Computational Statistics and Data Analysis 56 (2012) 4312–4319

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Induced smoothing for the semiparametric accelerated hazards model

Haifen Li [a,b], Jiajia Zhang [b,*], Yincai Tang [a]

[a] School of Finance and Statistics, East China Normal University, Shanghai, 200241, PR China

[b] Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Article history:
Received 21 July 2011
Received in revised form 1 April 2012
Accepted 1 April 2012
Available online 9 April 2012

Keywords:
Accelerated hazards model
Rank estimation
Induced smoothing

ABSTRACT

Compared to the proportional hazards model and accelerated failure time model, the accelerated hazards model has a unique property in its application, in that it can allow gradual effects of the treatment. However, its application is still very limited, partly due to the complexity of existing semiparametric estimation methods. We propose a new semiparametric estimation method based on the induced smoothing and rank type estimates. The parameter estimates and their variances can be easily obtained from the smoothed estimating equation; thus it is easy to use in practice. Our numerical study shows that the new method is more efficient than the existing methods with respect to its variance estimation and coverage probability. The proposed method is employed to reanalyze a data set from a brain tumor treatment study.

1. Introduction

The proportional hazards (PH) model (Cox, 1972) and accelerated failure time (AFT) model (Cox and Oakes, 1984) are the most popular survival models in practice. Unlike the PH model and AFT model, there is little discussion of the accelerated hazards (AH) model (Chen and Wang, 2000) in the literature, partly due to its complex estimation methods.

Let us denote $h(\cdot)$ as the hazard risk, t as the event time and $\mathbf{x}$ as the possible covariate vector. The AH model can be written as

$$h(t|\mathbf{x}) = h_0(te^{\boldsymbol{\beta}'\mathbf{x}}), \tag{1}$$

where $h_0(\cdot)$ is an arbitrary unknown baseline hazard function and $\boldsymbol{\beta}$ is a vector of unknown parameters of interest. When $t = 0$, $h(0|\mathbf{x}) = h_0(0)$, so the main characteristic of the AH model is that it allows the treatment effect to release gradually, and a lag period may exist before the treatment is fully effective (Zucker and Lakatos, 1990). It is worthwhile pointing out that neither the PH model nor the AFT model can allow the same hazard risk at time 0 because $h(0|\mathbf{x}) = h_0(0)e^{\boldsymbol{\beta}'\mathbf{x}}$ in both models, which means both the PH model and the AFT model implicitly expect the treatment to take a significant effect immediately following its application (Zhang and Peng, 2009a). Furthermore, another difference among the PH, AFT and AH models is the crossing status of the hazard function. As mentioned in Zhang and Peng (2009b), the hazard function from the AFT model and AH model may have crossing point which is not allowed in the PH model.

A typical application of the AH model is to evaluate the effectiveness of biodegradable carmustine (BCNU) polymers to the recurrent brain malignant gliomas (Brem et al., 1995). In order to reach a higher local drug concentration, the BCNU polymer was supposed to gradually release BCNU in the first 3-week period following placement. The kernel smoothed

* Corresponding author. Tel.: +1 803 777 4474; fax: +1 803 777 2524.
E-mail address: jzhang@mailbox.sc.edu (J. Zhang).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.04.001

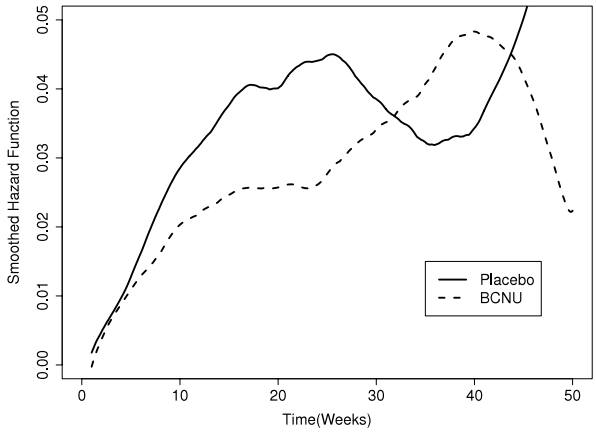


Fig. 1. Fitted kernel-smoothed hazard functions for the placebo and BCNU trial.

hazard curve with 8-week bandwidth based on the Nelson–Aalen estimation method for each group is investigated in Fig. 1, which clearly displays the pattern of hazard risk between the placebo and BCNU treatment group. The BCNU treatment does not have any effects during the first three weeks. It has a tendency to increase the hazard risk between 3 and 40 weeks, and decrease thereafter. Furthermore, the hazard curves of the placebo and BCNU treatment have two crossing points around the 31st and 42nd weeks. That is, the BCNU group has lower risk between 3 and 32 and after 43 weeks, while it has higher risk between 33 and 43 weeks. The gradual effects and crossing points provide evidence to apply the AH model to this data set. More discussions can be found in Chen (2001), Chen and Wang (2000) and Zhang et al. (2011).

The computational challenges in estimating both regression parameters and the covariance matrix of the AH model may limit its application in practice. Chen and Wang (2000) proposed a semiparametric estimation method based on a rank-type estimation method. The variance of an estimated parameter will depend on the nonparametric estimator of the baseline hazard function and its numerical derivative and the parameter estimates must be obtained through a root finding procedure, which may increase computational difficulties in practice. Chen (2001) improved the variance estimation procedure and Chen and Jewell (2001) extended the estimation method to a more general model. Zhang et al. (2011) proposed an efficient semiparametric estimation method in the AH model based on a kernel-smoothed approximation of the profile likelihood function. The covariance matrix can be consistently estimated by the second derivative of the approximation of the profile likelihood function. However, as they illustrated, the estimated variances from the Hessian matrix of the kernel-smoothed profile likelihood approximation may underestimate the true variances when the sample size is small, or even moderate.

The purpose of this paper is to address the nonsmoothed rank type estimating equation by the induced smoothing approach, which is an advanced technique used in the AFT model (Brown and Wang, 2007), the clustered AFT model (Johnson and Strawderman, 2009; Fu et al., 2010), and the quantile regression model (Pang et al., 2012). We successfully develop an estimation method for the regression parameters and variance estimation for the AH model based on the induced smoothing technique. The rest of this article is organized as follows. In Section 2, we describe the induced smoothing method for the AH model based on the rank-type estimating equation and establish its asymptotic properties. Simulation studies are conducted in Section 3 to investigate the finite-sample behavior of the proposed method. We revisit the brain malignant gliomas example in Section 4 and make conclusions in Section 5.

2. Estimation method

Let $\boldsymbol{O}_i = (t_i, \delta_i, \mathbf{x}_i)$ denote the observed data for the ith individual, $i = 1, \ldots, n$, where t_i is the observed survival time, δ_i is the censoring indicator with $\delta_i = 1$ for an uncensored t_i and $\delta_i = 0$ for a censored t_i, and $\mathbf{x}_i$ is the corresponding $p \times 1$ covariate vector. We assume that censoring is independent and noninformative.

2.1. Rank-type estimation method

Chen and Wang (2000) proposed to estimate $\boldsymbol{\beta}$ using a weighted rank-type estimating equation, which can be written as

$$\boldsymbol{U}_n^w(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i(\boldsymbol{\beta})\delta_i \left(\mathbf{x}_i - \frac{\sum_{j=1}^{n} I\left(r_j(\boldsymbol{\beta}) \geq r_i(\boldsymbol{\beta})\right) e^{-\boldsymbol{\beta}'\mathbf{x}_j}\mathbf{x}_j}{\sum_{j=1}^{n} I\left(r_j(\boldsymbol{\beta}) \geq r_i(\boldsymbol{\beta})\right) e^{-\boldsymbol{\beta}'\mathbf{x}_j}} \right) = \mathbf{0}, \tag{2}$$

where $r_i(\boldsymbol{\beta}) = \log t_i + \boldsymbol{\beta}'\mathbf{x}_i$, $i = 1, \ldots, n$ and $w_i(\cdot)$ are nonnegative weight functions. Assuming $w_i^g(\cdot)$ is the Gehan-type weight function, which is $w_i^g(\boldsymbol{\beta}) = \frac{1}{n}\sum_{j=1}^n I(r_j(\boldsymbol{\beta}) \geq r_i(\boldsymbol{\beta}))e^{-\boldsymbol{\beta}'\mathbf{x}_j}$, the Gehan-rank estimating equation is

$$\boldsymbol{U}_n^g(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \delta_i(\mathbf{x}_i - \mathbf{x}_j)I\left(r_j(\boldsymbol{\beta}) \geq r_i(\boldsymbol{\beta})\right)e^{-\boldsymbol{\beta}'\mathbf{x}_j} = \mathbf{0}. \tag{3}$$

Estimation of regression parameters can be obtained by solving Eq. (2) or (3) with respect to $\boldsymbol{\beta}$. As mentioned in Chen et al. (2002), the estimated $\boldsymbol{\beta}$ asymptotically follows a normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \boldsymbol{D}_n^{-1}\boldsymbol{B}_n\boldsymbol{D}_n^{-1}$, $\boldsymbol{B}_n = \mathrm{Var}\{\boldsymbol{U}_n^w(\boldsymbol{\beta})\}$, and $\boldsymbol{D}_n = [\partial E\{\boldsymbol{U}_n^w(\boldsymbol{\beta})\}/\partial\boldsymbol{\beta}]_{\boldsymbol{\beta}_0}$. $\boldsymbol{B}_n$ can be approximated by

$$\hat{\boldsymbol{B}}_n = \sum_{i=1}^n\left[w_i(\boldsymbol{\beta})\delta_i\left(\mathbf{x}_i - \frac{\sum_{j=1}^n I\left(r_j(\boldsymbol{\beta}) \geq r_i(\boldsymbol{\beta})\right)e^{-\boldsymbol{\beta}'\mathbf{x}_j}\mathbf{x}_j}{\sum_{j=1}^n I\left(r_j(\boldsymbol{\beta}) \geq r_i(\boldsymbol{\beta})\right)e^{-\boldsymbol{\beta}'\mathbf{x}_j}}\right)\right]^{\otimes 2}$$

where $\boldsymbol{v}^{\otimes 2} = \boldsymbol{v}\boldsymbol{v}'$ for any vector $\boldsymbol{v}$. But, the approximation of $\boldsymbol{D}_n$ is a challenging task since the closed form of partial derivatives of $E\{\boldsymbol{U}_n^w(\boldsymbol{\beta})\}$ does not exist.

Therefore, Chen (2001) presented an alternative method to estimate the variance. He first decomposed the empirical estimate $\hat{\boldsymbol{B}}_n$ as $\boldsymbol{\sigma}\boldsymbol{\sigma}'$, where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_p)'$, and computed the solutions of $\boldsymbol{U}_n^w(b_i) = \sigma_i$, for $i = 1, \ldots, p$. The variance estimator of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is then given by $(b_1 - \hat{\beta}_1, \ldots, b_p - \hat{\beta}_p)(b_1 - \hat{\beta}_1, \ldots, b_p - \hat{\beta}_p)'$. This algorithm is applied in our simulations for the rank-type method.

2.2. Induced smoothing method

Based on the similar idea of Brown and Wang (2005), we consider the smoothed estimating function $\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, which is constructed by adding the random perturbation $(\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}$ to $\boldsymbol{\beta}$ in the score function $\boldsymbol{U}_n^g(\boldsymbol{\beta})$ and then taking the expectation of the nonsmoothed estimating function with respect to $\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a $N(\mathbf{0}, \boldsymbol{I_p})$ random vector independent of the data ($\boldsymbol{I_p}$ is the $p \times p$ identity matrix), and $\boldsymbol{\Sigma} = \boldsymbol{O}(\mathbf{1})$ is some symmetric positive definite matrix. Let $u_{ij} = \frac{1}{n}(\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}(\mathbf{x}_i - \mathbf{x}_j)$, $v_{ij} = \frac{1}{n}\mathbf{x}_j'\boldsymbol{\Sigma}(\mathbf{x}_i - \mathbf{x}_j)$, $u_j = \frac{1}{n}\mathbf{x}_j'\boldsymbol{\Sigma}\mathbf{x}_j$, the smoothed version of $\boldsymbol{U}_n^g(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is

$$\begin{aligned}\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= E_{\boldsymbol{Z}}\left\{\boldsymbol{U}_n^g\left(\boldsymbol{\beta} + (\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}\right)\right\} \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \delta_i(\mathbf{x}_i - \mathbf{x}_j)e^{\frac{u_j}{2} - \boldsymbol{\beta}'\mathbf{x}_j}\Phi\left(\frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{\sqrt{u_{ij}}} + \frac{v_{ij}}{\sqrt{u_{ij}}}\right),\end{aligned} \tag{4}$$

where $\Phi(\cdot)$ be the cumulative distribution function of the standard normal distribution. The above smoothed estimating equation is a monotone decreasing function with respect to each element of $\boldsymbol{\beta}$. Appendix gives more detail for deriving $\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ and its monotonicity. When $\boldsymbol{\Sigma}$ is given, the smoothed regression coefficient estimator $\tilde{\boldsymbol{\beta}}$ can be obtained by solving the equation $\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \mathbf{0}$. When $\boldsymbol{\Sigma}$ is unknown, we can get its estimate by iterating the estimating equation, which uses a symmetric positive definite $p \times p$ matrix $\tilde{\boldsymbol{\Sigma}}_0 = O(\mathbf{1})$ as its initial value. In fact, as long as the matrix used for the initial smoothing is positive definite with order $\boldsymbol{O}(\mathbf{1})$, the corresponding smoothed coefficient estimator $\tilde{\boldsymbol{\beta}}$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}$.

According to Brown and Wang (2007), the estimation of the asymptotic variance is mainly based on the sandwich form of the covariance matrix of estimated parameters. The partial derivatives of the smoothed estimating function (4) can be explicitly expressed as

$$\begin{aligned}\tilde{\boldsymbol{D}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \frac{\partial\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial\boldsymbol{\beta}} \\ &= -\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \delta_i(\mathbf{x}_i - \mathbf{x}_j)e^{u_j/2 - \boldsymbol{\beta}'\mathbf{x}_j}\left[\phi(d_{ij})\frac{(\mathbf{x}_i - \mathbf{x}_j)'}{\sqrt{u_{ij}}} + \Phi(d_{ij})\mathbf{x}_j'\right],\end{aligned}$$

where $d_{ij} = \frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{\sqrt{u_{ij}}} + \frac{v_{ij}}{\sqrt{u_{ij}}}$, $\phi(\cdot)$ is the probability density function of the standard normal distribution, which is negative when d and $\mathbf{x}$ are bounded. Then, its variance can be similarly developed as

$$\hat{\boldsymbol{B}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \frac{1}{n^2}\sum_{i=1}^n\left[\delta_i\sum_{j=1}^n(\mathbf{x}_i - \mathbf{x}_j)e^{u_j/2 - \boldsymbol{\beta}'\mathbf{x}_j}\Phi(d_{ij})\right]^{\otimes 2}.$$

Thus, the asymptotic variance of the estimated parameter can be consistently estimated via $\tilde{\boldsymbol{D}}_n^{-1}\hat{\boldsymbol{B}}_n\tilde{\boldsymbol{D}}_n^{-1}$.

As suggested in Pang et al. (2012), an iterative procedure can be used to simultaneously estimate the regression parameters and their covariance matrix. The estimation procedure consists of the following steps:

Step 1. Let $\tilde{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}}$, the estimator solved from the Gehan-rank estimating Eq. (3), and $\tilde{\boldsymbol{\Sigma}}_0 = \boldsymbol{I_p}$.

Step 2. Given $\tilde{\boldsymbol{\beta}}_{k-1}$ and $\tilde{\boldsymbol{\Sigma}}_{k-1}$ from the $(k-1)$th step, update $\tilde{\boldsymbol{\beta}}_k$ and $\tilde{\boldsymbol{\Sigma}}_k$ as:

$$\tilde{\boldsymbol{\beta}}_k = \tilde{\boldsymbol{\beta}}_{k-1} + \left\{-\tilde{\boldsymbol{D}}_n(\tilde{\boldsymbol{\beta}}_{k-1}, \tilde{\boldsymbol{\Sigma}}_{k-1})\right\}^{-1} \tilde{\boldsymbol{U}}_n(\tilde{\boldsymbol{\beta}}_{k-1}, \tilde{\boldsymbol{\Sigma}}_{k-1}), \quad \text{and}$$
$$\tilde{\boldsymbol{\Sigma}}_k = \tilde{\boldsymbol{D}}_n^{-1}(\tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\Sigma}}_{k-1})\hat{\boldsymbol{B}}_n(\tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\Sigma}}_{k-1})\tilde{\boldsymbol{D}}_n^{-1}(\tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\Sigma}}_{k-1}).$$

Step 3. Repeat *Step* 2 until a selected stopping criterion, $\max(\max|\tilde{\boldsymbol{\beta}}_k - \tilde{\boldsymbol{\beta}}_{k-1}|, \max|\tilde{\boldsymbol{\Sigma}}_k - \tilde{\boldsymbol{\Sigma}}_{k-1}|) < \tau_0$, is reached. The convergence is controlled by $\tau_0 = 10^{-4}$ in the simulation and real data analysis.

Denote the coefficient estimates and their associated covariance estimate at convergence as $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\Sigma}}$, respectively. When $\tilde{\boldsymbol{\beta}}$ is available, the Breslow-type estimator of the baseline cumulative hazard function (Chen et al., 2002) is

$$\hat{\boldsymbol{H}}_0(t, \tilde{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \frac{\delta_i I(\log(t) \geq r_i(\tilde{\boldsymbol{\beta}}))}{\sum_{j=1}^{n} I(r_j(\tilde{\boldsymbol{\beta}}) \geq r_i(\tilde{\boldsymbol{\beta}}))e^{-\tilde{\boldsymbol{\beta}}'\mathbf{x}_j}}.$$

Based on $\hat{\boldsymbol{H}}_0(t, \tilde{\boldsymbol{\beta}})$, we can obtain the smoothed hazard curve by the kernel smoothing method. For an arbitrary time point t, which satisfies $b \leq t \leq \max(t_i) - b$, similar to Klein and Moeschberger (2003), the kernel-smoothed hazard estimator of $\hat{\boldsymbol{H}}_0(t)$ is

$$\hat{\boldsymbol{H}}_0(t) = b^{-1} \sum_{i=1}^{n} K\left(\frac{t - t_i}{b}\right) \Delta\hat{\boldsymbol{H}}_0(t_i, \tilde{\boldsymbol{\beta}})$$

where $K(\cdot)$ is the kernel function and b denotes the bandwidth. The choice of the kernel function can be found in Klein and Moeschberger (2003), and we use the Epanechnikov kernel, which is $K(x) = 0.75(1 - x^2)$ for $-1 \leq x \leq 1$, in the real data analysis.

3. Simulation study

We conduct simulation studies to investigate the performance of the proposed estimation method. We generate the data from the AH model (1) with $\mathbf{x} = (x_1, x_2)$, where x_1 is generated from the standard normal distribution, and x_2 is generated from the binary distribution with the probability of 0.5. The regression coefficient $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ is set to $(1, -1)'$. The baseline hazard function $h_0(t)$ comes from the log-normal distribution (LN(0, 1)), log–logistic distribution with shape parameter 2 and scale parameter 1, and Weibull distribution with both the shape and scale parameters equal to 0.5. The censoring times are generated independently from the uniform distribution $U(0, a)$ with proper values of a so that the resulting censoring rates are about 15% and 30%. The sample sizes are 500 and 1000, with 1000 simulation replications.

We fit the data with the AH model using the induced smoothing method. For comparison purposes, we also list results from the Gehan rank-type method (Chen and Wang, 2000) and profile likelihood method (Zhang et al., 2011). The Gehan rank-type method is described in Section 2.1. The profile likelihood method (Zhang et al., 2011) is based on a kernel-smoothed approximation of the limit of the profile likelihood function, which is

$$l(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \left\{-r_i(\boldsymbol{\beta}) + \log\left[\frac{\frac{1}{na_n}\sum_{j=1}^{n} \delta_j K\left(\frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{a_n}\right)}{\frac{1}{n}\sum_{j=1}^{n} \int_{-\infty}^{\frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{a_n}} e^{-\boldsymbol{\beta}'\mathbf{x}_j} K(s)ds}\right]\right\} \tag{5}$$

where $r_i(\boldsymbol{\beta}) = \log t_i + \boldsymbol{\beta}'\mathbf{x}_i$, $K(\cdot)$ is the kernel function and a_n is the bandwidth. The estimation of $\boldsymbol{\beta}$ can be obtained through maximizing (5). Similar to Zhang et al. (2011), we choose the kernel function $K(\cdot)$ to be the standard normal density function and the bandwidth to be the optimal bandwidths for convenience, which are $(8\sqrt{2}/3)^{1/5}\sigma_1 n^{-1/5}$ for the kernel density and $4^{1/3}\sigma_2 n^{-1/3}$ for the cumulative kernel density, where σ_1 and σ_2 are the sample standard deviations of $r_i(\boldsymbol{\beta})$ using uncensored individuals and all individuals, respectively, based on the initial value of $\boldsymbol{\beta}$.

The biases (Bias), empirical standard deviations (StDev), average of the standard error estimates (StErr), and coverage probability (CP) of β_1 and β_2 are computed and summarized in Table 1.

From Table 1, the performance of all methods are quite similar for the different distributions. With respect to the bias, the proposed method and Gehan rank-type method are similar, which are smaller than that from the profile likelihood method; while these three methods are comparable with respect to the StDev. For the induced smoothing method, the standard error estimates (StErr) are close to the empirical standard deviations (StDev), which shows that the proposed variance

Table 1
Simulation results for 1000 replications.

Cen.	n	$\hat{\boldsymbol{\beta}}$	Induced smoothing				Profile likelihood				Gehan rank-type			
			Bias	StDev	StErr	CP	Bias	StDev	StErr	CP	Bias	StDev	StErr	CP
Log-normal baseline distribution														
15%	500	$\hat{\beta}_1$	−0.013	0.112	0.112	0.924	0.039	0.117	0.109	0.915	−0.013	0.113	0.088	0.852
		$\hat{\beta}_2$	0.009	0.228	0.225	0.934	−0.037	0.208	0.202	0.945	0.009	0.226	0.205	0.915
	1000	$\hat{\beta}_1$	−0.003	0.076	0.079	0.942	0.023	0.077	0.074	0.926	−0.003	0.076	0.065	0.897
		$\hat{\beta}_2$	−0.001	0.161	0.158	0.942	−0.022	0.145	0.139	0.944	0.000	0.160	0.145	0.917
30%	500	$\hat{\beta}_1$	0.000	0.125	0.118	0.917	0.050	0.131	0.119	0.907	0.000	0.125	0.090	0.834
		$\hat{\beta}_2$	−0.007	0.226	0.227	0.942	−0.054	0.233	0.214	0.929	−0.007	0.225	0.201	0.912
	1000	$\hat{\beta}_1$	−0.005	0.083	0.083	0.933	0.027	0.084	0.080	0.920	−0.005	0.083	0.066	0.879
		$\hat{\beta}_2$	0.002	0.159	0.162	0.948	−0.016	0.160	0.145	0.915	0.003	0.159	0.144	0.919
Log–logistic baseline distribution														
15%	500	$\hat{\beta}_1$	−0.008	0.104	0.107	0.936	0.071	0.118	0.105	0.880	−0.009	0.104	0.079	0.846
		$\hat{\beta}_2$	0.004	0.193	0.192	0.939	−0.073	0.197	0.184	0.918	0.006	0.191	0.167	0.907
	1000	$\hat{\beta}_1$	0.000	0.070	0.075	0.952	0.041	0.074	0.071	0.916	0.000	0.070	0.058	0.891
		$\hat{\beta}_2$	−0.003	0.136	0.135	0.938	−0.040	0.121	0.123	0.953	−0.003	0.136	0.116	0.905
30%	500	$\hat{\beta}_1$	0.006	0.123	0.119	0.935	0.094	0.177	0.122	0.857	0.006	0.122	0.085	0.829
		$\hat{\beta}_2$	−0.013	0.203	0.205	0.948	−0.099	0.244	0.206	0.917	−0.012	0.201	0.171	0.907
	1000	$\hat{\beta}_1$	−0.002	0.081	0.082	0.945	0.043	0.086	0.079	0.907	−0.002	0.081	0.062	0.867
		$\hat{\beta}_2$	0.000	0.140	0.143	0.951	−0.042	0.142	0.134	0.923	0.000	0.140	0.120	0.897
Weibull baseline distribution														
15%	500	$\hat{\beta}_1$	0.027	0.164	0.163	0.945	−0.079	0.108	0.113	0.879	0.028	0.165	0.161	0.944
		$\hat{\beta}_2$	−0.021	0.275	0.272	0.944	0.093	0.182	0.192	0.936	−0.021	0.273	0.239	0.911
	1000	$\hat{\beta}_1$	0.013	0.109	0.112	0.957	−0.047	0.080	0.082	0.896	0.013	0.110	0.105	0.941
		$\hat{\beta}_2$	−0.006	0.191	0.188	0.942	0.054	0.141	0.139	0.929	−0.006	0.190	0.163	0.906
30%	500	$\hat{\beta}_1$	0.013	0.165	0.166	0.948	−0.056	0.121	0.120	0.911	0.015	0.172	0.165	0.947
		$\hat{\beta}_2$	0.004	0.274	0.278	0.951	0.071	0.200	0.205	0.942	0.004	0.275	0.246	0.917
	1000	$\hat{\beta}_1$	0.008	0.111	0.114	0.958	−0.028	0.087	0.087	0.936	0.008	0.111	0.108	0.956
		$\hat{\beta}_2$	−0.006	0.196	0.194	0.947	0.037	0.145	0.149	0.950	−0.007	0.198	0.171	0.913

estimation method performs well in the simulation settings, and the coverage probability is close to its nominal level (95%). Compared to the profile likelihood method, the proposed method is more reliable with respect to its bias, variance and CP. The StDev is closer to StErr for the proposed approach than for the profile likelihood method, and the CP is closer to its nominal level. Compared to the Gehan rank-type method, the bias and StDev for the proposed method are almost the same, while the StErr and CP are significantly improved. For example, under the log-normal baseline distribution with 15% censoring, $\beta_1 = -0.003, 0.076, 0.079, 0.942$ when $n = 1000$ from the induced smoothing method, 0.023, 0.077, 0.074, 0.926 from the profile likelihood method, and −0.003, 0.076, 0.065, 0.897 from the Gehan rank-type method. When the sample size increases, the bias or variance of the estimated parameters will decrease. With an increase in the censoring rate, the bias or variance of the estimated parameters will increase.

From our computation experience, the iterative procedure will converge in 5 iterations for 99% of data sets. However, one must be cautious with a small sample size (<200) due to the possible convergence issue. We ran the simulations using 8 parallel 2300 MHZ CPU of a Quad-Core AMD Opteron processor. When the sample size is 500, the total computing time for completing 1000 simulation replications is about 7.8 h for the proposed method, 2.7 h for the profile likelihood method, and 11.6 h for the Gehan rank-type method. The proposed method is not as fast as the profile likelihood method, but it is acceptable considering its overall performance.

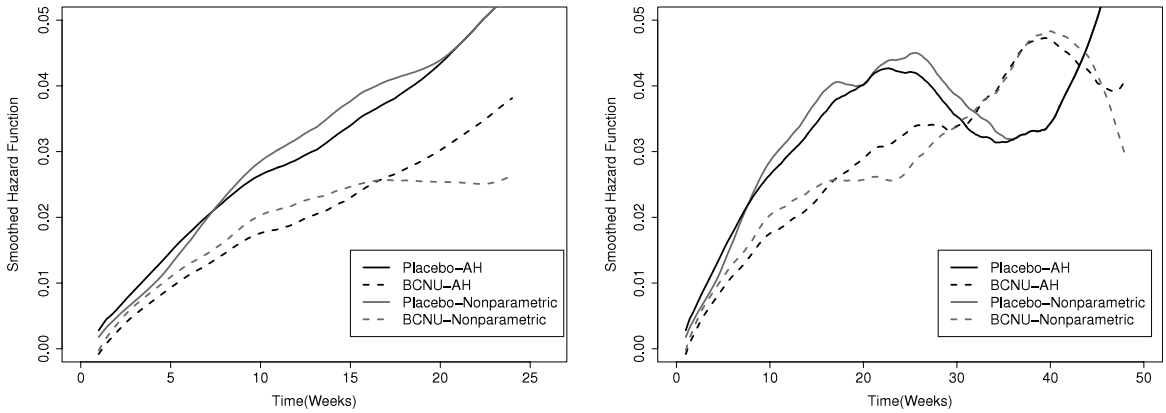
4. Brain tumor treatment

The proposed method will be used to evaluate the effectiveness of BCNU polymers. In this trial, 110 of 222 participants were randomized to the BCNU polymer treatment group, and 112 were randomized to the placebo polymer control group. In analyzing the BCNU trial, we are interested in estimating the treatment effect in the early period. Therefore, similar to Chen and Wang (2000), we conducted separate analyses for the first 26-week period, which has 53.6% censoring, and for the first 52-week period, which has 20.7% censoring. The results are summarized in Table 2.

The results are consistent with the results reported in Chen and Wang (2000). From Table 2, we can see that the BCNU is significant in both periods. The BCNU extends the time scale of hazard risk. This means to reach the same level of risk, the patients in the BCNU will need $e^{0.5883} = 1.80$ times of the time of the placebo group in the first 26 weeks and $e^{0.6054} = 1.83$ times of the time needed by the placebo group in the first 52 weeks. For comparison purposes, we plot the fitted kernel

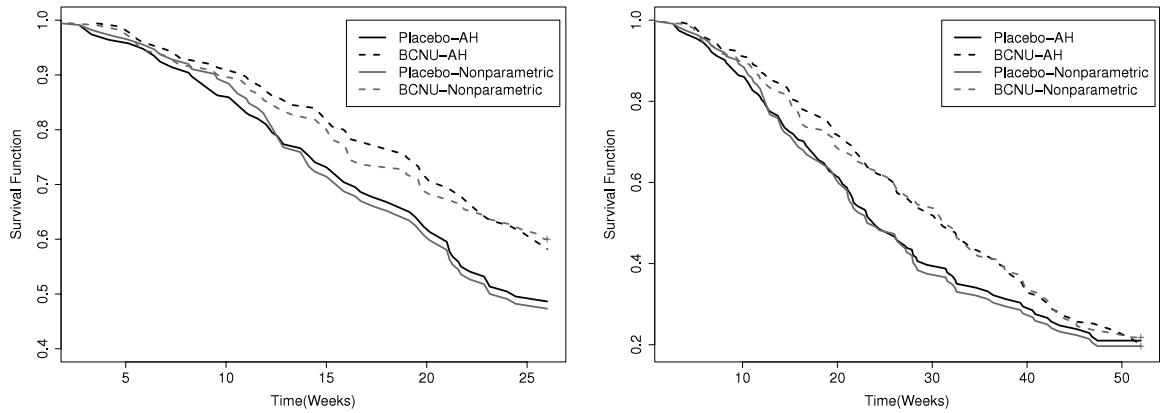
Table 2
Results for analysis of the BCNU trial.

		Induced smoothing	Profile likelihood	Gehan rank-type
First 26 weeks	$\hat{\beta}$	−0.5883	−0.4243	−0.5889
	Var($\hat{\beta}$)	0.0378	0.0118	0.0257
	95% CI	(−0.9693, −0.2074)	(−0.6371, −0.2114)	(−0.9033, −0.2947)
First 52 weeks	$\hat{\beta}$	−0.6054	−0.4785	−0.6086
	Var($\hat{\beta}$)	0.0314	0.0259	0.0601
	95% CI	(−0.9527, −0.3030)	(−0.7937, −0.1632)	(−1.0892, −0.1697)



(a) 26 Weeks: 8-week bandwidth. (b) 52 Weeks: 8-week bandwidth.

Fig. 2. Fitted kernel-smoothed hazard functions for placebo and BCNU trial.



(a) 26 Weeks. (b) 52 Weeks.

Fig. 3. Fitted survival functions for placebo and BCNU trial.

smoothed hazard function along with the kernel smoothed hazard function based on the Nelson–Aalen estimates in Fig. 2, where both of them use the bandwidth = 8 weeks. This displays the good performance of this method, since the fitted curves are close to the ones from the nonparametric approach. From Fig. 2(a), we can see that the hazard in the BCNU group is smaller than that in the placebo group in the first 26 weeks. In Fig. 2(b), we can see the increasing and decreasing trend of hazard in the BCNU group in the first 52 weeks. Furthermore, the same starting point of the hazard function in each group clearly indicates the gradual effects of BCNU, which can only be captured by the AH model.

In order to reveal the effects of BCNU to the recurrent brain malignant gliomas, we also plot the estimated survival curves along with the one from the Kaplan Meier estimator in Fig. 3. The closeness of the survival curves from both methods demonstrates good performances of the proposed method, and also shows that the BCNU can delay the recurrence of brain malignant gliomas.

5. Discussions

In this paper, we propose a new semiparametric estimation method of the AH model, which is based on the induced smoothing technique and Gehan rank-type estimation. From simulation studies, we show that the proposed method is more efficient compared to existing methods with regard to its parameter estimates and variance estimation. The sandwich estimation method avoids the extra computational burden of the bootstrap method. From computation experience, the proposed method always converges with moderate sample size. It is worthwhile expanding the proposed method to more general cases such as the multivariate case and the general class model (Chen and Jewell, 2001) in the future work.

Acknowledgements

The authors thank the reviewers and associate editor for insightful comments, which have led to a great improvement in the presentation of the work. The project is supported by an award from the National Cancer Institute (NCI, Award Number R03CA137790). The content is solely the responsibility of the authors and does not necessarily represent the official views of NCI.

Appendix

Deriving of $\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})$

We assume $\boldsymbol{Z}$ is a $N(\mathbf{0}, \boldsymbol{I_p})$ random vector independent of the data ($\boldsymbol{I_p}$ is the $p \times p$ identity matrix), and $\boldsymbol{\Sigma} = \boldsymbol{O}(\mathbf{1})$ is some symmetric positive definite matrix.

$$\begin{aligned}\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= E_{\boldsymbol{Z}}\left\{\boldsymbol{U}_n^g\left(\boldsymbol{\beta} + (\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}\right)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i(\mathbf{x}_i - \mathbf{x}_j)E_{\boldsymbol{Z}}\left\{I\left(r_j\left(\boldsymbol{\beta} + (\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}\right) \geq r_i\left(\boldsymbol{\beta} + (\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}\right)\right)e^{-\left(\boldsymbol{\beta}+(\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}\right)'\mathbf{x}_j}\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i(\mathbf{x}_i - \mathbf{x}_j)e^{-\boldsymbol{\beta}'\mathbf{x}_j}E_{\boldsymbol{Z}}\left\{I\left((\mathbf{x}_i - \mathbf{x}_j)'(\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z} \leq r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})\right)e^{-\mathbf{x}_j'(\boldsymbol{\Sigma}/n)^{1/2}\boldsymbol{Z}}\right\}. \end{aligned} \tag{6}$$

Let $\boldsymbol{a} = \mathbf{x}_j'(\boldsymbol{\Sigma}/n)^{1/2}$ and $\boldsymbol{b} = (\mathbf{x}_i - \mathbf{x}_j)'(\boldsymbol{\Sigma}/n)^{1/2}$. We define new random variables $Y_1 = \frac{\boldsymbol{a}}{\sqrt{\boldsymbol{aa}'}}\boldsymbol{Z}$ and $Y_2 = \frac{\boldsymbol{b}}{\sqrt{\boldsymbol{bb}'}}\boldsymbol{Z}$, which satisfy

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \boldsymbol{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\rho = \frac{\boldsymbol{ab}'}{\sqrt{\boldsymbol{aa}'}\sqrt{\boldsymbol{bb}'}}$. According to the character of binary normal distribution, random variables $W_1 = \frac{Y_1 - \rho Y_2}{\sqrt{1-\rho^2}}$ and $W_2 = Y_2$ are independent and follow the standard normal distribution. Then, the expectation part of (6) can be rewritten as

$$\begin{aligned} &E_{(W_1,W_2)}\left\{I\left(W_2 \leq \frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{\sqrt{\boldsymbol{bb}'}}\right)e^{-\sqrt{\boldsymbol{aa}'}\left(\sqrt{1-\rho^2}W_1 + \rho W_2\right)}\right\} \\ &= E_{W_1}\left\{e^{-\sqrt{\boldsymbol{aa}'}\sqrt{1-\rho^2}W_1}\right\}E_{W_2}\left\{I\left(W_2 \leq \frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{\sqrt{\boldsymbol{bb}'}}\right)e^{-\sqrt{\boldsymbol{aa}'}\rho W_2}\right\} \\ &= e^{\frac{\boldsymbol{aa}'}{2}}\Phi\left(\frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{\sqrt{\boldsymbol{bb}'}} + \sqrt{\boldsymbol{aa}'}\rho\right) \end{aligned} \tag{7}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. Let $u_{ij} = \frac{1}{n}(\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}(\mathbf{x}_i - \mathbf{x}_j)$, $v_{ij} = \frac{1}{n}\mathbf{x}_j'\boldsymbol{\Sigma}(\mathbf{x}_i - \mathbf{x}_j)$, $u_j = \frac{1}{n}\mathbf{x}_j'\boldsymbol{\Sigma}\mathbf{x}_j$, we can obtain (4) by substituting (7) into the expectation part of (6).

Monotonicity of $\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})$

When $\boldsymbol{\Sigma}$ is given, we will prove $\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is a monotone decreasing function when d and $\mathbf{x}$ are bounded, by showing its partial derivative with respect to β_i, $i = 1, \ldots, p$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, is non-positive. We will show the detail proof for one dimension, which can be extended to the p-dimension case.

First $\phi(d)/\Phi(d)$ is a monotone decreasing function since the first derivative of $\phi(d)/\Phi(d)$ is less than 0. Thus, for any $\varepsilon = o(n^{-1/2}) > 0$, there exists $|d| \leq M$, which satisfies $\phi(d)/\Phi(d) \geq \varepsilon$. Assuming that $|x|$ is bounded by $(\varepsilon/\sigma)\sqrt{n}$, where $\varepsilon/\sigma = o(n^{-1/2})$, the derivative of $\tilde{U}_n(\beta, \sigma)$ can be explicitly expressed by

$$\tilde{D}_n(\beta, \sigma) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i(x_i - x_j)e^{u_j/2 - \beta x_j}\left[\phi(d_{ij})\frac{x_i - x_j}{\sqrt{(\sigma^2/n)(x_i - x_j)^2}} + \Phi(d_{ij})x_j\right]$$

$$\leq \sum_i \sum_{(x_i - x_j)x_j < 0} \delta_i |x_i - x_j| e^{u_j/2 - \beta x_j} \left[\Phi(d_{ij}) |x_j| - \phi(d_{ij}) \frac{1}{\sqrt{\sigma^2/n}} \right]$$

$$= \sum_i \sum_{(x_i - x_j)x_j < 0} \delta_i |x_i - x_j| e^{u_j/2 - \beta x_j} \Phi(d_{ij}) \left[|x_j| - \frac{\phi(d_{ij})}{\Phi(d_{ij})} \frac{\sqrt{n}}{\sigma} \right]$$

$$\leq \sum_i \sum_{(x_i - x_j)x_j < 0} \delta_i |x_i - x_j| e^{u_j/2 - \beta x_j} \Phi(d_{ij}) \left[|x_j| - (\varepsilon/\sigma)\sqrt{n} \right] \leq 0.$$

Therefore, the estimating equation $\tilde{U}_n(\beta, \sigma)$ is a monotone decreasing function when d and $\mathbf{x}$ are bounded.

References

Brem, H., Piantadosi, S., Burger, P., Walker, M., Selker, R., Vick, N., Black, K., Sisti, M., Brem, S., Mohr, G., et al., 1995. Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas. Lancet 345 (8956), 1008–1012.

Brown, B., Wang, Y., 2005. Standard errors and covariance matrices for smoothed rank estimators. Biometrika 92 (1), 149–158.

Brown, B., Wang, Y., 2007. Induced smoothing for rank regression with censored survival times. Statistics in Medicine 26 (4), 828–836.

Chen, Y., 2001. Accelerated hazards regression model and its adequacy for censored survival data. Biometrics 57 (3), 853–860.

Chen, Y., Jewell, N., 2001. On a general class of semiparametric hazards regression models. Biometrika 88 (3), 687–702.

Chen, Y., Jewell, N., Yang, J., 2002, Accelerated hazards model: method, theory and applications, UC Berkeley Division of Biostatistics Working Paper Series, 117.

Chen, Y., Wang, M., 2000. Analysis of accelerated hazards models. Journal of the American Statistical Association 95 (450), 608–618.

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

Cox, D., Oakes, D., 1984. Analysis of Survival Data. Chapman & Hall.

Fu, L., Wang, Y., Bai, Z., 2010. Rank regression for analysis of clustered data: a natural induced smoothing approach. Computational Statistics & Data Analysis 54 (4), 1036–1050.

Johnson, L., Strawderman, R., 2009. Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. Biometrika 96 (3), 577–590.

Klein, J., Moeschberger, M., 2003. Survival Analysis: Techniques for Censored and Truncated Data. Springer Verlag.

Pang, L., Lu, W., Wang, H., 2012. Variance estimation in censored quantile regression via induced smoothing. Computational Statistics & Data Analysis 56, 785–796.

Zhang, J., Peng, Y., 2009a. Accelerated hazards mixture cure model. Lifetime Data Analysis 15 (4), 455–467.

Zhang, J., Peng, Y., 2009b. Crossing hazard functions in common survival models. Statistics & Probability Letters 79 (20), 2124–2130.

Zhang, J., Peng, Y., Zhao, O., 2011. A new semiparametric estimation method for accelerated hazard model. Biometrics 67 (4), 1352–1360.

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