HHS Public Access

Author manuscript

Lifetime Data Anal. Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:

Lifetime Data Anal. 2015 April ; 21(2): 259–279. doi:10.1007/s10985-014-9301-0.

Quantifying the Average of the Time-varying Hazard Ratio via a Class of Transformations

QINGXIA CHEN,
Department of Biostatistics, Vanderbilt University, Nashville, TN 37232
cindy.chen@vanderbilt.edu

DONGLIN ZENG,
Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

JOSEPH G. IBRAHIM,
Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

MING-HUI CHEN,
Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269

ZHIYING PAN, and
Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320

XIAODONG XUE
Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320

Abstract

The hazard ratio derived from the Cox model is a commonly used summary statistic to quantify a treatment effect with a time-to-event outcome. The proportional hazards assumption of the Cox model, however, is frequently violated in practice and many alternative models have been proposed in the statistical literature. Unfortunately, the regression coefficients obtained from different models are often not directly comparable. To overcome this problem, we propose a family of weighted hazard ratio measures that are based on the marginal survival curves or marginal hazard functions, and can be estimated using readily available output from various modeling approaches. The proposed transformation family includes the transformations considered by [18] as special cases. In addition, we propose a novel estimate of the weighted hazard ratio based on the maximum departure from the null hypothesis within the transformation family, and develop a Kolmogorov–Smirnov type of test statistic based on this estimate. Simulation studies show that when the hazard functions of two groups either converge or diverge, this new estimate yields a more powerful test than tests based on the individual transformations recommended in [18], with a similar magnitude of power loss when the hazards cross. The proposed estimates and test statistics are applied to a colorectal cancer clinical trial.

Keywords

Average hazard ratios; Crossing hazards; Non-proportional hazards; Survival analysis; Weighted estimation

1 Introduction

Since its publication in 1972, the Cox proportional hazards regression model [4] has been widely used in practice for analyzing time-to-event data. Its key assumption is the proportional hazards assumption, which is often violated in practice when, for example, the effect of the treatment or a prognostic factor changes over time. Several modifications to the Cox model have been proposed to relax this assumption, including the stratified Cox model [20], the Cox model with artificial time-dependent covariates [20], or the Cox model with time-depedent effects based on spline functions [2, 22]. In particular, [7] used smoothing splines and allowed time-varying coefficients and some interaction terms; [8] introduced the varying-coefficient model, which allows one to fit an additive model with time-varying coefficients of the covariates; [9] used cubic spline functions to model the time-by-covariate interactions; and [11] used linear splines and their tensor products to estimate the conditional log-hazard function. When the proportional hazards assumption is violated, other alternatives include separate modeling for the different time periods, weighted estimation for Cox regression, or non-Cox type models including the accelerated failure time model, the proportional odds model, the parametric log-logistic model, or the generalized odds-rate hazards model, among many others [3, 14, 16, 17, 21]. We refer the reader to [18] for a detailed discussion and comparison of those approaches.

The problem of the aforementioned approaches is that the regression coefficients based on the different models are not always directly comparable. For example, [1] used the Cox proportional hazards model, the accelerated failure time model, the generalized Gamma regression model, and the log-logistic regression for the same breast cancer dataset. Since the estimates obtained from these various models are not directly comparable, the authors had to rely completely on p-values for comparisons of the treatment effects, although they could obtain the marginal survival curves for all their models. The lack of direct comparisons in the estimated regression coefficients is in fact one of the barriers in using models other than the Cox model in the medical literature. On the other hand, it is important to note that one should never build artificial models in order to force arbitrary comparisons of regression coefficients between models. There lies the more important goal of building models that are scientifically meaningful for the application at hand, are interpretable, and fit the data well. In this sense, the aforementioned models can be quite useful for inference. Our intention here is to build such a broad class of models with these primary goals in mind.

Another motivation for our methods comes from recent work on complex data with treatment switching and competing risk events [23]. Due to the complexity of the data, the method developed in [23] includes multiple modeling components in the estimation procedure where each component allows for different coefficients of the treatment effect. Therefore, the overall treatment effect cannot be summarized by an individual coefficient in any of the sub-models and has to be evaluated using marginal survival functions. In addition to the marginal survival functions, it is of interest to compute a hazard ratio estimate to summarize the treatment effect so that one can carry out sample size calculations for future studies and to compare the hazard ratio estimate to other studies using the same treatment. Such a methodology has not been developed or formally studied in the statistical literature. In particular, it is not at all clear how to define the concept of a hazard ratio for the purposes

of assessing a treatment effect in these situations. To address these needs, we propose a family of weighted estimates to quantify the average hazard ratio based on the marginal survival curves or marginal hazard functions. The proposed estimates are useful in situations when the marginal survival function or marginal hazard function is available from models other than the Cox model.

Using a different motivation, [18] proposed three types of average hazard ratios to appropriately deal with non-proportional hazards. The paper by [18] also serves as a motivation for our work. In this present paper, we define a general and novel class of transformations on the hazard ratio to obtain a general measure of the hazard ratio. We propose a family of weighted estimates that include the transformations considered by [18] as special cases and develop a novel test statistic based on the maximum departure from the null hypothesis within the transformation family to test the hypothesis of no association with the survial outcome. Note that unlike [18], who extended their transformed average hazard ratio measures to weighted estimation for Cox regression in order to adjust for additional covariates, we assume here that the marginal survival functions or marginal hazard functions used as the input information for our estimates have already been properly adjusted for potential confounders, and hence, we do not extend our proposed estimates to adjust for additional covariates. An example of this setting, the panitumumab study, is discussed in detail in section 4. The methods in [18] were proposed as a solution to extend the Cox proportional hazards model when the proportional hazards assumption is violated, while the estimates proposed in this paper are intended to bridge together and unify different survival modeling techniques.

The rest of this paper is organized as follows. In Section 2, we first introduce the one-parameter transformation family of weighted estimates. In Section 2.1, we discuss its connections with the three types of average hazard ratios described in [18], and in Section 2.2, we develop a new test statistic to test the hypothesis of no association based on the maximum departure from the null hypothesis. We state the assumptions needed for the inference procedure and establish asymptotic properties of the estimates in Section 2.3. In Section 3, we evaluate the proposed estimates through extensive simulation studies. We illustrate the proposed approach with a colorectal cancer clinical trial in Section 4. We conclude the paper in Section 5 with some comments and remarks. Proofs of the theorems are given in the Appendix.

2 Quantifying Time-Varying Hazards Ratios

2.1 Estimation and Inference

Let $h_1(t)$ and $h_0(t)$ denote the hazard functions in the treatment arm and control arm, respectively. A general methodology for quantifying the average of time-varying hazard ratios is given by the following family of transformations,

$$\theta \equiv G^{-1}\left\{\int G\left(\frac{h_1(t)}{h_0(t)}\right)\Omega(t)\,dt\right\}, \quad (1)$$

where $G(\cdot)$ is a strictly increasing transformation and $\Omega(t)$ is a weight function so that $\int \Omega(t)dt = 1$. In particular, we allow $\Omega(t)$ to depend on the underlying survival functions $S_1(t)$ and $S_0(t)$, and therefore, we write $\Omega(t)$ as $\Omega(t; S_0(t), S_1(t))$.

An interesting one-parameter transformation family containing the three average hazard ratios considered in [18] as special cases is

$$G(x;a) = \left\{1 - (a+x)^{-a}\right\}/a, \quad (2)$$

where a is an unknown parameter. Since $\Omega(t)$ is a weight function, the transformation family can be written as

$$\theta_a = \left[\int \left\{a + \frac{h_1(t)}{h_0(t)}\right\}^{-a} \Omega\left(t; S_0(t), S_1(t)\right) dt\right]^{-1/a} - a. \quad (3)$$

In particular, when $a = -1$, $G(x; -1) = x - 2$ yields the identity transformation with

$$\theta_{iden} = \int \frac{h_1(t)}{h_0(t)} \Omega\left(t; S_0(t), S_1(t)\right) dt; \quad (4)$$

when $a = 0$, $G(x; 0) = \log(x)$ yields the logarithmic transformation with

$$\theta_{log} = exp\left[\int log\left\{\frac{h_1(t)}{h_0(t)}\right\} \Omega\left(t; S_0(t), S_1(t)\right) dt\right]; \quad (5)$$

when $a = 1$, $G(x; 1) = x/(1 + x)$ yields the ratio transformation with

$$\theta_{ratio} = \frac{\int \left\{h_1(t)/h(t)\right\} \Omega\left\{t; S_0(t), S_1(t)\right\} dt}{\int \left\{h_0(t)/h(t)\right\} \Omega\left\{t; S_0(t), S_1(t)\right\} dt}, \quad (6)$$

where $h(t) = h_0(t) + h_1(t)$.

Note that under the proportional hazards assumption, i.e., when $h_1(t) = h_0(t)e^{\beta}$, $\theta_a = \theta_{iden} = \theta_{log} = \theta_{\text{ratio}} = e^{\beta}$. We also note that the identity transformation with $a = -1$ yields what [18] called the 'simple average hazard ratio', the logarithmic transformation with $a = 0$ yields the 'geometric average hazard ratio', and the ratio transformation with $a = 1$ yields the 'average hazard ratio', respectively. The 'average hazard ratio' estimate was originally defined in [10]. We note that the logarithmic and ratio transformations are the only two transformations among the one-parameter transformation family in (3) that are symmetric in $h_0(t)$ and $h_1(t)$. In other words, $\theta_a(h_0/h_1) = \{\theta_a(h_1/h_0)\}^{-1}$ when $a = 0$ or $a = 1$.

When the marginal survival functions can be estimated using survival models other than the Cox model, for example, as in the context of [23], we can estimate $h_1(t)$ and $h_0(t)$ using the kernel estimates

$$\hat{h}_k(t) = -\int K_{a_n}(t - s)\, d\, log\, \hat{S}_k(s), k=0, 1,$$

where $\hat{S}_k(t)$ is the estimated survival function in treatment arm k and $K_{a_n}(x) = a_n^{-1}K(x/a_n)$ for some non-negative and symmetric kernel function $K(x)$ with a_n being a bandwidth. We also estimate $\Omega(t; S_0(t), S_1(t))$ using $\Omega(t; \hat{S}_0(t), \hat{S}_1(t))$. Thus, we obtain the estimator

$$\hat{\theta}_G = G^{-1}\left[\int G\left\{\frac{\hat{h}_1(t)}{\hat{h}_0(t)}\right\}\Omega\left\{t, \hat{S}_0(t), \hat{S}_1(t)\right\}dt\right]. \quad (7)$$

We use the notation $\hat{\theta}_G$ to emphasize that this proposed estimator is valid for the general transformation $G(\cdot)$, but is not restricted to the one-parameter transformation in (2).

In the case when $-\log \hat{S}_k(t)$ is obtained nonparametrically using the Breslow estimator, we obtain

$$\hat{h}_k(t) = \int \frac{\sum_{i=1}^{n_k} K_{a_n}(s-t)dN_{ik}(t)}{\sum_{i=1}^{n_k} Y_{ik}(s)}, \quad k=0,1,$$

where $N_{ik}(t)$ and $Y_{ik}(t)$ denote the observed counting process and at-risk process for subject i in treatment arm k, respectively. However, in our motivating treatment switching example [23], whose details are given in Section 4.2, the Breslow estimator may not be appropriate due to the systematic differences in the drop-in patterns and disease susceptibility in the two treatment arms; instead, $-\log \hat{S}_k(t)$ is obtained from the prediction using the model proposed in [23]. Estimation of the asymptotic variance of $\hat{\theta}_G$ is discussed in Section 3.2.

The choice of weight function, $\Omega(t)$, reflects the inferential goals in the study. In this paper, we focus on a weight function proportional to $\{S_0(t)S_1(t)\}^{1/2}$. Recall that the comparison of two survival curves can be done using the log-rank test and its various extensions with different weight functions to emphasize the survival differences at early or late follow-up periods. In particular, the weight functions of the log-rank, Wilcoxon, Peto-Prentice, Fleming-Harrington, and Tarone-Ware tests are proportional to 1, $R(t)$, $S(t)$, $S(t)^p(1 - S(t))^q$, and $\sqrt{R(t)}$, respectively, where $R(t)$ is the number of patients at risk at time t, $S(t)$ is the overall survival function at time t, and p and q are predetermined numbers in (0, 1). The choice of weight function is a broad research area and has been studied in many papers including [13,15,19] among others. Therefore, we only focus on the transformation family itself in this paper. However, we do emphasize that similar to the log-rank test and its various extensions, the inference may vary depending on the chosen weight, and sometimes inferences are sensitive to the choice of weights. To avoid this possible dilemma, the choice of weight function should be pre-specified based on the research goals at hand (cf., e.g. [14])

2.2 Testing the Hypothesis of No Association

In this section, we develop a hypothesis testing procedure for testing a value of a in [0,1]. We propose a novel estimate of θ_a based on the maximum departure from the null hypothesis. Since the proposed transformation class is monotonic for a in [0,1], we can also develop asymptotic theory for the estimate of θ_a, and hence obtain the asymptotic distribution of the test statistic. Recall that $a = 0$ corresponds to the logarithmic

transformation, and $a = 1$ corresponds to the ratio transformation. Although, $a = -1$ is an interesting and interpretable transformation, the theory given below does not cover this case.

We define the estimate based on the maximum departure from the null as $\hat{\theta}_{sup}=\hat{\theta}_{\tilde{a}}$, where

$$\tilde{a}=aregsup_{a\in[0,1]}\left\{a:|\hat{\theta}_a-1|\right\}, \quad (8)$$

and

$$\hat{\theta}_a=\left[\int\left\{a+\frac{\hat{h}_1(t)}{\hat{h}_0(t)}\right\}^{-a}\Omega\left(t;\hat{S}_0(t),\hat{S}_1(t)\right)dt\right]^{-1/a}-a. \quad (9)$$

Another motivation for restricting the parameter a to be in [0, 1] is that estimates with $a < 0$ tend to be numerically unstable and estimates with $a > 1$ impose large weights on local regions. The test statistic based on the estimate with maximum departure from the null is a Kolmogorov–Smirnov type test statistic, which is given by

$$T_{sup}=sup_{a\in[0,1]}|\hat{\theta}_a-1|. \quad (10)$$

Theoretical properties of the maximum departure-based estimate and its corresponding test statistic are established in Section 2.3. It is difficult to solve for the a that maximizes $|\hat{\theta}_a - 1|$ explicitly. To calculate $\hat{\theta}_{sup}$ and T_{sup}, we recommend a grid search on $a \in [0, 1]$. Many resampling approaches, like the bootstrap, can be used to construct confidence intervals for $\hat{\theta}_{sup}$ and also to compute p-values.

2.3 Asymptotic Results

Under the conditions given in the Appendix, we establish consistency of $\hat{\theta}_G$ and derive its asymptotic distribution. The following theorem states our main result regarding the asymptotic behavior of $\hat{\theta}_G$.

Theorem 1 *Under Conditions (C.1)–(C.5), $\hat{\theta}_G$ is consistent and $\sqrt{n}\left(\hat{\theta}_G-\theta_G\right)$* converges in distribution to a mean-zero normal distribution N(0, σ^2), *where σ^2 is defined in (11).*

We let

$$\begin{aligned} Q_0(t)&= G\{h_{10}(t)/h_{00}(t)\}, Q_1(t)=\Omega\{t;S_{00}(t),S_{10}(t)\},\\ Q_2(t)&= \partial\Omega\{t;S_{00}(t),S_{10}(t)\}/\partial S_0(t), Q_3(t)=\partial\Omega\{t;S_{00}(t),S_{10}(t)\}/\partial S_1(t), \end{aligned}$$

and $H = (G^{-1})'\{\int Q_0(t)Q_1(t)dt\}$. As shown in the Appendix,

$$\sigma^2=var\left[\sqrt{n}\int A_1(t)\,d\left\{log\,\hat{S}_1(t)-log\,S_{10}(t)\right\}+\sqrt{n}\int A_0(t)\,d\left\{log\,\hat{S}_0(t)-log\,S_{00}(t)\right\}+\quad\sqrt{n}\int B_1(t)\left\{\hat{S}_1(t)-S_{10}(t)\,dt+\sqrt{n}\right.\right.$$

where

$$A_0(t) = HG'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\} Q_1(t)\, h_{10}(t)/h_{00}(t)^2, \quad A_1(t) = -HG'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\} Q_1(t)/h_{00}(t),$$
$$B_0(t) = HQ_0(t)\, Q_2(t), \quad \text{and} \quad B_1(t) = HQ_0(t)\, Q_3(t).$$

We can use expression (11) to estimate σ^2 and to conduct inference on $\hat{\theta}_G$. First, we replace h_{k0} and S_{k0} by their corresponding estimates to obtain $\hat{A}_k(t)$ and $\hat{B_k}(t)$. Then, the asymptotic variance of $\sqrt{n}\left(\hat{\theta}_G - \theta_G\right)$ can be estimated by the sample variance of

$$\begin{aligned}
&\sqrt{n}\int \hat{A}_1(t)\, d\left\{log\, \hat{S}_1^*(t) - log\, \hat{S}_1(t)\right\} \\
&\quad + \sqrt{n}\int \hat{A}_0(t)\, d\left\{log\, \hat{S}_0^*(t) - log\, \hat{S}_0(t)\right\} \\
&\quad + \sqrt{n}\int \hat{B}_1(t)\, \hat{S}_1^*(t) \\
&\qquad - \hat{S}_1(t)\Big\}\, dt \\
&\quad + \sqrt{n}\int \hat{B}_0(t)\left\{\hat{S}_0^*(t) - \hat{S}_0(t)\right\} dt,
\end{aligned} \tag{12}$$

where $\left(\hat{S}_1^*, \hat{S}_0^*\right)$ are resampled statistics for ($\hat{S}_1$, $\hat{S}_0$), for example, by using a bootstrap sample or other resampling techniques. The derivation of equation (12) is given in the Appendix. Since $\hat{A}_k$ and $\hat{B_k}$ are uniformly consistent for A_k and B_k, the consistency of this variance estimator follows if $\left(\hat{S}_1^*, \hat{S}_0^*\right)$ actually converge to the true distribution. In other words, we assume condition (C.6).

An alternative procedure for obtaining the asymptotic variance of $\hat{\theta}_G$ is to adopt a bootstrap method and estimate θ_G directly in each bootstrap sample. The validity of using the bootstrap for variance estimation follows by the results of [12] Chapter 10. The full development of the asymptotic properties of the bootstrapped variance estimate will require a new paper and detract from the focus here. Therefore, we will investigate it in future work. Our experience shows that the bootstrap appears to be more accurate with small sample sizes and, therefore, will be used in the subsequent numerical studies.

For the class of transformation $G_a(x)$ as defined in (2) with $a \in [-1, 1]$. Theorem 2 shows that the transformation that yields maximum local power is $a = 1$.

Theorem 2 Under Conditions (C.1)–(C.5), for the transformation family in (2) with a ∈ [–1, 1], $\hat{\theta}_a$ *achieves its maximum local power at* $a = 1$, the ratio transformation, when the weight function $\Omega(t)$ *is independent of* $S_0(t)$ *and* $S_1(t)$.

Theorem 2 supports the recommendation made by [18] in using the ratio transformation, that is, the "average hazard ratio" estimator.

Theorem 3 Under Conditions (C.1)–(C.5) and the null hypothesis of equal hazards, $\sqrt{n}T_{sup}$ *converges in distribution to* $sup_{a\in[0,1]}|\mathscr{G}_a|$, *where* $\mathscr{G}_a$ *is a Gaussian process with mean 0.*

The proofs of the Lemma and the Theorems are given in the Appendix.

3 Simulation Studies

3.1 Simulation Study I

We examine the numerical performance of the estimates including the estimate based on the maximum departure from the null and the estimates with the identity, logarithmic, and ratio transformations. These estimates were compared to the estimates of the Cox proportional hazards model. To do this, we carried out a simulation study assuming $h_0(t) = 1$ and (a) $h_1(t) = h_0(t)$ (identical hazards), (b) $h_1(t)=1{\cdot}2h_0(t)$ (proportional hazards), (c) $h_1(t)=0{\cdot}25\exp(2t)$ (crossing hazards), (d) $h_1(t) = 0{\cdot}5+0{\cdot}9/(1+0{\cdot}5t)\}h_0(t)$ (converging hazards), and (e) $h_1(t)=(1+0{\cdot}45t)h_0(t)$ (diverging hazards).· The true underlying hazard and survival functions of the aforementioned scenarios are plotted in Figure 1. The simulated survival data are then censored by τ=1·5, resulting in censoring rates of 24%, 21%, 18·6%, 21·3%, and 19·6%, respectively, for scenarios (a)-(e) in Figure 1. In the simulation and the application, we set the bandwidth to be a_n=$0.9min\,(\hat{\sigma}, IQR/1.34)\,n^{-1/5}$, where $\hat{\sigma}$ and IQR are the standard deviation and the inter-quartile range of the variable in the kernel estimation, respectively. More discussion on choosing the bandwidth can be found in [5] Chapter 5.

In simulation study I, we first consider the weight function $\Omega(t) \propto \{S_0(t)S_1(t)\}^{1/2}$, and sample sizes of n = 800 and n = 400 with 1000 replicates. Bootstrap samples of 1000 are used to construct 95% confidence intervals for the weighted hazard ratio estimates and the maximum departure-based estimate. In particular, the 95% confidence intervals of the weighted hazard ratio estimates were based on the normal approximation according to Theorem 1. For the maximum departure-based estimate, the normal approximation is not valid and, therefore, we constructed the 95% confidence interval using the 2.5% and 97.5% quantiles of the bootstrap samples. For all the estimates, we show the true value calculated from the corresponding formulae, the average bias of the parameter estimates, the sample standard deviation of the estimates, the average of the standard errors, the coverage probability of the 95% confidence interval, the type I error under identical hazards, and the power to reject the null hypothesis of identical hazards under various alternatives. Since the calculation of the p-value for $\hat{\theta}_{sup}$ involves resampling, to avoid intensive numerical computation, the grid search for a was conducted in the interval of [0, 1] with a step size of 0·1 as in Table 1. We also conducted the simulation with a finer grid search and the differences in the results are negligible. The results in Table 1 with n = 800 and Table 2 with n = 400 show that all the methods provide estimates with small biases and 95% confidence intervals with nominal coverage rates. For the individual weighted estimates, when the alternative is crossing hazards, the ratio and logarithmic transformations yield substantial power gains compared to the identity transformation and the Cox model. For example, 86% and 82% verses 6% and 5%, respectively, for n = 800. When the alternatives are proportional but not unit, or converging/diverging hazards, the ratio and logarithmic transformations yield little loss of power compared to the identity transformation and the Cox model. For example, when n = 800, the biggest loss was observed for converging hazards where the powers of the ratio and logarithmic transformations are 58% and 57%, respectively, compared to 63% and 66% powers for the identity transformation and the Cox model. When comparing the maximum departure-based estimate to the weighted hazard ratio estimates based on individual transformations, the test based on the maximum

departure-based estimate is more powerful than the tests based on individual transformations when the hazard functions of the two groups either cross or diverge, with a similar magnitude of power loss when the hazards cross. The advantages of the maximum departure-based approach are in balance with its potential disadvantages and its properties depend on the shape of the underlying changes in the hazard ratio functions over time.

3.2 Simulation Study II

In the second simulation study, we consider $n = 400$ and take the weight function to be proportional to 1 in order to evaluate the influence of the weight function in comparing the maximum departure-based estimate to the estimates based on individual transformations. The rest of the simulation setup is the same as simulation study I. In comparing the results in Table 3 and Table 2, we see that the choice of weight functions has a great influence on the estimates as studied in [13, 15, 19]. For example, in Table 2, we were surprised to observe superior power of $\hat{\theta}_{sup}$ over the standard Cox model under proportional hazards. This is no longer observed in Table 3 when the weight function is proportional to 1. Furthermore, for the weight function proportional to 1, comparisons with the Cox model, identity transformation, logarithmic transformation, and ratio transformation are not as clear as the weight $\propto \{S_0(t)S_1(t)\}^{1/2}$. However, when comparing the maximum departure-based estimate to the estimates based on the logarithmic and ratio transformations, the maximum departure-based estimate outperforms the other two estimates in terms of power in all settings.

4 Panitumumab Study

As mentioned in Section 1, [23] presented a novel class of semi-parametric semi-competing risks transition survival models to accommodate partial treatment crossover at the progression time for assessing the treatment effect on overall survival in a phase III colorectal cancer clinical trial. In this phase III multi-center clinical trial, overall survival is the primary endpoint and the patients were randomized to receive panitumumab plus best supportive care or best supportive care alone. During the trial, the patients receiving best supportive care alone were allowed to switch to the experimental treatment if they experience disease progression. In order to properly adjust for treatment switching, [23] proposed a statistical model with four submodels: a logistic regression model for the distribution of the progression status, a proportional hazards model for time to death for the no-progression population, a proportional hazards model for time to progression in the progression population, and another proportional hazards model for time to death in the progression population.

To describe the model in detail, we need the following notation: U is a binary variable denoting the lifetime disease progression status of the subjects, with $U = 1$ indicating the subject has disease progression before death and 0 otherwise; T_D denotes the time to death for the no-progression subjects with $U = 0$; for the other subjects with $U = 1$, T_U denotes their time to disease progression and G denotes the time from disease progression to death. Let $h_D(t|\cdot)$, $h_U(t|\cdot)$, and $h_G(t|\cdot)$ be the conditional hazard function of T_D, T_U, and T_G, correspondingly, given the covariates. The following four sub-models are used to model the complex data:

$$
\begin{aligned}
&logit\left\{pr\left(U{=}1|R,X\right)\right\}{=}\alpha_0{+}\alpha_1\,R{+}\alpha_2 X,\\
&h_D\left(t|R,X,U{=}0\right){=}h_0\left(t\right)exp\left(\beta_0\,R{+}\gamma_0 X\right),\\
&h_U\left(t|R,X,U{=}1\right){=}h_1\left(t\right)exp\left(\beta_1\,R{+}\gamma_1 X\right),\\
&h_G\left(t|R,Z,V,U{=}1,T_U\right){=}h_2\left(t\right)exp\left\{\beta_{21}\,R{+}\beta_{22}\,V\left(1-R\right){+}\gamma_2^T\left(Z^T,T_U\right)^T\right\},
\end{aligned}
\tag{13}
$$

where $R = 1$ is the treatment indicator, X denotes the baseline covariates, V is the treatment switching indicator, Z contains the covariates collected at baseline and at disease progression, $h_0(t)$, $h_1(t)$, and $h_2(t)$ are the unknown baseline hazard functions, and α's, βs, and γ's are unknown regression coefficients. The marginal survival function $S_a\left(t\right){=}P\left(T_D^*\left(a\right){>}t\right)$, where $T_D^*\left(a\right)$ is the potential survival time when a subject receives treatment a and never switches treatment, can be written as

$$
\begin{aligned}
&\int_x P\left(T_D{>}t|X,U{=}0,R{=}a\right)P\left(U{=}0|X,R{=}a\right)f_X\left(x|R{=}a\right)dx\\
&\quad{+}\int_{x,z}\left\{\int_0^t P\left(G{>}t-s|T_u{=}s,V{=}0,Z,U{=}1,R{=}a\right)dP\left(T_U\leq s|Z,U{=}1,R{=}a\right){+}P\left(T_U{>}t|Z,U{=}1,R{=}1\right)\right\}f_Z\left(z|X,U{=}1,\right.
\end{aligned}
\tag{14}
$$

Therefore, $S_a(t)$ can be expressed in terms of the parameters in sub-models (13) and the distributions of X and Z given (X, $U = 1$, R). Hence, by plugging the estimates of these parameters into the above expression, we can estimate $S_a(t)$, and thus obtain the causal effect of treatment. We refer the reader to [23] for the detailed estimation algorithm and inference procedure for this complex model.

In this paper, we are interested in quantifying the hazard ratio for the subset of patients with the wild type K-Ras gene. In the panitumumab study, there are 238 patients with the wild type K-Ras gene, among which 35 patients died without disease progression, 177 had disease progression and died later, 19 had disease progression and censored without death, and 7 were censored without disease progression or death. In the submodels for the time of disease progression and time of death, we controlled for all the baseline covariates including age in years at screening, baseline performance status (0 or 1 versus ≥ 2), primary tumor diagnosis type (rectal versus colon), gender, and region at two levels (western Europe and the rest of the world). The covariates in the submodel for time from disease progression to death include progonostic factors for the switching decision in addition to the baseline covariates. The prognostic factors include progression time, best tumor response (partial response or stable disease versus progressive disease) according to tumor response assessments based on a central radiologist's review, last performance status, and grade 2 or above adverse events.

In order to estimate the hazard ratio of best supportive care alone versus panitumumab plus best supportive care, we applied the pre-specified weight function $\Omega(t) \propto \{S_0(t)S_1(t)\}^{1/2}$ to the maximum departure-based estimate as well as the estimates with the logarithmic and ratio transformations. The marginal survival functions $S_a(t)$ were used in the calculation. The estimates, standard errors, 95% confidence intervals, and p-values based on 1000 bootstrap samples were reported in Table 4. Note that the bootstrap sampling procedure was conducted on the original dataset and the marginal survival curves were constructed for each bootstrap sample. Therefore, the reported 95% confidence intervals and p-values have

incorporated the variation associated with the estimation of the marginal survival functions. Furthermore, due to the complexity of the data, survival models such as the Cox proportional hazards model or weighted Cox proportional hazards model are no longer valid, and hence, were not included in Table 4. As shown in the table, when comparing the best supportive care to panitumub plus best supportive care, the hazard ratio estimates based on different transformations are very stable, ranging from 2.90 to 2.92, with highly significant p-values. Judging from the p-values, the ratio-transformed estimate and the maximum departure-based estimate are a little more powerful than the logarithm-transformed estimate, although the difference may be of limited practical relevance in this study (as both are highly significant).

5 Concluding Remarks

We have proposed a class of transformations to quantify the average hazard ratio and established asymptotic properties of their estimates. Our numerical studies results suggested that when the weight function is proportional to $\{S_0(t)S_1(t)\}^{1/2}$, the ratio transformation tended to provide a large power in distinguishing between two treatment arms compared to the identity and logarithmic transformations. We also demonstrated in Theorem 2 that ratio transformation achieves the maximum local power within the transformation family with $a \in [-1, 1]$ when the weight function is independent of $S_0(t)$ and $S_1(t)$. Unfortunately, extending Theorem 2 to a general situation is challenging and will require much more development to fully investigate it. We will investigate this issue in the future. Note that ratio transformation estimator was also the recommendation of [18]. One direct implication is that when the hazard ratio is not expected to be constant, a powerful clinical trial design may be based on the ratio transformation of the hazard ratios. Furthermore, we have also developed a novel test statistic to test the hypothesis of no association based on the maximum departure from the null hypothesis within the transformation family. Simulation studies show that when the hazard functions of the two groups either converge or diverge, the proposed test statistic is more powerful than the test statistic based on the individual transformations recommended in [18], with a similar magnitude of power loss when the hazards cross, demonstrating the importance of correctly identifying the shape of the time-varying hazard ratio function for estimation.

We have only presented a class of transformations for $G(x)$, but clearly there are infinitely many choices of $G(x)$. In practice, choosing $G(x)$ should not only depend on the resulting statistical power under the alternative, but it also needs to yield a good interpretation of the parameters. For reasons of interpretability, we recommend using the ratio-transformed estimator when estimation is of primary interest, and to use the Kolmogorov-Smirnov type test statistic T_{sup} when hypothesis testing is of primary interest.

Although we considered a class of transformations to quantify the average hazard ratios, the same concept can be generalized to quantify many other time-varying comparison measures. These include time-varying intensities for recurrent events, time varying treatment effects over time, and time-sensitive diagnostic measures. Future work will investigate which transformations are appropriate for such comparison measures.

Appendix

We state here the conditions needed to establish consistency of $\hat{\theta}_G$ and to derive its asymptotic distribution.

(C.1) G is thrice-continuously differentiable and is strictly increasing. Additionally, $\Omega(t; s, v)$ is twice-continuously differentiable in (t, s, v).

(C.2) $\sqrt{n}\left(\hat{S}_1 - S_{10}, \hat{S}_0 - S_{00}\right)$ converges in distribution to a bivariate mean-zero Gaussian process, denoted by $(\mathscr{G}_1, \mathscr{G}_0)$ in $BV[0, \tau] \times BV[0, \tau]$, where $BV[0, \tau]$ denotes the spaces consisting of functions that have finite total variation in $[0, \tau]$ and S_{k0} is the true survival function in treatment arm k. Here, τ is the study duration.

(C.3) S_{k0} is strictly decreasing and thrice-continuously differentiable in $[0, \tau]$. Moreover, $S_{k0}(\tau) > 0$.

(C.4) The kernel function $K(x)$ is differentiable, symmetric with respect to 0, and has compact support on [–1, 1].

(C.5) The bandwidth a_n satisfies $na_n^2 \to \infty$ and $na_n^4 \to 0$.

(C.6) Conditional on the data, $\sqrt{n}\left(\hat{S}_1^* - \hat{S}_1, \hat{S}_0^* - \hat{S}_0\right)$ converges in distribution to $(\mathscr{G}_1, \mathscr{G}_0)$ where $\left(\hat{S}_1^*, \hat{S}_0^*\right)$ are resampled statistics for $(\hat{S}_1, \hat{S}_0)$.

Condition (C.6) is an assumption regarding the consistency and asymptotic distribution of the estimates generated by the boostrap procedure. Chapter 20 of [12] validates this assumption for several survival modeling techniques including the Cox proportional hazards model.

Lemma 1 *Under Conditions (C.1)–(C.5),* $\sup_{t\in[0,\tau]} |\hat{h}_k(t) - h_{k0}(t)| \to_p 0$, $k = 0, 1$, *where* h_{k0} denotes the true hazard function in treatment arm k.

Proof (of Lemma 1) First, we note that $\hat{h}_k(t) = -a_n^{-1}\int K(x)\, d\log \hat{S}_k(t+a_n x)$. By carrying out an integration by parts, we can rewrite $\hat{h}_k(t)$ as $\hat{h}_k(t) = \int a_n^{-1} \log \hat{S}_k(t+a_n x)\, K'(x)\, dx$. Moreover, we can continuously extend defining $\hat{S}_k$ and S_{k0} to $[-a_n, \tau + a_n]$ so that (C.2) still holds. Thus, (C.2) implies $\sup_{t\in[-a_n, \tau+a_n]} |\hat{S}_k(t) - S_{k0}(t)| = O_p(n^{-1/2})$. (C.3) further gives $\sup_{t\in[-a_n, \tau+a_n]} |\log \hat{S}_k(t) - S_{k0}(t)| = O_p(n^{-1/2})$. Therefore,

$$\sup_{t\in[0,\tau]} |\hat{h}_k(t) - h_{k0}(t)| \le \int a_n^{-1} \sup_{t\in[-a_n,\tau+a_n]} |\log \hat{S}_k(t) - \log S_{k0}(t)|\, |K'(x)|\, dx \le O_p\left(1/\sqrt{n}a_n^2\right)$$

goes to 0 by condition (C.5). This proves the lemma.

Proof (of Theorem 1) Using Lemma 1 and noting $\inf_{t\in[0,\tau]} h_0(t) > 0$, we obtain that uniformly in $t \in [0, \tau]$, as $n \to \infty$, $G\left\{\frac{\hat{h}_1(t)}{\hat{h}_0(t)}\right\} \to_p G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}$ and $\Omega\{t; \hat{S}_0(t), \hat{S}_1(t)\} \to_p \Omega\{t; S_{00}(t), S_{10}(t)\}$. Thus, it is clear that $\hat{\theta}_G \to_p \theta_G$ as $n \to \infty$. This establishes consistency.

To derive the asymptotic distribution of $\hat{\theta}_G$, by the mean-value theorem, we obtain

$$\begin{aligned} & n^{1/2}\left(\hat{\theta}_G-\theta_G\right) \\ & =n^{1/2}\left(\left(G^{-1}\right)'\left[\int G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}dt\right]+o_p(1)\right) \\ & \times\left[\int G\left\{\frac{\hat{h}_1(t)}{\hat{h}_0(t)}\right\}\Omega\left\{t;\hat{S}_0(t),\hat{S}_1(t)\right\}dt-\int G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}dt\right]. \end{aligned}$$

Using the mean-value theorem again, we have

$$\begin{aligned} & G\left\{\frac{\hat{h}_1(t)}{\hat{h}_0(t)}\right\}\Omega\left\{t;\hat{S}_0(t),\hat{S}_1(t)\right\} \\ & -G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\} \\ & =\left[G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}+o_p(1)\right]\left[\frac{\hat{h}_1(t)-h_{10}(t)}{h_{00}(t)}-\frac{h_{10}(t)\left\{\hat{h}_0(t)-h_{00}(t)\right\}}{h_{00}(t)^2}\right] \\ & +\left[G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}\frac{\partial}{\partial S_0(t)}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}+o_p(1)\right]\left\{\hat{S}_0(t)-S_{00}(t)\right\} \\ & +\left[G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}\frac{\partial}{\partial S_1}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}+o_p(1)\right]\left\{\hat{S}_1(t)-S_{10}(t)\right\}, \end{aligned}$$

where $o_p(1)$ is a random element converging in probability to zero uniformly in $t \in [0, \tau]$.

For convenience, we denote $H=\left(G^{-1}\right)'\left\{\int Q_0(t)\,Q_1(t)\,dt\right\}$, $Q_0(t)=G\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}$, $Q_1(t)=\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}$, $Q_2(t)=\frac{\partial}{\partial S_0(t)}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}$, and $Q_3(t)=\frac{\partial}{\partial S_1(t)}\Omega\left\{t;S_{00}(t),S_{10}(t)\right\}$. Then, after combining the above results, we obtain

$$\begin{aligned} & n^{1/2}\left(\hat{\theta}_G-\theta_G\right) \\ & =n^{1/2}\left\{H+o_p(1)\right\}\left[\int G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}Q_1(t)\left\{\hat{h}_2(t)-h_{10}(t)\right\}/h_{00}(t)\,dt\right] \\ & -n^{1/2}\left\{H+o_p(1)\right\}\left[\int G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}Q_1(t)\,h_{10}(t)\left\{\hat{h}_0(t)-h_{00}(t)\right\}/h_{00}(t)^2dt\right] \\ & +n^{1/2}\left\{H+o_p(1)\right\}\left[\int Q_0(t)\,Q_3(t)\left\{\hat{S}_1(t)-S_{10}(t)\right\}dt\right] \\ & +n^{1/2}\left\{H+o_p(1)\right\}\left[\int Q_0(t)\,Q_2(t)\left\{\hat{S}_0(t)-S_{00}(t)\right\}dt\right]. \end{aligned}$$

The last two terms on the right-hand side both take the form $n^{1/2}\int[A(t)+o_p(1)](\hat{S}_k(t)-S_{k0}(t))dt$ for some bounded function $A(t)$ so that they converge to a normal distribution by condition (C.2). Thus, we only focus on the first two terms, namely (I) and (II), on the right-hand side. Using the definition of $\hat{h}_1(t)$, we can rewrite the first term (I) as

$$n^{1/2}\{H+o_p(1)\}\int G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}Q_1(t)\left\{-\int K_{a_n}(s-t)\,d\log\hat{S}_1(s)+\int K_{a_n}(s-t)\,d\log S_{10}(s)-\int K_{a_n}(s-t)\,d\log S_{10}(s)-h\right.$$
$$\left./h_{00}(t)\,dt.\right.$$

Since

$$-\int K_{a_n}(s-t)\,d\log S_{10}(s)-h_{10}(t)=\int K_{a_n}(s-t)\,h_{10}(s)\,ds-h_{10}(t)=\int K(x)\{h_{10}(t+a_nx)-h_{10}(t)\}\,dx=O(a_n^2)$$

and $na_n^4\to 0$, (I) becomes

$$n^{1/2}\{H+o_p(1)\}\int G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}Q_1(t)\times\left[-\int K_{a_n}(s-t)\,d\left\{\log\hat{S}_1(s)-\log S_{10}(s)\right\}\right]/h_{00}(t)\,dt+o_p(1),$$

which is also equal to

$$n^{1/2}\{H+o_p(1)\}\int_s d\left\{\log\hat{S}_1(s)-\log S_{10}(s)\right\}\times\int_t G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}Q_1(t)\,K_{a_n}(s-t)/h_{00}(t)\,dt+o_p(1).$$

However, since

$$\int_t G'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\}Q_1(t)\,K_{a_n}(s-t)/h_{00}(t)\,dt=G'\left\{\frac{h_{10}(s)}{h_{00}(s)}\right\}Q_1(s)/h_{00}(s)+O\left(a_n^2\right),$$

it follows that

$$(I)=-n^{1/2}H\int_s G'\left\{\frac{h_{10}(s)}{h_{00}(s)}\right\}Q_1(s)/h_{00}(s)\,d\left\{\log\hat{S}_1(s)-\log S_{10}(s)\right\}+o_p(1).$$

Similarly, we obtain

$$(II)=n^{1/2}H\int_s G'\left\{\frac{h_{10}(s)}{h_{00}(s)}\right\}Q_1(s)\,h_{10}(s)/h_{00}(s)^2d\left\{\log\hat{S}_0(s)-\log S_{00}(s)\right\}+o_p(1).$$

Thus, we have

$$\begin{aligned}&n^{1/2}\left(\hat{\theta}_G-\theta_G\right)\\&\quad=n^{1/2}\int A_1(t)\,d\left\{\log\hat{S}_1(t)-\log S_{10}(t)\right\}\\&\quad+n^{1/2}\int A_0(t)\,d\left\{\log\hat{S}_0(t)-\log S_{00}(t)\right\}\\&\quad+n^{1/2}\int B_1(t)\left\{\hat{S}_1(t)-S_{10}(t)\right\}dt\\&\quad+n^{1/2}\int B_0(t)\left\{\hat{S}_0(t)-S_{00}(t)\right\}dt+o_p(1),\end{aligned}\qquad(15)$$

where $A_0(t) = HG'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\} Q_1(t)\, h_{10}(t)/h_{00}(t)^2$, $A_1(t) = -HG'\left\{\frac{h_{10}(t)}{h_{00}(t)}\right\} Q_1(t)/h_{00}(t)$, $B_0(t) = HQ_0(t)Q_2(t)$, and $B_1(t) = HQ_0(t)Q_3(t)$. Hence, Theorem 1 follows from condition (C. 2).

Proof (of Theorem 2) Let the weight function $\Omega(t)$ be independent of $S_0(t)$ and $S_1(t)$ and satisfy $\int \Omega(t)dt = 1$. We write the hazard function of the treatment arm as $h_1(t)/h_0(t) = 1+\varepsilon\lambda(t)$, where $\lambda(t)$ is a function of t. When $h_1(t)$ is in the neighborhood of the null hypothesis with $h_1(t) = h_0(t)$, i.e., ε is close to zero, the Taylor's series expansion of θ_a is given by

$$\begin{aligned}\theta_a = & G_a^{-1}\left\{\int_0^\tau G_a(1)\,\Omega(t)\,dt\right\} + \left(G_a^{-1}\right)'\left\{\int_0^\tau G_a(1)\,\Omega(t)\,dt\right\}\int_0^\tau G_a'(1)\,\Omega(t)\,\lambda(t)\,dt\times\epsilon + o(\epsilon)\\ = & 1+\left(G_a^{-1}\right)'\left\{G_a(1)\right\}\times G_a'(1)\int_0^\tau \Omega(t)\,\lambda(t)\,dt\times\epsilon + o(\epsilon).\end{aligned}$$

Moreover, according to (15), the variance of $\hat{\theta}_a$ around the null hypothesis can be written as $\left(G_a'(t)\right)^2 B$, where B is a positive value independent of a. Therefore, the local power of θ_a can be written as

$$P\left(\left|\frac{\hat{\theta}_a - 1}{sd\left(\hat{\theta}_a\right)}\right| > Z_{1-\alpha/2} \,|\, h_0(t) = h_1(t)\right) = P\left(\left|\left(G_a^{-1}\right)'\left\{G_a(1)\right\}\epsilon\int_0^\tau \Omega(t)\,\lambda(t)\,dt + o(\epsilon)\right| > BZ_{1-\alpha/2}\right).$$

For the transformation family in (2) with $a \in [-1, 1]$, we can optimize the local power by maximizing $\left|\left(G_a^{-1}\right)'\left\{G_a(1)\right\}\right| = \left|(1+a)^{1+a}\right|$ for $a \in [-1, 1]$, resulting in an optimal value at $a = 1$.

When $\int \Omega(t)\lambda(t)dt = 0$ for the crossing hazards case, we need a higher order Taylor's series expansion, given by

$$\begin{aligned}\theta_a = & G_a^{-1}\left\{\int_0^\tau G_a(1)\,\Omega(t)\,dt\right\}\\ & + \left(G_a^{-1}\right)'\left\{\int_0^\tau G_a(1)\,\Omega(t)\,dt\right\}\\ & \times \int_0^\tau G_a'\Omega(t)\,\lambda(t)\,dt\\ & \times \epsilon + \left(G_a^{-1}\right)''\left\{\int_0^\tau G_a(1)\,\Omega(t)\,dt\right\}\\ & \times \left\{\int_0^\tau G_a'(1)\,\Omega(t)\,\lambda(t)\,dt\right\}^2\\ & \times \epsilon^2 + \left(G_a^{-1}\right)'\left\{\int_0^\tau G_a(1)\,\Omega(t)\,dt\right\}\\ & \times \left\{\int_0^\tau G_a''(1)\,\Omega(t)\,\lambda^2(t)\,dt\right\}\\ & \times \epsilon^2 + o\left(\epsilon^2\right) = 1 + \left(G_a^{-1}\right)'\left\{G_a(1)\right\}\times G_a''(1)\times\int_0^\tau \Omega(t)\,\lambda^2(t)\,dt\\ & \times \epsilon^2 + o\left(\epsilon^2\right)\end{aligned}$$

In this case, the local power of θ_a is given by

$$P\left(\left|\left(G_a^{-1}\right)'\{G_a(1)\}\times\frac{G_a''(1)}{G_a'(1)}\times\int_0^\tau\Omega(t)\lambda^2(t)\,dt\times\epsilon^2+o\left(\epsilon^2\right)\right|>BZ_{1-\alpha/2}\right)$$

where $G_a''(x)=-\left(\frac{a+1}{a+x}\right)G_a'(x)$. Again, this local power is maximized at $a = 1$.

Proof (of Theorem 3) The proof is based on the same linearization given in equation (15) but on the right hand side of equation (15), expressions $A_1(t)$, $A_0(t)$, $B_1(t)$, and $B_0(t)$ are indexed by $a \in [0, 1]$. Additionally, $o_p(1)$ on the right hand side of (15) converges in probability to zero uniformly in a. It is easy to check from the explicit expressions of A_1, A_0, B_1, B_0 that they all belong to a bounded set in $BV[0, \tau]$ for any $a \in [0, 1]$. Thus, condition (C.2) and the results in [6] yield that $n^{1/2}\left(\hat{\theta}_a-\theta_a\right)$, as a stochastic process indexed by $a \in [0, 1]$, converges weakly to a Gaussian process. Theorem 3 thus follows from the continuity theorem.

References

1. Abadi A, Amanpour F, Bajdik C, Yavari P. Breast Cancer Survival Analysis: Applying the Generalized Gamma Distribution under Different Conditions of the Proportional Hazards and Accelerated Failure Time Assumptions. International journal of preventive medicine. 2012; 3(9): 644. [PubMed: 23024854]
2. Abrahamowicz M, Mackenzie T, Esdaile JM. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. Journal of the American Statistical Association. 1996; 91(436):1432–1439.
3. Banerjee T, Chen MH, Dey DK, Kim S. Bayesian analysis of generalized odds-rate hazards models for survival data. Lifetime data analysis. 2007; 13(2):241–260. [PubMed: 17401683]
4. Cox D. Regression models and life-tables (with Discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:187–220.
5. Fan, J.; Yao, Q. Nonlinear time series. Vol. 2. Springer; 2003.
6. Gill RD, van der Vaart AW. Non- and semi-parametric maximum likelihood estimators and the von Mises method: II. Scand. J. Statist. 1993; 20(4):271–288.
7. Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. Journal of the American Statistical Association. 1992; 87(420):942–951.
8. Hastie T, Tibshirani R. Varying-coefficient models. Journal of the Royal Statistical Society. Series B (Methodological). 1993:757–796.
9. Hess KR. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. Statistics in medicine. 1994; 13(10):1045–1062. [PubMed: 8073200]
10. Kalbfleisch JD, Prentice RL. Estimation of the average hazard ratio. Biometrika. 1981; 68:105–112.
11. Kooperberg C, Stone CJ, Truong YK. Hazard regression. Journal of the American Statistical Association. 1995; 90(429):78–94.
12. Kosorok, MR. Introduction to empirical processes and semiparametric inference. Springer; 2007.
13. Lininger L, Gail MH, Green SB, Byar DP. Comparison of four tests for equality of survival curves in the presence of stratification and censoring. Biometrika. 1979; 66:417–428.
14. Moeschberger, ML.; Klein, JP. Survival analysis: Techniques for censored and truncated data. Springer; 2003.
15. Pepe MS, Fleming TR. Weighted Kaplan-Meier Statistics: A Class of Distance Tests for Censored Survival Data. Biometrics. 1989; 45(2):497–507. [PubMed: 2765634]

16. Robins J, Tsiatis A. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. Biometrika. 1992; 79:311–20.
17. Royston P, Parmar M. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine. 2002; 21:2175–97. [PubMed: 12210632]
18. Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. Statistics in Medicine. 2009; 28:2473–2489. [PubMed: 19472308]
19. Shen Y, Fleming TR. Weighted Mean Survival Test Statistics: a Class of Distance Tests for Censored Survival Data. ournal of the Royal Statistical Society: Series B (Statistical Methodology). 1997; 59(1):269–280.
20. Therneau, T.; Grambsch, P. Modeling Survival Data: Extending the Cox Model. Springer; New York: 2000.
21. Wei L. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. Statistics in Medicine. 1992; 11:1871–79. [PubMed: 1480879]
22. Xu R, O'Quigley J. Estimating average regression effect under non-proportional hazards. Biostatistics. 2000; 1:423–439. [PubMed: 12933565]
23. Zeng D, Chen Q, Chen MH, Ibrahim J, Group AR. Estimating Treatment Effects with Treatment Crossovers via Semi-Competing Risks Models: An Application to a Colorectal Cancer Study. Biometrika. 2012; 99:167–184. [PubMed: 23049136]

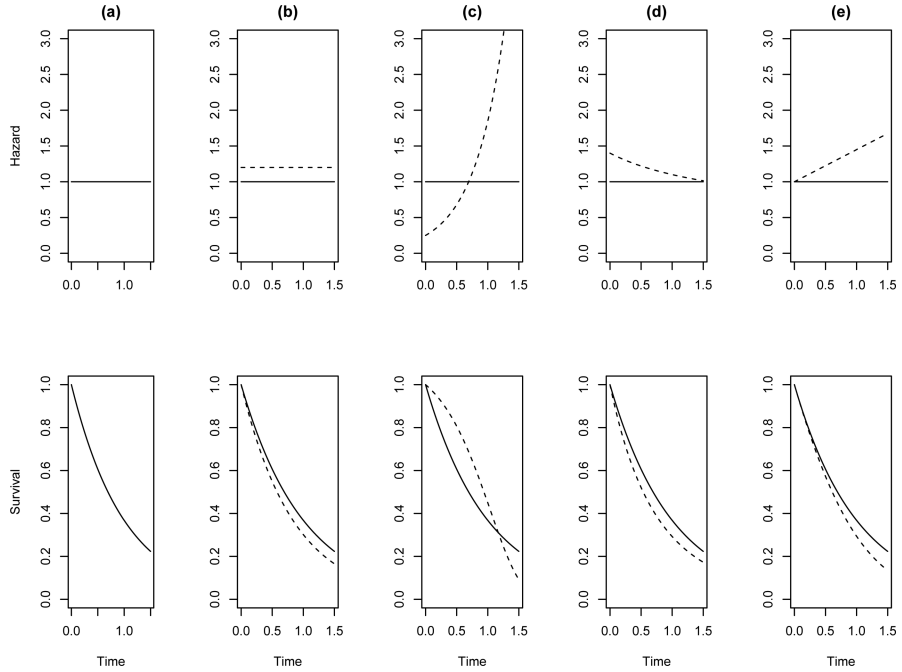

Fig. 1.

Survival and hazard functions of two treatments with five typical treatment effects. The curves in the first row are the true hazard functions and those in the second row are the true survival functions. In each panel, the solid curve is the true hazard/survival function in the control arm and the dashed curve is the true hazard/survival function in the treatment arm. The hazard function of the control arm is defined by a constant hazard $h_0(t) = 1$ in each column. The hazard functions for the treatment arm are characterized by (a) $h_1(t) = h_0(t)$ as identical hazards, by (b) $h_1(t) = 1 \cdot 2h_0(t)$ as proportional hazards, by (c) $h_1(t) = 0 \cdot 25 \exp(2t)h_0(t)$ as crossing hazards, by (d) $h_1(t) = \{0 \cdot 5 + 0 \cdot 9/(1 + 0 \cdot 5t)\}h_0(t)$ as converging hazards, and by (e) $h_1(t) = (1 + 0 \cdot 45t)h_0(t)$ as diverging hazards.

Table 1

Simulation Study I with $\Omega(t) \propto \{S_0(t)S_1(t)\}^{1/2}$ and sample size $n = 800$.

Method	TRUE	BIAS	SD	ESE	CP%	Type I
	Figure 1 (a) Identical Hazards; CR=24%					
Cox	1·00	5	8·2	8·2	95	5
Identity ($a = -1$)	1·00	15	8·4	8·7	96	4
Logarithm ($a = 0$)	1·00	5	8·3	8·5	96	4
Ratio ($a = 1$)	1·00	5	8·3	8·4	96	4
$\hat{\theta}_{sup}$	1·00	5	8·3	8·5	95	5

Method	TRUE	BIAS	SD	ESE	CP%	Power
	Figure 1 (b) Proportional Hazards; CR=21%					
Cox	1·20	6	9·7	9·6	94	58
Identity ($a = -1$)	1·20	19	10·0	10·4	96	57
Logarithm ($a = 0$)	1·20	7	9·9	10·1	96	52
Ratio ($a = 1$)	1·20	6	9·8	10·0	96	52
$\hat{\theta}_{sup}$	1·20	7	9·9	10·1	95	62
	Figure 1 (c) Crossing Hazards; CR=18·6%					
Cox	1·00	6	8·1	8·0	95	5
Identity ($a = -1$)	1·05	33	10·7	11·1	97	6
Logarithm ($a = 0$)	0·76	29	7·2	7·0	93	82
Ratio ($a = 1$)	0·77	25	6·4	6·2	94	86
$\hat{\theta}_{sup}$	0·76	28	7·1	6·8	92	81
	Figure 1 (d) Converging Hazards; CR=21·3%					
Cox	1·21	21	9·9	9·8	95	66
Identity ($a = -1$)	1·22	15	10·2	10·5	96	63
Logarithm ($a = 0$)	1·22	3	10·1	10·4	96	57
Ratio ($a = 1$)	1·22	3	10·0	10·2	96	58
$\hat{\theta}_{sup}$	1·22	4	10·1	10·4	95	65
	Figure 1 (e) Diverging Hazards; CR=19·6%					
Cox	1·24	5	10·0	9·9	94	72
Identity ($a = -1$)	1·25	24	10·4	10·9	96	73
Logarithm ($a = 0$)	1·23	12	10·0	10·3	96	68
Ratio ($a = 1$)	1·23	11	9·9	10·2	96	69
$\hat{\theta}_{sup}$	1·23	13	10·0	10·4	95	75

"TRUE" is the true value of θ_a calculated from the corresponding formulae; "BIAS" (×1000) is the average bias of the parameter estimates; "SD" (×100) is the sample standard deviation of the estimates; "ESE" (×100) is the average of the standard error estimates; "CP%" is the coverage probability of the 95% confidence interval based on a normal approximation; "Type I" is the type I error under the null hypothesis of identical hazards; "Power"(×100) is the power to reject the null hypothesis of identical hazards; "CR" is the censoring rate.

Table 2

Simulation Study I with $\mathcal{X}(t) \propto \{S_0(t)S_1(t)\}^{1/2}$ and sample size $n = 400$. Notation is the same as Table 1.

Method	TRUE	BIAS	SD	ESE	CP%	Type I
	Figure 1 (a) Identical Hazards; CR=24%					
Cox	1·00	8	11·3	11·6	96	4
Identity ($a = -1$)	1·00	25	11·7	12·8	97	3
Logarithm ($a = 0$)	1·00	7	11·4	12·2	97	4
Ratio ($a = 1$)	1·00	7	11·3	12·0	97	4
$\hat{\theta}_{sup}$	1·00	7	11·4	12·2	96	5

Method	TRUE	BIAS	SD	ESE	CP%	Power
	Figure 1 (b) Proportional Hazards; CR=21%					
Cox	1·20	10	13·4	13·7	96	27
Identity ($a = -1$)	1·20	30	13·7	15·3	97	23
Logarithm ($a = 0$)	1·20	8	13·4	14·5	97	21
Ratio ($a = 1$)	1·20	7	13·3	14·3	97	22
$\hat{\theta}_{sup}$	1·20	9	13·4	14·5	97	33
	Figure 1 (c) Crossing Hazards; CR=18·6%					
Cox	1·00	13	11·5	11·4	95	5
Identity (a = –1)	1·05	62	15·4	17·5	97	3
Logarithm ($a = 0$)	0·75	43	10·0	10·1	95	52
Ratio ($a = 1$)	0·77	38	8·8	8·9	95	58
$\hat{\theta}_{sup}$	0·76	41	9·8	10·0	93	49
	Figure 1 (d) Converging Hazards; CR=21·3%					
Cox	1·21	26	13·7	14·0	96	34
Identity ($a = -1$)	1·22	26	14·0	15·4	97	28
Logarithm ($a = 0$)	1·22	5	13·7	14·9	97	23
Ratio ($a = 1$)	1·22	3	13·5	14·6	97	25
$\hat{\theta}_{sup}$	1·22	5	13·7	14·8	96	36
	Figure 1 (e) Diverging Hazards; CR=19·6%					
Cox	1·24	9	13·7	14·0	95	39
Identity ($a = -1$)	1·25	38	14·3	16·2	98	36
Logarithm ($a = 0$)	1·23	16	13·7	14·9	97	34
Ratio ($a = 1$)	1·23	14	13·5	14·5	97	34
$\hat{\theta}_{sup}$	1·23	17	13·6	14·9	97	46

Table 3

Simulation Study II with $\Omega(t) \propto 1$ and sample size $n = 400$. Notation is the same as Table 1.

Method	TRUE	BIAS	SD	ESE	CP%	Type I
Figure 1 (a) Identical Hazards; CR=24%						
Cox	1·00	8	11·3	11·6	96	4
Identity ($a = -1$)	1·00	31	13·5	14·9	97	3
Logarithm ($a = 0$)	1·00	8	12·9	13·7	96	4
Ratio ($a = 1$)	1·00	7	12·7	13·3	95	5
$\hat{\theta}_{sup}$	1·23	8	12·9	13·7	95	5

Method	TRUE	BIAS	SD	ESE	CP%	Power
Figure 1 (b) Proportional Hazards; CR=21%						
Cox	1·20	10	13·4	13·7	96	27
Identity ($a = -1$)	1·20	37	16·2	18·3	98	12
Logarithm ($a = 0$)	1·20	9	15·5	16·7	96	14
Ratio ($a = 1$)	1·20	6	15·2	16·2	96	15
$\hat{\theta}_{sup}$	1·20	9	15·5	16·7	96	25
Figure 1 (c) Crossing Hazards; CR=18·6%						
Cox	1·00	13	11·5	11·4	95	5
Identity ($a = -1$)	1·59	79	31·4	36·4	97	44
Logarithm ($a = 0$)	1·12	48	14·3	15·2	97	12
Ratio ($a = 1$)	1·10	38	11·5	12·0	97	15
$\hat{\theta}_{sup}$	1·12	50	14·3	15v2	95	22
Figure 1 (d) Converging Hazards; CR=21·3%						
Cox	1·21	26	13·7	14·0	96	34
Identity ($a = -1$)	1·17	36	15·6	17·4	97	10
Logarithm ($a = 0$)	1·17	9	15·5	16·5	96	9
Ratio ($a = 1$)	1·17	8	15·3	16·0	95	10
$\hat{\theta}_{sup}$	1·17	9	15·5	16·5	95	18
Figure 1 (e) Diverging Hazards; CR=19·6%						
Cox	1·24	9	13·7	14·0	95	39
Identity ($a = -1$)	1·34	46	19·2	21·9	98	32
Logarithm ($a = 0$)	1·32	13	16·7	18·2	96	41
Ratio ($a = 1$)	1·32	8	16·0	17·4	96	44
$\hat{\theta}_{sup}$	1·32	13	16·6	18·2	96	58

Table 4

The estimates of $\theta(h_0/h_1; a)$ in the Panitumumab Study with the outcome of overall survival and weight function proportional to $\{S_0(t)S_1(t)\}^{1/2}$.

Transformation	Estimate	SD	95% CI	P-value
Logarithm	2·90	0·84	(1·48, 4·73)	0·008
Ratio	2·91	0·81	(1·61, 4·69)	< 0·001
$\hat{\theta}_{sup}$	2·92	0·84	(1·61, 4·80)	< 0·001