Published in final edited form as:

Biometrics. 2018 March ; 74(1): 77–85. doi:10.1111/biom.12727.

Simple and Fast Overidentified Rank Estimation for Right-Censored Length-Biased Data and Backward Recurrence Time

Yifei Sun^{1,*}, Kwun Chuen Gary Chan^{2,**}, and Jing Qin^{3,***}

¹Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A

²Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A

³Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892, U.S.A

Summary

Length-biased survival data subject to right-censoring are often collected from a prevalent cohort. However, informative right censoring induced by the sampling design creates challenges in methodological development. While certain conditioning arguments could circumvent the problem of informative censoring, related rank estimation methods are typically inefficient because the marginal likelihood of the backward recurrence time is not ancillary. Under a semiparametric accelerated failure time model, an overidentified set of log-rank estimating equations is constructed based on the left-truncated right-censored data and backward recurrence time. Efficient combination of the estimating equations is simplified by exploiting an asymptotic independence property between two sets of estimating equations. A fast algorithm is studied for solving non-smooth, non-monotone estimating equations. Simulation studies confirm that the overidentified rank estimator can have a substantially improved estimation efficiency compared to just-identified rank estimators. The proposed method is applied to a dementia study for illustration.

Keywords

Backward and forward recurrence time; Generalized method of moments; Weighted log-rank estimating equation

1. Introduction

The accelerated failure time (AFT) model is an important alternative to Cox's proportional hazards model and is particularly appealing to medical investigators due to its straightforward interpretation. In an ideal situation, prospective follow-up studies are conducted by sampling incident cases over a possibly long period, and the subsequent survival time of interest is usually subject to right censoring. Methods for AFT model for

* ysun26@jhu.edu. ** kcgchan@u.washington.edu. *** jingqin@niaid.nih.gov.

Supplementary Materials

The proof of Lemma 1, Theorem 1, and Theorem 2 referenced in Section 2, and the R program for data analysis are available with this article at the *Biometrics* website on Wiley Online Library.

traditional right-censored survival data has been extensively studied by many authors, see Buckley and James (1979), Tsiatis (1990), Ying (1993) among others. In practice, due to constraints on cost and time, studies on incident cohorts are often unavailable, and data on a prevalent cohort of diseased individuals, who have experienced the disease incidence before recruitment but not the failure event, are collected and analyzed. For example, in the Canadian Study of Health and Aging (CSHA), survival data were collected from a prevalent cohort of dementia patients who were alive at the time of recruitment. In many applications, including the CSHA, it is reasonable to assume that the incidence of disease onset is stable over time, and the survival time in the prevalent cohort is length-biased (Wang, 1991; Asgharian et al., 2002).

Semiparametric estimation of the AFT model for length-biased and right-censored data has been studied by Shen et al. (2009); Ning et al. (2011, 2014a,b). Specifically, Shen et al. (2009) proposed an inverse weighted estimating equation approach with a closed-form expression. Ning et al. (2011) generalized a Buckley–James type of estimator to length-biased and right-censored data. Given the feature that observed failure time data can be transformed to identically and independently distributed random variables without covariate effects, Ning et al. (2014a) proposed a class of estimating equations based on the score functions for the transformed data. Ning et al. (2014b) proposed two rank-based estimators, one based on modified risk-sets, and another based on inverse weighting and ranking. As shown in Ning et al. (2014b), there is no uniformly best estimation method regarding statistical efficiency in the current literature, and the authors provide decision guidelines on how to choose an estimation method only for scenarios with a few symmetric error distributions. Moreover, although well-established statistically, some of the existing approaches may suffer from unstable computational properties. Hence, it is desirable to develop efficient, computationally fast and stable estimation procedures under the AFT model for right-censored length-biased data.

In this article, we introduce a simple and efficient rank-based method for the estimation and inference of the AFT model under length-biased sampling. In addition to the rank-based estimating equations for left-truncated and right-censored data (Lai and Ying, 1991), we construct an additional set of estimating equations based on an induced model of the backward recurrence time. To improve efficiency, the overidentified sets of estimating equations are combined, in the spirit of the generalized method of moments (Hansen, 1982). The estimation and inference are greatly simplified by the fact that the two sets of estimating functions are asymptotically independent, even though they are constructed from correlated survival times. A further advantage of the proposed estimator is that the AFT model can be estimated using only the backward recurrence time data, which means that one can obtain a consistent estimator after recruitment even without follow-up; most of the existing works dealing with semiparametric AFT model under length-biased sampling (Shen et al., 2009; Ning et al., 2011, 2014a,b) require some failure events to be observed and cannot handle this case. Furthermore, a computationally efficient algorithm is given to provide a solution of the estimating equations which are neither continuous nor monotone.

We note that Li and Yin (2009) proposed an overidentified rank estimator for clustered survival data. Our estimator is sufficiently different in a few key aspects. The construction of

overidentified rank estimator of Li and Yin (2009) was motivated by efficiency improvement from multiple working correlation structures, extending the work of Qu et al. (2000) for uncensored data. The survival times, as well as the estimating functions, are correlated in that setting. We consider univariate length-biased survival data but decompose the survival time into two correlated portions to construct overidentified estimating equations, while the two sets of estimating functions are asymptotically independent and can be easily combined by exploiting the independence structure.

The content of the article is organized as follows: In Section 2.1, we introduce the overidentified weighted log-rank estimating equations and propose an efficient combination. To further improve efficiency, we derive and incorporate the optimal weight functions in the estimating equations in Section 2.2. Moreover, in the absence of censoring, we show the proposed estimator with correctly estimated weight function achieves the semiparametric efficiency bound. In Section 3, a fast algorithm for parameter and variance estimation is developed. Simulation studies and an application to a dementia study are presented in Section 4 and 5 for illustration. We conclude with a discussion in Section 6.

2. Estimation

2.1. Over-Identified Estimating Equations

For individuals in the target population, let $\tilde{T}$ denote the time from the disease onset to the failure event of interest, and let $\tilde{X}$ denote a $p \times 1$ vector of covariates. We assume that the survival time in the target population follows the AFT model

$$\log \tilde{T} = \beta^T \tilde{X} + \tilde{\varepsilon}, \quad (1)$$

where β is a $p \times 1$ vector of parameters, and $\tilde{\varepsilon}$ follows an unspecified distribution. We denote by $\tilde{A}$ the time between disease onset and study enrollment, and assume that $\tilde{A}$ is independent of $\tilde{T}$. In a prevalent cohort study, a diseased subject would be qualified to be sampled if the failure event does not occur before the sampling time, that is, $\tilde{T} \geq \tilde{A}$. In other words, $\tilde{T}$ is left truncated by $\tilde{A}$. Denote by T , A , and X the survival time, truncation time, and the covariates for individuals in the prevalent cohort. Then (T, A, X) has the same joint distribution as $(\tilde{T}, \tilde{A}, \tilde{X})$ conditional on $\tilde{T} \geq \tilde{A}$. When prospective follow-up is present, the observation of the survival time in the prevalent cohort is usually subject to right censoring. Instead of the actual value of T , we observe possibly censored survival time $Y = \min(T, A + C)$ and censoring indicator $\delta = \mathbb{I}(T \leq A + C)$. In many applications, it is reasonable to assume that the censoring time after enrollment, C , is independent of (T, A) given X . Note, however, that the survival time T and the total censoring time $A + C$ are typically correlated given X , as they share the same A . Thus the survival time T is subject to informative censoring. We assume that the observed data $\{(Y_i, A_i, X_i, \delta_i), i = 1, \dots, n\}$ are independent and identically distributed replicates of (Y, A, X, δ) .

Let $f(t)$ and $S(t)$ denote the density and survival function of the random variable $\exp(\tilde{\varepsilon})$, and $\mu(x, \beta) = e^{\beta^\top x} \int_0^\infty S(u) du$ be the mean of $\tilde{T}$ given $\tilde{X} = x$. Under length-biased sampling, the observed data likelihood, conditioning on X , is $L = L_C \times L_M$ (Wang, 1991), where we have

$$L_C \propto \prod_{i=1}^n \left[\frac{f(Y_i e^{-\beta^\top X_i}) e^{-\beta^\top X_i}}{S(A_i e^{-\beta^\top X_i})} \right]^{\Delta_i} \left[\frac{S(Y_i e^{-\beta^\top X_i})}{S(A_i e^{-\beta^\top X_i})} \right]^{1-\Delta_i}$$

$$\text{and } L_M = \prod_{i=1}^n \left[\frac{S(A_i e^{-\beta^\top X_i})}{\mu(X_i, \beta)} \right].$$

Based on the conditional likelihood function L_C (i.e., likelihood function of the observed failure time conditioning on truncation time and X), rank estimation for model (1) was proposed by Lai and Ying (1991), treating the data as left-truncated and right-censored. Note that inference based on the conditional likelihood L_C is not fully efficient for length-biased sampling, as evidenced by Vardi (1989), Wang (1991), Asgharian et al. (2002), Shen et al. (2009) among others. The reason is that the marginal likelihood L_M (i.e., likelihood function of the truncation time A given X) contains β and is not ancillary. Therefore, full likelihood inference will be more efficient than conditional likelihood inference. However, even under the simplest case of one-sample estimation, the maximum likelihood estimator based on the full likelihood does not have a closed form expression as discussed in Vardi (1989). Moreover, there is a thorny issue of informative censoring that invalidates risk set methods to be directly extended based on the full likelihood, because T and $A + C$ are correlated given covariates X . In what follows, we propose an estimator that combines information from L_C and L_M to improve efficiency.

To estimate β , weighted log-rank estimating equation was proposed in Lai and Ying (1991) based on inverting a class of linear rank test statistics constructed from L_C . We define $N_i^Y(t, \beta) = I(\log Y_i - \beta^\top X_i \leq t)$ and $R_i^Y(t, \beta) = I(\log A_i - \beta^\top X_i \leq t \leq \log Y_i - \beta^\top X_i)$. Let $\phi_1(t, \beta)$ denote a weight function that possibly depends on data. A system of weighted log-rank estimating functions can be constructed as

$$\Psi_1(\beta) = n^{-1} \sum_{i=1}^n \int_{-\infty}^{\infty} \phi_1(u, \beta) \left\{ X_i - \frac{\sum_{j=1}^n X_j R_j^Y(u, \beta)}{\sum_{j=1}^n R_j^Y(u, \beta)} \right\} dN_i^Y(u, \beta). \quad (2)$$

We denote $\hat{\beta}_{\text{WLR}, 1}$ to be the solution of $\Psi_1(\beta) = o_p(n^{-1/2})$. The right-hand side of the equation may not be identical to 0 because Ψ_1 is discontinuous and the solution is typically defined as the zero-crossing of $\Psi_1(\beta)$.

Since $\Psi_1(\beta)$ is based on L_C , we can improve estimation efficiency by considering L_M , the marginal likelihood of A given X . Under length-biased sampling, we have

$$\begin{aligned} L_M &= \prod_{i=1}^n \frac{S(A_i e^{-\beta^\top X_i})}{\mu(X_i, \beta)} = \prod_{i=1}^n \frac{S(A_i e^{-\beta^\top X_i})}{\int_0^\infty S(A_i e^{-\beta^\top X_i - \beta^\top X_i E(e^{\tilde{\varepsilon}})}) dE(e^{\tilde{\varepsilon}})} \\ &\propto \prod_{i=1}^n \frac{S(A_i e^{-\beta^\top X_i}) A_i e^{-\beta^\top X_i}}{E(e^{\tilde{\varepsilon}})} \\ &\stackrel{\text{def}}{=} \prod_{i=1}^n f_\eta(\log A_i - \beta^\top X_i), \end{aligned}$$

where $f_\eta(u) = S(e^u) e^u / E(e^{\tilde{\varepsilon}})$ is a density function. Thus L_M is equivalent to the likelihood based on the following induced model on the truncation time A :

$$\log A_i = \beta^\top X_i + \eta_i, \quad i = 1, \dots, n \quad (3)$$

where η is a random variable with density function $f_\eta(\cdot)$. Model (3) was first discussed by Yamaguchi (2003), where the author considered parametric AFT models when follow-up is not present. Define, $N_i^A(t, \beta) = I(\log A_i - \beta^\top X_i \leq t)$ and $R_i^A(t, \beta) = I(\log A_i - \beta^\top X_i \geq t)$. Based on the induced model (3), a weighted log-rank estimating function is given by

$$\Psi_2(\beta) = n^{-1} \sum_{i=1}^n \int_{-\infty}^{\infty} \phi_2(u, \beta) \left\{ X_i \frac{\sum_{j=1}^n X_j R_j^A(u, \beta)}{\sum_{j=1}^n R_j^A(u, \beta)} \right\} dN_i^A(u, \beta), \quad (4)$$

where $\phi_2(t, \beta)$ is a weight function that possibly depends on data. We denote $\hat{\beta}_{\text{WLR}, 2}$ to be a solution of $\Psi_2(\beta) = o_p(n^{-1/2})$.

To estimate the parameter β , we have two sets of estimating equations. Combining $\Psi_1(\beta)$ and $\Psi_2(\beta)$ yields an overidentified set of estimating equations for β , and a question arises as for how to combine the estimating equations to attain optimal efficiency. One possible way is the generalized method of moments (GMM) (Hansen, 1982). Define $\Psi(\beta) = (\Psi_1(\beta)^\top, \Psi_2(\beta)^\top)^\top$, and let W be a $2p \times 2p$ positive-definite weight matrix. A consistent estimator of β can be obtained by $\hat{\beta}_{\text{GMM}} = \arg \min_{\beta} \Psi(\beta)^\top W \Psi(\beta)$. Moreover, the optimal matrix W that yields an efficient estimator is the inverse of asymptotic covariance matrix of $\sqrt{n} \Psi(\beta_0)$, where β_0 is the true value of β . The following lemma implies that the optimal weight matrix is a block diagonal matrix.

Lemma 1—Under Assumptions (A1)–(A4) in the Appendix, $\sqrt{n}\Psi_1(\beta_0)$ and $\sqrt{n}\Psi_2(\beta_0)$ are asymptotically independent.

Lemma 1 is a non-trivial result because T and A , the outcomes used to construct Ψ_1 and Ψ_2 , are positively correlated. The proof of Lemma 1 is given in the Supplementary Materials. The independence of estimation functions can also be rationalized from a likelihood perspective. It is easy to see that the β -score functions from conditional likelihood L_C and marginal likelihood L_M are orthogonal. Moreover, by projecting the score functions to the space orthogonal to the nuisance tangent space, the efficient score functions are still orthogonal. Since the weighted log-rank estimating functions are constructed based on the efficient score functions (Ritov and Wellner, 1988), the asymptotic independence of $\sqrt{n}\Psi_1(\beta_0)$ and $\sqrt{n}\Psi_2(\beta_0)$ can be proved.

It can be verified that $\sqrt{n}(\hat{\beta}_{\text{WLR},1} - \beta_0)$ and $\sqrt{n}(\hat{\beta}_{\text{WLR},2} - \beta_0)$ are asymptotically normal with covariance-variance matrices V_1 and V_2 (V_1, V_2 are given in the Supplementary Materials). By applying Lemma 1, the optimal GMM type estimator has asymptotic variance $(V_1^{-1} + V_2^{-1})^{-1}$. However, the computation of $\hat{\beta}_{\text{GMM}}$ requires to minimize a quadratic form, which can be computationally intensive, particularly because $\Psi(\beta)$ is neither continuous nor monotone.

Based on Lemma 1, we can construct a simpler estimator that is asymptotically equivalent to the optimal GMM estimator. It is shown in the Supplementary Materials that $\sqrt{n}(\hat{\beta}_{\text{WLR},1} - \beta_0)$ and $\sqrt{n}(\hat{\beta}_{\text{WLR},2} - \beta_0)$ are asymptotically orthogonal. This suggests us to consider a linearly weighted estimator, $(V_1^{-1} + V_2^{-1})^{-1}(V_1^{-1}\hat{\beta}_{\text{WLR},1} + V_2^{-1}\hat{\beta}_{\text{WLR},2})$, whose asymptotic variance equals that of the optimal GMM estimator. In practice, V_1 and V_2 are usually unknown and need to be estimated. Suppose $(\hat{V}_1, \hat{V}_2)$ are consistent estimators of (V_1, V_2) , we propose to use the following weighted estimator,

$$\hat{\beta}_W = (\hat{V}_1^{-1} + \hat{V}_2^{-1})^{-1}(\hat{V}_1^{-1}\hat{\beta}_{\text{WLR},1} + \hat{V}_2^{-1}\hat{\beta}_{\text{WLR},2}).$$

A detailed computation procedure to obtain $(\hat{\beta}_{\text{WLR},1}, \hat{\beta}_{\text{WLR},2})$ and $(\hat{V}_1, \hat{V}_2)$ is given in Section 3. Let β_0 be the true regression coefficient, Theorem 1 summarizes the asymptotic properties of $\hat{\beta}_W$, with a proof given in the Supplementary Materials.

Theorem 1—Under assumptions (A1)–(A5) in the Appendix, $\sqrt{n}(\hat{\beta}_W - \beta_0)$ converges weakly to a zero mean normal random vector with covariance matrix $(V_1^{-1} + V_2^{-1})^{-1}$.

From Theorem 1, the proposed estimator $\hat{\beta}_W$ is more efficient than the estimators using just identified estimating equations, because $V_1 \geq (V_1^{-1} + V_2^{-1})^{-1}$ and $V_2 \geq (V_1^{-1} + V_2^{-1})^{-1}$, where $V \geq U$ if $V - U$ is positive semi-definite for matrix V, U .

The above discussion and theoretical results are based on unspecified weight functions $\phi_1(\cdot, \beta)$ and $\phi_2(\cdot, \beta)$. For instance, setting $\phi_1(\cdot, \beta) = \phi_2(\cdot, \beta) = 1$ yields the log-rank estimating equations. Moreover, because Model (3) is the standard semi-parametric linear regression model, a natural choice to estimate β is the least square estimator $\hat{\beta}_{LS}$, defined as the solution of the following estimating equation,

$$\Psi_{LS}(\beta) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(\log A_i - X_i \beta) = 0,$$

where $\bar{X} = \sum_{i=1}^n X_i/n$. By setting $\phi_2(t, \beta) = t - \frac{\int_t^\infty u dF_\eta(u)}{1 - F_\eta(t)}$, we have

$\sqrt{n}\Psi_{LS}(\beta_0) = \sqrt{n}\Psi_2(\beta_0) + o_p(1)$, where F_η is the cumulative distribution function of η (Ritov, 1990). Therefore the asymptotically independence result of $\sqrt{n}\Psi_1(\beta_0)$ and $\sqrt{n}\Psi_{LS}(\beta_0)$ also holds, and one can linearly combine $\hat{\beta}_{1, \text{WLR}}$ and $\hat{\beta}_{LS}$ to improve efficiency. Without additional assumptions, it is not clear whether $\hat{\beta}_{2, \text{WLR}}$ is more efficient than $\hat{\beta}_{LS}$. Although rank estimation in (4) is not the standard way to handle uncensored data, it is used because of the independence property that leads to a simple combined estimator. In Section 2.2, we explore the weight functions $\phi_1(t, \beta)$ and $\phi_2(t, \beta)$, so that $\hat{\beta}_{2, \text{WLR}}$ could be more efficient than $\hat{\beta}_{LS}$ with properly chosen weight functions.

2.2. Efficient Adaptive Rank Estimators

To further improve the efficiency, we derive the optimal weight functions $\phi_1(\cdot, \beta)$ and $\phi_2(\cdot, \beta)$ for the two sets of estimating equations. Define $\phi_1^0(u, \beta)$ to be the limit of $\phi_1(u, \beta)$ as $n \rightarrow \infty$, and let $\lambda_{\tilde{\varepsilon}}(\cdot)$ denote the hazard function of $\tilde{\varepsilon}$. For the first set of estimating function $\Psi_1(\beta)$, it is shown that random vector $\sqrt{n}(\hat{\beta}_{\text{WLR}, 1} - \beta_0)$ is asymptotically normal with covariance matrix $V_1 = \Gamma_1(\beta_0)^{-1} \sum_1(\beta_0) \Gamma_1(\beta_0)^{-1}$, where

$$\Gamma_1(\beta_0) = E \int_{-\infty}^{\infty} \phi_1^0(u, \beta_0) \frac{\dot{\lambda}_{\tilde{\varepsilon}}(u)}{\lambda_{\tilde{\varepsilon}}(u)} \left[X - \frac{E\{R^Y(u, \beta_0)X\}}{E\{R^Y(u, \beta_0)\}} \right]^{\otimes 2} dN^Y(u, \beta_0),$$

and

$$\sum_1(\beta_0) = E \int_{-\infty}^{\infty} \phi_1^0(u, \beta_0)^2 \left[X - \frac{E\{R^Y(u, \beta_0)X\}}{E\{R^Y(u, \beta_0)\}} \right]^{\otimes 2} dN^Y(u, \beta_0).$$

By Cauchy–Schwartz inequality, the optimal weight is

$$\phi_1^{\text{opt}}(u) = \dot{\lambda}_{\tilde{\varepsilon}}(u)/\lambda_{\tilde{\varepsilon}}(u) = e^u \dot{\lambda}(e^u)/\lambda(e^u) + 1, \quad (5)$$

where $\dot{\lambda}(u) = d\lambda(u)/du$ and $\dot{\lambda}_{\tilde{\varepsilon}}(u) = d\lambda_{\tilde{\varepsilon}}(u)/du$. Similarly, for $\Psi_2(\beta)$, let λ_{η} be the hazard function of η with $\dot{\lambda}_{\eta}(u) = d\lambda_{\eta}(u)/du$, then the optimal weight function is

$$\phi_2^{\text{opt}}(u) = \dot{\lambda}_{\eta}(u)/\lambda_{\eta}(u) = -\lambda(e^u)e^u + 1 + \frac{S(e^u)e^u}{\int_u^{\infty} S(e^x)e^x dx}. \quad (6)$$

There are a few options to estimate the weight functions $\phi_1^{\text{opt}}(\cdot)$ and $\phi_2^{\text{opt}}(\cdot)$: for example, kernel smoothing techniques have been applied in Lai and Ying (1991) and Lin and Chen (2013). However, substituting such nonparametric type smoothing estimators into equations (2) and (4) could lead to estimators for β that perform poorly with moderate sample sizes, due to the unstableness of the kernel estimators. As an alternative, we can assume a flexible working parametric model for $\tilde{\varepsilon}$. For instance, $e^{\tilde{\varepsilon}}$ can be assumed to follow the generalized gamma distribution (Cox et al., 2007), which is an extensive family that contains nearly all of the most commonly-used survival distributions. Then the unknown parameter involved in the distribution of $\tilde{\varepsilon}$ can be estimated through score equation of the conditional likelihood using rescaled survival times. Even in the case where the working model is mis-specified, the proposed estimator is consistent and asymptotically normal.

In the absence of censoring, if the error term $\tilde{\varepsilon}$ follows the working model distribution, the combined estimator with consistently estimated optimal weights achieves the

semiparametric efficiency bound. Define $M_1(t, \beta) = N^Y(t, \beta) - \int_{-\infty}^t R^Y(u, \beta) \lambda_{\tilde{\varepsilon}}(u) du$ and

$M_2(t, \beta) = N^A(t, \beta) - \int_{-\infty}^t R^A(u, \beta) \lambda_{\eta}(u) du$. Theorem 2 states the efficiency score of the AFT

model with length-biased survival data, and the proof is given in the Supplementary Materials.

Theorem 2—In the absence of censoring, the efficient score of model (1) with length-biased data $\{(A_i, T_i, X_i), i = 1, \dots, n\}$ is

$$S_{\text{eff}}(A, T, X) = \int_{-\infty}^{\infty} \frac{\dot{\lambda}_{\tilde{\varepsilon}}(u)}{\lambda_{\tilde{\varepsilon}}(u)} \{X - E(X)\} dM_1(u, \beta_0) \\ + \int_{-\infty}^{\infty} \frac{\dot{\lambda}_{\eta}(u)}{\lambda_{\eta}(u)} \{X - E(X)\} dM_2(u, \beta_0).$$

Remark 1—When the optimal weight function is correctly estimated, $\hat{\beta}_W$ is asymptotically equivalent to $\hat{\beta}_S$, which is the solution of $\Psi_1(\beta) + \Psi_2(\beta) = o_p(n^{-1/2})$. However, when the

user-specified weight function is different from the optimal choice, then $\hat{\beta}_W$ is asymptotically more efficient than $\hat{\beta}_S$ in general.

Remark 2—The following induced models hold in the absence of censoring,

$$\log T_i = X_i\beta + \varepsilon_i, \quad (7)$$

$$\log A_i = X_i\beta + \eta_i, \quad (8)$$

where the joint density function of (ε, η) is $f_{(\varepsilon, \eta)}(u, v) = f(e^u)e^{u+v}/E(e^{\tilde{\varepsilon}})$ for $u < v$. Model (7) has been studied in Chen (2010) and Mandel and Ritov (2010). In this case, T_i 's are sufficient for estimating β , and only (7) is needed for estimation. Moreover, it can be shown that our proposed estimator, with consistently estimated optimal weight, is asymptotically equivalent to the efficient estimator based on marginal likelihood of model (7). However, the rank estimator of Chen (2010) cannot handle length-biased right-censored data because of induced informative censoring. To improve efficiency in the presence of right censoring, we need to consider (7) and (8) jointly.

Remark 3—It has been shown in Ritov and Wellner (1988) that the efficient score function for model (3) is

$$\int_{-\infty}^{\infty} \frac{\dot{\lambda}_{\eta}(t)}{\bar{\lambda}_{\eta}(t)} \{X - EX\} dM_2(t),$$

where $M_2(t) = I(A \leq t) - \int_0^t (A \geq t)\lambda_{\eta}(t)dt$, and the efficiency bound is

$$I_2 = \int_{-\infty}^{\infty} \left[\frac{\dot{f}_{\eta}(t)}{f_{\eta}(t)} \right]^2 f_{\eta}(t) dt \cdot \text{Cov}(X).$$

When the weight function ϕ_2^{opt} is consistently estimated, the estimator $\hat{\beta}_{\text{WLR}, 2}$ will achieve the semi-parametric efficiency bound I_2 , and thus asymptotically will be more efficient than the least square estimate $\hat{\beta}_{LS}$.

3. Fast Computation

The computation of rank estimators is typically challenging, because the weighted log-rank estimating equation is usually neither continuous nor monotone, and it may have inconsistent roots in addition to a consistent root (Fygenson and Ritov, 1994). In such cases, the estimator needs to be defined in a shrinking neighborhood of the true value β_0 , and

iterative methods require a consistent initial value. However, finding a consistent initial estimate is usually as computationally challenging as directly finding the root of the estimating equation. This computational challenge is a major obstacle for applying the rank estimation techniques in practice even for the standard right-censored data. In what follows, a computationally simple approach is given for computing $\hat{\beta}_{\text{WLR},1}$ by borrowing strength from two algorithms proposed by Huang (2002) and Huang (2013). A parallel argument applies to $\hat{\beta}_{\text{WLR},2}$ and is thus omitted.

Although methodologies for length-biased and right-censored data is usually thought as more complicated than that for right-censored data, a rather surprising fact is that a simple consistent initial estimator of β can be obtained from the induced model (3). Specifically, based on model (7) and Yamaguchi (2003), the least square estimate $\hat{\beta}_{\text{LS}}$ by regressing the backward recurrent time $\log A$ against X is a $\sqrt{n}$ -consistent estimate of β and thus can serve as an initial value for an iterative algorithm.

To compute $\hat{\beta}_{\text{WLR},1}$, we consider a modified Newton's method, following the arguments of Huang (2013). Under regularity conditions (A1)–(A5) in the Appendix, an asymptotic local linearity condition holds. Specifically, let $\|\cdot\|$ denote the Euclidean norm, for every sequence $d_n > 0$ and d_n converges to 0 in probability,

$$\sup_{\beta: \|\beta - \beta_0\| \leq d_n} \frac{\|\Psi_1(\beta) - \Psi_1(\beta_0) - \hat{\Gamma}_1(\beta - \beta_0)\|}{n^{-1/2} + \|\beta - \beta_0\|} = o_p(1), \quad (9)$$

where $\hat{\Gamma}_1$ is a consistent estimate of matrix $\Gamma_1(\beta_0)$, the derivative at β_0 of the limiting $\Psi_1(\beta)$ when $n \rightarrow \infty$. Based on (9), a Newton-type algorithm can be made iteratively,

$$\hat{\beta}^{(k)} = \hat{\beta}^{(k-1)} - \hat{\Gamma}_1 \Psi_1(\hat{\beta}^{(k-1)}), \quad k \geq 1 \quad (10)$$

where $\hat{\beta}^{(0)} = \hat{\beta}_{\text{LS}}$. Since $\hat{\beta}^{(0)}$ is a $\sqrt{n}$ -consistent estimate of β_0 , it can be shown that the one-step estimator $\hat{\beta}^{(1)}$ satisfies $\sqrt{n}\Psi_1(\hat{\beta}^{(1)}) = o_p(1)$. Moreover, to avoid the problem of over-shooting, we halve the step size repeatedly until the new estimate leads to a decrease in the quadratic score $\Psi_1(\beta)^\top \hat{\Sigma}_1(\beta)^{-1} \Psi_1(\beta)$, where $\hat{\Sigma}_1(\beta)$ is defined as

$$\hat{\Sigma}_1(\beta) = n^{-1} \sum_{i=1}^n \int_{-\infty}^{\infty} \phi_1^2(u, \beta) \left[X_i - \frac{\sum_{j=1}^n X_j R_j^Y(u, \beta)}{\sum_{j=1}^n R_j^Y(u, \beta)} \right]^{\otimes 2} dN_i^Y(u, \beta).$$

In order to apply the algorithm in (10), a consistent estimate of $\Gamma_1(\beta_0)$ is needed. Note that for a $p \times 1$ vector h , we have

$$\Psi_1(\hat{\beta}^{(0)} + n^{-1/2}h) - \Psi_1(\hat{\beta}^{(0)}) = n^{-1/2}\Gamma_1(\beta_0)h + o_p(n^{-1/2}). \quad (11)$$

Let H_1 be a $p \times p$ non-singular matrix with $\|H_1\|_{\max} = O_p(1)$ and $\|H_1^{-1}\|_{\max} = O_p(1)$, where $\|\cdot\|_{\max}$ denotes the maximum absolute value of the matrix elements. Let $h_{11}, \dots, h_{1p}$ be the column vectors of H_1 , that is, $H_1 = (h_{11}, \dots, h_{1p})$. Define the matrix $A_1 = \sqrt{n}\{\Psi_1(\hat{\beta}^{(0)} + n^{-1/2}h_{11}) - \Psi_1(\hat{\beta}^{(0)}), \dots, \Psi_1(\hat{\beta}^{(0)} + n^{-1/2}h_{1p}) - \Psi_1(\hat{\beta}^{(0)})\}$, it follows from (11) that $A_1H_1^{-1}$ is a consistent estimate of $\Gamma_1(\beta_0)$, thus we estimate $\Gamma_1(\beta_0)$ by

$$\hat{\Gamma}_1 = A_1H_1^{-1}.$$

One possible choice of $n^{-1/2}H_1$ is the Cholesky factorization of the estimated covariance matrix of $\hat{\beta}^{(0)}$. Given $\hat{\Gamma}_1$, $\hat{\beta}_{\text{WLR},1}$ can be obtained by the Newton type algorithm in (10). Moreover, the asymptotic variance estimate of $\sqrt{n}(\hat{\beta}_{\text{WLR},1} - \beta_0)$ is readily available as

$$\hat{V}_1 = \hat{\Gamma}_1^{-1} \sum_1 (\hat{\beta}_{\text{WLR},1})(\hat{\Gamma}_1^{-1})^{-1}, \quad (12)$$

which converges in probability to V_1 . The variance estimation is simpler than many other existing methods that either require kernel smoothing or resampling (Tsiatis, 1990; Parzen et al., 1994; Jin et al., 2003).

The above algorithm is similar in flavor to the algorithm in Huang (2002), but with certain important differences. The algorithm of Huang (2002) approximates the inverse of estimating function, which requires solution-finding and may be computationally intensive. Moreover, due to the lack of a consistent initial estimate, Huang (2002) uses a recursive bisection algorithm. Our algorithm is also similar to the algorithm in Huang (2013), which requires an initial value obtained from a censored quantile regression model (Huang, 2010). Our problem structure permits us to use a least square estimate as the initial estimation, which is much simpler. Also, the method of Huang (2013) may not be readily used for finding the solution of $\Psi_1(\beta) = o_p(n^{-1/2})$, since it is unclear how a computationally simple and consistent initial value is obtained from censored quantile regression for left-truncated and right-censored data.

4. Simulations

Simulation studies are conducted to examine the finite-sample performance of the proposed inference procedures. We generate failure times from the following model

$$\log \tilde{T} = \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \tilde{\varepsilon}$$

where $\tilde{X}_1$ is generated from a Bernoulli distribution with success probability 0.5, and $\tilde{X}_2$ is a continuous variable from the uniform distribution on $[0,1]$. We set $\beta_1 = 0.5$ and $\beta_2 = 1$. The error distribution were generated from (i) $e^{\tilde{\epsilon}}$ follows Weibull distribution with shape parameter 2, scale parameter 0.5; (ii) $\tilde{\epsilon}$ follows extreme value distribution with scale parameter 0.2; (iii) $e^{\tilde{\epsilon}}$ follows gamma distribution with mean one and variance 0.25; and (iv) $\tilde{\epsilon}$ follows normal distribution with mean zero and variance 1/12. The truncation times and residual censoring times were generated in the original time scale (not log-scale). Specifically, the truncation times were generated from a uniform distribution with a large enough upper bound to ensure the stationarity assumption, and we kept only the pairs satisfying $\tilde{A} < \tilde{T}$. The residual censoring times, C , were independently generated from a uniform distribution over $[0,c]$, where c was chosen to yield the censoring percentage of 0, 25, and 50%. For each specified set of parameters, sample size of 200 and 800 are chosen, and each scenario was repeated 1000 times. The results are summarized in Tables 1 and 2. We denote the proposed estimator with log-rank weight by $\hat{\beta}_W^{lr}$ and the proposed estimator with estimated optimal weight using generalized gamma family as the working model by $\hat{\beta}_W^{opt}$. We compare our estimators with the estimator $\hat{\beta}_{LT}$ by solving log-rank estimation equation for left-truncated and right-censored data, and the weighted log-rank estimator $\hat{\beta}_M$ based on the marginal likelihood with estimated ϕ_2 using the working model. We also present the results of parametric maximum likelihood estimator by assuming $\tilde{\epsilon}$ follows generalized gamma distribution ($\hat{\beta}_{MLE}$) and normal distribution ($\hat{\beta}_{MLE}^{normal}$).

It can be seen from the table that all the estimators perform well in finite sample studies, and the proposed estimators substantially outperform $\hat{\beta}_{LT}$ and $\hat{\beta}_M$ in all the scenarios. In Scenario (i)–(iii), the distributions of $e^{\tilde{\epsilon}}$ belong to generalized gamma family, and $\hat{\beta}_W^{opt}$ has similar standard error as $\hat{\beta}_W^{lr}$. Note that $\phi_1^{opt} \equiv 1$ in Scenario (i) and (ii). In Scenario (iv), general gamma distribution approaches normal distribution (Cox et al., 2007), and $\hat{\beta}_W^{opt}$ have smaller standard error than $\hat{\beta}_W^{lr}$. The improvement of our estimator is mainly due to combination of the two sets of estimating equations, and improvement from estimating the optimal weight function is less notable. When the parametric model is correctly specified, the MLE is slightly more efficient than the proposed estimators; however, MLE can be less efficient when the parametric model is wrongly specified, for example, $\hat{\beta}_{MLE}^{normal}$ has relatively large variance in Scenario (i)–(iii).

5. Data Analysis

We illustrate the proposed estimation procedure by analyzing the CSHA data. As discussed in Wolfson et al. (2001), the CSHA was a prevalent cohort where the survival data were collected from a cohort of dementia patients at recruitment. Thus, patients who died before the recruitment period were not qualified to enter the cohort. CSHA recruited a prevalent cohort of individuals aged 65 and older with dementia during the period between February

1991 and May 1992. The survival time of interest is the time from onset to death, and the truncation time in the prevalent cohort is the duration from the onset of dementia to study enrollment. The goal of our analysis is to estimate the relative survival following the onset of dementia among subcategories of dementia, which is an important scientific question studied by Mölsä et al. (1986) and Roberson et al. (2005). We considered a subset of the study data by excluding those with missing date of onset or classification of dementia subtype. Moreover, as in Wolfson et al. (2001), patients with observed survival time of 20 or more years were excluded because these subjects are considered unlikely to have Alzheimers disease or vascular dementia. A total of 807 subjects were analyzed; among them, 249 were diagnosed with possible Alzheimers disease, 388 had probable Alzheimers disease, and 170 had vascular dementia. The observation of the residual survival time after recruitment is censored by end of the follow up period. The constant disease incidence assumption was checked in Huang and Qin (2012) with the Kolmogorov–Smirnov test, based on the fact that under mild conditions, the truncation time A and the residual lifetime after enrollment $T - A$ have identical distributions if and only if the incidence of disease is constant over time (Asgharian et al., 2006). The applicability of the AFT time to the application was checked using QQ-plots Ning et al. (2011).

We consider the following AFT model,

$$\log(\tilde{T}) = \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \tilde{\varepsilon},$$

where $\tilde{X}_1$ and $\tilde{X}_2$ are binary variables that indicate whether the patients is probable Alzheimer and vascular dementia, respectively. The proposed estimator of β_1 is -0.107 , with a 95% confidence interval $(-0.216, -0.001)$, and β_2 is -0.166 , with a 95% confidence $(-0.289, -0.044)$. Our analysis suggests that the survival time for probable Alzheimer and vascular dementia patients are significantly shorter than that of the possible Alzheimer patients. For comparison, we also applied the two rank-based methods in Ning et al. (2014b). Using the first method in Ning et al. (2014b), based on modified risk sets, the estimated β_1 is -0.138 (CI: $-0.361, 0.085$) and β_2 is -0.152 (CI: $-0.375, 0.071$). Using their second method based on inverse weighting and ranking, the estimated β_1 is -0.102 (CI: $-0.214, 0.010$) and β_2 is -0.156 (CI: $-0.319, 0.007$). All estimators have similar point estimates, but our proposed estimator has the smallest standard error estimates and detect significant effects of probable Alzheimer and vascular dementia on survival time.

6. Discussion

In this article, we propose an estimator to efficiently combine overidentified sets of estimating equations resulting from the follow-up data as well as the backward recurrence time data for a length-biased prevalent cohort. The proposed estimator is simple to implement, but is asymptotically equivalent to the optimal GMM estimator. A computationally fast and stable procedure is also presented for estimation and inference.

Rank-based estimating equation can be regarded as the inversion of weighted log-rank statistics. In our case, the estimating equations can be regarded as the inversion of the log-

rank test of Ying (1990) for left-truncated and right-censored data and the log-rank test of Chan and Qin (2015) for backward recurrence data. However, in terms of estimation, the proposed method for estimating regression parameter is much simpler than directly inverting the combined log-rank test of Chan and Qin (2015).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the editor, an associate editor, and a reviewer for their helpful comments that greatly improve the article. The first and second authors are partially supported by US National Institutes of Health grant R01-HL122212.

Appendix A

We adopt the following regularity conditions:

- (A1) The random variable $\tilde{\varepsilon}$ has a bounded density function with bounded derivative.
- (A2) The censoring time C is independent of T conditioning on the truncation time A and covariates X . The density function of C is bounded.
- (A3) The vector of covariates X is bounded.
- (A4) Denote the compact parameter space by $\mathcal{B}$, with $\beta_0 \in \mathcal{B}$. The nonnegative weight functions $\phi_1(t, \beta)$ and $\phi_2(t, \beta)$ have bounded variation and converges almost surely to $\phi_1^0(t, \beta)$ and $\phi_2^0(t, \beta)$ uniformly for $\beta \in \mathcal{B}$ respectively. Let $\|\cdot\|_0$ denote the supremum norm in a neighborhood $\mathcal{B}_0 \subset \mathcal{B}$ of β , we assume $\|\phi_1(t, \beta) - \phi_1^0(t, \beta)\|_0 = O_p(n^{-1/2})$ and $\|\phi_2(t, \beta) - \phi_2^0(t, \beta)\|_0 = O_p(n^{-1/2})$. Furthermore, $\phi_1^0(t, \beta)$ and $\phi_2^0(t, \beta)$ are differentiable in β , and the derivatives are continuous and uniformly bounded for $t \in (-\infty, \infty)$ and $\beta \in \mathcal{B}$.
- (A5) The matrices $\Gamma_1(\beta_0)$ and $\Gamma_2(\beta_0)$ are nonsingular, where,

$$\Gamma_1(\beta) = E \int_{-\infty}^{\infty} \phi_1^0(u, \beta) \frac{\dot{\lambda}_{\tilde{\varepsilon}}(u)}{\lambda_{\tilde{\varepsilon}}(u)} \left[X - \frac{E\{R^Y(u, \beta)X\}}{E\{R^Y(u, \beta)\}} \right]^{\otimes 2} dN^Y(u, \beta),$$

and

$$\Gamma_2(\beta) = E \int_{-\infty}^{\infty} \phi_2^0(u, \beta) \frac{\dot{\lambda}_{\eta}(u)}{\lambda_{\eta}(u)} \left[X - \frac{E\{R^A(u, \beta)X\}}{E\{R^A(u, \beta)\}} \right]^{\otimes 2} dN^A(u, \beta)$$

References

- Asgharian M, M'Lan CE, Wolfson DB. Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association*. 2002; 97:201–209.
- Asgharian M, Wolfson DB, Zhang X. Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in medicine*. 2006; 25:1751–1767. [PubMed: 16220462]
- Buckley J, James I. Linear regression with censored data. *Biometrika*. 1979; 66:429–436.
- Chan KCG, Qin J. Rank-based testing of equal survivorship based on cross-sectional survival data with or without prospective follow-up. *Biostatistics*. 2015; 16:772–784. [PubMed: 25813647]
- Chen YQ. Semiparametric regression in size-biased sampling. *Biometrics*. 2010; 66:149–158. [PubMed: 19432792]
- Cox C, Chu H, Schneider MF, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*. 2007; 26:4352–4374. [PubMed: 17342754]
- Fygenon M, Ritov Y. Monotone estimating equations for censored data. *The Annals of Statistics*. 1994; 22:732–746.
- Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*. 1982; 50:1029–1054.
- Huang CY, Qin J. Composite partial likelihood estimation under length-biased sampling, with application to a prevalent cohort study of dementia. *Journal of the American Statistical Association*. 2012; 107:946–957. [PubMed: 24000265]
- Huang Y. Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association*. 2002; 97:318–327.
- Huang Y. Quantile calculus and censored regression. *Annals of Statistics*. 2010; 38:1607–1637. [PubMed: 20592942]
- Huang Y. Fast censored linear regression. *Scandinavian Journal of Statistics*. 2013; 40:789–806.
- Jin Z, Lin D, Wei L, Ying Z. Rank-based inference for the accelerated failure time model. *Biometrika*. 2003; 90:341–353.
- Lai TL, Ying Z. Rank regression methods for left-truncated and right-censored data. *The Annals of Statistics*. 1991:531–556.
- Li H, Yin G. Generalized method of moments estimation for linear regression with clustered failure time data. *Biometrika*. 2009; 96:293–306.
- Lin Y, Chen K. Efficient estimation of the censored linear regression model. *Biometrika*. 2013; 100:525–530.
- Mandel M, Ritov Y. The accelerated failure time model under biased sampling. *Biometrics*. 2010; 66:1306–1308. [PubMed: 19995351]
- Mölsä PK, Marttila R, Rinne U. Survival and cause of death in alzheimer's disease and multi-infarct dementia. *Acta Neurologica Scandinavica*. 1986; 74:103–107. [PubMed: 3776457]
- Ning J, Qin J, Shen Y. Buckley-james-type estimator with right-censored and length-biased data. *Biometrics*. 2011; 67:1369–1378. [PubMed: 21385160]
- Ning J, Qin J, Shen Y. Score estimating equations from embedded likelihood functions under accelerated failure time model. *Journal of the American Statistical Association*. 2014a; 109:1625–1635. [PubMed: 25663727]
- Ning J, Qin J, Shen Y. Semiparametric accelerated failure time model for length-biased data with application to dementia study. *Statistica Sinica*. 2014b; 24:313–333. [PubMed: 24478570]
- Parzen M, Wei L, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika*. 1994; 81:341–350.
- Qu A, Lindsay BG, Li B. Improving generalised estimating equations using quadratic inference functions. *Biometrika*. 2000; 87:823–836.
- Ritov Y. Estimation in a linear regression model with censored data. *The Annals of Statistics*. 1990:303–328.
- Ritov Y, Wellner JA. Censoring, martingales, and the cox model. *Contemporary Mathematics*. 1988; 80:191–219.

- Roberson E, Hesse J, Rose K, Slama H, Johnson J, Yaffe K, et al. Frontotemporal dementia progresses to death faster than alzheimer disease. *Neurology*. 2005; 65:719–725. [PubMed: 16157905]
- Shen Y, Ning J, Qin J. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*. 2009; 104:1192–1202. [PubMed: 21057599]
- Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*. 1990; 18:354–372.
- Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*. 1989; 76:751–761.
- Wang MC. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*. 1991; 86:130–143.
- Wolfson C, Wolfson DB, Asgharian M, M'Lan CE, Østbye T, Rockwood K, Hogan DF. A reevaluation of the duration of survival after the onset of dementia. *New England Journal of Medicine*. 2001; 344:1111–1116. [PubMed: 11297701]
- Yamaguchi K. Accelerated failure–time mover–stayer regression models for the analysis of last-episode data. *Sociological Methodology*. 2003; 33:81–110.
- Ying Z. Linear rank statistics for truncated data. *Biometrika*. 1990; 77:909–914.
- Ying Z. A large sample study of rank estimation for censored regression data. *The Annals of Statistics*. 1993; 21:76–99.

Table 1

Simulation summary statistics (n = 200)

$\hat{\beta}_W^{\text{opt}}$				$\hat{\beta}_W^{\text{lr}}$				$\hat{\beta}_{LT}$			
Cen	Bias	SE	SEE	RE	Bias	SE	SEE	RE	Bias	SE	
Scenario I											
0	(0,-1)	(63,107)	(59,104)	(69,72)	(-1,-2)	(63,107)	(60,104)	(69,72)	(-1,-4)	(76,126)	
25	(0,-2)	(68,118)	(65,114)	(64,65)	(0,-3)	(68,118)	(66,115)	(64,65)	(1,2)	(85,146)	
50	(-2,-11)	(75,133)	(73,127)	(51,50)	(-2,-11)	(74,131)	(74,128)	(50,48)	(2,-4)	(105,189)	
Scenario II											
0	(1,-3)	(28,49)	(27,47)	(86,89)	(1,-1)	(30,50)	(28,49)	(99,92)	(-3,-1)	(30,52)	
25	(2,-1)	(30,52)	(29,51)	(94,86)	(1,1)	(30,51)	(30,52)	(94,83)	(1,3)	(31,56)	
50	(0,0)	(34,58)	(33,57)	(84,79)	(-1,0)	(34,59)	(34,59)	(84,82)	(-2,3)	(37,65)	
Scenario III											
0	(1,2)	(68,119)	(63,108)	(66,68)	(2,2)	(69,122)	(66,114)	(67,72)	(0,2)	(84,144)	
25	(-1,-2)	(73,124)	(69,117)	(63,63)	(0,2)	(73,126)	(71,123)	(63,65)	(-2,-7)	(92,156)	
50	(2,3)	(80,146)	(75,129)	(53,56)	(2,3)	(81,147)	(77,134)	(55,57)	(6,8)	(110,195)	
Scenario IV											
0	(0,2)	(42,73)	(39,63)	(63,66)	(0,1)	(46,79)	(45,78)	(75,77)	(0,3)	(53,90)	
25	(-1,0)	(47,82)	(42,69)	(62,66)	(0,0)	(51,88)	(49,85)	(72,76)	(1,2)	(60,101)	
50	(-1,-8)	(53,89)	(47,77)	(64,56)	(-2,-8)	(58,95)	(54,95)	(77,64)	(3,1)	(66,119)	
$\hat{\beta}_M$				$\hat{\beta}_{MLE}$				$\hat{\beta}_{MLE}^{\text{norm}}$			
Cen	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE		
Scenario I											
0	(3,6)	(106,185)	(195,216)	(0,-3)	(60,105)	(62,69)	(-2,-9)	(72,130)	(90,107)		
25	(2,-5)	(107,186)	(158,162)	(-1,-2)	(65,114)	(58,61)	(-1,-7)	(78,136)	(84,87)		
50	(0,-3)	(109,185)	(108,96)	(0,-8)	(71,125)	(46,44)	(-5,-23)	(83,145)	(63,60)		
Scenario II											
0	(0,8)	(71,122)	(555,553)	(1,1)	(28,47)	(86,82)	(1,-4)	(33,61)	(120,138)		
25	(6,0)	(71,123)	(528,481)	(1,0)	(30,52)	(94,86)	(0,-3)	(39,74)	(158,174)		

Cen	$\hat{\beta}_M$			$\hat{\beta}_{MLE}$			$\hat{\beta}_{MLE}^{norm}$		
	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE
Scenario III									
50	(2.3)	(70,122)	(357,352)	(1,2)	(35,62)	(89,91)	(-1,-3)	(42,73)	(129,126)
Scenario IV									
0	(0,2)	(112,194)	(178,181)	(-5,-15)	(66,110)	(62,59)	(0,-2)	(69,121)	(67,71)
25	(-2,-7)	(113,195)	(150,156)	(2,5)	(70,127)	(58,66)	(-3,-5)	(74,125)	(65,64)
50	(6,8)	(110,201)	(100,106)	(3,4)	(77,140)	(49,52)	(-1,-4)	(79,142)	(51,53)
Scenario IV									
0	(0,-11)	(85,162)	(257,325)	(2,-5)	(42,72)	(63,64)	(1,-6)	(43,76)	(66,72)
25	(1,7)	(90,164)	(225,264)	(-2,-1)	(45,81)	(56,64)	(-2,-1)	(45,83)	(56,68)
50	(7,3)	(95,158)	(207,177)	(3,1)	(50,85)	(57,51)	(3,-2)	(51,89)	(60,56)

Note: Cen is the censoring rate (%); Bias is the empirical bias ($\times 1000$); SE is the empirical standard error ($\times 1000$); SEE is the empirical mean of the standard error estimates ($\times 1000$); RE is the relative efficiency ($\times 100$) compared to $\hat{\beta}_{LT} \cdot \hat{\beta}_W^{opt}$; $\hat{\beta}_{LT}$ is the combined estimator with estimated weight function as in Section 2.2; $\hat{\beta}_W^{lr}$ is the combined estimator with $\phi_1 = \phi_2 = 1$; $\hat{\beta}_{LT}$ is the estimator from log-rank estimating equations based on LC ; $\hat{\beta}_M$ is the rank-based estimator based on LM with estimated ϕ_2 by assuming $\tilde{\mathcal{E}}$ follows a generalized gamma distribution; $\hat{\beta}_{MLE}$ and $\hat{\beta}_{MLE}^{normal}$ are the parametric maximum likelihood estimators assuming generalized gamma and normal distribution for $\tilde{\mathcal{E}}$. RE of $\hat{\beta}_{LT}$ is 100 and is omitted in the table.

Table 2

Simulation summary statistics (n = 800)

Cen	$\hat{\beta}_W^{\text{opt}}$			$\hat{\beta}_W^{\text{lr}}$			$\hat{\beta}_{LT}$		
	Bias	SE	SEE	RE	Bias	SE	SEE	RE	SE
Scenario I									
0	(1,1)	(30,52)	(30,51)	(66,64)	(0,0)	(30,52)	(30,51)	(66,64)	(1,1) (37,65)
25	(1,0)	(31,54)	(32,56)	(60,56)	(1,-1)	(31,54)	(32,56)	(60,56)	(2,-2) (40,72)
50	(0,-2)	(35,63)	(36,63)	(51,50)	(-1,-2)	(35,63)	(36,63)	(51,50)	(-2,-2) (49,89)
Scenario II									
0	(1,1)	(14,24)	(13,23)	(87,85)	(1,1)	(14,24)	(13,23)	(87,85)	(1,1) (15,26)
25	(1,0)	(14,26)	(14,24)	(77,86)	(1,0)	(14,26)	(14,25)	(77,86)	(1,1) (16,28)
50	(0,-1)	(16,26)	(16,27)	(88,70)	(0,-1)	(16,27)	(16,28)	(89,76)	(0,0) (17,31)
Scenario III									
0	(3,0)	(34,57)	(32,55)	(53,64)	(3,0)	(34,57)	(33,57)	(53,64)	(-2,-1) (47,71)
25	(-1,2)	(36,60)	(35,59)	(59,59)	(-1,1)	(36,61)	(35,61)	(59,61)	(1,-1) (47,78)
50	(0,3)	(38,66)	(38,66)	(48,51)	(0,2)	(38,67)	(39,67)	(48,53)	(-2,5) (55,92)
Scenario IV									
0	(1,0)	(21,37)	(20,34)	(65,65)	(1,1)	(22,39)	(23,39)	(72,72)	(1,1) (26,46)
25	(0,2)	(23,40)	(22,37)	(63,64)	(0,1)	(25,42)	(24,42)	(74,71)	(0,1) (29,50)
50	(-2,0)	(25,43)	(24,41)	(51,55)	(-1,0)	(27,45)	(27,47)	(60,60)	(-1,1) (35,58)

Cen	$\hat{\beta}_M$			$\hat{\beta}_{MLE}$			$\hat{\beta}_{MLE}^{\text{norm}}$		
	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE
Scenario I									
0	(-1,-1)	(51,90)	(190,192)	(-1,-3)	(30,50)	(66,59)	(-3,-7)	(37,67)	(100,107)
25	(0,-2)	(50,90)	(156,156)	(1,1)	(31,54)	(60,56)	(-3,-9)	(41,72)	(105,101)
50	(-1,3)	(51,86)	(108,93)	(0,-2)	(35,62)	(51,49)	(-7,-19)	(43,79)	(79,83)
Scenario II									
0	(1,3)	(34,56)	(512,465)	(0,-1)	(13,22)	(75,72)	(0,-1)	(17,32)	(128,151)
25	(1,-2)	(33,60)	(424,459)	(1,1)	(15,26)	(88,86)	(-2,-5)	(21,37)	(173,178)

Cen	$\hat{\beta}_M$			$\hat{\beta}_{MLE}$			$\hat{\beta}_{MLE}^{norm}$		
	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE
Scenario III									
50	(0,-2)	(35,58)	(424,350)	(0,-1)	(17,29)	(100,88)	(0,-1)	(22,43)	(167,192)
Scenario IV									
0	(-1,2)	(45,77)	(299,280)	(-1,1)	(21,37)	(65,65)	(-2,-2)	(22,38)	(72,68)
25	(0,0)	(44,79)	(230,250)	(0,-1)	(23,39)	(63,61)	(-2,-4)	(23,40)	(63,65)
50	(2,2)	(44,81)	(158,195)	(2,-3)	(25,43)	(51,55)	(1,-3)	(25,45)	(51,75)

Note: Cen is the censoring rate (%); Bias is the empirical bias ($\times 1000$); SE is the empirical standard error ($\times 1000$); SEE is the empirical mean of the standard error estimates ($\times 1000$); RE is the relative efficiency ($\times 100$) compared to $\hat{\beta}_{LT}^{opt} \cdot \hat{\beta}_W^{opt}$ is the combined estimator with estimated weight function as in Section 2.2; $\hat{\beta}_W^{lr}$ is the combined estimator with $\phi_1 = \phi_2 = 1$; $\hat{\beta}_{LT}$ is the estimator from log-rank estimating equations based on LC ; $\hat{\beta}_M$ is the rank-based estimator based on LM with estimated ϕ_2 by assuming $\tilde{\mathcal{E}}$ follows a generalized gamma distribution; $\hat{\beta}_{MLE}$ and $\hat{\beta}_{MLE}^{normal}$ are the parametric maximum likelihood estimators assuming generalized gamma and normal distribution for $\tilde{\mathcal{E}}$. RE of $\hat{\beta}_{LT}$ is 100 and is omitted in the table.