Statistica Sinica **24** (2014), 313-333
doi:http://dx.doi.org/10.5705/ss.2011.197

SEMIPARAMETRIC ACCELERATED FAILURE TIME MODEL FOR LENGTH-BIASED DATA WITH APPLICATION TO DEMENTIA STUDY

Jing Ning[1], Jing Qin[2] and Yu Shen[1]

[1] *The University of Texas M. D. Anderson Cancer Center*
[2] *National Institute of Allergy and Infectious Diseases*

Abstract: A semiparametric accelerated failure time (AFT) model is proposed to evaluate the effects of risk factors on the unbiased failure times for the target population given the observed length-biased data. The analysis of length-biased data is complicated by informative right censoring due to the biased sampling mechanism, and consequently the techniques for conventional survival analysis are not applicable. We propose estimating equation methods for estimation and show the asymptotic properties of the proposed estimators. The small sample performance of the estimating methods are investigated and compared with that of existing methods under various underlying distributions and censoring mechanisms. We apply the proposed model and estimating methods to a prevalent cohort study, the Canadian Study of Health and Aging (CSHA), to evaluate the survival duration according to diagnosis of subtype of dementia.

Key words and phrases: Accelerated failure time model, dementia, dependent censoring, estimating equation, length-biased sampling, prevalent cohort.

1. Introduction

Dementia is a common geriatric syndrome that affects more than 5 million Americans (Lee and Chodosh (2009); Hebert et al. (2003)). Among people older than 65 years in the United States, dementia is the fifth leading cause of death (Minino et al. (2007)). Evaluation of survival duration from dementia onset has become an important public health issue, since life expectancy is increasing and consequently more older adults are being diagnosed with dementia. Although it is generally agreed that dementia shortens life expectancy, the estimated survival duration from the onset of dementia has a large variation, ranging from 3 to 9 years (Fitzpatrick et al. (2003); Helzner et al. (2003)). Most of these estimates were based on prevalent cohort studies, in which subjects who were identified at recruitment to the studies were followed to record their time to death. Although prevalent sampling has been widely used in prospective cohort studies due to its convenience and economic considerations, estimation of survival duration could be inflated without adjustment for prevalent sampling bias. Wolfson et al. (2001) reported that the median survival estimate after adjustment for biased sampling is much shorter than the previously reported estimates that ignored the bias.

Prevalent sampling has been widely used in prospective cohort studies, which involves the observance of left-truncated failure times with potential right censoring. Typically, subjects who have experienced the initiating event but have not experienced the failure event at the time of examination (or recruitment) are sampled into the cohort and followed prospectively for the failure event. The initiating event may be the onset or diagnosis of a disease, and the failure event may be death/recurrence of the disease. The time to the failure event may be right censored. It is clear that the subjects under observation comprise a subset of the original target population, and the subjects who did not survive to the examination times are left truncated. If the initial event (e.g. the onset of a disease) follows a stationary Poission process, i.e., the truncation time follows a uniform distribution, these left-truncated data are called length-biased data. The assumption of a uniform truncation distribution can be examined by formal goodness-of-fit tests (Addona and Wolfson (2006); Mandel and Betensky (2007)).

Length-biased data are often encountered in studies of epidemiologic cohorts, cancer prevention, and labor economy (Zelen and Feinleib (1969); Lancaster (1979); Vardi (1989); Nowell and Stanley (1991); De Una-Alvarez, Otero-Giraldez, and Alvarez-Llorente (2003); Zelen (2004); Greenberg et al. (2005); Song et al. (2006)), in which the probability of an individual being selected from the target population is proportional to the duration from the first event to the failure event. Although the issues of length bias and the need to correct for bias in estimation and inference in various applications have been well recognized for decades in the epidemiology and statistics literature, considerable methodological gaps remain. Previous work has largely focused on one-sample estimates for the length-biased failure time distribution, either conditional on the observed truncation times (Turnbull (1976); Lagakos, Barraj, and De Gruttola (1988); Wang (1991); Kalbfleisch and Lawless (1991); Wang, Brookmeyer, and Jewell (1993); Luo and Tsai (2009)), or with an unconditional approach (Vardi (1982, 1989); Asgharian, M'Lan, and Wolfson (2002); Asgharian and Wolfson (2005)). There is little work considering regression analysis for modeling the association between risk factors and population failure times. Recently, under the proportional hazards model, Tsai (2009) extended the pseudo-partial likelihood approach of Wang (1996), and Qin and Shen (2010) proposed inverse weighted estimating equation approaches for right-censored length-biased data.

The accelerated failure time (AFT) model is an important alternative to Cox's proportional hazards model and appeals to investigators because of its straightforward interpretation. Large sample properties and methods of inference have been extensively investigated over the last twenty years for traditional right-censored survival data (Prentice (1978); Buckley and James (1979); Miller and Halpern (1982); Ritov (1990); Tsiatis (2009); Lai and Ying (1991a); Ying (1993); Lin and Ying (1995); Jin et al. (2003)). However, the length-biased sampling and the resultant violation of the assumption for independent censoring complicate the use of the AFT model on right-censored length-biased data. Recently, Chen (2010) and Mandel and Ritov (2010) utilized the invariant property of the covariate effects in the AFT model to propose estimating methods for length-biased data not subject to right censoring. When length-biased failure times are subject to right censoring, the aforementioned methods cannot be easily extended to accommodate the induced dependent censoring. Shen, Ning, and Qin (2006) proposed an estimating equation approach under the transformation model as well as the AFT model. The advantage of the estimating equation approach is its ease in calculation because of the closed form of the solution for the estimating equations. However, the equations are based on the first moment property of the observed failure times, which may not be efficient. In a subsequent paper, Ning, Qin, and Shen (2011) generalized a Buckley-James type of estimator in traditional survival analysis of length-biased data. The proposed estimators under the AFT model have pros and cons under different scenarios. Hence, it is desirable to develop more efficient estimation methods under the AFT model given right-censored length-biased data and to compare their performance for various censoring distributions and underlying distributions.

Our goal is to use the semiparametric AFT model to infer the relationship between the unbiased survival time and the diagnosis of a subtype of dementia using data from the Canadian Study of Health and Aging (CSHA). The CSHA is a large national epidemiology study of dementia and other health problems of people aged 65 years and over in Canada (Lindsay et al. (2004); Wolfson et al. (2001)). In the first phase of the study, conducted from 1991 to 1992, 14,026 subjects aged 65 or older were randomly selected from 36 urban and surrounding rural areas in 10 Canadian provinces. Among them, 10,263 agreed to participate and were screened for dementia with the Modified Mini-Mental State Examination. Researchers identified 1,132 subjects with dementia and classified them into subcategories of dementias. Our analysis is limited to subjects with a diagnosis of probable Alzheimer's disease, possible Alzheimer's disease, or vascular dementia. While there is no consensus regarding the association between dementia subtype and survival outcome, it is of great interest to estimate the survival distribution by each diagnosis type and to investigate the association under a flexible model structure. Information about the date of onset of dementia was collected in a hierarchical fashion from the answers to three questions from the Cambridge Examination for Mental Disorders of the Elderly (Wolfson et al. (2001)). In the second phase of the CSHA, conducted from 1996 to 1997, all participants who could be contacted were re-evaluated for dementia; subjects with dementia who died before and during the follow-up study were identified and the dates and causes of death were recorded. The third phase of the CSHA from 2001 to 2002 was generally the same as the second phase of the study. As expected, selection bias occurred under the described sampling circumstances: the observed time intervals from dementia diagnosis to death tended to be longer for subjects in the CSHA compared to those from subjects in the general population. An analytic approach that ignores such selection bias could result in an overestimation of the survival duration for subjects with dementia, and consequently an underestimation of the deleterious effect of dementia on life expectancy. Moreover, the estimation of the association between dementia diagnoses and disease prognosis could be biased as well.

The paper is organized as follows. Section 2 introduces the notation and the semi-parametric AFT model. In Section 3, we propose two types of semi-rank-based regression methods to model length-biased right-censored data under the AFT model and describe the associated large sample properties of the estimators. We examine the performance of the proposed estimators through a simulation study in Section 4. We revisit the motivating example and present the analysis results of the CSHA in Section 5. We conclude with a discussion in Section 6, and provide details of the proofs in the Appendix.

2. Notation and Models

Consider the CSHA, in which subjects who were identified with dementia but had survived up to the time of recruitment were sampled and followed prospectively until death or censoring. Let $\widetilde{T}$ be the unbiased time measured from onset of dementia to death in the target population, T be the observed length-biased time in the sample population, A be the time of recruitment measured from onset of dementia, V be the time from recruitment to death, and C be the residual censoring time measured from recruitment. Denote the covariate of interest as the p-vector $\mathbf{X}$. Conditional on $\mathbf{X}$, the residual censoring time C and (A, V) are assumed to be independent. Observed data are represented by n independent samples, $(Y_i, A_i, \delta_i, \mathbf{X}_i)$, where $Y_i = \min\{T_i, A_i + C_i\}, \delta_i = I(V_i \leq C_i)$, and $I(.)$ is the indicator function. Given the covariates, $\mathbf{X} = \mathbf{x}$, let $f(.|\mathbf{x})$ be the density function for the unbiased time $\widetilde{T}$. Under the sampling constraint that the value of $\widetilde{T}$ is observed only when $\widetilde{T} > A$, the density function of the length-biased time T is

$$f_{LB}(t|\mathbf{x}) = \frac{tf(t|\mathbf{x})}{\int_0^\infty uf(u|\mathbf{x})du} = \frac{tf(t|\mathbf{x})}{\mu(x)},$$

where $\mu(x) \equiv \int_0^\infty uf(u|\mathbf{x})du$.

The AFT model (Kalbfleisch and Prentice (1980); Cox (1984)) relates the logarithms of the survival time to the covariate of interest as,

$$\log \widetilde{T}_i = \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{2.1}$$

where $\boldsymbol{\beta}$ is a $P \times 1$ parameter vector and $\epsilon_i, i = 1, \cdots, n$ are independently and identically distributed (i.i.d.) random errors with an unspecified distribution.

3. Semi-rank-based Estimation Methods

For classical survival data, Tsiatis (2009) proposed a rank-based estimating equation by considering the transformed time scale under the AFT model. For left-truncated and right-censored data, Lai and Ying (1991b) introduced a class of rank-based estimators under the AFT model. Under length-biased sampling, the censoring time $A + C$ is dependent on the failure time $A + V$ even if the censoring time C is independent of (A, V). Hence, the generalizations of the above methods entail new challenges in adjusting for non-informative censoring. The generalized methods not only depend on the rank, but also rely on the magnitude of failure time for length-biased data.

3.1. Estimation method based on modified risk set

For general left-truncated and right-censored data, Lai and Ying (1991b) took the at-risk set at time t as $R^*(t, \mathbf{b}) = \{i \leq n : A_i \exp(-\mathbf{X}_i^T \mathbf{b}) < t \leq Y_i \exp(-\mathbf{X}_i^T \mathbf{b})\}$, and proposed the rank-based estimating equations for $\boldsymbol{\beta}$ based on the at-risk sets,

$$\mathbf{U}_{LT}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\mathbf{X}_i - \frac{\sum_{j=1}^{n} \mathbf{X}_j I\{j \in R^*(Y_i e^{-\mathbf{X}_i^T \boldsymbol{\beta}}, \boldsymbol{\beta})\}}{\sum_{j=1}^{n} I\{j \in R^*(Y_i e^{-\mathbf{X}_i^T \boldsymbol{\beta}}, \boldsymbol{\beta})\}} \right]. \tag{3.1}$$

This estimating equation, originally proposed for general left-truncated data, does not effectively utilize information on the truncation time and is consequently inefficient for length-biased data. Under the stationarity assumption for length-biased data, Tsai (2009) proposed the pseudo-partial likelihood method under Cox's proportional hazards model by including the information for the stationarity process. Under the AFT model assumption for $\widetilde{T}$, we include the information on truncation time for length-biased data and modify this "at risk" set to increase the estimation efficiency. Given the observed data $(Y = y, \delta, \mathbf{X} = \mathbf{x})$, the conditional expectation of the indicator function of the "at risk" set is

$$\begin{aligned} R(y,t) &= E[I\,(A < t \leq Y)\,|\delta, Y = y, \mathbf{X} = \mathbf{x}] \\ &= \delta I(y \geq t) \frac{P(A < t, T = y, C \geq y - A | \mathbf{X} = \mathbf{x})}{P(T = y, C \geq y - A | \mathbf{X} = \mathbf{x})} \\ &\quad + (1 - \delta) I(y \geq t) \frac{P(A < t, T \geq y, C = y - A | \mathbf{X} = \mathbf{x})}{P(T \geq y, C = y - A | \mathbf{X} = \mathbf{x})}. \end{aligned}$$

Using the fact that the bivariate distribution of (T, A) given $\mathbf{X} = \mathbf{x}$ is $f(t|\mathbf{x})/\mu(\mathbf{x})$ for $t > a > 0$,

$$\begin{aligned} R(y,t) = \delta I(y \geq t) & \frac{f(y|\mathbf{x}) \int_0^t \bar{G}(y-a|\mathbf{x})da/\mu(\mathbf{x})}{f(y|\mathbf{x}) \int_0^y \bar{G}(y-a|\mathbf{x})da/\mu(\mathbf{x})} \\ & +(1-\delta)I(y \geq t) \frac{\int_y^\infty f(s|\mathbf{x})/\mu(\mathbf{x})ds \int_0^t g(y-a|\mathbf{x})da}{\int_y^\infty f(s|\mathbf{x})/\mu(\mathbf{x})ds \int_0^y g(y-a|\mathbf{x})da}; \end{aligned}$$

one has

$$R(y,t) = \delta I(y \geq t) \frac{W(y|\mathbf{x}) - W(y-t|\mathbf{x})}{W(y|\mathbf{x})} + (1-\delta) I(y \geq t) \frac{G(y|\mathbf{x}) - G(y-t|\mathbf{x})}{G(y|\mathbf{x})},$$

where $G(.|\mathbf{x})$, $\bar{G}(.|\mathbf{x})$, and $g(.|\mathbf{x})$, respectively, are the cumulative distribution function, survival function, and density function for the residual censoring time given $\mathbf{X} = \mathbf{x}$, and $W(t|\mathbf{x})$ is the integral of the survival function of residual censoring time from 0 to t, $W(t|\mathbf{x}) = \int_0^t \bar{G}(s|\mathbf{x})ds$. For brevity, we remove $\mathbf{x}$ from the functions of G and W in the derivations.

By modifying (3.1), we replace the indicator function for the at-risk set with the conditional expectation of the indicator function in (3.2). Note that information of the stationary process is utilized in the equation:

$$\begin{aligned} \widetilde{\mathbf{U}}_I(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n \delta_i \left\{ \mathbf{X}_i - \frac{\sum_{j=1}^n \mathbf{X}_j R(Y_j e^{-\mathbf{X}_j^T \boldsymbol{\beta}}, Y_i e^{-\mathbf{X}_i^T \boldsymbol{\beta}})}{\sum_{j=1}^n R(Y_j e^{-\mathbf{X}_j^T \boldsymbol{\beta}}, Y_i e^{-\mathbf{X}_i^T \boldsymbol{\beta}})} \right\} \qquad (3.2) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left\{ \mathbf{X}_i - \frac{\sum_{j=1}^n \mathbf{X}_j R(Y_j e^{-\mathbf{X}_j^T \boldsymbol{\beta}}, t)}{\sum_{j=1}^n R(Y_j e^{-\mathbf{X}_j^T \boldsymbol{\beta}}, t)} \right\} dN_i(\boldsymbol{\beta}, t), \end{aligned}$$

where $N_i(\boldsymbol{\beta}, t) = I\left(Y_i e^{-\mathbf{X}_i^T \boldsymbol{\beta}} \geq t, \delta_i = 1\right)$. Note that $\mathbf{X}$ is a categorical variable in the motivating example and the survival function of the censoring variable C can be estimated consistently for each of the three categories if the censoring distributions are not the same. Otherwise, G and W can be estimated using the pooled data. After plugging in the consistent estimators of the unknown quantities into $R(y,t)$,

$$\widehat{R}(y,t) = \delta I(y \geq t) \frac{\widehat{W}(y) - \widehat{W}(y-t)}{\widehat{W}(y)} + (1-\delta) I(y \geq t) \frac{\widehat{G}(y) - \widehat{G}(y-t)}{\widehat{G}(y)},$$

where $\widehat{G}(y)$ is the Kaplan-Meier estimator of the cumulative distribution function for the residual censoring variable and $\widehat{W}(y) = \int_0^y \widehat{\bar{G}}(s)ds$. With the estimated $\widehat{R}(y,t)$, we obtain the estimating equation for $\boldsymbol{\beta}$,

$$\mathbf{U}_I(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left\{ \mathbf{X}_i - \frac{\sum_{j=1}^n \mathbf{X}_j \widehat{R}(Y_j e^{-\mathbf{X}_j^T \boldsymbol{\beta}}, t)}{\sum_{j=1}^n \widehat{R}(Y_j e^{-\mathbf{X}_j^T \boldsymbol{\beta}}, t)} \right\} dN_i(\boldsymbol{\beta}, t). \qquad (3.3)$$

Here (3.3) is not a continuous function of $\boldsymbol{\beta}$, so it is not always possible to obtain an exact solution. We take a solution of (3.3), $\widehat{\boldsymbol{\beta}}_I$, to be the minimizer of the Euclidean norm of $\mathbf{U}_I(\boldsymbol{\beta})$ (Wei, Ying, and Lin (1993)). Let $\boldsymbol{\beta}_0$ be the true value of $\boldsymbol{\beta}$. The consistency and weak convergence of $\widehat{\boldsymbol{\beta}}_I$ can be established under regularity conditions listed in the Appendix.

Theorem 1. *If A.1-A.5 in the Appendix hold, $\widehat{\boldsymbol{\beta}}_I$ is a consistent estimator of $\boldsymbol{\beta}_0$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}_I$.*

The variance-covariance matrix $\boldsymbol{\Sigma}_I$ in Theorem 1 is given in the Appendix. The asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$ is derived by using the asymptotic linearity of the estimating equation $\mathbf{U}_I(\boldsymbol{\beta})$ and empirical process theory. The detailed proof of Theorem 1 is in the Appendix. The estimation of the variance-covariance matrix $\boldsymbol{\Sigma}_I$ is not straightforward because of the unknown hazard function in $\boldsymbol{\Sigma}_I$. Given the established weak convergence of $\widehat{\boldsymbol{\beta}}_I$, we use the bootstrap resampling method to approximate the variance of $\widehat{\boldsymbol{\beta}}_I$.

3.2. Estimation equation based on inverse weighting and ranking

In the absence of right censoring, Wang (1996) derived a score estimating equation from the pseudo-likelihood function under Cox's proportional hazards model:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{X}_i - \frac{\sum_{j=1}^{n}\mathbf{X}_j e^{\mathbf{X}_j^T\boldsymbol{\beta}} I(T_j \geq T_i)/T_j}{\sum_{j=1}^{n} e^{\mathbf{X}_j^T\boldsymbol{\beta}} I(T_j \geq T_i)/T_j}\right\} = 0. \tag{3.4}$$

Under the AFT model assumption for $\widetilde{T}$ in (2.1), the transformed data $\widetilde{T}_0 \equiv \widetilde{T}e^{-X^T\boldsymbol{\beta}}$ are i.i.d. and the AFT model (2.1) can be equivalently expressed by Cox's proportional hazards model with no covariate effects (Tsiatis (2009)),

$$\lambda_{\widetilde{T}_0}(t) = \lambda_0(t),$$

where $\lambda_0(t)$ is an unspecified hazard function of e^{ϵ}. We can thus generalize (3.4) under Cox's model to the AFT model with the transformed length-biased data,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{X}_i - \frac{\sum_{j=1}^{n}\mathbf{X}_j I(T_j e^{-\mathbf{X}_j^T\boldsymbol{\beta}} \geq T_i e^{-\mathbf{X}_i^T\boldsymbol{\beta}})/(T_j e^{-\mathbf{X}_j^T\boldsymbol{\beta}})}{\sum_{j=1}^{n} I(T_j e^{-\mathbf{X}_j^T\boldsymbol{\beta}} \geq T_i e^{-\mathbf{X}_i^T\boldsymbol{\beta}})/(T_j e^{-\mathbf{X}_j^T\boldsymbol{\beta}})}\right\} = 0. \tag{3.5}$$

In the presence of potential right censoring to length-biased data, only $Y = \min\{T, A + C\}$ and the corresponding censoring indicator are observed. Qin and Shen (2010) proposed an estimating equation to accommodate the dependent right censoring under Cox's proportional hazards model:

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\left\{\mathbf{X}_i - \frac{\sum_{j=1}^{n}\mathbf{X}_j e^{\mathbf{X}_j^T\boldsymbol{\beta}}\delta_j I(Y_j \geq Y_i)/\widehat{W}(Y_j)}{\sum_{j=1}^{n} e^{\mathbf{X}_j^T\boldsymbol{\beta}}\delta_j I(Y_j \geq Y_i)/\widehat{W}(Y_j)}\right\},$$

where $\widehat{W}(t) = \int_0^t \widehat{\widehat{G}}(s)ds$ is a consistent estimator of $W(t)$. Using a technique parallel to that of Qin and Shen (2010) under the proportional hazards model, we take an estimating equation under the AFT model as,

$$\begin{aligned}\mathbf{U}_{II}(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\left\{\mathbf{X}_i - \frac{\sum_{j=1}^{n}\mathbf{X}_j\delta_j I(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}} \ge Y_ie^{-\mathbf{X}_i^T\boldsymbol{\beta}})/\widehat{W}(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}})}{\sum_{j=1}^{n}\delta_j I(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}} \ge Y_ie^{-\mathbf{X}_i^T\boldsymbol{\beta}}\}/\widehat{W}(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}})}\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{X}_i - \frac{\sum_{j=1}^{n}\mathbf{X}_j\delta_j I(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}} \ge t)/\widehat{W}(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}})}{\sum_{j=1}^{n}\delta_j I(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}} \ge t)/\widehat{W}(Y_je^{-\mathbf{X}_j^T\boldsymbol{\beta}})}\right\}dN_i(\boldsymbol{\beta}, t). \quad (3.6)\end{aligned}$$

As (3.6) is a step function of $\boldsymbol{\beta}$, we take a solution, $\widehat{\boldsymbol{\beta}}_{II}$, to be the minimizer of the Euclidean norm of $\mathbf{U}_{II}(\boldsymbol{\beta})$. The consistency and weak convergence of $\widehat{\boldsymbol{\beta}}_{II}$ can be established under regularity conditions listed in the Appendix.

Theorem 2. *If A.1-A.5 in the Appendix hold, $\widehat{\boldsymbol{\beta}}_{II}$ is a consistent estimator of $\boldsymbol{\beta}_0$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}_{II}$.*

4. Simulation

We explored the finite sample properties of the two semi-rank-based proposed estimators ($\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\boldsymbol{\beta}}_{II}$) via simulation studies. We compared the two proposed estimators with estimators by three existing methods: the inverse weighted estimating equation $\mathbf{U}_{IW}$ of Shen, Ning, and Qin (2006), the Buckley-James-type estimating equation $\mathbf{U}_{BJ}$ of Ning, Qin, and Shen (2011), and the rank-based estimating equation (3.1) ofLai and Ying (1991b) for general left-truncated data, specified as:

$$\mathbf{U}_{IW}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\mathbf{X}_i}{\widehat{W}(Y_i)}(\log Y_i - \mathbf{X}_i^T\boldsymbol{\beta}) = 0, \quad (4.1)$$

$$\mathbf{U}_{BJ}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\Big\{\delta_i\frac{\log Y_i - \mathbf{X}_i^T\boldsymbol{\beta}}{Y_i\exp(-\mathbf{X}_i^T\boldsymbol{\beta})} + (1-\delta_i)\frac{\int_{Y_{i0}}^{\infty}u^{-1}\log u d\widehat{F}_0(u;\boldsymbol{\beta})}{1-\widehat{F}_0(Y_{i0};\boldsymbol{\beta})}\Big\}, \quad (4.2)$$

where

$$\widehat{F}_0(t;\boldsymbol{\beta}) = \frac{\int_0^t s^{-1}d\widehat{G}_0(s;\boldsymbol{\beta})}{\int_0^{\infty} s^{-1}d\widehat{G}_0(s;\boldsymbol{\beta})}$$

is the consistent estimate of the unbiased distribution of transformed data $\widetilde{T}_0$ derived from Vardi's nonparametric estimate for the biased distribution of T_0. We further investigated these semiparametric estimators and maximum likelihood estimator(MLE) in terms of the efficiency and the robustness, with respect to the distribution assumption of the random error, of the MLE. Specifically, we derived the MLE under the assumption that the distribution of the random error is normal, then evaluated all estimators including the MLE under the normal assumption by using data generated from normal or uniform distributions.

We generated the unbiased failure times $\widetilde{T}_i$ from an AFT model with two covariates:

$$\log(\widetilde{T}_i) = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \epsilon_i,$$

where X_{1i} is a binary covariate with $P(X_{1i} = 1) = 0.5$, and X_{2i} is a continuous covariate with a uniform(0,1) distribution. We set $\alpha_0 = 1, \alpha_1 = 0.5$, and $\alpha_2 = 1$. The $\epsilon_i s$ were generated either from a uniform $(-0.5, 0.5)$ distribution or from a Normal(0, 1/12) distribution. After an exponential transformation, we have unbiased survival times. The truncation times and residual censoring times were generated in the original time scale (not log-scale); the truncation times generated from a uniform distribution with an upper boundary bigger than the upper bound of $\widetilde{T}$ to ensure the stationarity assumption. We kept only the pairs satisfying $A_i < \widetilde{T}_i$. The residual censoring times, C_i, were independently generated from a uniform distribution over $(0, c)$, where c was chosen to yield the censoring percentage of 15%, 30%, or 50%. For each specified set of parameters, a sample size of 100 or 200 was chosen and each scenario was repeated 1,000 times.

Table 1 displays the means and empirical standard errors of each estimator. We make some observations: (i) All estimators were close to the true values with mild or moderate censoring (15-30%), the biases of $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\boldsymbol{\beta}}_{II}$ were slightly larger than those of the others, but all were in a reasonable range with heavy censoring (50%). (ii) $\widehat{\boldsymbol{\beta}}_I$ performed as well as $\widehat{\boldsymbol{\beta}}_{II}$ with mild or moderate censoring (15-30%) and was less efficient with heavy censoring (50%), in particular when the random errors were normal. (iii) The two proposed estimators were consistently more efficient than the estimators from $\mathbf{U}_{LT}$ for general left-truncated data regardless of the underlying distribution for the random errors. The unknown censoring distribution has to be estimated in the proposed estimating methods due to the unique data structure of length-biased data. The uncertainty from the estimated censoring distribution could result in a relative large bias for the proposed methods compared to the LT method. (iv) Under normal errors, the estimators by $\mathbf{U}_{IW}$ performed as well as the two proposed estimators and better for heavy censoring, and $\mathbf{U}_{BJ}$ outperformed all other semiparametric estimators with mild or moderate censoring (15-30%) in terms of efficiency. (iiv) Under uniform errors, the estimators from $\mathbf{U}_I$ and $\mathbf{U}_{II}$ consistently outperformed the other semiparametric estimators, whereas the estimators from $\mathbf{U}_{IW}$ and $\mathbf{U}_{BJ}$ were less efficient than the estimator from $\mathbf{U}_{LT}$.

Table 2 compares the estimated relative efficiency of the six estimators; the estimated relative efficiency was calculated as the ratio of two mean squared

Table 1: Summary statistics for six estimators.

Size	Cen%	$\mathbf{U}_I$ Estimator		$\mathbf{U}_{II}$ Estimator		$\mathbf{U}_{IW}$ Estimator		$\mathbf{U}_{BJ}$ Estimator		$\mathbf{U}_{LT}$ Estimator		Parametric MLE	
								$(\widehat{\alpha}_1, \widehat{\alpha}_2)$					
		Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
								$\epsilon \sim U(-0.5, 0.5)$					
100	15%	(.006, .012)	(.044,.079)	(-.004,-.006)	(.044,.079)	(-.007,-.005)	(.070,.122)	(.000, .002)	(.068,.121)	(.002, .001)	(.051,.092)	(-.011,-.018)	(.064,.113)
100	30%	(-.008, -.029)	(.052,.094)	(.008, .027)	(.051,.095)	(-.007,-.006)	(.073,.136)	(-.005,-.001)	(.067,.120)	(.001,-.006)	(.059,.117)	(-.028,-.053)	(.071,.125)
100	50%	(.029, .039)	(.073,.164)	(-.018,-.033)	(.070,.127)	(-.008,-.005)	(.087,.157)	(.015, .036)	(.106,.190)	(.000,-.060)	(.078,.187)	(-.043,-.086)	(.082,.151)
200	15%	(.005, .011)	(.027,.049)	(-.003,-.007)	(.027,.049)	(-.003, .000)	(.050,.087)	(.001,-.001)	(.048,.080)	(.001,-.001)	(.031,.054)	(-.011,-.020)	(.047,.082)
200	30%	(.014,.028)	(.032,.059)	(-.011,-.025)	(.032,.062)	(-.003,-.005)	(.050,.094)	(.000, .002)	(.048,.086)	(.001,-.003)	(.037,.067)	(-.026,-.051)	(.050,.087)
200	50%	(.030, .057)	(.043,.098)	(-.013,-.028)	(.042,.078)	(.000, .000)	(.058,.107)	(.015, .036)	(.066, .120)	(.001,-.014)	(.046,.102)	(-.046,-.080)	(.059,.102)
								$\epsilon \sim Normal(0, 1/12)$					
100	15%	(.014, .019)	(.072,.124)	(-.006,-.019)	(.073,.121)	(-.002,-.011)	(.073,.124)	(-.001,-.002)	(.072,.118)	(.004,-.004)	(.086,.138)	(-.012,-.023)	(.067,.115)
100	30%	(.025, .045)	(.079,.133)	(-.018,-.043)	(.076,.136)	(-.002,-.013)	(.076,.134)	(.001,-.005)	(.067,.123)	(.001,-.014)	(.093,.151)	(-.027,-.052)	(.070,.124)
100	50%	(.043, .050)	(.099,.198)	(-.029,-.044)	(.087,.162)	(-.003,-.012)	(.087,.160)	(-.030,-.065)	(.097,.186)	(-.002,-.060)	(.104,.183)	(-.044,-.081)	(.083, .140)
200	15%	(.010, .019)	(.051,.089)	(-.010,-.024)	(.049,.088)	(-.002,-.011)	(.051,.086)	(-.001,-.001)	(.049,.081)	(.000,-.003)	(.060,.097)	(-.009,-.022)	(.047,.084)
200	30%	(.024, .044)	(.056,.098)	(-.021,-.044)	(.051,.096)	(-.002,-.012)	(.054,.093)	(-.001, .002)	(.050,.085)	(-.001,-.006)	(.063,.106)	(-.025,-.053)	(.051, .089)
200	50%	(.039, .084)	(.072,.135)	(-.028,-.031)	(.060,.116)	(-.002,-.009)	(.059,.111)	(-.030,-.052)	(.065,.121)	(-.003,-.012)	(.070,.125)	(-.040,-.084)	(.059, .105)

Table 2. Relative efficiency for six estimators.

Cohort Size	Cen% Cen%	$\mathbf{U}_I$ Estimator	$\mathbf{U}_{IW}$ Estimator	$\mathbf{U}_{BJ}$ Estimator	$\mathbf{U}_{LT}$ Estimator	Parametric MLE
			$\epsilon \sim U(-0.5, 0.5)$			
100	15%	(1.01, 1.02)	(2.54, 2.38)	(2.37, 2.33)	(1.33, 1.35)	(2.16, 2.09)
100	30%	(1.04, 0.99)	(2.02, 1.90)	(1.69, 1.48)	(1.31, 1.41)	(2.18, 1.89)
100	50%	(1.18, 1.65)	(1.46, 1.43)	(2.19, 2.17)	(1.16, 2.24)	(1.64, 1.77)
200	15%	(1.02, 1.03)	(3.40, 3.09)	(3.12, 2.61)	(1.30, 1.19)	(3.16, 2.91)
200	30%	(1.07, 0.95)	(2.19, 1.98)	(2.01, 1.66)	(1.20, 1.01)	(2.77, 2.27)
200	50%	(1.42, 1.87)	(1.74, 1.67)	(2.37, 2.29)	(1.10, 1.54)	(2.90, 2.45)
			$\epsilon \sim Normal(0, 1/12)$			
100	15%	(1.00, 1.05)	(0.99, 1.03)	(0.97, 0.93)	(1.38, 1.27)	(0.86, 0.92)
100	30%	(1.13, 0.97)	(0.95, 0.89)	(0.74, 0.74)	(1.42, 1.13)	(0.92, 0.89)
100	50%	(1.47, 1.48)	(0.96, 0.91)	(1.30, 1.38)	(1.36, 1.32)	(1.11, 0.93)
200	15%	(1.08, 1.00)	(1.04, 0.90)	(0.96, 0.79)	(1.44, 1.13)	(0.91, 0.91)
200	30%	(1.22, 1.03)	(0.96, 0.79)	(0.82, 0.65)	(1.31, 1.01)	(1.06, 0.96)
200	50%	(1.53, 1.75)	(0.79, 0.86)	(1.17, 1.20)	(1.12, 1.09)	(1.16, 1.25)

errors (MSEs), the MSE of the estimator from $\mathbf{U}_{II}$ being used as a reference. From Table 2, we find that (i) when the censoring percentage increases, the efficiency gain of $\widehat{\boldsymbol{\beta}}_{II}$ relative to $\widehat{\boldsymbol{\beta}}_I$ increases; (ii) when the error is uniform, the two proposed estimators are much more efficient than the estimators from $\mathbf{U}_{IW}$ or $\mathbf{U}_{BJ}$ even with mild censoring, and the estimators from $\mathbf{U}_{LT}$ can be more efficient than those from $\mathbf{U}_{IW}$ or $\mathbf{U}_{BJ}$; (iii) when the error is normal, using the Buckley-James method is more efficient than using the proposed methods except under heavy censoring, and then $\mathbf{U}_{LT}$ is the least efficient method. It is not surprising that the parametric MLE is the most efficient estimator when the error distribution is correctly specified in the likelihood. However, we also find that (i) compared to the MLE under correctly specified distribution assumption, the efficiency loss of the five semiparametric estimators is mild for length-biased data; (ii) the two proposed semiparametric estimators are much robust than the MLE with respect to the distribution assumption.

There is no uniform best estimation method in terms of statistical efficiency among the estimation methods considered, that depends on the degree of censoring and the underlying distribution. When the distribution of $\log \widetilde{T}$ is close to normal, the inverse weighted and Buckley-James methods, derived from the least squares principle, are more efficient than the two proposed semi-rank-based estimation methods; when the underlying distribution is uniform, the proposed semi-rank-based estimation methods are better. Figure 1 summarizes a decision guideline on the estimation methods for various combinations of underlying distribution and censoring patterns.

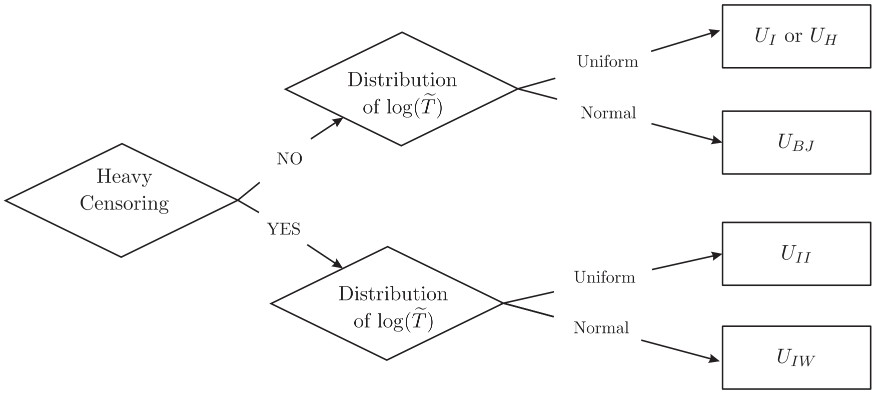

Figure 1. Decision tree plot of the estimation method selection for length-biased data.

5. Data Application

Occurrence of dementia shortens life expectancy, and the accurate estimation of survival time could empower clinicians and patients with dementia to make timely decisions about their treatment. The CSHA is a national longitudinal study of Canadian people aged 65 years and over, focusing on prevalence and incidence of dementia. In the study, subjects with dementia were identified and classified into subcategories of dementia. Our goal was to estimate and compare survival following the onset of dementia according to the subcategory of dementia (Alzheimer's disease, possible Alzheimer's disease, or vascular dementia). The available data were collected from 818 subjects with dementia. Among them, 393 subjects were diagnosed as having probable Alzheimer's disease, 252 as having possible Alzheimer's disease, and 173 as having vascular dementia. For each subject, the date of dementia onset was collected during the clinical examination in the first stage of the CSHA (Wolfson et al. (2001)), and the date of censoring or death was collected prospectively during the second or third phase of the CSHA. Subjects who died quickly after dementia onset were more likely to be excluded (left truncated) from the study and, in turn, the observed survival duration from dementia onset in the prevalent cohort tended to be longer than that in the target population. Hence, adjustment of biased sampling is important to avoid an overestimation of survival for patients with dementia.

Wolfson et al. (2001) provided a figure that included three estimated survival functions according to diagnosis of subcategory of dementia, using the Expectation Maximization(EM) algorithm of Vardi (1989) for length-biased and right-censored data. The figure suggested some differences among the three subgroups; however, the one sample estimation could not answer whether differences were statistically significant or not. We applied the proposed inference methods under

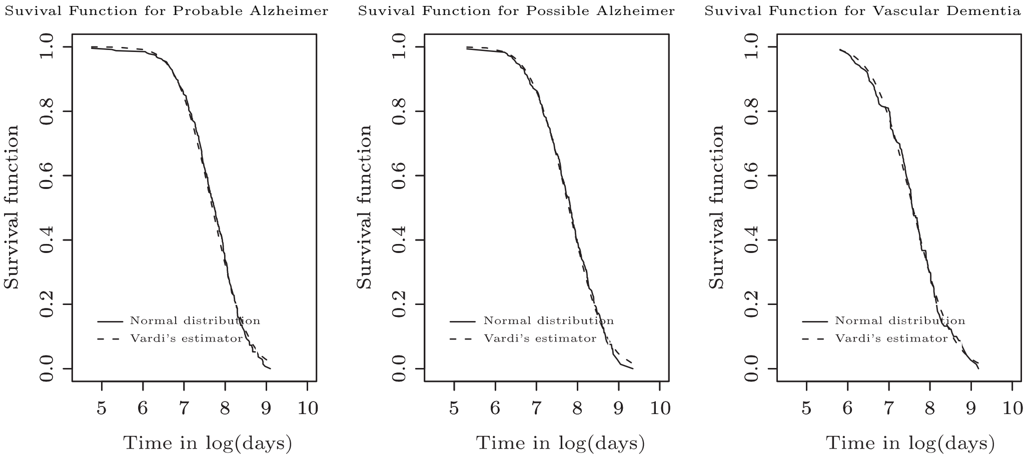

Figure 2. Plot of estimated survival functions according to the subcategory of dementia.

the accelerated failure time model $\log \widetilde{T} = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \epsilon$ to evaluate the effects of different diagnostic subcategories of dementia on survival, where X_1 and X_2 indicate whether the subject had vascular dementia or probable Alzheimer's disease, respectively.

The stationarity assumption that the truncation time is uniformly distributed was examined using the formal test given by Addona and Wolfson (2006). The test gave a p-value of 0.94, suggesting that the observed left-truncated data were length-biased data. The applicability of the AFT time to the application was checked using QQ-plots (Ning, Qin, and Shen (2011)). We checked the underlying distributions of $\log \widetilde{T}$ for three subgroups by comparing the nonparametric estimators of survival distributions (Vardi (1989)) and normal survival distributions. The plots shown in Figure 2 suggest that the underlying distributions of survival duration from dementia onset are close to normal. Considering the censoring degrees among the three groups ($20\% \sim 24\%$), the Buckley-James method should be a good choice among the estimation methods considered. However, for comparison, all estimated results, including the proposed two estimators, the three estimators by equations $\mathbf{U}_{IW}$, $\mathbf{U}_{BJ}$, and $\mathbf{U}_{LT}$, and the corresponding bootstrap standard errors, are listed in Table 3.

The estimators and inference from the five estimating methods with adjustment for length-biased sampling were comparable. The results indicate a trend for shorter survival time for the category of vascular dementia, though the difference was not statistically significant. Among the estimation methods applied, the estimator from the proposed estimating equation $\mathbf{U}_{II}$ had a smaller variation, standardized by dividing the size of the estimator, than the estimators from the

Table 3. Estimates (Est) and standard errors (SE) of regression coefficients for dementia data.

	$\mathbf{U}_I$ Estimator	$\mathbf{U}_{II}$ Estimator	$\mathbf{U}_{IW}$ Estimator	$\mathbf{U}_{BJ}$ Estimator	$\mathbf{U}_{LT}$ Estimator
	Est(SE)	Est(SE)	Est(SE)	Est(SE)	Est(SE)
Vascular Dementia	-0.152(0.114)	-0.156(0.083)	-0.210(0.112)	-0.187(0.128)	-0.106(0.076)
Probable Alzheimer	-0.138(0.118)	-0.102(0.057)	-0.135(0.151)	-0.114(0.150)	-0.033(0.063)

other estimating equations. Estimators obtained from the estimating equations $\mathbf{U}_I$, $\mathbf{U}_{IW}$, and $\mathbf{U}_{BJ}$ were fairly comparable.

6. Discussion

In this article, we have introduced two estimation methods for right-censored length-biased data under the AFT model. Besides the straightforward interpretation of the parameters, another appealing feature of the AFT model is that the observed failure time data can be transformed to a different time scale so that the transformed samples are i.i.d. without the covariate effect. This facilitates the use of existing estimating equations proposed for Cox's proportional hazards model or other rank-based estimating equations after some modifications. In contrast to the existing rank-based approaches (Lai and Ying (1991b); Tsiatis (2009)), the proposed semi-rank-based methods rely on both the magnitude and the rank of failure times in the estimating equations. In the two proposed methods, the intercept in the AFT model is not estimated and the inference focus is on the slope parameters. If it is of interest to estimate the intercept in the AFT model, the inverse weighted method of (Shen, Ning, and Qin (2006)) can be easily adapted for this purpose.

The censoring distribution is often considered a nuisance quantity when analyzing traditional survival data assuming non-informative censoring. Due to the informative censoring caused by biased sampling, the distribution of the censoring variable plays an important role in the construction of both estimating equations.

We here compared the performance of the two proposed methods with three existing methods in empirical studies. The inverse weighted estimating equation approach (Shen, Ning, and Qin (2006)) is the most computationally efficient with a simple closed form of the solution, but it is not the most statistically efficient in general. There is no uniformly most efficient estimation method among the five investigated methods; estimation efficiency depends on the underlying distribution and censoring patterns. In general, if the log-transformed failure times is normal, the least squares or weighted least squares methods should be more efficient; if the underlying distribution of log-transformed times is uniform or exponential, rank-based methods should be better.

We have restricted attention to covariate-independent censoring. The Buckley-James estimator is robust to this assumption. The proposed semi-rank-based estimators can be easily generalized by including covariate-dependent weights in the estimating equations when C depends on X, given that the conditional distribution $(G(t|X))$ can be estimated consistently using nonparametric or semiparametric methods. One advantage of the proposed estimators over the Buckley-James estimator is that there is no need to use a consistent estimator as an initial value in the iterative estimation algorithms. As a result, the variances of the proposed estimators do not depend on the the initial estimator.

Our work is based on the length-biased density function conditional on the covariates X. There are some discussions on efficiency loss in the literature when using likelihood conditional on the covariates for length-biased data (Bergeron, Asgharian, and Wolfson (2008); Mandel and Ritov (2010)). Specifically, there can be a loss of efficiency for the proposed estimator conditional on the covariates when the marginal distribution of covariates is known (Bergeron, Asgharian, and Wolfson (2008)). However, there is no loss of efficiency for the estimators based on the likelihood conditional on X if the marginal distribution of the covariates is unknown, which is common in applications. Extending the AFT model to include time dependent covariates is nontrivial for length-biased data, due in part to the sampling mechanism. The necessary developments for this important extension are worthy of future research but are beyond the scope of this paper.

Acknowledgement

This work was supported in part by grant CA0769466 from the National Institutes of Health. The authors are grateful to Professor M. Asgharian and investigators from the Canadian Study of Health and Aging for providing us with the dementia data collected as part of the CSHA. The core study was funded by the Seniors' Independence Research Program through the National Health Research and Development Program of Health Canada (Project no.6606-3954-MC(S)). Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/Pharmaceutical Manufacturers Association of Canada Health Activity Program, NHRDP Project 6603-1417-302(R), Bayer Incorporated, and the British Columbia Health Research Foundation Projects 38 (93-2) and 34 (96-1). The study was coordinated through the University of Ottawa and the Division of Aging and Seniors, Health Canada.

Appendix

We use certain regularity conditions.

(A.1) $\mathbf{X}$ is uniformly bounded by M_X, and if there exists a constant vector $\boldsymbol{c}$ such that $\boldsymbol{c}^T\boldsymbol{X} = 0$ with probability one, then $\boldsymbol{c} = 0$.

(A.2) The parameter space of $\boldsymbol{\beta}$, $\mathbf{B}$, is a compact set including the true value of parameter $\boldsymbol{\beta}_0$.

(A.3) $\tau = \inf\{t : F_0(t) = 1\} < \infty$, where $F_0(t)$ is the cumulative density function of $\widetilde{T}_0$.

(A.4) The density function of $\widetilde{T}_0$, f_0, is a continuous and differentiable distribution function over $(0, \tau)$ with bounded derivative f_0'.

(A.5) The residual censoring time has a uniformly bounded density g.

Following the arguments for asymptotic properties of linear rank-based estimators (Ying (1993)) for right-censored data, we show the consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I}}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{II}}$.

Proof of Theorem 1. Let $\bar{N}^x(\boldsymbol{b},t) = (1/n)\sum_{i=1}^n \boldsymbol{X}_i N_i(\boldsymbol{b},t)$, $\bar{N}(\boldsymbol{b},t) = (1/n)\sum_{i=1}^n N_i(\boldsymbol{b},t)$, $\bar{R}(\boldsymbol{b},t) = (1/n)\sum_{i=1}^n \widehat{R}(Y_i e^{-\mathbf{X}_i^T\boldsymbol{b}},t)$, and $\bar{R}^x(\boldsymbol{b},t) = (1/n)\sum_{i=1}^n \mathbf{X}_i \widehat{R}(Y_i e^{-\mathbf{X}_i^T\boldsymbol{b}},t)$.

For the estimating equation $\mathbf{U}_I(b)$, let

$$\mathbf{m}_I(\boldsymbol{b}) = E\Big[\int_0^\infty \Big\{\boldsymbol{X} - \frac{E\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)}{E\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)}\Big\} dN(\boldsymbol{b},t)\Big].$$

For any $N_b > 0$,

$$\begin{aligned}
&\sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\le N_b} ||\mathbf{U}_I(\boldsymbol{b}) - \mathbf{m}_I(\boldsymbol{b})|| \\
&\quad \le \sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\le N_b} \Big\| E\int_0^\infty \boldsymbol{X} dN(\boldsymbol{b},t) - \int_0^\infty d\bar{N}^x(\boldsymbol{b},t) \Big\| \\
&\quad\quad + \sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\le N_b} \Big\| \int_0^\infty \frac{\bar{R}^x(\boldsymbol{b},t)}{\bar{R}(\boldsymbol{b},t)} d\left\{EN(\boldsymbol{b},t) - \bar{N}(\boldsymbol{b},t)\right\} \Big\| \\
&\quad\quad + \sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\le N_b} \Big\| \int_0^\infty \left\{ \frac{E\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)}{E\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)} - \frac{\bar{R}^x(\boldsymbol{b},t)}{\bar{R}(\boldsymbol{b},t)} \right\} dEN(\boldsymbol{b},t) \Big\|. \qquad (A.1)
\end{aligned}$$

For any $\epsilon > 0$, applying Lemma 1 in Ying (1993),

$$\sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\le N_b} ||E\int_0^\infty \boldsymbol{X} dN(\boldsymbol{b},t) - \int_0^\infty d\bar{N}^x(\boldsymbol{b},t)|| = o(n^{-1/2+\epsilon}) \quad a.s..$$

The second term on the right side of (A.1) is $o(n^{-1/2+\epsilon})$ by applying integration by parts and that fact that the total variation

$$\sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\le N_b} \int_0^\infty |d\frac{\bar{R}^x(\boldsymbol{b},t)}{\bar{R}(\boldsymbol{b},t)}| \le pM_x \int_0^\infty \frac{-d\bar{R}(\boldsymbol{b},t)}{\bar{R}(\boldsymbol{b},t)} = O(logn),$$

where $\int |dv(t)|$ of a vector function $v(t)$ denotes the sum of the total variations of all its components and p is the dimension of $\mathbf{X}$.

Since

$$\Big\| \frac{E\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)}{E\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)} - \frac{\bar{R}^x(\boldsymbol{b},t)}{\bar{R}(\boldsymbol{b},t)} \Big\|$$
$$\leq \frac{||E\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t) - \bar{R}^x(\boldsymbol{b},t)||}{E\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)} + \frac{||E\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t) - \bar{R}(\boldsymbol{b},t)||}{E\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{b}},t)},$$

the third term on the right side of (A.1) is also $o(n^{-1/2+\epsilon})$. Hence, for any $\epsilon > 0$

$$\sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\leq N_b} ||U_I(\boldsymbol{b}) - \mathbf{m}(\boldsymbol{b})|| = o(n^{-1/2+\epsilon}) \;\; a.s.. \tag{A.2}$$

This implies that the estimating equation $U_I(\boldsymbol{b})$ can be uniformly approximated by the nonrandom function $\boldsymbol{m}_I(\boldsymbol{b})$ up to the order of $n^{-1/2+\epsilon}$. If the function $\boldsymbol{m}_I(\boldsymbol{b})$ has a unique solution given a compact region $\mathbf{B}$ containing $\boldsymbol{\beta}_0$ as an interior point, the estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I}}$, that satisfies $\mathbf{U}_I(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I}}) = \min_{\boldsymbol{b}\in C_{\boldsymbol{b}}} \|\mathbf{U}_I(\boldsymbol{b})\|$, is strongly consistent. As in Ying (1993), this assumption can be evaluated for any given joint distribution of $(\tilde{T}, C, \boldsymbol{X})$. Let the slope of $\boldsymbol{m}_I(\boldsymbol{b})$ be

$$\mathbf{\Gamma}_I(\boldsymbol{b}) = -E\Big[\int_0^\infty \boldsymbol{X}^T\Big\{\boldsymbol{X} - \frac{E\boldsymbol{X}\widehat{R}(\eta(\boldsymbol{b}),t)}{E\widehat{R}(\eta(\boldsymbol{b}),t)}\Big\}e^{-\boldsymbol{X}^T\boldsymbol{b}}$$
$$\times\{f_0'(t(\boldsymbol{b}))W(t(\boldsymbol{b})) + f_0(t(\boldsymbol{b}))\bar{G}(t(\boldsymbol{b}))g(t(\boldsymbol{b}))\}dt\Big],$$

where $\eta(\boldsymbol{b}) = Ye^{-\boldsymbol{X}^T\boldsymbol{b}}$ and $t(\boldsymbol{b}) = te^{-\boldsymbol{X}^T\boldsymbol{b}}$.

The function $\mathbf{m}_{\boldsymbol{I}}(\boldsymbol{b})$ is a continuous function, so by a Taylor expansion, we have

$$\mathbf{m}_I(\boldsymbol{b}) = \mathbf{\Gamma}_I(\boldsymbol{b} - \boldsymbol{\beta}_0) + o(\boldsymbol{b} - \boldsymbol{\beta}_0), \tag{A.3}$$

where $\mathbf{\Gamma}_I = \mathbf{\Gamma}_I(\boldsymbol{\beta}_0)$. Following (A.2) and (A.3), the estimating equation $\mathbf{U}_I$ is asymptotically linear, for $\boldsymbol{d}_n \to 0^+$ in probability,

$$\sup_{||\boldsymbol{b}-\boldsymbol{\beta}_0||\leq \boldsymbol{d}_n} \| \mathbf{U}_I(\boldsymbol{b}) - \mathbf{U}_I(\boldsymbol{\beta}_0) - \mathbf{\Gamma}(\boldsymbol{b} - \boldsymbol{\beta}_0) \| = o_p(n^{-1/2} + \| \boldsymbol{b} - \boldsymbol{\beta}_0 \|).$$

As $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I}} \to \boldsymbol{\beta}_0 \;\; a.s.$, we have

$$\sqrt{n}\mathbf{U}_I(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I}}) = \sqrt{n}\mathbf{U}_I(\boldsymbol{\beta}_0) + \mathbf{\Gamma}_I\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I}} - \boldsymbol{\beta}_0) + o_p(1). \tag{A.4}$$

We next study the asymptotic properties of the estimating equations, $\mathbf{U}_I(\boldsymbol{\beta}_0)$. Let $\widehat{\mathcal{E}}$ and $\mathcal{E}$ represent the sample empirical mean and the limit of average expectation. Using these notations, the estimating equation $\mathbf{U}_I(\boldsymbol{\beta})$ can then be expressed as

$$\boldsymbol{U}_I(\boldsymbol{\beta}) = \widehat{\mathcal{E}}[\boldsymbol{X}N(\boldsymbol{\beta},\tau)] - \int_0^\tau \frac{\widehat{\mathcal{E}}[\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}},t)]}{\widehat{\mathcal{E}}[\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}},t)]} d\widehat{\mathcal{E}}[N(\boldsymbol{\beta},t)]. \qquad (A.5)$$

The function $R(y,t)$ is estimated by using the Kaplan-Meier estimator of the survival function for the censoring. The estimators $\widehat{\mathcal{E}}[\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]$ and $\widehat{\mathcal{E}}[\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]$ converge to the limits $\mathcal{E}[\boldsymbol{X}R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]$ and $\mathcal{E}[R(Y e^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]$. By the Central Limit Theorem,

$$\sqrt{n}\left\{\widehat{\mathcal{E}}[\boldsymbol{X}N(\boldsymbol{\beta}_0,\tau)] - \mathcal{E}[\boldsymbol{X}N(\boldsymbol{\beta}_0,\tau)]\right\}$$

converges weakly to a normal variable, W_1, with mean zero, and

$$\sqrt{n}\left\{\widehat{\mathcal{E}}[N(\boldsymbol{\beta}_0,t)] - \mathcal{E}[N(\boldsymbol{\beta}_0,t)]\right\}$$

converges to a Gaussian process, $\mathcal{W}_2$. Furthermore,

$$\sqrt{n}\left\{\widehat{\mathcal{E}}[\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)] - \mathcal{E}[R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]\right\} = n^{-1/2}\sum_{i=1}^n \psi_{i1}(t;\boldsymbol{\beta}_0) + o_p(1),$$

$$\sqrt{n}\left\{\widehat{\mathcal{E}}[\boldsymbol{X}\widehat{R}(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)] - \mathcal{E}[\boldsymbol{X}R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]\right\} = n^{-1/2}\sum_{i=1}^n \psi_{i2}(t;\boldsymbol{\beta}_0) + o_p(1).$$

By the Central Limit Theorem and the i.i.d. representations, the two processes converge weakly to Gaussian processes $\mathcal{W}_3$ and $\mathcal{W}_4$, respectively. Under the regularity conditions, the mapping of $\mathbf{U}_I(\boldsymbol{\beta}_0)$ from the four processes is compactly differentiable with respect to the supremum norm. We therefore apply the functional delta method and establish the asymptotic i.i.d. representation of equation $\mathbf{U}_I(\boldsymbol{\beta}_0)$, $n^{1/2}\mathbf{U}_I(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \psi_{Ii}(\boldsymbol{\beta}_0) + o_p(1)$, where

$$\begin{aligned}\psi_{Ii}(\boldsymbol{\beta}_0) = {} & \mathbf{X}_i N_i(\boldsymbol{\beta}_0,\tau) - \mathcal{E}[\boldsymbol{X}N(\boldsymbol{\beta}_0,\tau)] \\ & + \int_0^\tau \frac{\psi_{i1}(t;\boldsymbol{\beta}_0)\mathcal{E}[\boldsymbol{X}R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]}{\mathcal{E}[R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]^2} d\mathcal{E}[N(\boldsymbol{\beta}_0,t)] \\ & - \int_0^\tau \frac{\psi_{i2}(t;\boldsymbol{\beta}_0)}{\mathcal{E}[R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]} d\mathcal{E}[N(\boldsymbol{\beta}_0,t)] \\ & - \int_0^\tau \frac{\mathcal{E}[\boldsymbol{X}R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]}{\mathcal{E}[R(Ye^{-\boldsymbol{X}^T\boldsymbol{\beta}_0},t)]} d\left(N_i(\boldsymbol{\beta}_0,\tau) - \mathcal{E}[N(\boldsymbol{\beta}_0,t)]\right).\end{aligned}$$

Then, with (A.4),

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\right) = \boldsymbol{\Gamma}_I^{-1} n^{-1/2}\sum_{i=1}^n \psi_{Ii}(\boldsymbol{\beta}_0) + o_p(1).$$

Thus the desired asymptotic normality of $\widehat{\boldsymbol{\beta}}_I$ follows from the classical Central Limit Theorem; $\sqrt{n}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$ converges to a normal variable with mean zero and variance $\boldsymbol{\Sigma}_I = \boldsymbol{\Gamma}_I^{-1}\boldsymbol{\Sigma}_{UI}\boldsymbol{\Gamma}_I^{-1}$, where $\boldsymbol{\Sigma}_{UI} = E[\psi_{Ii}(\boldsymbol{\beta}_0)]^{\otimes 2}$.

Proof of Theorem 2. By arguments similar to those in the proof of Theorem 1, we get the asymptotic behavior of $\widehat{\boldsymbol{\beta}}_{II}$; give only the key steps here. The estimating equation $\mathbf{U}_{II}(\boldsymbol{b})$ can be shown to be uniformly approximated by the nonrandom function $\boldsymbol{m}_{II}(\boldsymbol{b})$ up to the order of $n^{-1/2+\epsilon}$, where

$$\boldsymbol{m}_{II}(\boldsymbol{b}) = E\bigg[\int_0^\infty \Big\{\boldsymbol{X} - \frac{E\boldsymbol{X}\delta I(\eta(\boldsymbol{b}) \geq t)/\widehat{W}(\eta(\boldsymbol{b})}{E\delta I(\eta(\boldsymbol{b}) \geq t)/\widehat{W}(\eta(\boldsymbol{b})}\Big\} dN(\boldsymbol{b}, t)\bigg].$$

If $\boldsymbol{m}_{II}(\boldsymbol{b})$ has a unique solution given a compact region $\mathbf{B}$ containing $\boldsymbol{\beta}_0$ as an interior point, the estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{II}}$ is strongly consistent. The slope function of $\boldsymbol{m}_{II}(\boldsymbol{b})$ is

$$\begin{aligned}\boldsymbol{\Gamma}_{II}(\boldsymbol{b}) = -\int_0^\infty \boldsymbol{X}^T\Big\{\boldsymbol{X} - \frac{E\boldsymbol{X}\delta I(\eta(\boldsymbol{b}) \geq t)/\widehat{W}(\eta(\boldsymbol{b})}{E\delta I(\eta(\boldsymbol{b}) \geq t)/\widehat{W}(\eta(\boldsymbol{b})}\Big\} e^{-\boldsymbol{X}^T\boldsymbol{b}} \\ \times \big\{f_0'(t(\boldsymbol{b}))W(t(\boldsymbol{b})) + f_0(t(\boldsymbol{b}))\bar{G}(t(\boldsymbol{b}))g(t(\boldsymbol{b}))\big\}\, dt.\end{aligned}$$

Furthermore, the estimating equation $\mathbf{U}_{II}$ is asymptotically linear and

$$\sqrt{n}\mathbf{U}_{II}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{II}}) = \sqrt{n}\mathbf{U}_{II}(\boldsymbol{\beta}_0) + \boldsymbol{\Gamma}_{II}\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{II}} - \boldsymbol{\beta}_0) + o_p(1), \tag{A.6}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{II}} = \boldsymbol{\Gamma}_{II}(\boldsymbol{\beta}_0)$. Using the functional delta method, we can establish the asymptotic i.i.d. representation of equation $\mathbf{U}_{II}(\boldsymbol{\beta}_0)$, denoted as $n^{1/2}\mathbf{U}_{II}(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \psi_{IIi}(\boldsymbol{\beta}_0) + o_p(1)$. Then the classical Central Limit Theorem implies that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ converges to a normal variable with mean zero and variance $\boldsymbol{\Sigma}_I I = \boldsymbol{\Gamma}_{II}^{-1}\boldsymbol{\Sigma}_{UII}\boldsymbol{\Gamma}_{II}^{-1}$, where $\boldsymbol{\Sigma}_{UII} = E[\psi_{IIi}(\boldsymbol{\beta}_0)]^{\otimes 2}$.

References

Addona, V. and Wolfson, D. B. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Anal.* **12**, 267-274.

Asgharian, M., M'Lan, C. E. and Wolfson, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Amer. Statist. Assoc.* **97**, 201-209.

Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.* **33**, 2109-2131.

Bergeron, P., Asgharian, M. and Wolfson, D. B. (2008). Covariate bias induced by length-biased sampling of failure times. *J. Amer. Statist. Assoc.* **103**, 737-742.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.

Chen, Y. Q. (2010), Semiparametric regression in size-biased sampling. *Biometrics*, doi:10.1111/j.1541-0420.2009.01269.x.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall/CRC Press, Boca Raton, FL.

De Una-Alvarez, J., Otero-Giraldez, M. S. and Alvarez-Llorente, G. (2003). Estimation Under Length-bias and Right-censoring: An Application to Unemployment Duration Analysis for Married Women *J. Appl. Statist.* **30**, 283-291.

Fitzpatrick, A. L., Kuller, L. H., Lopez, O. L., Kawas, C. H. and Jagust, W. (2005). Survival following dementia onset: Alzheimers disease and vascular dementia. *J. Neurol. Sci.* **15**, 229-230.

Greenberg, R. S., Daniels, S. R., Flanders, W. D., Eley, J. W. and Boring, J. R. (2005). *Medical Epidemiology.* McGraw-Hill Medical, 101-104.

Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A. and Evans, D. A. (2003). Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch. Neurol.* **60**, 1119-1122.

Helzner, E. P., Scarmeas, N., Cosentino, S., Tang, M. X., Schupf, N. and Stern, Y. (2008). Survival in Alzheimer disease: a multiethnic, population-based study of incident cases. *Neurology*, **71**, 1489-1495.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time models. *Biometrika* **90**, 341-353.

Kalbfleisch, J. D. and Lawless, J. F. (1991). Regression models for right truncated data with applications to AIDS incubation times and reporting lags. *Statist. Sinica* **1**, 19-32.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Lagakos, S. W. and Barraj, L. M. and De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with applications to AIDS. *Biometrika* **75**, 515-523.

Lai, T. L. and Ying, Z. (1991a). Large sample theory of a modified *B*uckley-*J*ames estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370-1402.

Lai, T. L. and Ying, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* **47**, 939-956.

Lee, M. and Chodosh, J. (2009). Dementia and life expectancy: what do we know? *J. Am. Med. Dir. Assoc.* **10**, 466-471.

Lin, D. Y. and Ying, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *J. Statist. Plann. Inference* **44**, 47-63.

Lindsay, J., Sykes, E., McDowell, I., Verreault, R. and Laurin, D. (2004). More than the epidemiology of Alzheimer's disease: contributions of the Canadian Study of Health and Aging. *Canad. J. Psychiatry* **49**, 83-91.

Luo, X. and Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika* **96**, 873-886.

Mandel, M. and Betensky, R. A. (2007). Testing goodness-of-fit of a uniform truncation model. *Biometrics* **63**, 405-412.

Mandel, M. and Ritov, Y. (2010). The accelerated failure time model under biased sampling, *Bimetrics* **66**, 1306-1308.

Miller, R. G. and Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521-531.

Minino, A. M. and Heron, M. P. and Murphy, S. L. and Kochanek, K. D. and Centers for Disease Control and Prevention National Center for Health Statistics National Vital Statistics System (2007). Deaths: Final data for 2004. *National Vital Statistics Reports* **55**, 1-119.

Ning, J., Qin, J. and Shen, Y. (2011). Buckley-James-type estimator with right-censored and length-biased data. *Biometrics* **67**, 1369-1378.

Nowell, C. and Stanley, L. R. (1991). Length-biased sampling in mall intercept surveys. *J. Market Res.* **28**, 475-479.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, **65**, 167-179.

Qin, J. and Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* **66**, 382-392.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303-328.

Shen, Y., Ning, J. and Qin, J. (2006). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *J. Amer. Statist. Assoc.* **104**, 1192-1202.

Song, R., Karon, J. M., White, E. and Goldbaum, G. (2006). Estimating the distribution of a renewal process from times at which events from an independent process are detected. *Biometrics* **62**, 838-846.

Tsai, W. Y. (2009). Pseudo-partial likelihood for proportional hazards models with biased-sampling data. *Biometrika* **96**, 601-615.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *J. Roy. Statist. Soc. Ser. B* **38**, 290-295.

Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10**, 616-620.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**, 751-761.

Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Amer. Statist. Assoc.* **86**, 130-143.

Wang, M. C. (1996). Hazards regression analysis for length-biased data. *Biometrika* **83**, 343-354.

Wang, M. C. and Brookmeyer, R. and Jewell, N. P. (1993). Statistical models for prevalent cohort data. *Biometrics* **49**, 1-11.

Wei, L. J., Ying, Z. and Lin, D. Y. (1993). TrustLinear regression analysis of censored survival data based on rank tests. *Biometrika*, **77**, 845-851.

Wolfson, C., Wolfson, D. B., Asgharian, M., M'Lan, C. E., Ostbye, T., Rockwood, K., Hogan, D. B. and the Clinical Progression of Dementia Study Group (2001). A Reevaluation of the Duration of Survival after the Onset of Dementia. *New Engl. J. Med.* **344**, 1111-1116.

Ying, Z. (1993). A Large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Zelen, M. (2004). Forward and backward recurrence times and length biased sampling: Age specific models. *Lifetime Data Anal.* **10**, 325-334.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56**, 601-614.

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

E-mail: jning@mdanderson.org

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA.

E-mail: jingqin@niaid.nih.gov

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

E-mail: yshen@mdanderson.org

(Received August 2011; accepted December 2012)