Score Estimating Equations from Embedded Likelihood Functions under Accelerated Failure Time Model

JING NING [Assistant Professor]
Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA

JING QIN [Mathematical Statistician]
Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, NIH Bethesda, Maryland 20892, USA

YU SHEN [Professor]
Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA

SUMMARY

The semiparametric accelerated failure time (AFT) model is one of the most popular models for analyzing time-to-event outcomes. One appealing feature of the AFT model is that the observed failure time data can be transformed to identically independent distributed random variables without covariate effects. We describe a class of estimating equations based on the score functions for the transformed data, which are derived from the full likelihood function under commonly used semiparametric models such as the proportional hazards or proportional odds model. The methods of estimating regression parameters under the AFT model can be applied to traditional right-censored survival data as well as more complex time-to-event data subject to length-biased sampling. We establish the asymptotic properties and evaluate the small sample performance of the proposed estimators. We illustrate the proposed methods through applications in two examples.

Keywords

Accelerated failure time model; Cox model; Length-biased data; Likelihood function; Proportional odds model; Score equation

1 Introduction

The accelerated failure time model (AFT), which relates covariates linearly to the logarithm of the survival time, has been one of the most commonly used regression models for analyzing right-censored survival data (Kalbfleisch and Prentice, 2002). The linear regression structure after the log-transformation of failure time and the straightforward interpretation of the regression coefficients are especially appealing to biomedical

jingqin@niaid.nih.gov. jning@mdanderson.org. yshen@mdanderson.org.

investigators. Since the parametric AFT is sensitive to misspecification of the distribution of survival time and to outliers, the literature has focused on conducting estimation and inference procedures under semiparametric AFT models, which does not assume any parametric model for the distribution of residuals (Buckley and James, 1979; Miller and Halpern, 1982; Ritov, 1990; Tsiatis, 1990; Lai and Ying, 1991a; Ying, 1993; Jones, 1997; Lin and Ying, 1995; Jin et al., 2003). The estimation approaches for traditional survival data under the AFT model include rank-based estimating equations, least squares estimators and the kernel-smoothed profile likelihood method (Zeng and Lin, 2007).

In this paper, we describe a unified estimation approach based on score equations derived from the likelihood functions under embedded models for transformed failure time data. Not surprisingly, the derived score estimating equation for traditional right-censored survival data under the proportional hazards embedded model coincides with the rank-based estimating equation originally proposed by Tsiatis (1990).

The unified estimating approach can be applied to estimate regression coefficients in the AFT model for both traditional right-censored and length-biased data. Length-biased survival data arise when the probability of observing the failure time in the target population is proportional to the length of the failure time. Such data are often observed when studying the natural history of a disease from prevalent cohorts. Statistical inference for length-biased right-censored data is generally different from that for traditional survival data, since the length-biased right-censored failure times are not random samples from the target population, and the right censoring times are informative due to the biased sampling scheme. Although several articles (Shen et al., 2009; Chen, 2010; Mandel and Ritov, 2010; Ning et al., 2011) are concerned with the AFT model for length-biased data using different estimating equations, or the least squares method, the estimation efficiency for the existing methods is less than ideal.

The outline of this article is as follows. Section 2 describes a class of score estimating equations derived from the full likelihood functions under embedded models of proportional hazards and proportional odds models. We address the computational issues and establish large sample properties of the estimators from the proposed estimating equations. Section 3 presents the finite sample performance through simulation studies, and is followed by the application of the methods to two real examples in Section 4. We provide a brief discussion in Section 5, and details of the proofs in the Appendix.

2 Full Likelihood-based Estimation Procedures

Let $\tilde{T}$ be the time measured from the onset of the initial event in the target population to failure, and X be a p-vector of the covariates. We consider the semiparametric AFT model (Kalbfleisch and Prentice, 2002), which relates the logarithms of the survival time to the covariate of interest through a linear form,

$$log\tilde{T} = \boldsymbol{X}^T \beta_0 + \epsilon, \quad (1)$$

where β_0 is a $p \times 1$ parameter vector and ε is independent of X with an unspecified distribution function. It is not necessary for the mean of ε to be zero, thus the regression parameter β_0 does not include an intercept. Let q, Q and $\tilde{Q}$ respectively denote the density, cumulative distribution, and survival function of $\exp(\varepsilon)$. One appealing feature of the AFT model is that the observed failure time data can be transformed to a residual time scale, so that the transformed samples are i.i.d. without the covariate effect. Specifically, under the AFT model assumption for $\tilde{T}$ in equation (1), the transformed time $\tilde{T} e^{-X^T \beta_0}$ is independent of the covariates X, thus its embedded likelihood function can be embedded under any commonly used semiparametric model, as we will describe in the following sections.

2.1 Full likelihood approach via Cox model

Consider a study cohort with n subjects. Let the observed data $\{(y_i, \delta_i, x_i), i = 1, \ldots, n\}$ be independent and identically distributed copies of (Y, δ, X), where $Y = \min(T, C)$, C is the censoring time measured from the onset of the initial event, and $\delta = I(T < C)$. We assume that the censoring time C is independent of $\tilde{T}$ conditional on X. Denote the transformed data as $\left\{ \left(s_i(\beta) = y_i e^{-x_i^T \beta}, \delta_i, x_i \right), i = 1, \cdots, n \right\}$ for a given β. We first consider traditional survival data to illustrate the embedded score equation for estimating the regression parameters β. The transformed time $T_0 = \tilde{T} e^{-x^T \beta_0}$, which has null effect for the covariates under the AFT model, can be assumed to follow the Cox proportional hazards model,

$$\lambda(t|X) = \lambda_0(t) \, exp\left(X^T \alpha\right), \quad (2)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\lambda(t|X)$ is the hazard function given covariate X. Note that any practical semiparametric model can be used in this framework as an embedded model. Under assumption (2), the full likelihood on the observed data is proportional to

$$L_{PH}(\Lambda_0, \alpha) = \prod_{i=1}^{n} \left\{ \lambda_0\left(s_i(\beta_0)\right) exp\left(x_i^T \alpha\right) \right\}^{\delta_i} exp\left\{ -\Lambda_0\left(s_i(\beta_0)\right) exp\left(x_i^T \alpha\right) \right\},$$

which leads to the log-likelihood function of

$$l_{PH}(\Lambda_0, \alpha) = \sum_{i=1}^{n} \left[\delta_i \left\{ log\lambda_0\left(s_i(\beta_0)\right) + x_i^T \alpha \right\} - \Lambda_0\left(s_i(\beta_0)\right) exp\left(x_i^T \alpha\right) \right]. \quad (3)$$

Taking the first derivative of the log-likelihood with respect to α, we have the score function of α,

$$S_{PH}(\Lambda_0, \alpha) = \sum_{i=1}^{n} x_i \left[\delta_i - \Lambda_0\left(s_i(\beta_0)\right) exp\left(x_i^T \alpha\right) \right].$$

The transformed residual times are independent and identically distributed under the true β_0. Its corresponding score function evaluated at the truth of the null covariate effect $S_{PH}(\Lambda_0, \alpha)|_{\alpha=0}$ is approximately centered around 0. Hence, the embedded score function can be treated as an estimating function of β:

$$\boldsymbol{E}_{PH}(\Lambda_0, \beta) = \sum_{i=1}^{n} x_i \{\delta_i - \Lambda_0(s_i(\beta))\}. \quad (4)$$

This estimating equation involves an unknown quantity, $\Lambda_0(.)$, which can be replaced by the Nelson-Aalen estimator using the transformed data as

$$\hat{\Lambda}_0(t; \beta) = \sum_{i=1}^{n} \frac{I(s_i(\beta) \leq t)\delta_i}{\sum_{j=1}^{n} I(s_j(\beta) > s_i(\beta))}.$$

Indeed, the estimating equation with the inserted Nelson-Aalen estimator for $\Lambda_0(t)$ is equivalent to the unweighted linear rank test statistic of Tsiastis (1990) and Jones (1997),

$$\sum_{i=1}^{n} \delta_i \left\{ x_i - \frac{\sum_{j=1}^{n} x_j I(s_j(\beta) > s_i(\beta))}{\sum_{j=1}^{n} I(s_j(\beta) > s_i(\beta))} \right\}.$$

Next we describe how a parallel score estimating equation can be proposed from the full likelihood function for length-biased data. Let $\tilde{T}$ and T respectively denote the unbiased failure time in the target population and the observed length-biased failure time, all measured from the initial event, and let $\tilde{A}$ and A respectively represent the truncation time in the target population and the observed truncation time measured from the initial event to the sampling time. Due to the potential residual censoring from the sampling time $(\tilde{C})$ on the residual failure time $(V = T - A)$, we can only observe $Y = \min(T, C)$ and $\delta = I(T < C)$, where $C = \tilde{C} + A$. We follow the standard assumption that the residual censoring time $(\tilde{C})$ is independent of (A, V) conditional on covariate $\boldsymbol{X}$. Data (T, A) can only be observed conditional on $T > \tilde{A}$ under the sampling schema. This sampling mechanism thus induces dependent censoring because $cov(T, C) = cov(A + V, A + \tilde{C}) = var(A) + cov(A, V) > 0$, except for in trivial cases.

It is worth noting that the model structure for T (length-biased) is generally different from that for $\tilde{T}$ (unbiased) in the target population when the time is subject to right-censoring. However, there is a unique feature for length-biased data under the AFT model: after the same transformation $e^{-\boldsymbol{X}^T \beta_0}$ is applied to the unbiased failure time $\tilde{T}$ and the biased failure time T, the transformed outcomes $\tilde{T}e^{-\boldsymbol{X}^T \beta_0}$ and $Te^{-\boldsymbol{X}^T \beta_0}$ are both independent of the covariate $\boldsymbol{X}$. Let $f(./\boldsymbol{x})$ and $F(./\boldsymbol{x})$ represent the probability density function (PDF) and the cumulated distribution function (CDF) for the unbiased time, $\tilde{T}$, given $\boldsymbol{X} = \boldsymbol{x}$. The PDF of the length-biased time, T, is related to the unbiased density of $\tilde{T}$ as follows,

$$f_{LB}(t|x) = \frac{tf(t|x)}{\int_0^\infty uf(u|x)\,du} \equiv \frac{tf(t|x)}{\mu(x)}, \quad (5)$$

where $\mu(x) = \int_0^\infty uf(u|\boldsymbol{x})\,du$. Due to the induced dependent censoring, the probability of observing an uncensored and censored time conditional on covariate are, respectively

$$Pr\left(Y{=}t, A{=}a, \delta{=}1|\boldsymbol{X}{=}x, \tilde{T}{>}\tilde{A}\right) = \frac{f(t|x)\,S_{\tilde{C}}(t-a|x)}{\mu(x)}$$

and

$$Pr\left(Y{=}t, A{=}a, \delta{=}0|\boldsymbol{X}{=}x, \tilde{T}{>}\tilde{A}\right) = \frac{\tilde{F}(t|x)\,f_{\tilde{C}}(t-a|x)}{\mu(x)}$$

where $\bar{F}(t|\boldsymbol{x}) = \int_t^\infty f(u|\boldsymbol{x})\,du$, $f_{\tilde{C}}(.|\boldsymbol{x})$ and $S_{\tilde{C}}(.|\boldsymbol{x})$ are PDF and survival function of the residual censoring time, given $\boldsymbol{X} = x$. Hence, the full likelihood function of the observed right-censored observation (y_i, a_i, δ_i) conditional on the covariates is proportional to

$$\prod_{i=1}^n \frac{f^{\delta_i}(y_i|x_i)\,\bar{F}^{(1-\delta_i)}(y_i|x_i)}{\mu(x)}I(y_i{>}a_i). \quad (6)$$

Under model (1), the survival function, density function and mean of the unbiased failure time, $\tilde{T}$, can be expressed by the survival and density functions of $\exp(\varepsilon)$ as

$$\bar{F}(t|x) = P\left(exp(\epsilon) \geq texp\left(-x^T\beta_0\right)\right) = \bar{Q}\left(texp\left(-x^T\beta_0\right)\right), t{>}0, \quad (7)$$

$$f(t|x) = exp\left(-x^T\beta_0\right)q\left(texp\left(-x^T\beta_0\right)\right), t{>}0, \quad (8)$$

and

$$\mu(x) = \int_0^\infty uexp\left(-x^T\beta_0\right)q\left(uexp\left(-x^T\beta_0\right)\right)du{=}exp\left(x^T\beta_0\right)\mu_0, \quad (9)$$

where $\mu_0 {=} \int_0^\infty uq(u)\,du$.

By equations (7), (8) and (9), the likelihood function (6) can be rewritten as

$$\prod_{i=1}^{n} \frac{\left\{exp\left(-x^T\beta_0\right) q\left(y_i exp\left(-x^T\beta_0\right)\right)\right\}^{\delta_i} \left\{\overline{Q}\left(y_i exp\left(-x^T\beta_0\right)\right)\right\}^{1-\delta_i}}{exp\left(x^T\beta_0\mu_0\right)}.$$

It implies that, by the Jacobian rule, the full likelihood for the transformed, observed data, $\left\{s_i\left(\beta_0\right) = y_i e^{-x_i^T\beta_0}, a_i e^{-x_i^T\beta_0}, i=1,\cdots,n\right\}$ can be expressed as follows

$$
\begin{aligned}
\tilde{L} &\propto \prod_{i=1}^{n} \frac{exp\left(2x^T\beta_0\right)\left\{exp\left(-x^T\beta_0\right)q(s_i(\beta_0))\right\}^{\delta_i}\overline{Q}(s_i(\beta_0))^{1-\delta_i}}{exp(x^T\beta_0)\mu_0} \\
&= \prod_{i=1}^{n} \left\{\frac{q(s_i(\beta_0))}{\mu_0}\right\}^{\delta_i} \left\{\frac{\overline{Q}(s_i(\beta_0))}{\mu_0}\right\}^{1-\delta_i} exp\left\{(1-\delta_i)\,x_i^T\beta_0\right\}.
\end{aligned}
\tag{10}
$$

Under the proportional hazards embedded model assumption with a null covariate effect,

$$q\left(t|x\right) / \overline{Q}\left(t|x\right) = \lambda_0\left(t\right) exp\left(x^T\alpha\right), \overline{Q}\left(t|x\right) = exp\left\{-\Lambda_0\left(t\right) exp\left(x^T\alpha\right)\right\},$$

thus the log-likelihood of the transformed data is

$$
\begin{aligned}
&\tilde{l}_{PH}\left(\Lambda_0,\alpha\right) \\
&= \sum_{i=1}^{n} \left[\delta_i\left\{log\lambda_0\left(s_i\left(\beta_0\right)\right) + x_i^T\alpha\right\} - \Lambda_0\left(s_i\left(\beta_0\right)\right) exp\left(x_i^T\alpha\right) - log\int exp\left\{-\Lambda_0\left(s\right) exp\left(x_i^T\alpha\right)\right\} ds\right] \\
&+ \sum_{i=1}^{n} \left(1-\delta_i\right) x_i^T\beta_0.
\end{aligned}
\tag{11}
$$

The corresponding score function of α evaluated at the null covariate effects ($\alpha = 0$) has a mean of zero (Ning et al., 2010) and can be used as an unbiased estimating equation for solving β,

$$\tilde{E}_{PH}\left(\Lambda_0,\beta\right) = \sum_{i=1}^{n} x_i \left\{\delta_i - \Lambda_0\left(s_i\left(\beta\right)\right) + \frac{\int \Lambda_0\left(s\right) exp\left\{-\Lambda_0\left(s\right)\right\} ds}{\int exp\left\{-\Lambda_0\left(s\right)\right\} ds}\right\}. \tag{12}$$

Note that the full likelihood function (10) for the transformed data is proportional to the Vardid's likelihood by a constant of proportionality. Hence the unknown baseline hazard function $\Lambda_0(t) = -\log(\tilde{Q}(t))$ can be estimated through

$$\tilde{Q}\left(t\right) = 1 - \frac{\int_0^t u^{-1} d\tilde{Q}_{LB}\left(u\right)}{\int_0^\infty u^{-1} d\tilde{Q}_{LB}\left(u\right)}, \tag{13}$$

where $\tilde{Q}_{LB}$ is the nonparametric maximum likelihood estimator of the distribution function (Vardi, 1989), using the observed transformed data solved from the following likelihood,

$$\prod_{i=1}^{n} dQ_{LB}\left(s_i\left(\beta\right)\right)^{\delta_i} \left(\int_{s_{i(\beta)}} u^{-1} dQ_{LB}\left(u\right)\right)^{1-\delta_i}, \quad and \quad dQ_{LB}\left(s\right)=sq\left(s\right)/\mu_0.$$

2.2 Full likelihood approach via the proportional odds model

Because the transformed data under the AFT model are i.i.d., any commonly used semiparametric model may be chosen as an embedded model to derive the score function to estimate the regression coefficients β. To illustrate this general framework, we fit the transformed failure time data using the semiparametric proportional odds model, as an alternative embedded model. Under the proportional odds model, the CDF of $\tilde{T}$ given covariate X is related to the CDF of $\tilde{T}_0$ given $X = 0$ by

$$\frac{1 - F_X\left(t\right)}{F_X\left(t\right)} = \frac{1 - F_0\left(t\right)}{F_0\left(t\right)} exp\left(\boldsymbol{X}^T \gamma\right), \quad (14)$$

where $F_X\left(t\right)$ is the CDF of $\tilde{T}_0$ given covariate X and $F_0(.)$ is the unspecified continuous baseline CDF of $\tilde{T}$ at $X = 0$. It is equivalent to having

$$F_X\left(t\right) = \left\{1 + H^{-1}\left(t\right) exp\left(\boldsymbol{X}^T \gamma\right)\right\}^{-1}, \quad H\left(t\right) = \frac{F_0\left(t\right)}{1 - F_0\left(t\right)}.$$

For traditional survival data, the full likelihood of the transformed data under the proportional odds model is

$$L_{PO}\left(F_0, \gamma\right) = \prod_{i=1}^{n} \left\{dF_X\left(s_i\left(\beta\right)\right)\right\}^{\delta_i} \left(\int_{s_{i(\beta)}\infty}^{\infty} dF_X\left(t\right)\right)^{1-\delta_i}.$$

The resulting log-likelihood function is

$$l_{PO}\left(F_0, \gamma\right) = \sum_{i=1}^{n} \left[x_i^T \gamma - log\left\{H\left(s_i\left(\beta_0\right)\right) + exp\left(x_i^T \gamma\right)\right\}\right] + \sum_{i=1}^{n} \delta_i \left[logdH\left(s_i\left(\beta_0\right)\right) - log\left\{H\left(s_i\left(\beta_0\right)\right) + exp\left(x_i^T \gamma\right)\right\}\right], \quad (15)$$

and its corresponding score function is

$$S_{PO}\left(F_0, \gamma\right) = \sum_{i=1}^{n} \left[x_i - \frac{x_i exp\left(x_i^T \gamma\right)}{\left\{H\left(s_i\left(\beta_0\right)\right) + exp\left(x_i^T \gamma\right)\right\}}\right] - \sum_{i=1}^{n} \delta_i \frac{x_i exp\left(x_i^T \gamma\right)}{\left\{H\left(s_i\left(\beta_0\right)\right) + exp\left(x_i^T \gamma\right)\right\}}.$$

Accordingly, the score function evaluated at the null covariate effects, $S_{PO}\left(F_0, \gamma\right)|_{\gamma=0}$ yields the estimating equations for β,

$$\boldsymbol{E}_{PO}\left(F_0, \beta\right) = \sum_{i=1}^{n} x_i \left\{\delta_i - (1+\delta_i) F_0 \left(s_i \left(\beta\right)\right)\right\}. \quad (16)$$

The unknown CDF, F_0, will be replaced by the nonparametric MLE, Kaplan-Meier estimator for the transformed data in the constructed embedded estimating equation to solve for β.

We can derive the log-likelihood of the observed length-biased survival data from equation (10) under the proportional odds embedded model,

$$\tilde{l}_{PO}\left(F_0, \gamma\right) = \sum_{i=1}^{n} \left[x_i^T \gamma - (1+\delta_i) \log \left\{H \; s_i \left(\beta_0\right)\right) + exp\left(x_i^T \gamma\right)\right\} + \delta_i \log dH \left(s_i \left(\beta_0\right)\right)\right] \\ - \sum_{i=1}^{n} \left[\log\left\{\int_0^\infty \frac{exp\left(x_i^T \gamma\right)}{H(s) + exp\left(x_i^T \gamma\right)} ds\right\}\right] + \sum_{i=1}^{n} \left(1 - \delta_i\right) x_i^T \beta_0. \quad (17)$$

The corresponding estimating function for β, which is constructed based on the score function under the proportional odds model at $\gamma = 0$, follows as

$$\tilde{E}_{PO}\left(F_0, \beta\right) = \sum_{i=1}^{n} x_i \left[\delta_i - (1+\delta_i) F_0 \left(s_i \left(\beta\right)\right) - \frac{\int F_0\left(s\right) \left\{1 - F_0\left(s\right)\right\} ds}{\int \left\{1 - F_0\left(s\right)\right\} ds}\right]. \quad (18)$$

Similar to the construction of the score estimating equation under the proportional hazards model, the unknown baseline CDF, F_0, will be estimated by $\tilde{Q}$ from Vardi's nonparametric MLE of equation (13) using the transformed data.

2.3 Computation procedure

In contrast to the score estimating equations for traditional survival data, note that the score estimating equations for length-biased survival data contain the constant terms from their mean μ_0 under two embedded models. Therefore, the two estimating equations for length-biased data can be further simplified by centering the covariates around their sample means to

$$\tilde{\boldsymbol{E}}_{PH}\left(\Lambda_0, \beta\right) = \sum_{i=1}^{n} x_i \left\{\delta_i - \Lambda_0 \left(s_i \left(\beta\right)\right)\right\} \quad (19)$$

and

$$\tilde{\boldsymbol{E}}_{PO}\left(F_0, \beta\right) = \sum_{i=1}^{n} x_i \left\{\delta_i - (1+\delta_i) F_0 \left(s_i \left(\beta\right)\right)\right\}. \quad (20)$$

An immediate consequence of this normalization is that the above estimating equations have the same expressions as the score estimating equations (4) and (16) for traditional survival data. However, the estimations of $\Lambda_0(t)$ and $F_0(t)$ are different for two types of data: Vardi's

estimator will be used for length-biased survival data and the Kaplan-Meier estimator will be used for traditional survival data.

Because the proposed estimating equations are not continuous functions of β, exact solutions to the estimating equations may not always exist. Using the principle similar to that in Jones (1997), the solutions to the estimating equations could be defined as the minimizers of the Euclidean norm of $\boldsymbol{E}_{PH}(\Lambda_0, \beta)$, $\boldsymbol{E}_{PO}(F_0, \beta)$, $\tilde{\boldsymbol{E}}_{PH}(\Lambda_0, \beta)$ and $\tilde{E}_{PO}(F_0, \beta)$, respectively. This type of minimization problem cannot be solved by using a standard optimization algorithm designed for continuous functions. One solution is to use a grid search method, which turns out to be impractical due to the intensive computation. We adopt the following iterative procedure to solve for β. Using the case of traditional survival data to illustrate, we fix an initial value $\hat{\beta}^{(0)}$ and transform the observed data $s_i\left(\hat{\beta}^{(0)}\right) = y_i exp\left(-\boldsymbol{x}_i^T \hat{\beta}^{(0)}\right)$. We then estimate $\Lambda_0(t)$ and $1 - F_0(t)$ using the Nelson-Aalen estimator $\hat{\Lambda}\left(t; \hat{\beta}^{(0)}\right)$ and Kaplan-Meier estimator $1 - \hat{F}\left(t; \hat{\beta}^{(0)}\right)$, respectively. Next, we search for $\hat{\beta}^{(1)}$ by minimizing the norm of the score estimating equations, as follows,

$$\boldsymbol{E}_{PH}\left(\hat{\Lambda}_0, \beta\right) = \sum_{i=1}^{n} x_i \left\{\delta_i - \hat{\Lambda}_0\left(y_i exp\left\{-x_i^T \beta\right\}; \hat{\beta}^{(0)}\right)\right\} \quad (21)$$

$$\boldsymbol{E}_{PO}\left(\hat{F}_0, \beta\right) = \sum_{i=1}^{n} x_i \left[\delta_i - (1+\delta_i) \hat{F}_0\left(y_i exp\left\{-x_i^T \beta\right\}; \hat{\beta}^{(0)}\right)\right]. \quad (22)$$

The iterative procedures continue until the pre-specified convergence criteria is met. We denote the final solutions of equations (4) and (16) by $\hat{\beta}_{PH}$ and $\hat{\beta}_{PO}$.

For length-biased survival data, the corresponding two sets of estimating equations in the mth step can be written as

$$\tilde{\boldsymbol{E}}_{PH}\left(\tilde{\Lambda}_0, \beta\right) = \frac{1}{n}\sum_{i=1}^{n} \int_0^\tau x_i \left\{\delta_i - \tilde{\Lambda}_0\left(t; \tilde{\beta}_{PH}^{(m-1)}\right)\right\} dN_i(\beta, t) \quad (23)$$

$$\tilde{\boldsymbol{E}}_{PO}\left(\tilde{F}_0, \beta\right) = \frac{1}{n}\sum_{i=1}^{n} \int_0^\tau x_i \left\{\delta_i - (1+\delta_i) \tilde{F}_0\left(t; \tilde{\beta}_{PO}^{(m-1)}\right)\right\} dN_i(\beta, t), \quad (24)$$

where $N_i(\beta, t) = I\left(Y_i e^{-x_i^T \beta} \leq t\right)$. Via the transformed times, we will use Vardi's nonparametric maximum likelihood (Vardi, 1989) to estimate the baseline CDF, $\tilde{F}_0\left(u; \tilde{\beta}_{PO}^{(m-1)}\right)$, or the baseline cumulative hazard function,

$$\tilde{\Lambda}_0\left(u;\tilde{\beta}_{PH}^{(m-1)}\right) = -log\left\{1 - \tilde{F}_0\left(u;\tilde{\beta}_{PH}^{(m-1)}\right)\right\}$$

under the assumption that the baseline CDF is continues function. Then the $m-th$ step estimator, $\tilde{\beta}_{PH}^{(m)}$ or $\tilde{\beta}_{PO}^{(m)}$ can be obtained by minimizing the norm of the above score estimating equations. It is worth noting that the nonparametric maximum likelihood estimator for length-biased data based on Vardi's estimator jumps at all unique failure and censoring time points, but does not have a closed-form expression and can be obtained via the EM algorithm.

When there are more than a few covariates, the aforementioned method to search the minimizers for the norm of the score estimating functions is computationally intensive. Alternatively, we can apply the L_1– minimization computational technique to find the solutions of the estimating equations (Jin et al., 2003; Shuang et al., 2012). Note that estimating equations (23) and (24) are monotone equations in each component of β, and are respectively the gradients of the L_1–type convex functions. For example, equation (23) equals the gradient of the following L_1–type convex function (Shuang et al., 2012),

$$L_1\left(\beta;\tilde{\beta}_{PH}^{(m-1)}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{h}\tilde{\lambda}_j^{(m-1)}\left(|logt_j - logy_i + x_i^T\beta| + |R^* - x_i^T\beta|\right) + |R^* - \sum_{i=1}^{n}\delta_i x_i^T\beta|, \quad (25)$$

where $\{t_1 < t_2 < \ldots < t_h\}$ are the ordered unique failure and censoring times for $\{y_1,\ldots,y_n\}$, $\left\{\tilde{\lambda}_{t_1},\cdots,\tilde{\lambda}_{t_h}\right\}$ are the corresponding positive masses of $\tilde{\Lambda}_0\left(u;\tilde{\beta}_{PH}^{(m-1)}\right)$ at the times $\{t_1 < t_2 < \ldots < t_h\}$, and R^* is a sufficiently large positive number (e.g., 10^6) which should be larger than $|x_i^T\beta|$ and $|\Sigma_{i=1}^n \delta_i x_i^T\beta|$ for all β's in the compact parameter space. Then, the updated estimator $\tilde{\beta}_{PH}^{(m)}$ can be readily obtained by using the Barrodale-Roberts algorithm (Barroda and Roberts, 1974), which is implemented in standard statistical software, such as the rq() function in R.

2.4 Asymptotic properties

For traditional survival data, the weak convergence of $\hat{\beta}_{PO}$ will be derived similarly to that of $\hat{\beta}_{PH}$ as in Jones (1997) and Tsiatis (1990). In this section, we focus on establishing asymptotic properties of the proposed score equations (23) and (24) in the mth step and their corresponding estimators $\tilde{\beta}_{PH}^{(m)}$ and $\tilde{\beta}_{PO}^{(m)}$ for length-biased survival data. The key step of the proof is to establish the asymptotic linearity of the estimating equations,

$\tilde{\boldsymbol{E}}_{PH}\left(\tilde{\Lambda}_0,\beta\right)$ and $\tilde{\boldsymbol{E}}_{PO}\left(\tilde{F}_0,\beta\right)$, for $\boldsymbol{\beta}$ close to the true value, denoted by β_0. Since the estimating equations derived from the score functions are not continuous functions of β, Taylor's expansion theorem is not applicable here. Following the arguments for asymptotic properties of linear rank-based estimators (Ying, 1993), and for large sample properties of nonparametric maximum likelihood estimators for length-biased and right-censored data

(Asgharian and Wolfson, 2005), we prove that, for any fixed m, both $\tilde{\beta}_{PH}^{(m)}$ and $\tilde{\beta}_{PO}^{(m)}$ are consistent and asymptotically normally distributed given a consistent and asymptotically normal initial estimator under the regularity conditions listed in the Appendix.

Theorem 1. Under regularity conditions (C1)-(C10) listed in the Appendix, both $\tilde{\beta}_{PH}^{(m)}$ and $\tilde{\beta}_{PO}^{(m)}$ are consistent estimators of β_0 and $\sqrt{n}\left(\tilde{\beta}_{PH}^{(m)} - \beta_0\right)$ and $\sqrt{n}\left(\tilde{\beta}_{PO}^{(m)} - \beta_0\right)$ converge weakly to normal distributions with mean zero and variance-covariance matrix $\Sigma_{PH}^{(m)}$ and $\Sigma_{PO}^{(m)}$, respectively.

The variance-covariance matrices, $\Sigma_{PH}^{(m)}$ and $\Sigma_{PO}^{(m)}$, are defined in the Appendix. The detailed proofs of Theorem 1 are provided in the Appendix. The estimation of the variance-covariance matrix is not straightforward because of the unknown density function involved in $\Sigma_{PH}^{(m)}$ and $\Sigma_{PO}^{(m)}$. Given the aforementioned weak convergence, we will use the bootstrap resampling method to approximate the variances of $\tilde{\beta}_{PH}^{(m)}$ and $\tilde{\beta}_{PO}^{(m)}$.

While both score estimating equations are valid under the assumed embedded models for large sample properties, an interesting question is which embedded model leads to a more efficient estimator of β. Indeed, there is not a uniformly better choice in general. We will conduct a series of simulation studies to compare the two estimating equation approaches for both traditional and length-biased survival data.

We next show the convergence of the proposed iterative algorithm for the estimation using estimating equation (23) as an illustration. From any initial point in the bounded closed n-dimensional rectangle $\mathbf{P} = \{(p_1, \cdots, p_n) \in (0, 1)^n : \Sigma_{i=1}^n p_i = 1\}$ (Davidov and Iliopoulos, 2013; Vardi, 1989), it can be shown that Vardi's algorithm using the transformed data, $y_i e^{-x_i^T \tilde{\beta}_{PH}^{(m-1)}}$, converges to the nonparametric maximum likelihood estimator (Vardi's estimator), denoted as $\tilde{\boldsymbol{P}}_{PH}^{(m)} = f^*\left(\tilde{\beta}_{PH}^{(m-1)}\right)$. Given the updated estimator of the cumulative hazard function via Vardi's estimator, we obtain the mth step estimator of the regression coefficients, $\tilde{\beta}_{PH}^{(m)}$, by minimizing the norm $\|\tilde{\boldsymbol{E}}_{PH}\left(\tilde{\Lambda}_0^{(m)}, \beta\right)\|$, denoted as $\tilde{\beta}_{PH}^{(m)} = q_n^*\left(\tilde{\beta}_{PH}^{(m-1)}, \tilde{\Lambda}_0^{(m)}\right)$. Here, function q_n^* a discontinuous step function. By using the fixed point theorem for discontinuous mappings (Cromme and Diener, 1991), there exists a point $\tilde{\beta}^* \in \mathbf{B}$, such that

$$\|q_n^*\left(\tilde{\beta}^*, \tilde{\Lambda}_0^{(m)}\right) - \tilde{\beta}^*\| \leq \sup_{b \in \mathbf{B}} \lim_{\epsilon \to 0} \sup \sup_{b' \in B_\epsilon(b)} \|q_n^*\left(b, \tilde{\Lambda}_0^{(m)}\right) - q_n^*\left(b', \tilde{\Lambda}_0^{(m)}\right)\|,$$

where $B_{(\varepsilon)}$ denotes the ε-neighborhood of b. After some simple algebra, we can show that

$$\|q_n^*\left(b,\tilde{\Lambda}_0^{(m)}\right)-q_n^*\left(b^{'},\tilde{\Lambda}_0^{(m)}\right)\|\leq\|q_n^*\left(b,\tilde{\Lambda}_0^{(m)}\right)-q^*(b)\|+\|q_n^*\left(b^{'}\tilde{\Lambda}_0^{(m)}\right)-q^*\left(b^{'}\right)\|+\|q^*(b)-q^*\left(b^{'}\right)\|,$$

where $q^*(b)=q_n^*\left(\tilde{\beta}_{PH}^{(m-1)},\Lambda_0\right)$. Thus, by the uniform consistency of Vardi's estimator (Asgharian and Wolfson, 2005) and the continuity of the underlying CDF, we have,

$$\underset{b\in\mathbf{B}}{suplim}\underset{\epsilon\to 0}{sup}\underset{b^{'}\in B_\epsilon(b)}{sup}\|q_n^*(b)-q_n^*\left(b^{'}\right)\|\to_{n\to\infty}=0.$$

Summarizing the previous arguments, we can see that the sequence $\tilde{\beta}_{PH}^{(m)}, m=0, 1, \cdots,$ converges to β_0 from any consistent initial point $\tilde{\beta}_{PH}^{(m)}$.

3 Simulation

We evaluated the finite sample performance of the two types of estimators for both traditional and length-biased survival data. We generated failure times from the AFT model with two covariates:

$$log\tilde{T}=0.5X_1+X_2+\epsilon, \quad (26)$$

where $X_1 \sim Binomial(0.5)$ and $X_2 \sim Uniform(0, 1)$. We considered two error distributions: a standard normal distribution with N(0,0.5) and a uniform distribution U(−0.5,0.5). Censoring times were generated independent of the covariates from uniform distributions, and the censoring percentages ranged from 15% to 50%. Each study comprised 1000 runs. A cohort size of 100 or 200 was used.

3.1 Traditional survival data

We first compare the finite sample performance of the estimators from two embedded score equations for traditional right-censored data. Simulation results for right-censored data are shown in Table 1. The true value β_0 is chosen to be (0.50, 1.00). A bias is calculated as the average of the differences between the estimators and the true value. The biases of the estimators from the estimating equation E_{PO} are slightly smaller than those from the estimating equation E_{PH}, although the biases from both methods are mostly close to zero. Interestingly, there is no uniform better estimating equations from the two embedded models in terms of statistical efficiency: when the random error follows the uniform distribution, the estimating equation E_{PH} is more efficient than the estimating equation E_{PO} ; when the random error follows the normal distribution, the estimating equation E_{PO} is more efficient than the estimating equation E_{PO}. However, the difference in efficiency decreases with the degree of censoring no matter what distribution the random error has.

We also compare the performances of two proposed score estimating equations with those of the inverse weighting estimating equation approach (Zhou, 1992) and the Buckley-James

method (Buckley and James, 1979). Generally, the inverse weighting estimating equation approach is least efficient with heavy censoring (50%). Compared with the two score estimating equations, the Buckely-James method is more efficient with normally distributed random errors, but is less efficient with uniformly distributed random errors.

3.2 Length-biased survival data

We next evaluate the performance of the two proposed estimators for length-biased data and compared them with the estimators from the inverse weighting estimating equation approach (Shen et al., 2009) and the Buckley-James method (Ning et al., 2011). To generate length-biased data with potential right-censoring, we first generate 2000 unbiased failure times from equation (26). We then sample the failure times from these unbiased times with weights proportional to their lengths. This procedure will be repeated until the desired sample size is reached. For each subject in the selected cohort, we generate its truncation time a_i from the uniform distribution, with range from 0 to t_i, where t_i is the observed failure time. The residual censoring time of this subject is independently generated from uniform distributions, and the censoring indicator is obtained by $\delta_i = I(t_i \quad c_i)$. The other aspects of the simulation settings are similar to those for traditional right-censored data.

Table 2 summarizes the empirical biases and empirical standard errors of four different methods for length-biased right-censored data: (I) the inverse weighting estimating equation method (Shen et al., 2009); (II) Buckley-James method (Ning et al., 2011); (III) the estimating equation $\tilde{E}_{PH}$ with proportional hazards embedded model and (IV) the estimating equation $\tilde{E}_{PO}$ with proportional odds embedded model. As expected, the two score estimating equation methods outperform the inverse weighting estimating equation and Buckley-James method. The proposed methods have negligible biases and smaller empirical standard errors compared with the other two existing methods, especially when there is heavy censoring (50%). For instance, the standard errors associated with the inverse estimating equation were 1.53 to 1.71 times greater, and the standard errors associated with the Buckley-James method were 1.39 to 1.63 times greater than those associated with the estimating equation $\tilde{E}_{PH}$ based on a sample size of 200 and uniformly distributed random errors. Similar to the performance for traditional survival data without length-biased sampling, there is no uniform better estimating equation from the two embedded models. Interestingly, the standard errors for the two proposed estimators seem to be more robust to the degree of right censoring for length-biased data than to those for traditional survival data. This phenomenon may be partly explained by the difference in Vardi's estimator and the Kaplan-Meier estimator for the nonparametric MLE of the survival distribution with and without length-biased sampling, respectively. In contrast to the Kaplan-Meier estimator, the former is estimated on both failure and censoring times.

We compare the computation time for the two computational algorithms in Section 2.3. We consider the scenario in which the length-biased data are generated from equation (26) with uniformly distributed random errors, and apply the score estimating equation (23) under the Cox model. When the sample size is 100, the number of covariates is 2 and the censoring rate is 50%, we make the following observations: (i) For 500 runs, the computational time of the L_1-minimization is 24 minutes using a 3.40GHz desktop CPU, while the time of directly

minimizing the norm of the score estimating equations (23) is around 60 minutes. (ii) When the number of covariates increases to four, two continuous covariates from Uniform(0,1) and two binary covariates from Binomial(0.5), the computational time of the L_1-minimization slightly increases to 30 minutes, and the time of directly minimizing the norm of the score estimating equations (23) increases to 77 minutes. (iii) The biases and empirical standard errors from the two computational algorithms are almost identical. For example, in the scenario of two covariates, the biases are (0.00, 0.00) versus (0.00, 0.01), and the empirical standard errors are (0.062, 0.113) vs (0.063, 0.112). Based on these simulation studies, both algorithms have acceptable performance, but the L_1-minimization algorithm outperforms the algorithm for minimizing the norm of the estimating equations in terms of computational time. Further, such an advantage would be more pronounced with increasing numbers of covariates.

4 Data Application

4.1 Bladder Cancer Data Example

We applied the proposed estimating methods to a bladder cancer study with traditional right-censored data (Sharma et al., 2007). Among 69 patients with bladder cancer, 33 had recurrence, and 15 died during the follow-up. The analysis goal is to investigate the association between the presence of CD8 tumor-infiltrating cytotoxic T lymphocytes (TILs), disease stage and the disease-free survival time. We used the AFT model to explore the association between the presence of CD8 TILs, pathologic disease stage and the time to disease progression or death. The presence of CD8 was defined as a binary variable: whether or not the number of CD8 TILs was greater than the median number (8) of CD8 TILs among all patients analyzed. The pathologic disease stage was also stratified as a binary variable, with early superficial disease (PT1 or PTa) compared with advanced, muscle-invasive disease (PT2, PT3, or PT4). We applied the proposed estimating equations derived from Cox proportional hazards and proportional odds ratio models, the inverse weighting method (Zhou, 1992), and the Buckley-James method (Buckley and James, 1979) to estimated regression coefficients. The estimated regression coefficients and their bootstrap standard errors are listed in Table 3.

The score estimating equation and Buckley-James methods all showed that the number of CD8 TILs and pathologic disease stage were significantly associated with disease-free survival time. The risk of disease progression or death is significantly higher for patients with lower numbers of CD 8 TILs and for patients with advanced muscle-invasive disease. Due to the high percentage of censoring (52%), the less efficient inverse weighting method did not show such an association.

4.2 Canadian Study of Health and Aging

We next illustrate the proposed likelihood-based methods for length-biased data by analyzing a prevalent cohort study, the Canadian Study of Health and Aging. The design of this prevalent cohort study and its main study results were fully described in Wolfson et al. (2001). In this cohort, subjects with dementia were identified and classified into three subcategories of dementia: Alzheimer's disease, possible Alzheimer's disease, or vascular

dementia. We aim to assess the relationship between the subtype of dementia and the survival time, from the dementia onset to death, by fitting an accelerated failure time model with two indicators of dementia subtypes. We considered a subset of the study data by excluding those with missing date of onset or classification of dementia subtype. A total of 818 patients with dementia were included in our analysis. Among them, 393 had a diagnosis of probable Alzheimer's disease, 252 had possible Alzheimer's disease, and 173 had vascular dementia. The observed event times were subject to length-biased sampling because subjects who died quickly after dementia onset were more likely to be excluded from the study.

The stationarity assumption was examined and confirmed for this cohort study by Addona and Wolfson (2006). We applied the proposed methods, the inverse weighting method and the Buckely-James approach under the AFT model, $\log \tilde{T} = a_1 X_1 + a_2 X_2 + \varepsilon$, to evaluate the association between different diagnostic subcategories of dementia and survival, where X_1 and X_2 indicate whether the subject had vascular dementia or probable Alzheimer's disease, respectively. The estimated regression coefficients and their bootstrap standard errors are listed in Table 4. Using the proposed score estimating equations, we found that the long-term survival distributions are significantly different between patients with vascular dementia and those with possible Alzheimer's dementia. In contrast, the less efficient methods (inverse weighting method and Buckley-James method) could not detect the difference in survival distributions between the patients with vascular dementia and those with possible Alzheimer's dementia.

5 Discussion

For traditional survival data under the AFT model, Tsiatis (1990) proposed an important rank-based weighted linear rank test using the transformed i.i.d failure times. Sharing the same spirit, we have constructed a unified class of estimation approaches by using the full likelihood function under two embedded semiparametric regression models based on the transformed times. The score function derived under the Cox embedded model with null covariate effects is reduced to the linear rank test statistic of Tsiatis (1990) for traditional survival data. One major advantage to viewing the estimation problem from this unified structure is that the estimating approach can be easily applied to both traditional and length-biased survival data for different embedded models. Additionally, the general estimation principle discussed in this paper can be applied to other types of survival data, such as interval censoring data.

In addition to the kernel-smoothed profile likelihood proposed by Zeng and Lin (2007), there has been considerable work on the estimation of covariate effects and on variance under the AFT model for traditional survival data by improving the selection of the weight in the weighted linear rank type of estimating equations (Prentice, 1978; Wei et al., 1990; Lai and Ying, 1991b; Robins and Tsiatis, 1992; Lin and Ying, 1995). There has been little work based on alternative semiparametric embedded models, such as using a proportional odds model with null effect for the rescaled data. In fact, under the AFT model for the failure times, this score-based estimating equation can be derived from a full likelihood embedded in any alternative semiparametric model with a null covariate effect for

transformed data as long as it is computationally feasible. Although we have focused on the proportional hazards and proportional odds models to illustrate the principle behind the method, more general types of linear transformation models can be used as well for transformed data under the AFT model.

Interestingly, this class of score-based embedded estimating equations shares the same expressions for length-biased data as for traditional survival data under the AFT model after normalizing the covariates; although the estimation of baseline hazards function or survival function is different for the two types of data. In contrast to the Nelson-Aalen and Kaplan-Meier estimators for traditional data, Vardi's estimator for length-biased data jumps on both failure and censoring times, which may lead to robust estimators under heavy censoring, as noted in our simulations. From the empirical studies, it is interesting to note that the score equation embedded in the Cox embedded model may not be the most efficient approach, but depends on the underlying error distribution. Compared to existing methods in the literature that are based on estimating equations, the proposed score-based estimating equations lead to more efficiency gain, as shown in the simulation studies (Tables 1 and 2). An efficiency gain is achieved because the proposed estimating equations are directly derived from the embedded full likelihood function. Another advantage for the proposed method for analyzing length-biased data is that the censoring distribution is not required to be estimated as in the other existing methods, e.g., Shen et al. (2009).

In this paper we have studied score estimating equations derived from embedding the underlying density or hazard in a larger semiparametric model. The basic idea is similar to Neyman's smooth goodness of fit test (Neyman, 1937), where an observed data set is tested against a specific density. A score test is obtained by embedding this density in a larger parameter family. A comprehensive discussion of this test can be found in the recent book by Rayner et al. (2011). In contrast to testing a specific density, we use the embedded likelihood method to deduce an unbiased estimating function for the regression parameter β in the AFT model. Based on our simulation studies, different enlarged models may yield different levels of estimation efficiency. It would be worthwhile to further investigate the choices for the enlarged models.

Acknowledgments

6 Appendix

Regularity conditions

To establish the large-sample properties of the proposed estimator with the embedded models, we impose the following conditions:

(C1) X is uniformly bounded, and if there exists a constant vector b such that $b^T X = 0$ with probability one, then $b = 0$.

(C2) The parameter space of β, $\mathbf{B}$, is a compact set including the true value of parameter β_0.

(C3) The density function, $q(.)$, and its derivative $q'(.)$ are bounded.

(C4) The residual censoring time has a uniformly bounded density $f_{\tilde{c}}$, that is there exists a real number such that $f_{\tilde{c}}(t) < M$ for all $t \in [0, \tau]$.

(C5) $Q(.)$ is a continuous and differentiable distribution function over $(0, \tau)$, where $\tau = \inf\{t : Q(t) = 1\} < \infty$.

(C6) $\left\{ \dfrac{2\tau}{\int_0^\tau S_{c_0}(u)\, du} - \dfrac{1}{S_{c_0}(0)} \right\} \{1 - S_{c_0}(0)\} < 1$, where S_{c0} is the survival function of $C_0 = \tilde{C} \exp\{-X^T \beta_0\}$.

(C7) $P(\tilde{T} < C) > 0$.

(C8) $E\left[\left\{ \delta \left(\log Y - X^T \beta_0 \right) \right\} / \left\{ \int_0^Y S_{\tilde{c}}(u)\, du \right\} \right]^2 < \infty$, where S_c is the survival function of the residual censoring time.

(C9)

$\int_0^\tau \{ \int_s^\tau S_{\tilde{c}}(u)\, du \}^2 / \left\{ S_{\tilde{c}}^2(s) S_v(s) \right\} dS_{\tilde{c}}(s) < \infty \quad and \quad \int_0^\tau \left\{ \mathbf{D}(s)^{\otimes 2} \right\} / \left\{ S_{\tilde{c}}^2(s) S_v(s) \right\} dS_{\tilde{c}}(s) < \infty$

, where $\mathbf{D}(s) = E\left[\left\{ \delta X I(Y \geq s) \left(\log Y - X^T \beta_0 \right) \int_s^Y S_{\tilde{c}}(u)\, du \right\} / \left\{ \int_0^Y S_{\tilde{c}}(u)\, du \right\} \right]$ and S_v is the survival function of V.

(C10) The functions $m_{PH}(\beta, t)$ and $m_{PO}(\beta, t)$ have a unique solution at a compact region containing β_0.

Assumptions (C5) to (C7) are to ensure uniform consistency and weak convergence of Vardi's estimator $\tilde{F}_0$ for all $0 < t \quad \tau$ (Asgharian et al., 2002; Asgharian and Wolfson, 2005). Assumptions (C8)-(C9) are for the consistency and asymptotic normality of the initial estimator (Shen et al., 2009).

Sketched proof of Theorem 1

Given a consistent and asymptotically normally distributed initial estimator, we need to prove that, for any fixed m, $\tilde{\beta}_{PH}^{(m)} \quad and \quad \tilde{\beta}_{PO}^{(m)}$ are consistent and asymptotically normally

distributed. Note that it suffices to show the asymptotic properties of $\tilde{\beta}_{PO}^{(m)}$ for $m = 1$. For notational simplicity, we assume covariate-independent censoring and use the estimator obtained from IWEE (Shen et al., 2009), which has a closed-form expression,

$$\tilde{\beta}_{PH}^{(0)} = \left\{ \sum_{i=1}^{n} \frac{\delta_i x_i x_i^T}{\int_0^{y_i} \hat{S}_{\tilde{c}}(u)\,du} \right\}^{-1} \sum_{i=1}^{n} \frac{\delta_i x_i \log y_i}{\int_0^{y_i} \hat{S}_{\tilde{c}}(u)\,du}, \quad (27)$$

as the initial value to show the asymptotic behavior of $\tilde{\beta}_{PH}^{(1)}$, where $\hat{S}_{\tilde{c}}(.)$ is the Kaplan-Meier estimator for the residual censoring survival function. Note that there is no required assumption that the censoring time C is independent of X. A consistent estimator for length-biased data always exists, whether the censoring depends on the covariates or not (Lai and Ying, 1991b), though this initial estimator is not an efficient estimator of β without adequately using the left-truncation time. The arguments here can be readily extended to the case of covariate-dependent censoring and the initial estimator from the left-truncation method (Lai and Ying, 1991b).

We first study the asymptotic behavior of $\tilde{E}_{PH}\left(\tilde{\Lambda}_0, \beta\right)$ at true value β_0. Let $\hat{\varepsilon}$ and ε represent the sample empirical mean and the limit of average expectation. Using these notations, estimating equation $\tilde{E}_{PH}$ at the first step can be expressed as

$$\tilde{E}_{PH}\left(\tilde{\Lambda}_0, \beta_0\right) = \hat{\epsilon}\left[X \beta N\left(\beta_0, \tau\right)\right] - \int_0^{\tau} \tilde{\Lambda}_0\left(t; \tilde{\beta}_{PH}^{(0)}\right) d\hat{\epsilon}\left[X N\left(\beta_0, t\right)\right], \quad (28)$$

where $N\left(\beta, t\right) = I\left(Y e^{-x^T\beta} \leq t\right)$. By the central limit theorem for variables and processes,

$$\sqrt{n}\left\{\hat{\epsilon}\left[X \delta N\left(\beta_0, \tau\right)\right] - \epsilon\left[X \delta N\left(\beta_0, \tau\right)\right]\right\}$$

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

$$\sqrt{n}\left\{\hat{\epsilon}\left[X N\left(\beta_0, t\right)\right] - \epsilon\left[X N\left(\beta_0, t\right)\right]\right\}$$

converges to a Gaussian process, W_2. The baseline cumulative hazard function is estimated through Vardi's estimator and the transferred data, $\tilde{\Lambda}_0\left(t; \tilde{\beta}_{PH}^{(0)}\right) = -\log\left\{1 - \tilde{Q}\left(u; \tilde{\beta}^{(0)}\right)\right\}$, where

$$\tilde{Q}\left(t; \tilde{\beta}^{(0)}\right) = 1 - \frac{\int_0^t u^{-1} d\tilde{Q}_{LB}\left(u; \tilde{\beta}^{(0)}\right)}{\int_0^{\infty} u^{-1} d\tilde{Q}_{LB}\left(u; \tilde{\beta}^{(0)}\right)}.$$

$$y_i\left(\beta\right) = y_i exp$$

$$-\left\{x_i^T\beta\right\}, Y_0$$

$$= Y exp\left\{-x_i^T\beta_0\right\}, T_0$$

$$= T exp\left\{-x_i^T\beta_0\right\}, G^*\left(t\right)$$

$$= P\left(Y_0 \le t|\delta=1\right), F^*\left(t\right)$$

$$= P\left(Y_0 \le t|\delta=0\right), h\left(t\right)$$

$$= \int_t^\infty z^{-1}dQ_{LB}\left(z\right), p$$

$$= P\left(\delta=1\right), \hat{p}$$

Define $=\Sigma_{i=1}^n \delta_i/n$ and $g^*(x)$ and $f^*(x)$ are the conditional density functions of $G^*(x)$ and $F^*(x)$, respectively. Let $n_1=\Sigma_{i=1}^n\delta_i, n_0=n-n_1$. Define

$$W_{1,\beta}\left(t\right) = \sqrt{n_1}\left\{\frac{1}{n_1}I\left(y_i\left(\beta\right)\le t, \delta_i=1\right) - G^*\left(t\right)\right\}$$

and

$$W_{0,\beta}\left(t\right) = \sqrt{n_0}\left\{\frac{1}{n_0}I\left(y_i\left(\beta\right)\le t, \delta_i=0\right) - F^*\left(t\right)\right\}.$$

Asgharian and Wolfson (2005) derived the expression of $\tilde{Q_{LB}}(t;\beta) - Q_{LB}(t)$ as a linear functional of $W_{1,\beta}, W_{0,\beta}$ and $\hat{p} - p$,

$$\tilde{Q}_{LB}\left(t;\beta\right) - Q_{LB}\left(t\right) = F_n^{-1}\left(\frac{V_\beta\left(t\right)}{\sqrt{n}}\right), \quad (29)$$

where

$$V_\beta\left(t\right) = p^{1/2}W_{1,\beta}\left(t\right) + \left(1-p\right)^{1/2}\hat{h}(t)\int_{0<z\le t}W_{0,\beta}\left(z\right)d\frac{1}{\hat{h}\left(z\right)} + \left(\frac{p}{1-p}\right)\left\{G^*\left(t\right) - Q_{LB}\left(t\right)\right\}\sqrt{n}\left(\hat{p}-p\right),$$

$$F_n\left(u\right)\left(t\right)$$

$$= p\left(\frac{1-\hat{p}}{1-p}\right)\int_{0<x\le t}\frac{g^*\left(x\right))}{qLB\left(x\right)}du\left(x\right)$$

$$+ \left(\frac{\hat{p}-p}{1-p}\right)u\left(t\right)$$

$$+ \left(1-p\right)\left(\frac{1-\hat{p}}{1-p}\right)\int_{0<x\le t}y\left(\int_{y\le z}\frac{u\left(z\right)}{z^2}dz\right)d\left\{\left(\frac{\hat{h}\left(t\right)}{\hat{h}\left(y\right)}-1\right)\frac{f^*\left(y\right)}{h\left(y\right)}\right\}.$$

In addition, it follows from the asymptotic properties of the empirical processes $W_{1,\tilde{\beta}_{PH}^{(0)}}$ and $W_{0,\tilde{\beta}_{PH}^{(0)}}$ (Lemma 19.24 of van der Vaart (1998)), that

$$W_{0,\tilde{\beta}_{PH}}(t) = W_{0,\beta_0}(t) + f^*(t)\left(\beta_0 - \tilde{\beta}_{PH}^{(0)}\right) + o_p(1)$$
$$W_{1,\tilde{\beta}_{PH}}(t) = W_{1,\beta_0}(t) + f^*(t)\left(\beta_0 - \tilde{\beta}_{PH}^{(0)}\right) + o_p(1).$$

Therefore, $\sqrt{n}\left\{\tilde{Q}_{LB}\left(t;\tilde{\beta}_{PH}^{(0)}\right) - \tilde{Q}_{LB}(t;\beta_0)\right\}$ can be shown to be asymptotically linear in $\tilde{\beta}_{PH}^{(0)}$,

$$
\begin{aligned}
&\sqrt{n}\left\{\tilde{Q}_{LB}\left(t;\tilde{\beta}_{PH}^{(0)}\right) - \tilde{Q}_{LB}(t;\beta_0)\right\}\\
=\ & F^{-1}\left[p^{1/2}\left\{W_{1,\tilde{\beta}_{PH}^{(0)}}(t) - W_{1,\beta_0}(t)\right\} + (1-p)^{1/2}h(t)\int_{0<z\le t}\left\{W_{0,\tilde{\beta}_{PH}^{(0)}}(z) - W_{0,\beta_0}(z)\right\}d\frac{1}{h(z)}\right] + o_p(1)\\
=\ & A_1(t)\sqrt{n}\left(\tilde{\beta}_{PH}^{(0)} - \beta_0\right) + o_p(1),
\end{aligned}
$$

where

$$\mathbf{A}_1(t) = F^{-1}\left\{p^{1/2}g^*(t)\,t\int x^T e^{x^T\beta_0}d\phi(x) + (1-p)^{1/2}h(t)\int_{0<z\le t}f^*(z)\,z\int x^T e^{x^T\beta_0}d\phi(x)\,d\frac{1}{h(z)}\right\},$$

and ϕ is the distribution function of $\mathbf{X}$.

Using Lemma 2 of Vardi and Zhang (1992) and Lemma 3 of Asgharian and Wolfson (2005), we have

$$\tilde{Q}_{LB}(t;\beta_0) - Q_{LB}(t) = \sum_{i=1}^{n}F^{-1}\left(\frac{V_{\beta_0,i}(t)}{n}\right) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where

$$F(u)(t) = p\int_{0<x\le t}\frac{g^*(x)}{qLB(x)}du(x) + (1-p)\int_{0<y\le t}y\left(\int_{y\le z}\frac{u(z)}{z^2}dz\right)d\left\{\left(\frac{h(t)}{h(y)}-1\right)\frac{f^*(y)}{h(y)}\right\},$$

and

$$V_{\beta_0,i}(t) = p^{1/2}\left\{I(y_{i0} \le t, \delta_i=1) - G^*(t)\right\}$$
$$+ \left(\frac{p}{1-p}\right)\left\{G^*(t) - Q_{LB}(t)\right\}\left\{\delta_i - p\right\}$$
$$+ (1-p)^{1/2}h(t)\int_{0<z\le t}\left\{I(y_{i0} \le z, \delta_i=0) - F^*(z)\right\}d\frac{1}{h(z)}.$$

As shown by Shen et al. (2009), the initial estimator has an independent and identically distributed (i.i.d.) representation,

$$\tilde{\beta}_{PH}^{(0)} - \beta_0 = n^{-1}\sum_{i=1}^{n}\eta_i + o_p\left(n^{-1/2}\right), \quad (30)$$

where

$$\eta_i = -\left[E\left\{\frac{\delta \boldsymbol{X}^{\otimes 2}}{\int_0^Y S_{\tilde{c}}(s)\,ds}\right\}\right]^{-1}\left\{\frac{\delta_i \mathrm{x}_i}{\int_0^{y_i} S_{\tilde{c}}(s)\,ds}\left(log\,y_i - \mathrm{x}_i^T\beta_0\right) + \int_0^\tau \frac{\mathrm{D}(t)\,dM_i^c(t)}{S_\upsilon(t)\,S_{\tilde{c}}(t)}\right\},$$

$$\mathbf{D}(s) = E\left[\left\{\delta\boldsymbol{X}I(Y \ge s)\left(log\,Y - \boldsymbol{X}^T\beta_0\right)\int_s^Y S_{\tilde{c}}(u)\,du\right\}\Big/\left\{\int_0^Y S_{\tilde{c}}(u)\,du\right\}\right],$$

$M_i^c(s) = I(y_i - a_i \le t, \delta_i=0) - \int_0^t I(y_i - a_i \ge u)\,d\Lambda_c(u)$, and $\Lambda_c(u)$ is the martingale and cumulative hazard function of the censoring times. Then we have the i.i.d. representation of

$\sqrt{n}\left(\tilde{Q}_{LB}\left(t;\tilde{\beta}_{PH}^{(0)}\right) - \tilde{Q}_{LB}(t)\right) = 1/\sqrt{n}\Sigma_{i=1}^n\psi_{1i}(t;\beta_0), \quad where \quad \psi_{1i}(t;\beta_0) = A_1(t)\eta_i - F^{-1}\left(V_{\beta_0,i}(t)\right)$

. It follows the i.i.d. representation of $\sqrt{n}\left(\tilde{Q}\left(t;\tilde{\beta}_{PH}^{(0)}\right) - \tilde{Q}(t)\right) = 1/\sqrt{n}\Sigma_{i=1}^n\psi_{2i}(t;\beta_0),$ where

$$\psi_{2i}(t;\beta_0) = \frac{\int_t^\tau s^{-1}dQ_{LB}(s)\int_0^\tau s^{-1}d\psi_{1i}(s;\beta_0) - \int_0^\tau s^{-1}d\psi_{1i}(s;\beta_0)\int_0^\tau s^{-1}dQ_{LB}(s)}{\left\{\int_0^\tau s^{-1}dQ_{LB}(s)\right\}^2}.$$

Then by the delta method, we have

$$\sqrt{n}\left(\tilde{\Lambda}_0\left(t;\tilde{\beta}_{PH}^{(0)}\right) - \Lambda_0(t)\right) = 1/\sqrt{n}\,exp\left\{\Lambda_0(t)\right\}\sum_{i=1}^{n}\psi_{2i}(t;\beta_0).$$

By the above i.i.d. representation of $\tilde{\Lambda}_0\left(t;\tilde{\beta}_{PH}^{(0)}\right)$ and Theorem 2 of Asgharian and Wolfson (2005), $\sqrt{n}\left(\tilde{\Lambda}_0\left(t;\tilde{\beta}_{PH}^{(0)}\right) - \Lambda_0(t)\right)$ converges weakly to Gaussian processes W_3. Under the regularity conditions (C2) and (C3), applying lemma 3 of Gill et al. (1988) and the chain

rule, we can show that the mapping of $\tilde{\boldsymbol{E}}_{PH}$ from the three 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

$n^{1/2}\tilde{\boldsymbol{E}}_{PH}\left(\tilde{\Lambda}_0,\beta_0\right)=n^{-1/2}\Sigma_{i=1}^n\psi_i\left(\beta_0\right)=o_p\left(1\right)$, where

$$
\begin{aligned}
\psi_i\left(\beta_0\right) &= \boldsymbol{X}_i\delta_i N_i\left(\beta_0,\tau\right) - \epsilon\left[\boldsymbol{X}\delta N\left(\beta_0,\tau\right)\right] \\
&- \int_0^\tau exp\left\{\Lambda_0\left(t\right)\right\}\psi_{2i}\left(t;\beta_0\right)d\epsilon\left[\boldsymbol{X}N\left(\beta_0,\tau\right)\right] \\
&- \int_0^\tau\Lambda_0\left(t\right)d\boldsymbol{X}_i N_i\left(\beta_0,\tau\right) - \epsilon\left[\boldsymbol{X}N\left(\beta_0,\tau\right)\right].
\end{aligned}
$$

Following this representation, $n^{1/2}\tilde{E}_{PH}\left(\tilde{\Lambda}_0,\beta_0\right)$ converges to a normal distribution with mean zero and variance $\Sigma_1^{(1)}=E[\psi_i\left(\beta_0\right)]^{\otimes 2}$.

Next, we show that estimating equation $\tilde{E}_{PH}\left(\tilde{\Lambda}_0,\beta\right)$ is asymptotically linear in a neighborhood of the true value β_0. Define the mean function for $\tilde{\boldsymbol{E}}_{PH}\left(\beta\right)$

$$
m_{PH}\left(\beta,t\right)=E\int_0^t x_i\left\{P_{ci}-\Lambda_0\left(t\right)\right\}dN_i\left(\beta,t\right),
$$

where $P_{ci}=P\left(\delta_i=1|\boldsymbol{x}_i\right)$. By applying the Theorem 3 of Ying (1993), we have for any $B>0$ and $\varepsilon>0$

$$
\begin{aligned}
&\sup_{\|\beta-\beta_0\|\leq B}\|m_{PH}\left(\beta,\tau\right)-\tilde{E}_{PH}\left(\tilde{\Lambda}_0,\beta\right)\| \\
\leq &\sup_{\|\beta-\beta_0\|\leq B}\|N_1^x\left(\beta,\tau\right)-EN_1^x\left(\beta,\tau\right)\| \\
+ &\sup_{\|\beta-\beta_0\|\leq B}\int_0^\tau\|\left\{\tilde{\Lambda}_0\left(t;\tilde{\beta}_{PH}^{(0)}\right)-\Lambda_0\left(t\right)\right\}dN^x\left(\beta,t\right)\| \\
+ &\sup_{\|\beta-\beta_0\|\leq B}\int_0^\tau\|\Lambda_0\left(t\right)d\left[N^x\left(\beta,t\right)-EN^x\left(\beta,t\right)\right]\|=o\left(n^{-1/2+\epsilon}\right),
\end{aligned}
$$

where $N_1^x\left(\beta,t\right)=\Sigma_{i=1}^n x_i\delta_i I\left(y_i e^{-X_i^T\beta}\leq t\right)$ and $N^x\left(\beta,t\right)=\Sigma_{i=1}^n x_i I\left(y_i e^{-X_i^T\beta}\leq t\right)$.

This implies that the estimating equation $\tilde{E}_{PH}\left(\tilde{\Lambda}_0,\beta\right)$ can be uniformly approximated by the nonrandom function $m_{PH}\left(\beta,\tau\right)$ up to the order of $n^{-1/2+\varepsilon}$. If the function $m_{PH}\left(\beta,\tau\right)$ has a unique solution given a compact region C_β containing β_0 as an interior point, the estimator, which satisfies $\tilde{E}_{PH}\left(\tilde{\beta}\right)=min_{\beta\in C_\beta}\|\tilde{E}_{PH}\left(\beta\right)\|$, is strongly consistent. As discussed in Ying (1993), this assumption can be evaluated for any given joint distribution of $(T,\tilde{C},\tilde{X})$. Denote the slope of function $m_{PH}\left(\beta,\tau\right)$ by Γ_{PH}, where

$$
\boldsymbol{\Gamma}_{PH}=-\boldsymbol{X}^{\otimes 2}e^{\boldsymbol{X}^T\beta_0}\int_0^\tau t\Lambda_0\left(t\right)\left\{q\left(t\right)S_{\tilde{c}}\left(te^{\boldsymbol{X}^T\beta_0}\right)+\overline{Q}\left(t\right)f_{\tilde{c}}\left(te^{\boldsymbol{X}^T\beta_0}\right)\right\}dt.
$$

Furthermore, the slope Γ_{PH} is nonsingular under the regularity assumption (C1), $\sqrt{n}\left(\beta_{PH}^{(1)} - \beta_0\right)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Sigma_{PH}^{(1)} = \Gamma_{PH}^{-1}\Sigma_1^{(1)}\Gamma_{PH}^{-1}$.

By arguments similar to those, we can show the asymptotic behavior of $\tilde{\beta}_{PO}^{(1)}$. For notational simplicity, the IWEE estimator,

$$\tilde{\beta}_{PO}^{(0)} = \left\{\sum_{i=1}^{n}\frac{\delta_i x_i x_i^T}{\int_0^{y_1}\hat{S}_{\tilde{c}}(u)\,du}\right\}^{-1}\sum_{i=1}^{n}\frac{\delta_1 x_i \quad log \quad y_i}{\int_0^{y_i}\hat{S}\tilde{c}(u)\,du}, \quad (31)$$

is used as the initial value. We obtain $\tilde{\beta}_{PO}^{(1)}$ by minimizing the norm of the non-continuous estimating equations via the transformed times $s_i\left(\tilde{\beta}_{PO}^{(0)}\right) = y_i exp\left(-x_i^T\tilde{\beta}_{PO}^{(0)}\right)$,

$$\tilde{\mathbf{E}}_{PO}\left(\tilde{F}_0,\beta\right) = \frac{2}{n}\sum_{i=1}^{n}\int_0^{\tau}x_i\left\{(1+\delta_i)\,\tilde{F}_0\left(t;\tilde{\beta}_{PO}^{(0)}\right) - \delta_i\right\}dN_i(\beta,t). \quad (32)$$

Given the asymptotic properties of the initial estimator and the nonparametric maximum likelihood estimator $\tilde{F}_0$, and analogous to the argument given for the asymptotic properties of $\tilde{\beta}_{PO}^{(1)}$, we can show that $n^{1/2}\tilde{E}_{PO}\left(\tilde{F}_0,\beta_0\right)$ converges to a normal distribution with mean zero and variance $\Sigma_2^{(1)}$. Furthermore, following the arguments in Ying (1993), the estimating equation $\tilde{E}_{PO}(\tilde{F}_0,\beta)$ can be shown to be uniformly approximated by the nonrandom function $\boldsymbol{m}_{PO}(\boldsymbol{\beta},\tau)$ up to the order of $n^{-1/2+\varepsilon}$, where

$$\mathbf{m}_{PO}(\beta,t) = E\int_0^t\boldsymbol{X}_i\left\{(1+P_{ci})\,F_0(t) - P_{ci}\right\}\quad dN_i(\beta,t).$$

Then under the assumption that the mean function has a unique solution at a compact region containing $\boldsymbol{\beta}_0$ and its slope function $\boldsymbol{\Gamma}_{PO}$ is nonsingular, $\sqrt{n}\left(\tilde{\beta}_{PO}^{(1)} - \beta_0\right)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Sigma_{PO}^{(1)} = \Gamma_{PO}^{-1}\Sigma_{21}^{(1)}\Gamma_{PO}^{-1}$. Then using such arguments iteratively, we can show that for any fixed $m\tilde{\beta}_{PH}^{(m)}$ and $\tilde{\beta}_{PO}^{(m)}$ are consistent and asymptotically normally distributed.

References

Addona V, Wolfson DB. A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. Lifetime Data Anal. 2006; 12:267–274. [PubMed: 16917734]

Asgharian M, M'Lan CE, Wolfson DB. Length-biased sampling with right censoring: An unconditional approach. J. Am. Statist. Assoc. 2002; 97:201–209.

Asgharian M, Wolfson DB. Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. Ann Statist. 2005; 33:2109–2131.

Barroda I, Roberts FDK. Solution of an overdetermined system of equations in the l_1 norm. Communications of the ACM. 1974; 17:319–320.

Buckley J, James I. Linear regression with censored data. Biometrika. 1979; 66:429–436.

Chen YQ. Semiparametric regression in size-biased sampling. Biometrics. 2010; 66:149–158. [PubMed: 19432792]

Cromme L, Diener J. Fixed point theorems for discontinuous mapping. Math. Programming. 1991; 2:257–267.

Davidov O, Iliopoulos G. Convergence of luo and tsai's iterative algorithm for estimation in proportional likelihood ratio models. Biometrika. 2013; 100:778–780.

Gill RD, Vardi Y, Wellner JA. Large-sample theory of empirical distributions in biased sampling models. Ann Statist. 1988; 16:1069–1112.

Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time models. Biometrika. 2003; 90:341–353.

Jones MP. A class of semiparametric regressions for the accelerated failure time model. Biometrika. 1997; 84:73–84.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. Wiley; New York: 2002.

Lai TL, Ying Z. Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. Ann Statist. 1991a; 19:1370–1402.

Lai TL, Ying Z. Rank regression methods for left-truncated and right-censored data. Ann Statist. 1991b; 19:531–556.

Lin DY, Ying Z. Semiparametric inference for the accelerated life model with time-dependent covariates. J. Statist. Plan. and Inference. 1995; 44:47–63.

Mandel M, Ritov Y. The accelerated failure time model under biased sampling. Bimetrics. 2010; 66:1306–1308.

Miller RG, Halpern J. Regression with censored data. Biometrika. 1982; 69:521–531.

Neyman J. 'Smooth' test for goodness of fit. Skand. Aktuar. 1937; 20:150–199.

Ning J, Qin J, Shen Y. Non-parametric tests for right-censored data with biased sampling. Journal of the Royal Statistical Society: Series B. 2010; 72:609–630.

Ning J, Qin J, Shen Y. Buckley-James-type estimator with right-censored and length-biased data. Biometrics. 2011; 67:1369–1378. [PubMed: 21385160]

Prentice RL. Linear rank tests with right censored data. Biometrika. 1978; 65:167–179.

Rayner, JCW.; Thas, O.; Best, DJ. Smooth Tests of Goodness of Fit: Using R. Wiley; Singapore: 2011.

Ritov Y. Estimation in a linear regression model with censored data. Ann Statist. 1990; 18:303–328.

Robins JM, Tsiatis AA. Semiparametric estimation of an accelerated failure time modelwith time-dependent covariates. Biometrika. 1992; 79:311–319.

Sharma P, Shen Y, Wen S, Yamada S, Jungbluth AA, Gnjatic S, Bajorin DF, Reuter VE, Herr H, Old LJ, Sato E. Cd8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. PNAS. 2007; 104:3967–3972. [PubMed: 17360461]

Shen Y, Ning J, Qin J. Analyzing length-biased data with semi-parametric transformation and accelerated failure time models. J. Am. Statist. Assoc. 2009; 104:1192–1202.

Shuang J, Peng L, Cheng Y, Lai H. Quantile regression for doubly censored data. Biometrics. 2012; 68:101–112. [PubMed: 21950348]

Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. Ann Statist. 1990; 18:354–372.

van der Vaart, AW. Asymptotic Statistics. Cambridge University Press; Cambridge: 1998.

Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. Biometrika. 1989; 76:751–761.

Vardi Y, Zhang CH. Large sample study of empirical distributions in a random-multiplicative censoring model. Ann. Statist. 1992; 20:1022–1039.

Wei LJ, Ying Z, Lin DY. Linear regression analysis of censored survival data based on rank tests. Biometrika. 1990; 77:845–851.

Wolfson C, Wolfson DB, Asgharian M, M'Lan CE, Ostbye T, Rockwood K, Hogan DB, Clinical Progression of Dementia Study Group. A reevaluation of the duration of survival after the onset of dementia. New Engl. J. Med. 2001; 344:1111–1116. [PubMed: 11297701]

Ying Z. A large sample study of rank estimation for censored regression data. Ann Statist. 1993; 21:76–99.

Zeng D, Lin DY. Efficient estimation for the accelerated failure time model. J. Am. Statist. Assoc. 2007; 102:1387–1396.

Zhou M. M-estimation in censored linear model. Biometrika. 1992; 79:837–841.

Table 1

Summary statistics of simulations for traditional right-censored data with mean bias and empirical standard errors (ESE)

Cohort size	Cen%	Inverse weighting Method		Buckley-James Method		E_{PH} Method		E_{PO} Method	
		Bias	ESE	Bias	ESE	Bias	ESE	Bias	ESE
				$(\alpha_1, \alpha_2) = (0.5,1), \varepsilon \sim U(-0.5, 0.5)$					
100	15%	(−.01,−.01)	(.060,.114)	(−.01, .00)	(.059,.113)	(.00, −.02)	(.054,.098)	(.00,−.01)	(.063,.111)
100	30%	(−.01,−.02)	(.070,.132)	(−.00,−.00)	(.066,.130)	(.01,−.03)	(.059,.113)	(.01,−.01)	(.064,.123)
100	50%	(−.06,−.12)	(.103,.165)	(−.00, .01)	(.085,.146)	(.02, −.05)	(.083,.154)	(.01,−.02)	(.083,.153)
200	15%	(−.00,−.01)	(.044,.079)	(−.00,−.01)	(.044,.078)	(.00, .01)	(.037,.060)	(.00,−.01)	(.046,.072)
200	30%	(−.01,−.01)	(.050,.087)	(−.00, .00)	(.049,.083)	(.01,−.02)	(.041,.077)	(.00,−.01)	(.049,.090)
200	50%	(−.05,−.10)	(.067,.118)	(.00, .00)	(.055,.095)	(.00,−.03)	(.056, .097)	(.00,−.02)	(.058, .097)
				$\varepsilon \sim Normal(0, 1/12)$					
100	15%	(−.01,−.00)	(.063,.110)	(.01, .01)	(.63,.107)	(.01,−.03)	(.068,.110)	(.01,−.01)	(.064,.108)
100	30%	(−.00,−.01)	(.066,.131)	(.01, .01)	(.65,.124)	(.01,−.04)	(.072,.119)	(.01,−.02)	(.068,.117)
100	50%	(−.05,−.11)	(.092,.180)	(−.00, .01)	(.075,.146)	(.02,−.05)	(.085,.141)	(.01,−.02)	(.082,.138)
200	15%	(−.00,−.00)	(.048,.073)	(.00,−.00)	(.046,.071)	(.00,−.02)	(.047,.082)	(.00, .00)	(.043,.081)
200	30%	(−.01,−.01)	(.051,.086)	(.00,−.00)	(.048,.082)	(.00,−.02)	(.050,.087)	(.00,−.01)	(.042,.075)
200	50%	(.05, .10)	(.067,.112)	(.00, .00)	(.055,.091)	(.01,−.02)	(.058,.103)	(.00,−.01)	(.054,.098)

Table 2

Summary statistics of simulations for length-biased right-censored data with mean bias and ESE

Cohort size	Cen%	Inverse weighting Method		Buckley-James Method		$\tilde{E}_{PH}$ Method		$\tilde{E}_{PO}$ Method	
		Bias	ESE	Bias	ESE	Bias	ESE	Bias	ESE
		$(\alpha_1, \alpha_2) = (0.5,1), \varepsilon \sim U(-0.5,0.5)$							
100	15%	(.00,−.01)	(.072,.127)	(.00, .00)	(.065,.116)	(.00, .00)	(.047,.082)	(−.01, .00)	(.067,.115)
	30%	(−.02,−.02)	(.077,.135)	(.00, .00)	(.068,.123)	(.00, .01)	(.051,.093)	(.00, .00)	(.071,.121)
	50%	(−.04,−.06)	(.096,.166)	(−.01,−.01)	(.087,.166)	(.00, .01)	(.063,.112)	(.01, .00)	(.078,.133)
200	15%	(.00,−.01)	(.047,.084)	(.00, .00)	(.045,.080)	(.00, .00)	(.030,.049)	(.00, .00)	(.044,.078)
	30%	(−.01,−.02)	(.052,.088)	(.00, .00)	(.050,.086)	(.00, .01)	(.034,.057)	(.00, .00)	(.046,.082)
	50%	(−.03,−.05)	(.068,.118)	(.00, .00)	(.057,.103)	(−.01, .00)	(.041,.070)	(.00,−.01)	(.053,.089)
		$\varepsilon \sim Normal(0, 1/12)$							
100	15%	(−.01,−.01)	(.072,.125)	(.00, .00)	(.072,.118)	(.00, .01)	(.065,.116)	(.00, .01)	(.065,.111)
	30%	(−.01,−.03)	(.078,.138)	(.00, .00)	(.067,.123)	(.00, .00)	(.070,.123)	(.00,−.01)	(.069,.116)
	50%	(−.03,−.06)	(.091,.166)	(−.03, −.06)	(.097,.186)	(.00, .01)	(.075,.134)	(.00, .01)	(.074,.125)
200	15%	(.00,−.01)	(.052,.085)	(.00, .00)	(.049,.081)	(.00, .00)	(.048,.080)	(.00,−.01)	(.045,.073)
	30%	(−.01,−.02)	(.057,.093)	(.00, .00)	(.050,.085)	(.00, .00)	(.049,.084)	(.00, .00)	(.047,.078)
	50%	(−.03,−.04)	(.067,.116)	(−.03,−.05)	(.065,.121)	(.00, .00)	(.053,.095)	(.00, .00)	(.049,.086)

Table 3

Estimates (Est) and Bootstrap standard errors (SE) of regression coefficients for bladder cancer data

	Inverse weighting Method Est(SE)	Buckley-James Method Est(SE)	E_{PH} Method Est(SE)	E_{PO} Method Est(SE)
CD8 8	0.08(.30)	0.87(.41)	1.03(.46)	1.27(.53)
P-stage	−0.45(.32)	−0.95(.34)	−1.19(.42)	−1.52(.48)

Table 4

Estimates (Est) and Bootstrap standard errors (SE) of regression coefficients for dementia data

Diagnosis	Inverse weighting Method Est(SE)	Buckley-James Method Est(SE)	$\tilde{E}_{PH}$ Method Est(SE)	$\tilde{E}_{PO}$ Method Est(SE)
Vascular Dementia	−0.21(.11)	−0.19(.13)	−0.17(.07)	−0.21(.07)
Probable Alzheimer	−0.14(.15)	−0.11(.15)	−0.11(.06)	−0.08(.08)