Published in final edited form as:

Lifetime Data Anal. 2012 October ; 18(4): 470–503. doi:10.1007/s10985-012-9225-5.

Imputation for semiparametric transformation models with biased-sampling data

Hao Liu,

Division of Biostatistics, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas, 77030, USA

Jing Qin, and

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, Maryland, 20892, USA

Yu Shen

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas, 77030, USA

Hao Liu: haol@bcm.edu; Jing Qin: jingqin@niaid.nih.gov; Yu Shen: yshen@mdanderson.org

Abstract

Widely recognized in many fields including economics, engineering, epidemiology, health sciences, technology and wildlife management, length-biased sampling generates biased and right-censored data but often provide the best information available for statistical inference. Different from traditional right-censored data, length-biased data have unique aspects resulting from their sampling procedures. We exploit these unique aspects and propose a general imputation-based estimation method for analyzing length-biased data under a class of flexible semiparametric transformation models. We present new computational algorithms that can jointly estimate the regression coefficients and the baseline function semiparametrically. The imputation-based method under the transformation model provides an unbiased estimator regardless whether the censoring is independent or not on the covariates. We establish large-sample properties using the empirical processes method. Simulation studies show that under small to moderate sample sizes, the proposed procedure has smaller mean square errors than two existing estimation procedures. Finally, we demonstrate the estimation procedure by a real data example.

Keywords

Biased sampling; Estimating equation; Imputation; Transformation models

1 Introduction

Biased sampling arises when the observations are not randomly selected from the target population (Smith 1993). When the sampling probability is proportional to the underlying outcome of interest, this is long known as length-biased sampling (Cox 1962, Chapter 5). Cox (1969) describes length-biased data where the wool fibers of different lengths are sampled cross-sectionally. Length-biased sampling and the needs for statistical corrections have been well recognized in a wide range of applied fields; some recent examples include industrial reliability (Gupta and Keating 1986), land economics (Nowell et al 1988), textile research (Zeidman et al 1991), marketing research (Nowell and Stanley 1991), wildlife management (Otis et al 1993), genome-wide linkage study (Terwilliger et al 1997), rock mechanics (Mauldon 1998), brain motoneurone study (Ventura et al 1998), labor economics (de Una-Álvarez et al 2003), and nanotechnology (Kvam 2008). As observational studies

with long-term follow-up provide a great opportunity to evaluate the benefit of disease prevention and treatment interventions at the population level, there is a rich literature on length-biased sampling in health science studies, including cancer screening trials (Zelen and Feinleib 1969; Zelen 2004), etiologic studies (Simon 1980), and general epidemiology follow-up studies (Varadhan and Frangakis 2004; Scheike and Keiding 2006).

Length-biased sampling generates biased and censored data but often provides the best information available for scientific investigation in these applications. Compared to traditional right-censored data (Kalbfleisch and Prentice 2002; Lin 2000), length-biased data have unique features that require different approaches for both nonparametric and semiparametric estimations (Vardi 1982, 1989; Wang 1991; Keiding 1992; Wang et al 1993; Asgharian et al 2002; Asgharian and Wolfson 2005). Under proportional hazards models, Wang (1996) constructed a pseudo-likelihood to estimate the covariate effects on outcomes without right censoring. Recently, Qin and Shen (2010) constructed an inversely-weighted estimating equation for proportional hazards models. Shen et al (2009) proposed a weighted U-statistic estimating equation approach for the linear transformation models. The existing approaches may be inefficient as the covariates of censored subjects are not included in the estimation procedure. They also require the censoring time to be independent of covariates unless a dependence structure is assumed. Despite recent progress, there is a lack of unified inference and prediction procedures under flexible semiparametric models for biased-sampling data.

Consider sampling cross-sectionally a subject with disease from the population and then following the subject up, who may be subject to right-censoring. As depicted by the following diagram, the observable data are the time $\tilde{A}$ from disease onset to enrollment, the time V from enrollment to death, and the censoring time C from enrollment to loss to follow-up.

With length-biased sampling, the population survival time $\tilde{T}$ from disease onset to death can only be observed among those with $\tilde{T} > \tilde{A}$. The observable survival time is length-biased and equal to $A + V$. Conditional on covariates Z , it is well known that under an assumption of stationarity on the underlying event processes, the length-biased time $T = A + V$ has the density function $tf(t|Z)/\mu(Z)$ for $t \geq 0$, where $f(\cdot|Z)$ is the density function of the population time $\tilde{T}$, and $\mu(Z) = \int tf(t|Z)dt$ is the mean of $\tilde{T}$, conditional on Z . Even with the assumption that the censoring time C is independent of (A, V) conditional on Z , note that the observed length-biased time $T = A + V$ is actually subject to dependent censoring $A + C$.

To analyze the covariate effects on the unbiased time $\tilde{T}$, a flexible semiparametric model is the following semiparametric transformation model specified through the cumulative hazard function

$$\Lambda(t|Z) \equiv Q \left\{ \int_0^t e^{\beta' Z} d\Lambda(s) \right\}, \quad (1)$$

where $\Lambda(t|Z)$ is the cumulative hazard function conditional on the covariates Z , $Q(\cdot)$ is a known increasing function, β is an unknown p -dimensional regression coefficient, and $\Lambda(\cdot)$ is an unspecified nondecreasing baseline function with $\Lambda(0) = 0$. The transformation models encompass a broad range of semiparametric models, including the proportional hazards model and the proportional odds model as special cases. An example is that for given η ,

$Q(x) = \log(1 + \eta x)/\eta$ for $x \geq 0$, which was called the Pareto model (Clayton and Cuzick 1985; Dabrowska and Doksum 1988). The case $\eta = 0$ is for the proportional hazards model and $\eta = 1$ for the proportional odds model. The class of semiparametric transformation models has been studied for traditional right-censored data when the covariates are time-dependent (Zeng and Lin 2007). When the covariates are time-independent, model (1) is reduced to $\Lambda(t|Z) = Q\{\exp(\beta'Z) \Lambda(t)\}$, which can be expressed equivalently to the linear transformation model (Cheng et al 1995, 1997), $\log \Lambda(\tilde{T}) = -\beta'Z + \varepsilon$, where ε is a random variable with a known survival function $\exp\{-Q(e^x)\}$ for $-\infty < x < \infty$.

The purpose of the paper is to develop a general estimation and inference procedure for right-censored length-biased data under the general semiparametric transformation model (1). We introduce an estimating equation method to impute the censored subjects, which can take advantage of the unique features of length-biased data. The imputation approach has its root back to the Buckley-James procedure under the accelerated failure time (AFT) model for traditional survival data (Buckley and James 1979), but is essentially different. Under the transformation model, the imputation for length-biased data needs to deal with the estimations of both the baseline function $\Lambda(\cdot)$ and the covariate effect β on the unbiased time $\tilde{T}$; this imposes significant technical challenges. We propose and show that imputation-based estimation procedures can consistently estimate both the regression coefficients and the baseline function. The proposed estimation procedure does not require the assumption of independence between the censoring time and the covariates.

We first introduce the principle of the imputation method for right-censored length-biased data. In Section 2, we describe the estimation procedure, the computation algorithm, and the asymptotic results under the semiparametric transformation models. This is followed by extensive simulation studies, for which the results are presented in Section 3 in addition to a real data analysis. The proofs in Appendix include sophisticated techniques for establishing the asymptotic properties.

2 Estimation procedures, computation algorithms and asymptotic results

2.1 Imputation methods for right-censored length-biased data

Define

$$N(t) = 1(T \leq t), \quad Y(t) = 1(T \geq t),$$

where $1(\cdot)$ is the indicator function, i.e., $1(\Omega) = 1$ if Ω is true; otherwise, $1(\Omega) = 0$. Let $\delta = 1(T \leq A + C)$ be the censoring indicator. Denote the hazard function, cumulative density function and survival function of the unbiased survival time $\tilde{T}$ by $\lambda(\cdot|Z)$, $F(\cdot|Z)$ and $S(\cdot|Z)$, respectively, conditional on Z . In the absence of right censoring, Wang (1996) developed a pseudo-likelihood estimation for the proportional hazards model based on the following process:

$$\tilde{M}(t) = N(t) - \int_0^t \frac{u}{T} Y(u) \lambda(u|Z) du.$$

Here, $\tilde{M}(t)$ does not involve censoring data.

In the presence of right censoring, we can impute $\tilde{M}(t)$ conditional on the censored data by

$$M(t) = \delta \tilde{M}(t) + (1 - \delta) E\{\tilde{M}(t) | 1 - \delta, C, Z\}. \quad (2)$$

Under the stationarity assumption, the joint density function of length-biased data (A, T) conditional on Z is $f(t|Z)/\mu(Z)$ for $t > a > 0$ (Vardi 1989; Asgharian and Wolfson 2005). We make the following assumption on the censoring mechanism: the distribution of (A, T) conditional on (C, Z) is the same as the distribution of (A, T) conditional on Z . This implies that the joint density function of (A, T) conditional on (C, Z) is $f(t|Z)/\mu(Z)$ for $t > a > 0$. These assumptions and the regularity conditions are listed in Appendix A.

Suppose that the distribution function of $\tilde{T}$ is on $[0, \tau]$, where τ is the upper bound of the support of $\tilde{T}$. We have the following key property on the imputation function $M(t)$:

Proposition 1—Under Assumptions (1) and (2), $E\{M(t)\} = 0$ for all t in $[0, \tau]$.

The detail of the proof is contained in Appendix B, where we note that the proposition does not require an independence assumption between Z and C .

As shown in equation (30) in the proof in Appendix B, the conditional expectation in (2) is

$$E\{d\tilde{M}(t) | 1 - \delta, C, Z\} = - \left\{ \int_{\tau}^{\tau} C \frac{S(u|Z)}{u^2} du \right\} t \lambda(t|Z) / \int_C S(u|Z) du, \quad (3)$$

where $x \vee y = \max\{x, y\}$. Thus, we can rewrite the imputation equation (2) as

$$M(t) = \delta N(t) - \int_0^t u W(u) \lambda(u|Z) du, \quad (4)$$

where

$$W(t) = \delta \frac{Y(t)}{T} + (1 - \delta) \int_t^{\tau} C 1(u \geq C) \frac{S(u|Z)}{u^2} du / \int_C S(u|Z) du.$$

While this may be regarded as a natural generalization of the imputation principle for traditional survival data (Buckley and James 1979; Lai and Ying 1991), our approach takes advantage of the special properties of length-biased data. In general, the process $M(t)$ is not a martingale. We will make use of its mean-zero property to construct unbiased estimating equations for the regression analysis of length-biased data under the semiparametric transformation models.

Remark 1: For length-biased data, the distribution of $\tilde{T}$ is identifiable on its entire support, as long as there is a minimal follow-up, i.e., there exists a $u_0 > 0$ such that $S_C(u_0) \equiv P(C > u_0) > 0$, where $S_C(\cdot)$ is the survival function of censoring time C . Without loss of generality, suppose that there is no covariate, and let $f(t)$ and $F(t)$ denote the density and distribution functions, respectively, for the unbiased time $\tilde{T}$. The observable data is $X = T \wedge (A + C)$, where $x \wedge y = \min\{x, y\}$. Given uncensored data,

$$P(X \leq t | \delta=1) = \frac{P(T \leq t, T \leq A+C)}{P(T \leq A+C)} = \frac{\int_0^t w(u) dF(u)}{\int_0^\tau w(u) dF(u)},$$

where $w(u)$ is the area under the survival curve of censoring time C ,

$$w(u) \equiv \int_0^u S_c(u-a) da = \int_0^u S_c(v) dv,$$

which is always positive for $u > 0$ by the minimal follow-up assumption. The conditional density function of X given $\delta = 1$ is then $w(t)f(t)/\int_0^\tau w(u)dF(u)$, which coincides with the density function of biased sampling with a weight function $w(t) > 0$. It follows that the distribution $F(t)$ is identifiable from the uncensored length-biased data on its entire support $[0, \tau)$.

Remark 2: We can estimate τ by $\hat{\tau}_n \equiv \max\{X_1, \dots, X_n\}$ consistently. If τ is infinite, then it is easy to show that $\hat{\tau}_n \rightarrow \infty$. If τ is finite, then we can show that under the minimal follow-up assumption, $\hat{\tau}_n$ converges to τ in probability in a rate faster than $n^{1/2}$.

Lemma 1—If $\tau < \infty$, then under Assumption (2) in Appendix A, $\hat{\tau}_n$ converges to τ in probability, and $n^\alpha(\hat{\tau}_n - \tau) = o_p(1)$ for $1 > \alpha > 1/2$.

2.2 Estimation procedures

Let the observed data $(X_i, A_i, \delta_i, Z_i)$ for $i = 1, \dots, n$ be independently and identically distributed, where $X_i = T_i \wedge (A_i + C_i)$ and $\delta_i = 1\{T_i < (A_i + C_i)\}$. Define for $t \geq 0$,

$$N_i(t) = 1(T_i \leq t), \quad Y_i(t) = 1(T_i \geq t).$$

We propose an estimation procedure for the semiparametric model (1) based on the imputation equation (4). By Proposition $E\{dM_i(t)|Z_i\} = 0$, where

$$M_i(t) = \delta_i N_i(t) - \int_0^t u W_i(u) d\Lambda(u|Z_i), \text{ and}$$

$$W_i(t, \beta, \Lambda) = \delta_i \frac{Y_i(t)}{X_i} + (1 - \delta_i) C_i \left\{ \int_t^\tau 1(u \geq C_i) \frac{S(u|Z_i)}{u^2} du \right\} / \int_{C_i}^\tau S(u|Z_i) du. \quad (5)$$

Hence, we can solve β and Λ jointly from the following joint estimating equations:

$$\sum_{i=1}^n \int_0^\tau Z_i dM_i(u) = 0, \quad \sum_{i=1}^n \int_0^t dM_i(u) = 0.$$

Under model (1), the survival function $S(t|Z_i)$ conditional on Z_i can be expressed as

$$S(t|Z_i) = \exp\{-\Lambda(t|Z_i)\} = \exp\left\{-Q\left(\int_0^t e^{\beta' Z_i} d\Lambda(s)\right)\right\}. \quad (6)$$

Suppose that the first derivative of the function $Q(\cdot)$ exists and is denoted by $q(x) = dQ(x)/dx$. Then, we can write the joint estimating equations of β and Λ as

$$U_{1n}(\beta, \Lambda) = \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} Z_i \left[\delta_i dN_i(s) - u W_i(u; \beta, \Lambda) e^{\beta' Z_i} q\{e^{\beta' Z_i} \Lambda(u)\} d\Lambda(u) \right] = 0, \quad (7)$$

$$U_{2n}(t, \beta, \Lambda) = \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \left[\delta_i dN_i(u) - u W_i(u; \beta, \Lambda) e^{\beta' Z_i} q\{e^{\beta' Z_i} \Lambda(u)\} d\Lambda(u) \right] = 0. \quad (8)$$

By Proposition 1, these estimating equations are unbiased. Define $\tilde{W}_i$ by equation (5) with τ being replaced with $\hat{\tau}_n = \max\{X_1, \dots, X_n\}$. Then, by Lemma 1, $\tilde{W}_i$ and W_i are asymptotically equivalent. We can thus replace τ with $\hat{\tau}_n$ in (7) and (8) to yield asymptotically equivalent estimating equations. We denote the estimating equations as $(\tilde{U}_{1n}, \tilde{U}_{2n})$ and refer their definition in (7) and (8) with τ replaced by $\hat{\tau}_n$.

The estimators $\hat{\beta}_n$ and $\hat{\Lambda}_n(\cdot)$ need to satisfy both the estimating equations (7) and (8). For each β in a compact set that contains the true value β_0 , let $\hat{\Lambda}_n(\cdot, \beta)$ be the solution to equation (8), i.e., $U_{2n}(t, \beta, \hat{\Lambda}_n(\cdot, \beta)) = 0$. It follows that $\hat{\Lambda}_n(\cdot, \beta)$ is a nondecreasing step function with jumps at the uncensored time points, continuous from the right and having limits from the left,

$$\hat{\Lambda}_n(t, \beta) = \int_0^t \left[\sum_{j=1}^n u \tilde{W}_j(u, \beta, \hat{\Lambda}_n(\cdot, \beta)) e^{\beta' Z_j} q\{e^{\beta' Z_j} \hat{\Lambda}_n(u, \beta)\} \right]^{-1} d \left\{ \sum_{i=1}^n \delta_i N_i(u) \right\}. \quad (9)$$

We assume that there is at least one uncensored length-biased observation so the estimator is well defined. This condition is asymptotically equivalent to the assumption $P(A_i + C_i > T_i) > 0$, which is listed in Assumption (2) in the Appendix.

Plugging $\hat{\Lambda}_n(\cdot, \beta)$ back to equation (7) to replace Λ , we can find the estimators $\hat{\beta}_n$ using the following estimating equations:

$$\sum_{i=1}^n \int_0^{\hat{\tau}_n} \left[Z_i - \frac{\sum_{j=1}^n Z_j \tilde{W}_j(u, \beta, \hat{\Lambda}_n(\cdot, \beta)) e^{\beta' Z_j} q\{e^{\beta' Z_j} \hat{\Lambda}_n(u, \beta)\}}{\sum_{j=1}^n \tilde{W}_j(u, \beta, \hat{\Lambda}_n(\cdot, \beta)) e^{\beta' Z_j} q\{e^{\beta' Z_j} \hat{\Lambda}_n(u, \beta)\}} \right] \delta_i dN_i(s) = 0. \quad (10)$$

The equations (9) and (10) are equivalent to the equations (7) and (8), and the estimators $\hat{\beta}_n$ and $\hat{\Lambda}_n(\cdot) = \hat{\Lambda}_n(\cdot, \hat{\beta}_n)$ satisfy jointly $\tilde{U}_{1n}(\hat{\beta}_n, \hat{\Lambda}_n(\cdot, \hat{\beta}_n)) = 0$ and $\tilde{U}_{2n}(t, \hat{\beta}_n, \hat{\Lambda}_n(\cdot, \hat{\beta}_n)) = 0$.

The estimator $\hat{\Lambda}_n(\cdot, \beta_0)$ is an asymptotically consistent estimator of $\Lambda_0(\cdot)$. Replacing $(\beta, \hat{\Lambda}_n)$ with the true parameter values (β_0, Λ_0) in the right-hand side of (9), we have a pseudo estimator asymptotically equivalent to $\hat{\Lambda}_n(\cdot, \beta_0)$, defined by

$$\Lambda_n^o(t) \equiv \int_0^t \left[\sum_{j=1}^n u W_j(u, \beta_0, \Lambda_0) e^{\beta_0' Z_j} q\{e^{\beta_0' Z_j} \Lambda_0(u)\} \right]^{-1} d \left\{ \sum_{i=1}^n \delta_i N_i(u) \right\}. \quad (11)$$

We can show that $\Lambda_n^o(t)$ converges to $\Lambda_0(t)$ almost surely and uniformly in t in $[0, \zeta]$ for all $\zeta < \tau$, using the functional law of large numbers and the continuous mapping theorem (van der

Vaart and Wellner 1996, page 54). More specifically by the Glivenko-Cantelli theorem (van der Vaart 1998, page 270), we can show that uniformly for t in $[0, \zeta]$,

$$n^{-1} \sum_{i=1}^n \delta_i N_i(t) \rightarrow \int_0^t E_0 \left[u W_i(u, \beta_0, \Lambda_0) e^{\beta_0' Z} q \{ e^{\beta_0' Z} \Lambda_0(u) \} \right] d\Lambda_0(u)$$

almost surely, where E_0 denotes the expectation under (β_0, Λ_0) .

The estimating equation (10) can be simplified for the Cox model with $q(x) = 1$:

$$\sum_{i=1}^n \int_0^{\tau_n} \left[Z_i - \frac{\sum_{j=1}^n Z_j \tilde{W}_j \{u, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j}}{\sum_{j=1}^n \tilde{W}_j \{u, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j}} \right] \delta_i dN_i(s) = 0.$$

When there is no censoring, $W_i(t) = Y_i(t)/T_i$ and this is exactly the same estimating equation derived from the pseudo-likelihood in Wang (1996).

2.3 Computational algorithms

Let $0 = T_{(0)} < T_{(1)} < \dots < T_{(K)}$ be the ordered uncensored times among $\{X_1, \dots, X_n\}$. The estimating equations (10) involving Stieltjes integral with respect to the counting processes can be written as

$$\sum_{i=1}^n \delta_i \left[Z_i - \frac{\sum_{j=1}^n Z_j \tilde{W}_j \{X_i, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j} q \{ e^{\beta' Z_j} \hat{\Lambda}_n(X_i, \beta) \}}{\sum_{j=1}^n \tilde{W}_j \{X_i, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j} q \{ e^{\beta' Z_j} \hat{\Lambda}_n(X_i, \beta) \}} \right] = 0. \quad (12)$$

If there are no ties among the uncensored failure times, then (12) can be written as

$$\sum_{k=1}^K \left[Z_{(k)} - \frac{\sum_{j=1}^n Z_j \tilde{W}_j \{T_{(k)}, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j} q \{ e^{\beta' Z_j} \hat{\Lambda}_n(T_{(k)}, \beta) \}}{\sum_{j=1}^n \tilde{W}_j \{T_{(k)}, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j} q \{ e^{\beta' Z_j} \hat{\Lambda}_n(T_{(k)}, \beta) \}} \right] = 0,$$

where $Z_{(k)}$ is the covariate vector corresponding to $T_{(k)}$, $k = 1, \dots, K$. Similarly, the formula (9) can be written as

$$\hat{\Lambda}_n(t, \beta) = \sum_{k=1}^K \left[\frac{1 \{T_{(k)} \leq t\}}{\sum_{j=1}^n T_{(k)} \tilde{W}_j \{T_{(k)}, \beta, \hat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j} q \{ e^{\beta' Z_j} \hat{\Lambda}_n(T_{(k)}, \beta) \}} \right].$$

Observe that the weight function W_j is always nonnegative, and the estimator $\hat{\Lambda}_n(\cdot, \beta)$ is a nondecreasing step function with positive jumps at $T_{(1)}, \dots, T_{(K)}$.

Denote the jump size of $\hat{\Lambda}_n(t, \beta)$ at time t by $\hat{\Lambda}_n(t, \beta) - \hat{\Lambda}_n(t-, \beta)$. Then for $k = 1, 2, \dots, K$,

$$\widehat{\lambda}_n(T_{(k)}, \beta) = \left[\sum_{j=1}^n T_{(k)} \tilde{W}_j \{T_{(k)}, \beta, \widehat{\Lambda}_n(\cdot, \beta)\} e^{\beta' Z_j} q \{e^{\beta' Z_j} \widehat{\Lambda}_n(T_{(k)}, \beta)\} \right]^{-1}. \quad (13)$$

As there is a one-to-one relationship between the estimator $\widehat{\Lambda}_n(\cdot, \beta)$ and $\widehat{\lambda}_n(T_{(k)}, \beta)$, $k = 1, \dots, K$ as $\widehat{\Lambda}_n(t, \beta) = \sum_{k=1}^K \widehat{\lambda}_n(T_{(k)}, \beta) 1(T_{(k)} \leq t)$, we consider the parameter set (β, λ) , where $\lambda = \{\lambda_1, \dots, \lambda_K\}$, in the computational algorithm.

The estimator $\{\widehat{\beta}_n, \widehat{\Lambda}_n(\cdot)\}$ defined as the solutions to the estimating equations (7) and (8) can be found computationally through (12) and (13) by the following iterative algorithm:

Step 0 Choose the initial value $\beta^{(0)}$ and $\lambda_k^{(0)}$ for $k = 1, \dots, K$. This can be the estimator, or a random perturbation of the estimator, obtained from the method for left-truncated and right-censored data. Set $\Lambda^{(0)}(t) = \sum_{k=1}^K \lambda_k^{(0)} 1(T_{(k)} \leq t)$.

For $m = 0, 1, \dots$, update $\beta^{(m+1)}$ and $\Lambda^{(m+1)}$ iteratively until convergence:

Step 1 Update $\Lambda^{(m+1)}$ from $\beta^{(m)}$ and $\Lambda^{(m)}(t) = \sum_{k=1}^K \lambda_k^{(m)} 1(T_{(k)} \leq t)$ via equation (13):

$$\lambda_k^{(m+1)} = \left[\sum_{j=1}^n T_{(k)} \tilde{W}_j \{T_{(k)}, \beta^{(m)}, \Lambda^{(m)}(\cdot)\} e^{\beta^{(m)'} Z_j} q \{e^{\beta^{(m)'} Z_j} \Lambda^{(m)}(T_{(k)})\} \right]^{-1}. \quad (14)$$

Step 2 Update $\beta^{(m+1)}$ from $\beta^{(m)}$ and $\Lambda^{(m)}$ via (12) by solving the following equation of β using the Newton–Raphson method:

$$\sum_{i=1}^n \delta_i \left[Z_i - \frac{\sum_{j=1}^n Z_j \tilde{W}_j \{X_i, \beta, \Lambda^{(m)}(\cdot)\} e^{\beta' Z_j} q \{e^{\beta' Z_j} \Lambda^{(m)}(X_i)\}}{\sum_{j=1}^n \tilde{W}_j \{X_i, \beta, \Lambda^{(m)}(\cdot)\} e^{\beta' Z_j} q \{e^{\beta' Z_j} \Lambda^{(m)}(X_i)\}} \right] = 0. \quad (15)$$

A predefined convergence criterion should ensure both $\beta^{(m)}$ and $\Lambda^{(m)}$ converges computationally. For example, a useful criterion is when

$\sum_{j=1}^p |\beta_j^{(m+1)} - \beta_j^{(m)}| + \sum_{k=1}^K |\lambda_k^{(m+1)} - \lambda_k^{(m)}|$ is smaller than a suitable precision error, where $\beta_j^{(m)}$ is the j th component of $\beta^{(m)}$. To update (14) and (15), note that the survival function $S(t|Z_i) = \exp\{-Q e^{\beta' Z_i} \Lambda(t)\}$, where $\Lambda(\cdot)$ is updated by the algorithm, is a nonincreasing step

function with jumps at $T_{(1)} < T_{(2)} < \dots < T_{(K)}$, so the integrals, $\int_{C_i}^{\tau_n} S(u|Z_i) du$ and $\int_{\tau_{C_i}}^{\tau_n} S(u|Z_i) u^{-2} du = -\int_{\tau_{C_i}}^{\tau_n} S(u|Z_i) d(u^{-1})$, are simply the summations.

A reasonable initial point for the algorithm is the values obtained from the method for the traditional left-truncated and right-censored (LTRC) data. In our simulation studies, we found that the algorithm almost always converged when using the LTRC estimator as the starting point. With this starting point, we can show the local convergence of the algorithm, i.e., the sequence $\{\beta^{(m)}, \Lambda^{(m)}, m = 0, 1, \dots\}$ defined by the algorithm converges to the solution to the equations (14) and (15), and hence to the solution to the estimating equations (7) and (8). The basic approach is to apply Ostrowski's Theorem for proving the local convergence of iterative solutions of nonlinear equations; see, for example, Ortega and Rheinboldt (2000, page 300) or Lange (2010, page 279). Similar approach has been applied by Chen (2009) to prove the local convergence of the computational algorithm for re-

weighted estimating equations for the transformation model under the traditional right-censored survival data. The sketch of the proof is in Supplementary Material.

2.4 Asymptotic properties

By the computational algorithm, the estimators $\{\hat{\beta}_n, \hat{\Lambda}_n(\cdot, \hat{\beta}_n)\}$ are the solutions to the joint estimating equations (7) and (8). For each β , the solution to the equation (8) depends on β so we denote it by $\hat{\Lambda}_n(t, \beta)$ for $t \geq 0$. With a slight abuse of notation, we write $\hat{\Lambda}_n(\cdot) = \hat{\Lambda}_n(\cdot, \hat{\beta}_n)$.

Under the regularity conditions listed in Appendix A, we show that $\hat{\beta}_n$ converges to β_0 , and that $\hat{\Lambda}_n(t)$ converges to $\Lambda_0(t)$ almost surely and uniformly in $t \in [0, \zeta]$ for each $\zeta < \tau$. The essence of the strong consistency proof is that if $\hat{\Lambda}_n(t, \beta)$ converges at all, then they must converge to the solution of the asymptotic version of integral equation (9) in Λ

$$\Lambda(t) = \int_0^t \frac{E_0[W_1(u, \beta_0, \Lambda_0)q\{e^{\beta_0' Z_1} \Lambda_0(t)\}e^{\beta_0' Z_1}]}{E_0[W_1(u, \beta, \Lambda)q\{e^{\beta' Z_1} \Lambda(t)\}e^{\beta' Z_1}]} d\Lambda_0(s). \quad (16)$$

This is a nonlinear integral equation of the Hammerstein type (Tricomi 1985, page 202).

Note by regularity Condition 5, $q\{e^{\beta_0' Z_1} \Lambda_0(\cdot)\} > 0$ is bounded uniformly on $[0, \zeta]$. It follows that by the classical existence and uniqueness theorems for Hammerstein integral equations, there exists one and at most one solution to the deterministic equation (16), which is denoted by $\Lambda_0(t, \beta)$. By the uniqueness of the solution, $\hat{\Lambda}_n(t, \beta)$ must converge to $\Lambda_0(t, \beta)$ if $\hat{\Lambda}_n$ converges at all, and by (16), $\Lambda_0(t, \beta)$ must be a monotone increasing function. If $\hat{\beta}_n$ converges to β_0 , then $\hat{\Lambda}_n(t, \hat{\beta}_n)$ must converge to $\Lambda_0(t, \beta_0)$. However, when $\beta = \beta_0$, $\Lambda_0(t)$ also satisfies (16). Thus by the uniqueness of the solution to (16), $\Lambda_0(t, \beta_0) = \Lambda_0(t)$. To prove the strong consistency uniformly in t and uniformly in β , we invoke the empirical processes techniques (van der Vaart and Wellner 1996), by applying the continuous mapping theorem and the functional law of large numbers. Note by the Borel-Cantelli lemma, $\hat{\tau}_n \rightarrow \tau$ in probability if and only if for every subsequence, there is a further subsequence $\{n'\}$ such that $\hat{\tau}_{n'} \rightarrow \tau$ almost surely (van der Vaart and Wellner 1996, Lemma 1.9.2). Hence by Lemma 1, we can replace $\hat{\tau}_n$ by τ in the asymptotic proof without loss of generality. The detail of consistency proof is contained in Supplementary Material.

We establish the asymptotic normality of $n^{1/2}(\hat{\beta}_n - \beta_0)$ by applying the Taylor's expansion with respect to β to the estimating function $U_{1n}(\beta, \hat{\Lambda}_n(\cdot, \beta))$. Define

$$\begin{aligned} M_i(t, \beta, \Lambda) &= \delta_i N_i(t) - \int_0^t u W_i(u, \beta, \Lambda) dQ\{e^{\beta' Z_i} \Lambda(u)\}, \\ M_i^o(t) &= M_i(t, \beta_0, \Lambda_0), \quad S_i^o(t) = \exp\{-Q(e^{\beta_0' Z_i} \Lambda_0(t))\}, \\ W_i^o(t) &= W_i(t, \beta_0, \Lambda_0), \quad B_i^o(t) = q(e^{\beta_0' Z_i} \Lambda_0(t))e^{\beta_0' Z_i}, \end{aligned} \quad (17)$$

where the superscript o emphasizes the terms under the true values $\{\beta_0, \Lambda_0(\cdot)\}$. Let $\dot{q}(t) = dq(t)/dt$ be the first derivative of $q(\cdot)$. Define for $l = 0, 1, 2$,

$$\begin{aligned} J_1^{(l)}(u) &= E_0 \left\{ u W_1^o(u) B_1^o(u) Z_1^{\otimes l} \right\}, \quad J_2^{(l)}(u) = E_0 \left\{ u W_1^o(u) \dot{q}(e^{\beta_0' Z} \Lambda_0(u)) e^{2\beta_0' Z} Z_1^{\otimes l} \right\}, \\ J_3^{(l)}(u, v) &= E_0 \left[u B_1^o(u) \left\{ \frac{1(v \geq u)}{v^2 \int_C S_1^o(s) ds} - \frac{\int_{u \vee C}^{\tau} s^{-2} S_1^o(s) ds}{\left(\int_C S_1^o(s) ds \right)^2} \right\} \bar{\delta}_1 C_1 1(v \geq C_1) S_1^o(v) B_1^o(v) Z_1^{\otimes l} \right], \end{aligned}$$

where $\tilde{\delta}_1 = 1 - \delta_1$, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa'$. Define for $l = 0, 1$ the bivariate functions

$$K_1^{(l)}(t, u) = \dot{J}_1^{(l)}(u) 1(t > u) - \int_0^t J_3^{(l)}(v, u) d\Lambda_0(v), \quad (18)$$

where $\dot{J}_1^{(l)}(u) = dJ_1^{(l)}(u)/du$. Define

$$\sum_0 = E_0 \left[\int_0^\tau \{Z_i dM_1^o(t) - \bar{z}^o(t) M_1^o(t) dt\} \int_0^\tau \{Z_i dM_1^o(t) - \bar{z}^o(t) M_1^o(t) dt\}' \right], \quad (19)$$

where

$$\bar{z}^o(t) = \{K_1^{(1)}(\tau, t) + \int_0^\tau H_0(v, t) K_1^{(1)}(\tau, v) dv\} / J_1^{(0)}(t), \quad (20)$$

and $H_0(v, t)$ is a deterministic bivariate function defined by the following integral equation

$$H_0(t, u) = \frac{K_1^{(0)}(t, u)}{J_1^{(0)}(t)} + \int_0^\tau \frac{K_1^{(0)}(t, v)}{J_1^{(0)}(t)} H_0(v, u) dv. \quad (21)$$

Let Γ_0 be the limiting form of Jacobian matrix of the estimating function $U_{1n}(\beta, \hat{\Lambda}_n(\cdot, \beta))$ evaluated at the true value β_0 , i.e., $\partial \beta U_{1n}(\beta, \hat{\Lambda}_n(\cdot, \beta))|_{\beta=\beta_0} \rightarrow \Gamma_0$. Under some mild regularity conditions listed in Appendix A, we establish Theorem 1 given below:

Theorem 1—Under the regularity conditions (1)–(8), $n^{1/2}(\hat{\beta}_n - \beta_0)$ converges in distribution to a mean zero normal distribution with variance matrix $\Gamma_0^{-1} \sum_0 \Gamma_0^{-1}$.

We assume that the matrix Γ_0 is nonsingular so it has an inverse. This is a reasonable assumption as $\partial \beta U_{1n}(\beta, \hat{\Lambda}_n(\cdot, \beta))|_{\beta=\beta_0}$ is asymptotically equivalent to $\partial \beta U_{10}(\beta, \Lambda_0(\cdot, \beta))|_{\beta=\beta_0}$, where $\Lambda_0(\cdot, \beta)$ is the solution to equation (16), and $U_{10}(\beta_0, \Lambda_0(\cdot, \beta_0)) = 0$. Hence, the non-singularity condition can be checked empirically by whether the equations $U_{1n}(\beta, \hat{\Lambda}_n(\cdot, \beta))$ is nonsingular at $\beta = \hat{\beta}_n$. In addition, we have the following characterization on Γ_0 :

$$\Gamma_0 = \int_0^\tau J_1^{(1)}(u) d\Lambda_{\beta_0}^o(u) + \int_0^\tau \left[\sum_{k=2}^3 J_k^{(1)}(u) \{\Lambda_{\beta_0}^o(u)\}' + \sum_{k=1}^3 J_k^{(2)}(u) \right] d\Lambda_0(u), \quad (22)$$

where $\Lambda_{\beta_0}^o(u)$ is the limiting form of $\partial \beta \hat{\Lambda}_n(u, \beta)|_{\beta=\beta_0}$, which is asymptotically equivalent to $\partial \beta \Lambda_0(u, \beta)|_{\beta=\beta_0}$, and its explicit form is given by (45) in Appendix C.

The variance $\Gamma_0^{-1} \sum_0 \Gamma_0^{-1}$ comprises not only the variation from the estimator $\hat{\beta}_n$ but also the variation from the estimator $\hat{\Lambda}_n(\cdot)$ for the baseline function. While the variance may be estimated by its empirical plug-in counterpart, the computation is complicated as it requires the numerical solution to the integral equation (21) and (45). As we have established the asymptotic normality properties, we can use the empirical bootstrap method to estimate the standard deviations for the regression parameters and baseline functions (van der Vaart and Wellner 1996, Chapter 3.6).

3 Numerical studies

3.1 Simulation studies

To evaluate the performance of the proposed imputation method, we generate length-biased data from various underlying distributions and under different censoring mechanisms. Using the Pareto model (Clayton and Cuzick 1985), $Q(x) = \log(1 + \eta x)/\eta$ for given η and for $x > 0$, we consider $\eta = 0, 0.5$ and 1 , where $\eta = 0$ corresponds to the proportional hazards model and $\eta = 1$ to the proportional odds model. For given covariates Z , we generate $T = A + V$ from the following density function by the acceptance-rejection algorithm (Devroye 1986):

$$\frac{tf(t|Z)}{\mu(Z)} = \frac{1}{\mu(Z)} t e^{\beta' Z} \lambda(t) \exp \left[-Q(e^{\beta' Z} \Lambda(t)) \right] q \left\{ \int_0^t e^{\beta' Z} d\Lambda(u) \right\}. \quad (23)$$

It follows that for U uniformly distributed on $(0,1)$, $A = TU$ and $V = T(1 - U)$ have the desired distributions according to length-biased sampling (Vardi 1989).

We study two censoring situations: covariate-independent and covariate-dependent censoring. For covariate-independent censoring, the censoring time C follows the exponential distribution. For covariate-dependent censoring, the censoring time $b_0 C$ is correlated with Z_1 by generating $(Z_1, \log(C))$ from a bivariate normal distribution with 0.5 correlation, where b_0 is a constant to control the censoring proportion. We consider two dependent covariates, (Z_1, Z_2) , generated from a bivariate normal distribution with correlation 0.5. For each simulation scenario, five censoring proportions are considered: about 10%, 20%, 30%, 50% and 70%. We choose the regression coefficient to be $(\beta_{10}, \beta_{20}) = (1, 0)$ or $(1, -1)$ and the underlying baseline function to be $\Lambda_0(t) = t^4$. Under the Cox model, (23) becomes $4e^{\beta' Z} \mu(Z)^{-1} x^4 \exp(-e^{\beta' Z} x^4)$, where $\mu(Z) = \Gamma(5/4) \exp(-\beta' Z/4)$, which can be simulated from the gamma distribution. All simulation replicates are 1,000. We consider small to moderate sample sizes, $n = 50, 100$, and 200 .

We report the simulation results for two special cases under the Pareto model with $\eta = 0$, which is the Cox model, and with $\eta = 1$, which is the proportional odds model. In a small simulation study on the sensitivity of model misspecification (results not shown), we find that the proposed method can yield a consistent estimator when η is approximately close to the true model for moderate or large sample sizes, but otherwise can be sensitive to the misspecification of the Q function.

Under the Cox model, we compare the proposed imputation estimating equation (IMEE) procedure with two existing estimation procedures: the inversely weighted estimating equation (IWEE) procedure studied in Qin and Shen (2010), and the weighted U-Statistic estimating equation (WUEE) procedure studied in Shen et al (2009). In the proportional odds model, we compare the IMEE procedure with the WUEE procedure, as the IWEE approach has not been extended to the proportional odds model. The IWEE procedure estimates the regression coefficients β using the estimating equation

$$\sum_{k=1}^K \left[Z_{(k)} - \frac{\sum_{i=1}^n \delta_i 1(X_i \geq X_{(k)}) \{w(X_i)\}^{-1} Z_i \exp(\beta' Z_i)}{\sum_{i=1}^n \delta_i 1(X_i \geq X_{(k)}) \{w(X_i)\}^{-1} \exp(\beta' Z_i)} \right] = 0, \quad (24)$$

where $w(u) = \int_0^u S_c(v) dv$, and $S_c(\cdot)$ can be estimated by the Kaplan-Meier method. The WUEE procedure estimates the regression coefficients β using the estimating equation

$$\sum_{i,j} Z_{ij} \delta_{ij} \left\{ \frac{1(X_i \geq X_j) - \xi(\beta' Z_{ij})}{w(X_i)w(X_j)} \right\} = 0, \quad (25)$$

where $Z_{ij} = Z_i - Z_j$, $\delta_{ij} = \delta_i \delta_j$, and $\xi(\cdot)$ is a known modeling function. For the Cox model, $\xi(x) = (1 + e^x)^{-1}$, and for the proportional odds model, $\xi(x) = \{1 + e^x(x-1)\}(e^x - 1)^{-2}$.

The simulation results for the bias, standard deviation, and relative efficiency are summarized in Figure 1 when the censoring is independent of covariates, and Table 1 when the censoring is dependent of covariates. We only report the results for case when the true regression coefficient is $(\beta_{10}, \beta_{20}) = (1, 0)$. The results for $(\beta_{10}, \beta_{20}) = (1, -1)$ are similar and are not presented here due to limited space.

Of all three procedures considered under the Cox model, the IMEE is the most efficient and consistently has smaller bias and smaller variance, especially in small sample situations. When the censoring is independent of covariates, the three procedures have similar patterns: the bias and variance reduce with the increase in sample sizes, but all get larger with the increase in censoring proportions. When the sample size is moderately large and the censoring is relatively light, the IMEE and the IWEE perform similarly under the Cox model.

When the censoring distribution is covariate dependent, the IMEE is more efficient than the IWEE and WUEE under the Cox model, and more efficient than the WUEE under the proportional odds model. In fact, only the IMEE approach provides an unbiased estimator, and the advantage of the IMEE estimator is apparent as shown in Table 1. Compared to those of IWEE and WUEE, the estimators obtained from the IMEE also had smaller standard derivations.

Another benefit of IMEE approach is that it can estimate the baseline function simultaneously. We hence present the simulation results for the estimation of $\Lambda(\cdot)$ in Figure 2 under the Pareto model for $\eta = 0, 0.5, 1$ and 1.0 with $\beta_0 = (1, 0)$ and $\Lambda_0(t) = t^4$ for a selected value $\Lambda_0(0.9) \approx 0.66$. Under both the covariate-independent and covariate-dependent censoring, the bias of baseline function estimated by the IMEE procedure is reasonable small. When the sample size is very small ($n=50$), the IMEE estimator underestimates slightly the true value. The bias starts to disappear when the sample size increases. The estimated standard deviation increases with the censoring rate and decreases with the sample size. Finally, the IMEE performs similarly under covariate-independent censoring and covariate-dependent censoring.

3.2 Data analysis

We apply the proposed inference method to analyze the dementia data from the Canadian Study of Health and Aging, which were previously studied in Asgharian et al (2006) and Shen et al (2009). This was a prevalent cohort study in which Canadians over 65 years old were randomly sampled and screened for prevalent cases of dementia. For those identified with symptoms of dementia, the dates of onset were retrieved and the subjects were followed until death or censoring. The scientific question of interest was whether survival time was associated with the diagnosis of dementia, classified as probable Alzheimer's disease, possible Alzheimer's disease, and vascular dementia. The validity of the stationarity assumption has been confirmed in Asgharian et al (2006). For our analysis, we use the data from 818 subjects and apply the Pareto model with the modeling function $Q(x) = \log(1 + \eta x)/\eta$ for $\eta = 0, 0.5, 1$, and 1.5 . The covariate is the diagnosis of dementia, with probable Alzheimer's disease as the baseline. We use 300 replicates for the bootstraps resampling

method with replacement. In a small simulation study to evaluate the performance of the bootstrapping method (results not shown), we confirmed that the bootstrap SD was approximately close to the Monte Carlo empirical SD under moderate or large sample sizes when the censoring was not very heavy.

Table 2 presents the estimates of regression coefficients and the corresponding bootstrap standard deviations. The results are similar to those in Shen et al (2009), but with higher absolute Z -values for each estimate. Compared to vascular dementia, probable and possible Alzheimer's diseases are associated with lower risk of death as reflected by all four models. In particular, the survival difference between possible Alzheimer's disease and vascular dementia are statistically different ($p < 0.05$). These results are consistent with the clinical observation.

Figure 2 shows the estimated survival curves with four η values in solid lines and their 95% pointwise confidence intervals in gray area. Specifically, we obtain the bootstrap standard deviation of the logarithm of the estimated cumulative hazard function,

$\log[Q\{\exp(\hat{\beta}_n Z)\hat{\Lambda}_n(t)\}]$, for given Z . The confidence interval for the estimated survival function $\hat{S}(t/Z)$ is then calculated by the Delta method. The estimated imputation-based curves are also compared with Vardi's nonparametric maximum likelihood estimator (NPMLE) plotted in dot-dash lines (Vardi 1989). Figure 2 suggests that in most cases the survival curves estimated by the IMEE are close to those estimated by Vardi's NPMLE, which are well within the 95% pointwise confidence intervals of the IMEE estimator, although the survival curves estimated by the IMEE seem to be consistently above those estimated by Vardi's NPMLE. Judging by the distance between the survival curves estimated by the IMEE and those by Vardi's NPMLE, it is interesting to note that for possible Alzheimer's disease, the proportional hazards model seems to fit the data reasonably well, while for vascular dementia the proportional odds model or the Pareto model with $\eta = 1.5$ seems to fit the data well. For probable Alzheimer's disease, all models seem to have a reasonable fit.

4 Discussion

The proposed imputation-based estimation procedure for semiparametric transformation models can be applied to analyze censored length-biased data in many applications. The proposed method is better than the existing IWEE and WUEE method in terms of efficiency and robustness to censoring assumptions. The reason for the improved efficiency is that the imputation approach fully utilizes the covariate information from all subjects, whereas the IWEE method does not use the covariate information from censored subjects. The imputation method replaces the censored observations with their conditional expectations in the estimating equations. It also has the desirable feature not requiring the independence assumption between the covariates and the censoring times. Optimal weight functions may exist theoretically for the IWEE method, but they can depend on the unknown underlying distribution that needs to be estimated. As a result, the optimal weight function may not necessarily provide better performing estimators than the unweighted function in small sample situations.

Our imputation method is different from the imputation methods that were developed for traditional survival data under AFT models (Buckley and James 1979; Lai and Ying 1991). The Buckley-James procedure utilizes a location shift transformation under the AFT model. Such transformation leads to independently and identically distributed (i.i.d.) failure time data so the Kaplan-Meier estimator can be directly plugged in to the imputation procedure. The imputation for length-biased data is more difficult as there is no corresponding i.i.d. location shift for the data under the semiparametric transformation models. As the

imputation method involves both baseline function and covariate coefficients, a new computational algorithm has been devised while some sophisticated techniques have been developed for large-sample properties. In contrast to the imputation method for parametric models where only a finite number of parameters need to be estimated, the proposed method needs to deal with the infinite dimensional baseline function under the semiparametric transformation models. As a result, our approach is technically more challenging than the parametric imputation method. Moreover, even though the imputation method has been used under the AFT model by Buckley and James (1979), we are not aware of its application in the Cox model and the more general transformation models. Current work provides an alternative to the full MLE approach, which is complex under the transformation models. The efficiency of the imputation approach depends on the original length-biased estimating function developed by Wang (1996) in the absence of censoring. In general, the proposed method is not fully efficient unless there is no covariate and no censoring. If a score estimating equation can be constructed, we expect an improved efficiency. It is a classic problem in statistical inference to balance efficiency and robustness, and future investigation is warranted.

The imputation approach as delineated in (2) may be extended to other semiparametric model settings for censored biased-sampling data. Of course, the corresponding computational algorithms and asymptotic properties should be established under the new models. While the variance for $\hat{\beta}_n$ may be obtained from the asymptotic result, the computation is intensive as it requires numerical solutions to some integral equations. With the established asymptotic normality, it is computationally more convenient to use the bootstrap method. The proposed estimation procedure for the baseline function can facilitate the prediction of covariate-specific survival probabilities for right-censored length-biased data. All models are only approximations in real data analyses. It is thus important to investigate methodology for assessing model misspecification. Checking model specifications such as the functional forms and the link functions may lead to a family of semiparametric transformation models that are adequate for data analysis. Techniques based on time-dependent residuals for traditional survival data may be extended to right-censored length-biased data (Lin et al 2000; Chen et al 2012), but it is challenging to study their theoretical properties and to find practical implementation procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work was supported in part by the U.S. NIH grants CA079466 and CA016672. The authors thank one Referee for his/her constructive comments. The authors also thank Professor Asgharian and the investigators from the Canadian Study of Health and Aging for generously sharing the dementia data. The data reported in this article were collected as part of the Canadian Study of Health and Aging. The core study was funded by the Seniors' Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada Project 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.

References

- Asgharian M, Wolfson DB. Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Annals of Statistics*. 2005; 33(5):2109–2131.

- Asgharian M, M'LAN CE, Wolfson DB. Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association*. 2002; 97(457):201–209.
- Asgharian M, Wolfson DB, Zhang X. Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine*. 2006; 25(10):1751–1767. [PubMed: 16220462]
- Buckley J, James L. Linear regression with censored data. *Biometrika*. 1979; 66(3):429–436.
- Chen L, Lin DY, Zeng D. Checking semiparametric transformation models with censored data. *Biostatistics*. 2012; 13(1):18–31. [PubMed: 21785165]
- Chen YH. Weighted breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika*. 2009; 96(3):591–600.
- Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82:835–845.
- Cheng SC, Wei LJ, Ying Z. Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association*. 1997; 92(437):227–235.
- Clayton DG, Cuzick J. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A*. 1985; 148(2):82–117.
- Cox, DR. *Renewal theory*. Methuen; London: 1962.
- Cox, DR. Some sampling problems in technology. In: Johnson, UL.; Smith, H., editors. *New Developments in Survey Sampling*. Wiley Interscience; 1969.
- Dabrowska DM, Doksum KA. Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*. 1988; 83:744–749.
- Devroye, L. *Non-Uniform Random Variate Generation*. Springer-Verlag; New York: 1986.
- Gupta RC, Keating JP. Relations for reliability measures under length biased sampling. *Scandinavian Journal of Statistics*. 1986; 13(1):49–56.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2. Wiley-Interscience; Hoboken, NJ: 2002.
- Keiding, N. Independent delayed entry. In: Klein, JP.; Goel, P., editors. *Survival Analysis: State of the Art*. Kluwer Academic Publishers; Boston: 1992. p. 309–326.
- Kvam P. Length bias in the measurements of carbon nanotubes. *Technometrics*. 2008; 50(4):462–467.
- Lai TL, Ying Z. Large sample theory of a modified buckley-james estimator for regression analysis with censored data. *Annals of Statistics*. 1991; 19:1370–1402.
- Lange, K. *Numerical Analysis for Statisticians*. 2. Springer; New York: 2010.
- Lin DY. On fitting cox's proportional hazards models to survey data. *Biometrika*. 2000; 87(1):37–47.
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2000; 62(4):711–730.
- Mauldon M. Estimating mean fracture trace length and density from observations in convex windows. *Rock Mechanics and Rock Engineering*. 1998; 31(4):201–216.
- Nowell C, Stanley LR. Length-biased sampling in mall intercept surveys. *Journal of Marketing Research*. 1991; 28(4):475–479.
- Nowell C, Evans MA, McDonald L. Length-biased sampling in contingent valuation studies. *Land Economics*. 1988; 64(4):367–371.
- Ortega, JM.; Rheinboldt, WC. *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics. Academic Press; Philadelphia: 2000. *Iterative Solution of Nonlinear Equations in Several Variables*. 1970
- Otis DL, McDonald LL, Evans MA. Parameter estimation in encounter sampling surveys. *The Journal of Wildlife Management*. 1993; 57(3):543–548.
- Qin J, Shen Y. Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics*. 2010; 66(2):382–392. [PubMed: 19522872]
- Scheike TH, Keiding N. Design and analysis of time-to-pregnancy. *Statistical methods in medical research*. 2006; 15(2):127–140. [PubMed: 16615653]

- Shen Y, Ning J, Qin J. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*. 2009; 104(487): 1192–1202. [PubMed: 21057599]
- Simon R. Length biased sampling in etiologic studies. *American Journal of Epidemiology*. 1980; 111(4):444–452. [PubMed: 7377187]
- Smith TMF. Populations and selection: Limitations of statistics. *Journal of the Royal Statistical Society Series A (Statistics in Society)*. 1993; 156(2):144–166.
- Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE. True and false positive peaks in genomewide scans: Applications of length-biased sampling to linkage mapping. *American Journal of Human Genetics*. 1997; 61(2):430–438. [PubMed: 9311749]
- Tricomi, FG. *Integral Equations*. Dover Publications; New York: 1985.
- de Una-Álvarez J, Otero-Giráldez MS, Álvarez-Llorente G. Estimation under length-bias and right-censoring: An application to unemployment duration analysis for married women. *Journal of Applied Statistics*. 2003; 30(3):283–291.
- van der Vaart, A. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press; Cambridge, UK: 1998. *Asymptotic Statistics*.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. Springer; New York: 1996.
- Varadhan R, Frangakis CF. Revealing and addressing length bias and heterogeneous effects in frequency case-crossover studies. *American Journal of Epidemiology*. 2004; 159(6):596–602. [PubMed: 15003964]
- Vardi Y. Nonparametric estimation in the presence of length bias. *Annals of Statistics*. 1982; 10(2): 616–620.
- Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*. 1989; 76:751–761.
- Ventura V, Davison AC, Boniface SJ. Statistical inference for the effect of magnetic brain stimulation on a motoneurone. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 1998; 47(1):77–94.
- Wang MC. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*. 1991; 86:130–143.
- Wang MC. Hazards regression analysis for length-biased data. *Biometrika*. 1996; 83(2):343–354.
- Wang MC, Brookmeyer R, Jewell NP. Statistical models for prevalent cohort data. *Biometrics*. 1993; 49(1):1–11. [PubMed: 8513095]
- Zeidman MI, Batra SK, Sasser PE. Determining short fiber content in cotton. Part I: Some theoretical fundamentals. *Textile Research Journal*. 1991; 61(1):21–30.
- Zelen M. Forward and backward recurrence times and length biased sampling: Age specific models. *Lifetime Data Analysis*. 2004; 10(4):325–334. [PubMed: 15690988]
- Zelen M, Feinleib M. On the theory of screening for chronic diseases. *Biometrika*. 1969; 56:601–614.
- Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2007; 69(4):507–564.

Appendix A: Regularity conditions

Let $|x|$ be the Euclidean norm of x in $\mathbb{R}^p$, where $\mathbb{R}$ is the set of real numbers and p is the dimension of the covariates Z . Let $\mathcal{A}$ be the parameter set for $\Lambda(\cdot)$, defined as the collection of all nondecreasing functions $\Lambda(\cdot)$ with $\Lambda(\zeta)$ bounded for all $\zeta < \tau$. We summarize the regularity conditions as follows:

1. The length-biased data satisfy the stationarity assumption.
2. The distribution of (A, T) conditional on (C, Z) is the same as the distribution of (A, T) conditional on Z . Also, $P(A + C > T) > 0$, and there exists $u_0 > 0$ such that $P(C - u_0) > 0$.

3. The parameter β is in a compact set $\mathcal{B}$ in $\mathbb{R}^p$, which contains an open neighborhood of the true value β_0 . The support of the covariate vector Z is bounded.
4. The true value $\lambda_0(\cdot) \in \mathcal{A}$ is continuous and strictly increasing with a derivative $\lambda_0(\cdot)$. Furthermore, $\lambda_0(\cdot)$ is of bounded variation on $[0, \zeta]$ for all $\zeta < \tau$.
5. The function $Q(\cdot)$ is continuous and increasing, with $Q(0) = 0$ and twice differentiable. Its first derivative, denoted by $q(\cdot)$, is bounded and has a derivative, denoted by $\dot{q}(\cdot)$. Both functions $q(\cdot)$ and $\dot{q}(\cdot)$ are of bounded variation on $[0, \zeta]$ for all $\zeta < \tau$.
6. The matrix F_0 defined in (22) is nonsingular.
7. $\int_0^\tau \int_0^\tau \{K_1^{(0)}(t, u)/J_1^{(0)}(t)\}^2 du dt < \infty$, and $\int_0^\tau \int_0^t \{J_2^{(0)}(u)\lambda_0(u)/J_1^{(0)}(t)\}^2 du dt < \infty$.
8. If $\tau = \infty$, $\int_{\max_{1 \leq i \leq n} T_i}^\tau S(u) du = o_p(n^{-1/2})$.

Appendix B: Proof of mean-zero estimating equation

We begin the proof by calculating $E\{d\tilde{M}(t)|1 - \delta, C, Z\}$ in (2). First,

$$E\{dN(t)|1 - \delta, Z, C\} = \frac{P\{T \in (t, t+dt), T > A+C|C, Z\}}{P\{T > A+C|C, Z\}} = \frac{(t-C)1(t > C)f(t|Z)}{\int_C^\tau S(v|Z)dv}, \quad (26)$$

where in the last step, using integration by parts

$$\begin{aligned} P\{T \in (t, t+dt), T > A+C|C, Z\} &= (t-C)1(t > C)f(t|Z)/\mu(Z), \\ P\{T > A+C|C, Z\} &= \int_C^\tau (v-C)f(v|Z)dv/\mu(Z) = \int_C^\tau S(v|Z)dv/\mu(Z). \end{aligned} \quad (27)$$

Second,

$$E\{Y(t)T^{-1}|1 - \delta, C, Z\} = \frac{E\{1(T \geq t)1(T > A+C)T^{-1}|C, Z\}}{P\{T > A+C|C, Z\}} = \frac{\int_{t \vee C}^\tau (1-C/v)f(v|Z)dv}{\int_C^\tau S(v|Z)dv}. \quad (28)$$

Using integration by parts,

$$\int_{t \vee C}^\tau (1-C/v)f(v|Z)dv = C \int_{t \vee C}^\tau \frac{S(v|Z)}{v^2} dv + S(t|Z)(1-C/t)1(t > C). \quad (29)$$

Combining (26), (28) and (29),

$$\begin{aligned} E\{d\tilde{M}(t)|1 - \delta, C, Z\} &= E\{dN(t) - Y(t)T^{-1}t\lambda(t|Z)|1 - \delta, C, Z\} \\ &= -C \left\{ \int_{t \vee C}^\tau \frac{S(v|Z)}{v^2} dv \right\} t\lambda(t|Z) / \int_C^\tau S(v|Z)dv. \end{aligned} \quad (30)$$

We thus have confirmed equation (3), and

$$dM(t) = \delta d\tilde{M}(t) - (1-\delta)C \left\{ \int_{\tau}^{\tau} \frac{S(v|Z)}{v^2} dv \right\} t\lambda(t|Z) / \int_c^{\tau} S(v|Z) dv.$$

We will show that

$$E\{\delta d\tilde{M}(t)|C, Z\} = \left\{ C \int_{\tau}^{\tau} S(v|Z) v^{-2} dv \right\} t\lambda(t|Z) / \mu(Z). \quad (31)$$

Note that by (27)

$$E(1-\delta|C, Z) = P\{T > A+C|C, Z\} = \int_c^{\tau} S(u|Z) du / \mu(Z). \quad (32)$$

Then combining (30), (31) and (32), we have $E\{dM(t)/C, Z\} = 0$. The proposition $E\{dM(t)\} = 0$ follows immediately by conditional expectation calculation.

It remains to show (31). We have

$$\begin{aligned} E\{\delta dN(t)|C, Z\} &= E\{1(T \leq A+C, T \in (t, t+dt))|C, Z\} \\ &= \{C1(t \geq C) + t1(C > t)\} f(t|Z) / \mu(Z), \end{aligned} \quad (33)$$

$$\begin{aligned} E\{\delta Y(t)T^{-1}|C, Z\} &= E\{1(A+C \geq T)1(T \geq t)T^{-1}|C, Z\} \\ &= \int_t^{\tau} \{C1(v \geq C) + v1(C > v)\} \frac{f(v|Z)}{v} dv / \mu(Z). \end{aligned} \quad (34)$$

Using integration by parts,

$$\int_t^{\tau} C1(v \geq C) \frac{f(v|Z)}{v} dv = C \frac{S(t|Z)}{t} 1(t \geq C) + S(C|Z) 1(\tau > C > t) - C \int_{\tau}^{\tau} \frac{S(v|Z)}{v^2} dv.$$

Noting $\int_t^{\tau} 1(C > v) f(v|Z) dv = S(t|Z) 1(C > t) - S(C|Z) 1(\tau > C > t)$, we have

$$E\{\delta Y(t)T^{-1}t\lambda(t|Z)|C, Z\} = \left[\{C1(t \geq C) + t1(C > t)\} f(t|Z) - \left(C \int_{\tau}^{\tau} \frac{S(v|Z)}{v^2} dv \right) t\lambda(t|Z) \right] / \mu(Z) \quad (35)$$

Combining (33) and (35), we have confirmed (31).

Appendix C: Large sample properties

Proof of Lemma 1

We will show that when $\tau < \infty$, $T_{(n)} \equiv \max_{1 \leq i \leq n} (\delta_i X_i)$ converges to τ in probability, under the modest condition that there is a minimal follow-up, i.e., there exists some $u_0 > 0$ such that $S_C(u_0) = P(C > u_0) > 0$. This also implies that $\hat{\tau}_n - T_{(n)}$ converges to τ in probability.

For a small $\varepsilon > 0$,

$$P(|T_{(n)} - \tau| > \varepsilon) = P(T_{(n)} < \tau - \varepsilon) = \{P(\delta_1 X_1 < \tau - \varepsilon)\}^n = \{1 - P(\delta_1 X_1 \geq \tau - \varepsilon)\}^n.$$

With loss of generality, consider the case that there is no covariate, and let $f(t)$ and $S(t)$ denote the density and survival function, respectively, for the unbiased time $\tilde{T}$ with mean μ . For length-biased data,

$$\begin{aligned} P(\delta_1 X_1 \geq \tau - \varepsilon) &= P(A_1 + C_1 \geq T_1 \geq \tau - \varepsilon) = \int_{\tau - \varepsilon}^{\tau} E \left\{ \int_0^t 1(a + C \geq t) da \right\} f(t) dt / \mu \\ &= \int_{\tau - \varepsilon}^{\tau} \left\{ \int_0^t S_c(t - a) da \right\} f(t) dt / \mu = \int_{\tau - \varepsilon}^{\tau} w(t) f(t) dt / \mu. \end{aligned} \quad (36)$$

By the mean value theorem, there exists $t - \varepsilon \leq \xi \leq \tau$ such that

$$\int_{\tau - \varepsilon}^{\tau} w(t) f(t) dt = w(\xi) f(\xi) \int_{\tau - \varepsilon}^{\tau} dt = w(\xi) f(\xi) \varepsilon.$$

It follows that by the minimal follow-up condition $(P(T_{(n)} - \tau > \varepsilon) = \{1 - w(\xi) f(\xi) \varepsilon\}^n \rightarrow 0$.

Furthermore, we can show that $n^\alpha \{T_{(n)} - \tau\} \rightarrow 0$ in probability even for $1 > \alpha > 1/2$. The proof is similar to the above arguments. Replacing ε with $n^{-\alpha} \varepsilon$ in (36), we have that there exists some $\tau - \varepsilon \leq \xi \leq \tau - n^{-\alpha} \varepsilon$,

$$\begin{aligned} P(n^\alpha |T_{(n)} - \tau| > \varepsilon) &= P(|T_{(n)} - \tau| > n^{-\alpha} \varepsilon) = \{1 - P(\delta_1 X_1 \geq \tau - n^{-\alpha} \varepsilon)\}^n \\ &= [1 - w(\xi) f(\xi) n^{-\alpha} \varepsilon]^n \approx \exp[-w(\xi) f(\xi) n^{1-\alpha} \varepsilon] \rightarrow 0. \end{aligned}$$

Notations

Let $W_i(t) = W_i(t, \beta, \Lambda)$ as defined in (5), $S_i(t, \beta, \Lambda) \equiv \exp[-Q\{\int_0^t e^{\beta' Z_i} d\Lambda(u)\}]$ as defined in (6), and $M_i(t) = M_i(t, \beta, \Lambda)$ as defined in (17). Denote

$$B_i(u) \equiv B_i(u, \beta, \Lambda) = q(e^{\beta' Z_i} \Lambda(u)) e^{\beta' Z_i}.$$

Denote these terms evaluated at (β_0, Λ_0) by

$$W_i^o(t) = W_i(t, \beta_0, \Lambda_0), S_i^o(t) = S_i(t, \beta_0, \Lambda_0), M_i^o(t) = M_i(t, \beta_0, \Lambda_0), B_i^o(t) = B_i(t, \beta_0, \Lambda_0).$$

Let $\hat{\Lambda}^o(\cdot) = \hat{\Lambda}(\cdot, \beta_0)$. Denote these terms evaluated at $(\beta_0, \hat{\Lambda}^o(\cdot))$ by

$$\widehat{W}_i^o(t) = W_i(t, \beta_0, \widehat{\Lambda}^o), \widehat{S}_i^o(t) = S_i(t, \beta_0, \widehat{\Lambda}^o), \widehat{M}_i^o(t) = M_i(t, \beta_0, \widehat{\Lambda}^o), \widehat{B}_i^o(t) = B_i(t, \beta_0, \widehat{\Lambda}^o).$$

We also denote $\tilde{\delta}_i \equiv 1 - \delta_i$ and $\tilde{C}_i(t) \equiv C_i 1(t - C_i)$.

Asymptotic normality

For notation simplicity, we suppress the subscript n from $\hat{\beta}_n$ and $\hat{\Lambda}_n$. We establish the asymptotic normality of $n^{1/2}(\hat{\beta} - \beta_0)$ by the Taylor's expansion of $\tilde{U}_{1n}\{\beta, \hat{\Lambda}(\beta)\}$ with respect to β in the neighborhood of β_0

$$0 = \tilde{U}_{1n}\{\hat{\beta}, \hat{\Lambda}(\hat{\beta})\} = \tilde{U}_{1n}\{\beta_0, \hat{\Lambda}(\beta_0)\} + \frac{\partial}{\partial \beta} \tilde{U}_{1n}\{\beta, \hat{\Lambda}(\beta)\} \Big|_{\beta=\beta_0} (\hat{\beta} - \beta_0) + o_p(n^{-1/2}).$$

By Lemma 1 and by Assumption (8), $\sqrt{n}\tilde{U}_{1n}\{\beta, \hat{\Lambda}(\beta)\}$ and $\sqrt{n}U_{1n}\{\beta, \hat{\Lambda}(\beta)\}$ are asymptotically equivalent. We will work with $U_{1n}\{\beta, \hat{\Lambda}(\beta)\}$ for the asymptotic proof. The proofs are divided into 4 steps: the independent and identically distributed representation of $\sqrt{n}\{\hat{\Lambda}(t, \beta_0) - \Lambda_0(t, \beta_0)\}$; the weak convergence of $U_{1n}\{\beta_0, \hat{\Lambda}(\cdot, \beta_0)\}$; the strong consistency of $\beta U_{1n}\{\beta, \hat{\Lambda}(\cdot, \beta)\}|_{\beta=\beta_0}$; and the strong consistency of $\beta \hat{\Lambda}(u, \beta)|_{\beta=\beta_0}$.

Step 1

Independent and identically distributed characterization of $\sqrt{n}\{\hat{\Lambda}(\cdot, \beta_0) - \Lambda_0(\cdot, \beta_0)\}$

Recall that $\hat{\Lambda}^o(\cdot) \equiv \hat{\Lambda}(\cdot, \beta_0)$ and $\Lambda_0(\cdot, \beta_0) = \Lambda_0(\cdot)$. By definition, $\hat{\Lambda}(t, \beta)$ satisfies equation (8) for each β , and in particular $U_{2n}(t, \beta_0, \hat{\Lambda}^o) = 0$. Then,

$$\begin{aligned} -U_{2n}(t, \beta_0, \Lambda_0) &= U_{2n}(t, \beta_0, \hat{\Lambda}^o) - U_{2n}(t, \beta_0, \Lambda_0) \\ &= -\frac{1}{n} \sum_{i=1}^n \int_0^t u \left[\{\hat{W}_i^o(u) - W_i^o(u)\} d \left\{ Q \left(e^{\beta_0' Z_i} \hat{\Lambda}^o(u) \right) - Q \left(e^{\beta_0' Z_i} \Lambda_0(u) \right) \right\} \right. \\ &\quad \left. + W_i^o(u) d \left\{ Q \left(e^{\beta_0' Z_i} \hat{\Lambda}^o(u) \right) - Q \left(e^{\beta_0' Z_i} \Lambda_0(u) \right) \right\} + \{\hat{W}_i^o(u) - W_i^o(u)\} d Q \left(e^{\beta_0' Z_i} \Lambda_0(u) \right) \right]. \end{aligned} \quad (37)$$

By the strong convergence of $\hat{\Lambda}^o(t) \rightarrow \Lambda_0(t)$ combined with the Taylor's expansion of $W_i(u; \beta_0, \Lambda)$ and $Q(e^{\beta_0' Z_i} \Lambda(u))$ with respect to Λ , the first term in (37) is $o_p(n^{-1/2})$, i.e.,

$$n^{-1/2} \sum_{i=1}^n \int_0^t u \{\hat{W}_i^o(u) - W_i^o(u)\} d \left\{ Q \left(e^{\beta_0' Z_i} \hat{\Lambda}^o(u) \right) - Q \left(e^{\beta_0' Z_i} \Lambda_0(u) \right) \right\} = o_p(1).$$

By Taylor's formula, the second term in (37) is

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \int_0^t u W_i^o(u) d \left\{ Q \left(e^{\beta_0' Z_i} \hat{\Lambda}^o(u) \right) - Q \left(e^{\beta_0' Z_i} \Lambda_0(u) \right) \right\} &= \frac{1}{n} \sum_{i=1}^n \int_0^t u W_i^o(u) q \left(e^{\beta_0' Z_i} \Lambda_0(u) \right) e^{\beta_0' Z_i} d \{\hat{\Lambda}^o(u) - \Lambda_0(u)\} + o_p(n^{-1/2}) \\ &= \int_0^t J_1^{(0)}(u) d \{\hat{\Lambda}^o(u) - \Lambda_0(u)\} + o_p(n^{-1/2}) \\ &= J_1^{(0)}(t) \{\hat{\Lambda}^o(t) - \Lambda_0(t)\} - \int_0^t \{\hat{\Lambda}^o(u) - \Lambda_0(u)\} \dot{J}_1^{(0)}(u) du + o_p(n^{-1/2}), \end{aligned}$$

where $\dot{J}_1^{(0)}(u) = dJ_1^{(0)}(t)/dt$, and the last line is obtained using integration by parts. The third term in (37) can be expressed as

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \int_0^t u \{W_i(u; \beta_0, \widehat{\Lambda}^o) - W_i(u; \beta_0, \Lambda_0)\} dQ(e^{\beta_0' Z_i} \Lambda_0(u)) \\
&= \int_0^t u \left[\frac{1}{n} \sum_{i=1}^n \tilde{\delta}_i \left\{ \frac{\int_u^\tau \tilde{C}_i(v) \{\widehat{S}_i^o(v) - S_i^o(v)\} v^{-2} dv}{\int_{C_i}^\tau \widehat{S}_i^o(s) ds} \right. \right. \\
&+ \left. \frac{\int_u^\tau \tilde{C}_i(v) S_i^o(v) v^{-2} dv}{\int_{C_i}^\tau \widehat{S}_i^o(s) ds} - \frac{\int_u^\tau \tilde{C}_i(v) S_i^o(v) v^{-2} dv}{\int_{C_i}^\tau \widehat{S}_i^o(v) dv} \right\} B_i^o(u) \Big] d\Lambda_0(u) \\
&= \int_0^t u \left[\frac{1}{n} \sum_{i=1}^n \frac{\tilde{\delta}_i B_i^o(u)}{\int_{C_i}^\tau \widehat{S}_i^o(s) ds} \left\{ \int_u^\tau v^{-2} \tilde{C}_i(v) S_i^o(v) B_i^o(v) \{\widehat{\Lambda}^o(v) - \Lambda_0(v)\} dv \right. \right. \\
&- \left. \left. \frac{\int_u^\tau \tilde{C}_i(v) S_i^o(v) v^{-2} dv}{\int_{C_i}^\tau \widehat{S}_i^o(s) ds} \int_{C_i}^\tau S_i^o(v) B_i^o(v) \{\widehat{\Lambda}^o(v) - \Lambda_0(v)\} dv \right\} \right] d\Lambda_0(u) + o_p(n^{-1/2}) \\
&= \int_0^t u \int_0^\tau \left[\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tilde{\delta}_i B_i^o(u)}{v^2 \int_{C_i}^\tau \widehat{S}_i^o(s) ds} \tilde{C}_i(v) S_i^o(v) B_i^o(v) 1(v \geq u) \right. \right. \\
&- \left. \left. \frac{\tilde{\delta}_i B_i^o(u) \int_{uvC_i}^\tau s^{-2} S_i^o(s) ds}{\left\{ \int_{C_i}^\tau S_i^o(s) ds \right\}^2} \tilde{C}_i(v) S_i^o(v) B_i^o(v) \right\} \{\widehat{\Lambda}^o(v) - \Lambda_0(v)\} dv \right] d\Lambda_0(u) + o_p(n^{-1/2}) \\
&= \int_0^t \left[\int_0^\tau J_3^{(0)}(u, v) \{\widehat{\Lambda}^o(v) - \Lambda_0(v)\} dv \right] d\Lambda_0(u) + o_p(n^{-1/2}) \\
&= \int_0^\tau \left\{ \int_0^t J_3^{(0)}(v, u) d\Lambda_0(v) \right\} \{\widehat{\Lambda}^o(u) - \Lambda_0(u)\} du + o_p(n^{-1/2}).
\end{aligned}$$

Now equation (37) becomes

$$\frac{1}{n} \sum_{i=1}^n M_i^o(t) = J_1^{(0)}(t) \{\widehat{\Lambda}^o(t) - \Lambda_0(t)\} - \int_0^t \left\{ J_1^{(0)}(u) 1(t > u) - \int_0^t J_3^{(0)}(v, u) d\Lambda_0(v) \right\} \{\widehat{\Lambda}^o(u) - \Lambda_0(u)\} du + o_p(n^{-1/2}).$$

Hence, $\widehat{\Lambda}^o(t) - \Lambda_0(t)$ is the solution of the following integral equation with respect to $\eta(\cdot)$:

$$\frac{1}{n} \sum_{i=1}^n \frac{M_i^o(t)}{J_1^{(0)}(t)} = \eta(t) - \int_0^t \frac{K_1^{(0)}(t, u)}{J_1^{(0)}(t)} \eta(u) du + o_p(n^{-1/2}). \quad (38)$$

where $K_1^{(0)}(t, u) = J_1^{(0)}(u) 1(t > u) - \int_0^t J_3^{(0)}(v, u) d\Lambda_0(v)$ as defined in (18).

This is the Fredholm integral equation of the second kind (Tricomi 1985, Chapter II). For length-biased data, $J_1^{(0)}(t) = E_0\{tW(t, \beta_0, \Lambda_0)q(e^{\beta_0' Z} \Lambda_0(t))e^{\beta_0' Z}\}$. By Proposition 1, $E\{M_i(t, \beta_0, \Lambda_0)\} = 0$. Hence, for $t > 0$,

$$J_1^{(0)}(t) \lambda_0(t) = E_0\{\delta dN(t)\} = E_0\left\{ \frac{f_0(t|Z)}{\mu_0(Z)} \int_0^t S_c(v|Z) dv \right\},$$

where $S_c(v|Z) = P(C > v|Z)$. For t in the support of unbiased time $\tilde{T}$, $\lambda_0(t)$ and $f_0(t|Z)/\mu_0(Z)$ are always positive all z in the support of Z . Assume a minimal follow-up: there exists $u_0 > 0$ such that $P(C > u_0) > 0$. Then, if $t > u_0$,

$$J_1^{(0)}(t) \lambda_0(t) \geq E_0\{\inf_{u > u_0} f_0(u|Z) \int_0^{u_0} S_c(v|Z) dv / \mu_0(Z)\} > 0. \text{ If } u_0 < t < \infty,$$

$J_1^{(0)}(t)\lambda_0(t) \geq E_0\{f_0(t|Z)\int_0^t S_c(u_0|Z)dv/\mu_0(Z)\}=tf_0(t|Z)S_c(u_0|Z)/\mu_0(z)>0$. It follows that $J_1^{(0)}(t)$ is always positive for $t>0$. As $M_i^o(0)=0$ and $K_1^{(0)}(0,u)=0$, we define $M_i^o(0)/J_1^{(0)}(0)=0$ and $K_1^{(0)}(0,u)/J_1^{(0)}(0)=0$ for all $u \neq 0$. Hence, the integrals in (38) are well defined and bounded. By the classical theorem of linear integral equations, there is a unique solution to the above integral equation (38) by the assumption (7) that the kernel function $K_1^{(0)}(t,u)/J_1^{(0)}(t)$ is square integrable. It follows that we have the following representation of $\hat{\Lambda}(t,\beta_0) - \Lambda_0(t,\beta_0)$:

$$\hat{\Lambda}^o(t) - \Lambda_0(t) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{M_i^o(t)}{J_1^{(0)}(t)} + \int_0^t H_0(t,u) \frac{M_i^o(u)}{J_1^{(0)}(u)} du \right\} + o_p(n^{-1/2}), \quad (39)$$

where the deterministic function $H_0(t,u)$ depends only on the true baseline function $\Lambda_0(\cdot)$ and its derivative $\lambda_0(\cdot)$, and it satisfies the equation given in (21):

$$H_0(t,u) = \frac{K_1^{(0)}(t,u)}{J_1^{(0)}(t)} + \int_0^t \frac{K_1^{(0)}(t,v)}{J_1^{(0)}(t)} H_0(v,u) dv.$$

Therefore by equation (39), $\hat{\Lambda}^o(t) - \Lambda_0(t)$ can be represented as the weighted summation of independent and identically distributed $\{M_i(t) : i = 1, \dots, n\}$. Note that by Proposition 1, the mean of the equation (39) is 0.

Step 2

Asymptotic normality of $n^{1/2}U_{1n}(\beta_0, \hat{\Lambda}(\cdot, \beta_0))$

We start with the following decomposition

$$\begin{aligned} U_{1n}(\beta_0, \hat{\Lambda}^o(\cdot)) &= U_{1n}(\beta_0, \Lambda_0) + U_{1n}(\beta_0, \hat{\Lambda}^o(\cdot)) - U_{1n}(\beta_0, \Lambda_0) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau Z_i \left[dM_i^o(u) - u \{ \widehat{W}_i^o(u) dQ(e^{\beta_0' Z_i} \hat{\Lambda}^o(u)) - W_i^o(u) dQ(e^{\beta_0' Z_i} \Lambda_0(u)) \} \right]. \end{aligned} \quad (40)$$

The last two terms in the above display can be written as

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau u \left[\{ \widehat{W}_i^o(u) - W_i^o(u) \} d \left\{ Q(e^{\beta_0' Z_i} \hat{\Lambda}^o(u)) - Q(e^{\beta_0' Z_i} \Lambda_0(u)) \right\} \right. \\ &\quad \left. + W_i^o(u) d \left\{ Q(e^{\beta_0' Z_i} \hat{\Lambda}^o(u)) - Q(e^{\beta_0' Z_i} \Lambda_0(u)) \right\} + \{ \widehat{W}_i^o(u) - W_i^o(u) \} dQ(e^{\beta_0' Z_i} \Lambda_0(u)) \right] Z_i. \end{aligned} \quad (41)$$

By the strong convergence of $\hat{\Lambda}^o(\cdot) \rightarrow \Lambda_0(\cdot)$, the first term in (41) is $o_p(n^{-1/2})$. The second term in (41) is

$$\begin{aligned} &\int_0^\tau u \left\{ \frac{1}{n} \sum_{i=1}^n W_i^o(u) B_i^o(u) Z_i \right\} d \{ \hat{\Lambda}^o(u) - \Lambda_0(u) \} + o_p(n^{-1/2}) \\ &= \int_0^\tau J_1^{(1)}(u) d \{ \hat{\Lambda}^o(u) - \Lambda_0(u) \} + o_p(n^{-1/2}) = \int_0^\tau \{ \hat{\Lambda}^o(u) - \Lambda_0(u) \} J_1^{(1)}(u) du + o_p(n^{-1/2}). \end{aligned}$$

Plugging in the characterization of $\hat{\Lambda}^o(u) - \Lambda_0(u)$, the above display becomes

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{M_i^o(u)}{J_1^{(0)}(u)} + \int_0^\tau H_0(u, v) \frac{M_i^o(v)}{J_1^{(0)}(v)} dv \right\} J_1^{(1)}(u) du + o_p(n^{-1/2}).$$

The third term in (41) is

$$\begin{aligned} & \int_0^\tau u \left[\frac{1}{n} \sum_{i=1}^n \{ \widehat{W}_i^o(u) - W_i^o(u) \} B_i^o(u) Z_i \right] d\Lambda_0(u) \\ &= \int_0^\tau u \left[\frac{1}{n} \sum_{i=1}^n \tilde{\delta}_i \left\{ \frac{\int_u^\tau v^{-2} \tilde{C}_i(v) \{ \widehat{S}_i^o(v) - S_i^o(v) \} dv}{\int_{C_i}^\tau \widehat{S}_i^o(s) ds} \right. \right. \\ & \quad \left. \left. + \frac{\int_u^\tau v^{-2} \tilde{C}_i(v) S_i^o(v) dv}{\int_{C_i}^\tau \widehat{S}_i^{(0)}(s) ds} - \frac{\int_u^\tau v^{-2} \tilde{C}_i(v) S_i^o(v) dv}{\int_{C_i}^\tau S_i^o(v) dv} \right\} B_i^o(u) Z_i \right] d\Lambda_0(u) \\ &= \int_0^\tau u \int_0^\tau \left[\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tilde{\delta}_i B_i^o(u)}{v^2 \int_{C_i}^\tau \widehat{S}_i^o(s) ds} \tilde{C}_i(v) S_i^o(v) B_i^o(v) 1(v \geq u) \right. \right. \\ & \quad \left. \left. - \frac{\tilde{\delta}_i B_i^o(u) \int_{u \vee C_i}^\tau s^{-2} S_i^o(s) ds}{\left\{ \int_{C_i}^\tau S_i^o(s) ds \right\}^2} \tilde{C}_i(v) S_i^o(v) B_i^o(v) \right\} Z_i \{ \widehat{\Lambda}^o(v) - \Lambda_0(v) \} dv \right] d\Lambda_0(u) + o_p(n^{-1/2}) \\ &= \int_0^\tau \left\{ \int_0^\tau J_3^{(1)}(v, u) d\Lambda_0(v) \right\} \{ \widehat{\Lambda}^o(u) - \Lambda_0(u) \} du + o_p(n^{-1/2}). \end{aligned}$$

Plugging in the characterization of $\widehat{\Lambda}^o(v) - \Lambda_0(v)$, the above display becomes

$$\int_0^\tau \left\{ \int_0^\tau J_3^{(1)}(v, u) d\Lambda_0(v) \right\} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{M_i^o(u)}{J_1^{(0)}(u)} + \int_0^\tau H_0(u, v) M_i^o(v) dv \right\} du + o_p(n^{-1/2}).$$

It follows that (40) is asymptotically equivalent to the summation of i.i.d. $M_i^o(\cdot)$, $i = 1, \dots, n$,

$$\frac{1}{n} \sum_{i=1}^n \left\{ \int_0^\tau \left(K_1^{(1)}(\tau, u) + \int_0^\tau K_1^{(1)}(\tau, v) H_0(v, u) dv \right) \frac{M_i^o(u)}{J_1^{(0)}(u)} du \right\} + o_p(n^{-1/2}),$$

where $K_1^{(1)}(\tau, u) = J_1^{(1)}(u) + \int_0^\tau J_3^{(1)}(v, u) d\Lambda_0(v)$ as defined in (18).

Finally, $U_{1n}(\beta_0, \widehat{\Lambda}(\cdot, \beta_0))$ is asymptotically equivalent to

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ Z_i dM_i^o(t) - \bar{z}^o(t) M_i^o(t) dt \right\} + o_p(n^{-1/2}),$$

where $\bar{z}^o(t) = \{ K_1^{(1)}(\tau, t) + \int_0^\tau H_0(v, t) K_1^{(1)}(\tau, v) dv \} / J_1^{(0)}(t)$ as defined in (20). Applying the central limit theorem, we have that $n^{1/2} U_{1n}(\beta_0, \widehat{\Lambda}(\beta_0))$ converges in distribution to a mean-zero normal distribution with a covariance given by (19).

Step 3

Strong consistency of $\hat{\Lambda}(\cdot, \beta)|_{\beta=\beta_0}$

Let $\hat{\Lambda}_\beta(u) = \int_0^u \hat{\Lambda}(u, \beta)$. Denote

$$\widehat{W}_i(t, \beta) \equiv W_i(t; \beta, \widehat{\Lambda}(\beta)), \widehat{S}_i(t, \beta) \equiv \exp \left\{ -Q \left(e^{\beta' Z_i} \widehat{\Lambda}(t, \beta) \right) \right\}, \widehat{B}_i(t, \beta) \equiv q \left\{ e^{\beta' Z_i} \widehat{\Lambda}(t, \beta) \right\} e^{\beta' Z_i}.$$

Differentiating both sides of the equation $U_{2n}(t, \beta, \hat{\Lambda}(\cdot, \beta)) = 0$ with respect to β , we have

$$0 = \frac{1}{n} \sum_{i=1}^n \int_0^t u \left[\widehat{W}_i(u, \beta) \widehat{B}_i(u, \beta) \left\{ Z_i d\widehat{\Lambda}(u, \beta) + d\widehat{\Lambda}_\beta(u) \right\} + \frac{\partial}{\partial \beta} \left\{ \widehat{W}_i(u, \beta) \widehat{B}_i(u, \beta) \right\} d\widehat{\Lambda}(u, \beta) \right], \quad (42)$$

where $\frac{\partial}{\partial \beta} \hat{S}_i(u, \beta) = -\hat{S}_i(u, \beta) \hat{q}_i(u, \beta) e^{\beta' Z_i} \{ \hat{\Lambda}_\beta(u) + Z_i \hat{\Lambda}(u, \beta) \}$

$$\begin{aligned} \frac{\partial}{\partial \beta} \widehat{B}_i(u, \beta) &= \dot{q} \left(e^{\beta' Z_i} \widehat{\Lambda}(u, \beta) \right) e^{2\beta' Z_i} \left[\widehat{\Lambda}_\beta(u) + Z_i \widehat{\Lambda}(u, \beta) \right], \\ \frac{\partial}{\partial \beta} \widehat{W}_i(t, \beta) &= \tilde{\delta}_i \left[\frac{\int_0^\tau u^{-2} \tilde{C}_i(u) \widehat{S}_i(u, \beta) \widehat{B}_i(u, \beta) (\widehat{\Lambda}_\beta(u) + Z_i \widehat{\Lambda}(u, \beta)) du}{\int_{C_i} \widehat{S}_i(u, \beta) du} - \frac{\int_0^\tau u^{-2} \tilde{C}_i(u) \widehat{S}_i(u, \beta) du}{\left\{ \int_{C_i} \widehat{S}_i(u, \beta) du \right\}^2} \int_{C_i} \widehat{S}_i(u, \beta) \widehat{B}_i(u, \beta) (\widehat{\Lambda}_\beta(u) + Z_i \widehat{\Lambda}(u, \beta)) du \right]. \end{aligned}$$

Rearranging the terms on both sides of equation (42) and evaluating at $\beta = \beta_0$, we observe that $\hat{\Lambda}_{\beta_0}(u) = \int_0^u \hat{\Lambda}(u, \beta)|_{\beta=\beta_0}$ satisfies the following integral equation

$$\begin{aligned} & \int_0^t u \left[\int_0^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u) \int_{u \vee C_i}^\tau \widehat{S}_i^o(s) s^{-2} ds}{\left(\int_{C_i}^\tau \widehat{S}_i^o(s) ds \right)^2} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i \right\} \widehat{\Lambda}^o(v) dv \right] d\widehat{\Lambda}^o(u) \\ & - \int_0^t u \left[\int_u^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u)}{v^2 \int_{C_i}^\tau \widehat{S}_i^o(s) ds} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i \right\} \widehat{\Lambda}^o(v) dv \right] d\widehat{\Lambda}^o(u) \\ & - \int_0^t u \left[\left\{ \frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \dot{q} \left(e^{\beta_0' Z_i} \widehat{\Lambda}^o(u) \right) e^{2\beta_0' Z_i} Z_i \right\} \widehat{\Lambda}^o(u) + \left\{ \frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \widehat{B}_i^o(u) Z_i \right\} \right] d\widehat{\Lambda}^o(u) \\ & = \int_0^t u \left[\frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \dot{q} \left\{ e^{\beta_0' Z_i} \widehat{\Lambda}^o(u) \right\} e^{2\beta_0' Z_i} \right] \widehat{\Lambda}_{\beta_0}(u) d\widehat{\Lambda}^o(u) + \int_0^t u \left[\frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \widehat{B}_i^o(u) \right] d\widehat{\Lambda}_{\beta_0}(u) \\ & + \int_0^t u \left[\int_u^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u)}{v^2 \int_{C_i}^\tau \widehat{S}_i^o(s) ds} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) \right\} \widehat{\Lambda}_{\beta_0}(v) dv \right] d\widehat{\Lambda}^o(u) \\ & - \int_0^t u \left[\int_0^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u) \int_{u \vee C_i}^\tau \widehat{S}_i^o(s) s^{-2} ds}{\left(\int_{C_i}^\tau \widehat{S}_i^o(s) ds \right)^2} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) \right\} \widehat{\Lambda}_{\beta_0}(v) dv \right] d\widehat{\Lambda}^o(u). \end{aligned} \quad (43)$$

We have re-expressed the formula in terms of the empirical processes, and moved the terms involving $\hat{\Lambda}_{\beta_0}(\cdot)$ to the right-hand side of the equation. By the functional law of large numbers and by the strong consistency of $\hat{\Lambda}(u, \beta_0)$ to Λ_0 , for $l = 0, 1$,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left\{ u \widehat{W}_i^o(u) \widehat{B}_i^o(u) Z_i^{\otimes l} \right\} \rightarrow J_1^{(l)}(u) \\
& \frac{1}{n} \sum_{i=1}^n \left\{ u \widehat{W}_i^o(u) \dot{q} \left(e^{\beta_0' Z_i} \widehat{\Lambda}^o(u) \right) e^{2\beta_0' Z_i} Z_i^{\otimes l} \right\} \rightarrow J_2^{(l)}(u) \\
& \frac{1}{n} \sum_{i=1}^n \left[u \widehat{B}_i^o(u) \left\{ \frac{1(v \geq u)}{v^2 \int_{C_i}^{\tau} \widehat{S}_i^o(s) ds} - \frac{\int_{u \vee C_i}^{\tau} \widehat{S}_i^o(s) s^{-2} ds}{\left(\int_{C_i}^{\tau} \widehat{S}_i^o(s) ds \right)^2} \right\} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i^{\otimes l} \right] \rightarrow J_3^{(l)}(u, v).
\end{aligned}$$

Note that the functions on the right-hand side of the above display are continuous, and involve only the true baseline function $\Lambda_0(\cdot)$ and its derivative $\lambda_0(\cdot)$.

Similar to the proof of strong consistency of $\hat{\Lambda}_n(\cdot, \beta_0)$, we can show that $\hat{\Lambda}_{\beta_0}(\cdot)$ converges to the solution of the population version of the nonlinear integral equation (43). First, as $\hat{\Lambda}_{\beta_0}(\cdot)$ is the derivative of $\hat{\Lambda}_n(\cdot, \beta)$ with respect to β and evaluated at β_0 , the family of the functions $\{\hat{\Lambda}_{\beta_0}(t) : t \in [0, \zeta]\}$ for all $\zeta < \tau$ is again uniformly bounded and equicontinuous at least for $n \rightarrow \infty$ for some large enough n . By the Arzelà-Ascoli theorem, we can find a subsequence of $\{\hat{\Lambda}_{\beta_0}(t)\}$ that converges to some function $\eta(t)$ uniformly on $t \in [0, \zeta]$ for all $\zeta < \tau$. As both $\Lambda_{\beta_0}(\cdot)$ and $\eta(\cdot)$ are nondecreasing functions, the integral equation (43) now converges to its population version

$$-\int_0^t \left\{ \int_0^\tau J_3^{(1)}(u, v) \Lambda_0(v) dv + J_2^{(1)}(u) \Lambda_0(u) + J_1^{(1)}(u) \right\} d\Lambda_0(u) = \int_0^t J_1^{(0)}(u) d\eta(u) + \int_0^t \left\{ \int_0^\tau J_3^{(0)}(u, v) \eta(v) dv + J_2^{(0)}(u) \eta(u) \right\} d\Lambda_0(u).$$

Using integration by parts, the last line can be expressed

$$J_1^{(0)}(t) \eta(t) - \int_0^t J_1^{(0)}(u) \eta(u) du + \int_0^t \left\{ \int_0^\tau J_3^{(0)}(v, u) d\Lambda_0(v) + J_2^{(0)}(u) \lambda_0(u) 1(t > u) \right\} \eta(u) du.$$

Hence, $\eta(\cdot)$ is the solution of a Fredholm integral equation of the second kind:

$$\frac{R_1(t)}{J_1^{(0)}(t)} = \eta(t) - \int_0^\tau \frac{R_2(t, u)}{J_1^{(0)}(t)} \eta(u) du, \quad (44)$$

where

$$\begin{aligned}
R_1(t) &= -\int_0^t \left\{ \int_0^\tau J_3^{(1)}(u, v) \Lambda_0(v) dv + J_2^{(1)}(u) \Lambda_0(u) + J_1^{(1)}(u) \right\} d\Lambda_0(u) \\
R_2(t, u) &= K_1^{(0)}(t, u) - J_2^{(0)}(u) \lambda_0(u) 1(t > u).
\end{aligned}$$

By the classical theorem of Integral equations (Tricomi 1985, Chapter II), there exists a unique solution to the integral equation (44) by the assumption that the kernel function $R_2(t, u)/J_1^{(0)}(t)$ is square integrable. provided that the kernel function $R_2(t, u)/J_1^{(0)}(t)$ in (44) is square integrable. This condition is satisfied by the assumption (7) that the functions $K_1^{(0)}(t, u)/J_1^{(0)}(t)$ and $J_2^{(0)}(u) \lambda_0(u)/J_1^{(0)}(t) 1(t > u)$ are square integrable. It follows that every convergent subsequence of $\hat{\Lambda}_{\beta_0}(\cdot)$ must converge to the unique solution of (44). By Helly's selection theorem, the whole sequence $\hat{\Lambda}_{\beta_0}(\cdot)$ must also converge to the unique solution.

We now show that the unique solution is $\Lambda_{\beta_0}^o(\cdot) \equiv \partial/\partial\beta\Lambda_0(\cdot, \beta)|_{\beta=\beta_0}$. By the strong consistency of $\hat{\Lambda}(\cdot, \beta)$ to $\Lambda_0(\cdot, \beta)$, we have $U_{2n}(t, \beta, \Lambda_0(\cdot, \beta)) = 0$. Differentiating both sides of the equation $U_{2n}(t, \beta, \Lambda_0(\cdot, \beta)) = 0$ with respect to β , we find that $\Lambda_{\beta_0}^o(\cdot)$ satisfies the integral equation (44). It follows that by the unique solution of the integral equation,

$$\widehat{\Lambda}_{\beta_0}(t) \equiv \frac{\partial}{\partial\beta}\widehat{\Lambda}(t, \beta)|_{\beta=\beta_0} \rightarrow \frac{\partial}{\partial\beta}\Lambda_0(t, \beta)|_{\beta=\beta_0} \equiv \Lambda_{\beta_0}^o(t)$$

almost surely and uniformly in $t \in [0, \zeta]$ for all $\zeta < \tau$. This solution can be characterized by the form

$$\Lambda_{\beta_0}^o(t) = \frac{R_1(t)}{J_1^{(0)}(t)} + \int_0^\tau H_1(t, u) \frac{R_1(u)}{J_1^{(0)}(u)} du, \quad (45)$$

where the Fredholm resolvent function $H_1(t, v)$ depends only on the true baseline function $\Lambda_0(\cdot)$ and its derivative $\lambda_0(\cdot)$ through the following equation

$$H_1(t, u) = \left\{ R_2(t, u) + \int_0^\tau R_2(t, v) H_1(v, u) dv \right\} \{J_1^{(0)}(t)\}^{-1}.$$

Step 4

Strong consistency of $\beta U_{1n}(\beta, \hat{\Lambda}(\cdot, \beta))$

Differentiating $U_{1n}(\beta, \hat{\Lambda}(\cdot, \beta))$ with respect to β , and by a straightforward calculation, we have

$$\begin{aligned} & -\frac{\partial}{\partial\beta} U_{1n}(\beta, \widehat{\Lambda}(\cdot)) \Big|_{\beta=\beta_0} \\ &= \int_0^\tau u \left\{ \frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \widehat{B}_i^o(u) Z_i' \right\} d\widehat{\Lambda}_{\beta_0}(u) + \int_0^\tau u \left\{ \frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \widehat{B}_i^o(u) Z_i Z_i' \right\} d\widehat{\Lambda}^o(u) \\ & \quad + \int_0^\tau u \left\{ \frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \dot{q} \left(e^{\beta_0' Z_i} \widehat{\Lambda}^o(u) \right) e^{2\beta_0' Z_i} Z_i \right\} \{ \widehat{\Lambda}_{\beta_0}(u) \}' d\widehat{\Lambda}^o(u) \\ & \quad + \int_0^\tau u \left\{ \frac{1}{n} \sum_{i=1}^n \widehat{W}_i^o(u) \dot{q} \left(e^{\beta_0' Z_i} \widehat{\Lambda}^o(u) \right) e^{2\beta_0' Z_i} Z_i Z_i' \right\} \widehat{\Lambda}^o(u) d\widehat{\Lambda}^o(u) \\ & \quad + \int_0^\tau u \left[\int_u^\tau \left(\frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u)}{\int_{C_i}^{\tau} \widehat{S}_i^o(s) ds} \right) \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i \right] \{ \widehat{\Lambda}_{\beta_0}(v) \}' v^{-2} dv \Big] d\widehat{\Lambda}^o(u) \\ & \quad - \int_0^\tau u \left[\int_0^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u) \int_{u \wedge C_i}^{\tau} \widehat{S}_i^o(s) s^{-2} ds}{\left(\int_{C_i}^{\tau} \widehat{S}_i^o(s) ds \right)^2} \right\} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i \right] \{ \widehat{\Lambda}_{\beta_0}(v) \}' dv \Big] d\widehat{\Lambda}^o(u) \\ & \quad + \int_0^\tau u \left[\int_u^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u)}{\int_{C_i}^{\tau} \widehat{S}_i^o(s) ds} \right\} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i Z_i' \right] v^{-2} \Lambda^o(v) dv \Big] d\widehat{\Lambda}^o(u) \\ & \quad - \int_0^\tau u \left[\int_0^\tau \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\widehat{B}_i^o(u) \int_{u \wedge C_i}^{\tau} \widehat{S}_i^o(s) s^{-2} ds}{\left(\int_{C_i}^{\tau} \widehat{S}_i^o(s) ds \right)^2} \right\} \tilde{\delta}_i \tilde{C}_i(v) \widehat{S}_i^o(v) \widehat{B}_i^o(v) Z_i Z_i' \right] \widehat{\Lambda}^o(v) dv \Big] d\widehat{\Lambda}^o(u) \end{aligned}$$

By the strong consistency of $\hat{\Lambda}_{\beta_0}(\cdot)$ to $\Lambda_{\beta_0}^o(\cdot)$ and $\hat{\Lambda}^o(\cdot)$ to $\Lambda_0(\cdot)$, we have that by the functional strong law of large numbers, $\int \beta U_{1n}(\beta, \hat{\Lambda}(\beta))|_{\beta=\beta_0} \rightarrow \Gamma_0$ almost surely, where as defined in (22),

$$\Gamma_0 = \int_0^\tau J_1^{(1)}(u) d\Lambda_{\beta_0}^o(u) + \int_0^\tau \left[\sum_{k=2}^3 J_k^{(1)}(u) \{\Lambda_{\beta_0}^o(u)\}' + \sum_{k=1}^3 J_k^{(2)}(u) \right] d\Lambda_0(u).$$

Furthermore, we can show that $\int \beta \hat{\Lambda}_n(u, \beta)|_{\beta=\beta_0}$ and $\int \beta \Lambda_0(u, \beta)|_{\beta=\beta_0}$ are asymptotically equivalent to each other. By the strong consistency of $\hat{\Lambda}(\cdot, \beta)$ to $\Lambda_0(\cdot, \beta)$, $U_{1n}\{\beta, \hat{\Lambda}_n(\cdot, \beta)\}$ is asymptotically equivalent to $U_{1n}\{\beta, \Lambda_0(\cdot, \beta)\}$, where

$$U_{1n}\{\beta, \Lambda_0(\cdot, \beta)\} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau Z_i \left[\delta_i dN_i(s) - u W_i\{u; \beta, \Lambda_0(\cdot, \beta)\} e^{\beta' Z_i} q\{e^{\beta' Z_i} \Lambda_0(u, \beta)\} d\Lambda_0(u, \beta) \right].$$

By the strong law of large numbers, $U_{1n}\{\beta, \Lambda_0(\cdot, \beta)\} \rightarrow U_{10}\{\beta, \Lambda_0(\cdot, \beta)\}$ almost surely, where

$$U_{10}\{\beta, \Lambda_0(\cdot, \beta)\} = \int_0^\tau u E_0 \left\{ Z_1 \left[e^{\beta_0' Z_1} W_1(u, \beta_0, \Lambda_0) q\{e^{\beta' Z_1} \Lambda_0(u)\} - e^{\beta' Z_1} W_1\{u, \beta, \Lambda_0(\cdot, \beta)\} q\{e^{\beta' Z_1} \Lambda_0(\cdot, \beta)\} \right] \right\} d\Lambda_0(u, \beta).$$

Taking derivative with respect to β ,

$$\frac{\partial}{\partial \beta} U_{10}\{\beta, \Lambda_0(\cdot, \beta)\} = -\frac{\partial}{\partial \beta} E_0 \left(\int_0^\tau u Z_1 e^{\beta' Z_1} W_1\{u, \beta, \Lambda_0(\cdot, \beta)\} q\{e^{\beta' Z_1} \Lambda_0(\cdot, \beta)\} d\Lambda_0(u, \beta) \right).$$

It follows that almost surely, $\int \beta U_{1n}\{\beta, \hat{\Lambda}_n(\cdot, \beta)\} \rightarrow \int \beta U_{10}\{\beta, \Lambda_0(\cdot, \beta)\}$. By a straightforward calculation, we can confirm that $\int \beta U_{10}\{\beta, \Lambda_0(\cdot, \beta)\}|_{\beta=\beta_0}$ equals to Γ_0 .

Fig. 1. Simulation results when the covariates are independent of the censoring time. Only the results for $\beta_{10} = 1$ are shown. IM is the imputation estimating equation; IW is the estimator from (24); WU is the estimator from (25); ESD is the empirical standard deviation; relative efficiency is the mean squared errors of IMEE over IWEE, and IMEE over WUEE, respectively.

Fig. 2. Simulation results for $\Lambda(0.9) \approx 0.656$ using the proposed IMEE procedure. ESD is the empirical standard deviation.

Fig. 3.

Fitted survival curves plotted separately for the category of dementia. The gray area is the 95% bootstrap pointwise confidence intervals. The survival within 10 years following disease onset is shown.

Table 1

Simulation results when the censoring time is dependent on the covariate.

<i>n</i>	$\gamma_c(\%)$	(a) Proportional hazards model						(b) Proportional odds model					
		Bias			ESD			Bias			ESD		
		IM	IW	WU	IM	IW	WU	IM	IW	WU	IM	IW	WU
							RE						
							$\gamma_c(\%)$						
50	13.8	0.050	0.138	0.197	0.37	0.38	0.50	87.1	50.0	13.5	0.056	0.318	0.66
													74.8
	20.4	0.049	0.144	0.205	0.39	0.40	0.51	83.1	50.9	19.5	0.048	0.355	0.67
													72
	33.0	0.043	0.156	0.218	0.43	0.46	0.56	81.1	52.8	32.9	0.031	0.355	0.74
100	51.6	0.041	0.165	0.227	0.54	0.61	0.71	75.1	53.4	53.8	0.026	0.261	0.92
													85.0
	69.8	0.108	0.221	0.259	0.85	0.98	1.08	72.4	58.7	70.0	0.072	0.237	1.16
													68.9
	13.8	0.020	0.102	0.143	0.25	0.26	0.34	82.3	47.0	13.7	0.020	0.281	0.43
200													65.2
	20.5	0.017	0.111	0.157	0.26	0.27	0.34	79.4	47.1	19.7	0.008	0.322	0.44
													60.3
	33.0	0.006	0.110	0.164	0.28	0.30	0.37	78.1	48.0	33.1	-0.012	0.317	0.47
													62.8
	51.2	0.011	0.101	0.143	0.33	0.37	0.43	77.9	53.4	54.2	-0.002	0.235	0.57
													78.0
	69.9	0.010	0.097	0.119	0.46	0.51	0.56	78.3	64.8	70.2	0.031	0.174	0.72
													81
	13.8	0.002	0.079	0.121	0.17	0.18	0.23	77.8	42.3	13.6	0.004	0.269	0.29
													32
	20.5	-0.001	0.086	0.136	0.18	0.19	0.24	74.9	40.8	19.6	-0.006	0.304	0.30
													33
	32.9	-0.007	0.087	0.142	0.20	0.20	0.26	77.8	43.2	33.0	-0.030	0.294	0.33
													54.4
	51.3	-0.016	0.064	0.107	0.23	0.24	0.30	83.1	51.8	54.4	-0.048	0.185	0.40
													41
	70.1	-0.003	0.065	0.090	0.30	0.32	0.36	84.5	66.1	70.2	-0.040	0.124	0.49
													54
													78.0

There are two covariates with $\beta_0 = (1, 0)$. Only the results for $\beta_0 = 1$ are shown. IM is the imputation estimating equation; IW is the estimator from (24); WU is the estimator from (25); ESD is the empirical standard deviation; RE is the relative efficiency, defined as the mean squared errors of IMEE over IWEE, and IMEE over WUEE, respectively; γ_c is the censoring percentage (%).

Estimates of regression coefficients for the Canadian dementia data. The baseline category is vascular dementia. BSD is the bootstrap standard deviation.

Table 2

	$\eta = 0$		$\eta = 0.5$		$\eta = 1$		$\eta = 1.5$	
	β	BSD	β	BSD	β	BSD	β	BSD
Probable Alzheimer	-0.111	0.078	-0.208	0.125	-0.327	0.212	-0.464	0.297
Possible Alzheimer	-0.187	0.081	-0.353	0.146	-0.558	0.225	-0.794	0.315