BIAS	ASE	ESD	CP
I	400	0.43	0.01	0.075	0.077	0.94	-0.86	0.01	0.174	0.179	0.94
	800	0.43	0.00	0.053	0.052	0.96	-0.86	0.00	0.123	0.122	0.94
	1200	0.43	0.00	0.043	0.044	0.95	-0.86	0.01	0.101	0.103	0.95
II	400	1.02	0.02	0.181	0.197	0.93	0.00	0.00	0.178	0.190	0.94
	800	1.00	0.01	0.128	0.131	0.95	0.00	0.00	0.127	0.129	0.95
	1200	1.01	0.01	0.105	0.108	0.94	0.00	0.00	0.105	0.107	0.95
III	400	0.51	0.02	0.098	0.103	0.94	-0.70	0.01	0.195	0.202	0.94
	800	0.50	0.01	0.070	0.072	0.95	-0.70	0.02	0.140	0.143	0.95
	1200	0.50	0.01	0.057	0.057	0.96	-0.70	0.01	0.115	0.113	0.95
IV	400	0.44	0.01	0.076	0.075	0.95	-0.82	0.02	0.170	0.168	0.95
	800	0.44	0.01	0.053	0.054	0.96	-0.83	0.01	0.12	0.123	0.94

Scenario	<i>n</i>	$\theta_{211}(\zeta_0)$					$\ln\{\theta_{211}(\zeta_0)\}$				
		EST	BIAS	ASE	ESD	CP	EST	BIAS	ASE	ESD	CP
	1200	0.44	0.01	0.044	0.045	0.95	-0.83	0.01	0.099	0.101	0.95
Scenario	<i>n</i>	$\theta_{2211}(\zeta_0)$					$\ln\{\theta_{2211}(\zeta_0)\}$				
		EST	BIAS	ASE	ESD	CP	EST	BIAS	ASE	ESD	CP
I	400	0.49	0.00	0.084	0.088	0.93	-0.72	0.01	0.171	0.182	0.93
	800	0.49	0.00	0.059	0.058	0.95	-0.72	0.01	0.122	0.119	0.96
	1200	0.49	0.00	0.049	0.049	0.93	-0.72	0.01	0.100	0.100	0.95
II	400	0.38	0.00	0.069	0.069	0.94	-0.99	0.01	0.182	0.181	0.94
	800	0.38	0.00	0.049	0.048	0.95	-0.99	0.01	0.130	0.129	0.95
	1200	0.37	0.00	0.040	0.039	0.94	-0.99	0.01	0.106	0.104	0.95
III	400	1.01	0.01	0.175	0.180	0.94	0.00	0.01	0.172	0.177	0.95
	800	1.01	0.00	0.125	0.129	0.94	0.00	0.01	0.124	0.127	0.95
	1200	1.00	0.00	0.102	0.100	0.95	0.00	0.00	0.102	0.099	0.95
IV	400	0.61	0.01	0.101	0.104	0.93	-0.51	0.00	0.165	0.172	0.95
	800	0.60	0.00	0.071	0.071	0.95	-0.51	0.01	0.118	0.117	0.95
	1200	0.60	0.00	0.058	0.058	0.94	-0.51	0.01	0.096	0.097	0.96

Table 3

Rejection rates for testing null hypotheses $H_{1211} : \ln\{\theta_{1211}(t)\} = 0$, $H_{2111} : \ln\{\theta_{2111}(t)\} = 0$, and $H_{2211} : \ln\{\theta_{2211}(t)\} = 0$ under different scenarios. Scenarios I, II and III respectively represent null hypotheses H_{1211} , H_{2111} and H_{2211} , while for scenario IV, all three null hypotheses are false

Scenario	n	H_{1211}	H_{2111}	H_{2211}
I	400	0.059	0.999	0.991
	800	0.053	1.000	1.000
	1200	0.055	1.000	1.000
II	400	0.992	0.064	1.000
	800	1.000	0.048	1.000
	1200	1.000	0.050	1.000
III	400	0.233	0.937	0.050
	800	0.444	1.000	0.050
	1200	0.603	1.000	0.046
IV	400	0.351	0.992	0.864
	800	0.587	1.000	0.994
	1200	0.753	1.000	0.999