Published in final edited form as:

Biometrics. 2019 March ; 75(1): 121–132. doi:10.1111/biom.12970.

Semiparametric regression analysis of length-biased interval-censored data

Fei Gao, **Kwun Chuen Gary Chan**

Department of Biostatistics, University of Washington, Seattle, Washington

Abstract

In prevalent cohort design, subjects who have experienced an initial event but not the failure event are preferentially enrolled and the observed failure times are often length-biased. Moreover, the prospective follow-up may not be continuously monitored and failure times are subject to interval censoring. We study the nonparametric maximum likelihood estimation for the proportional hazards model with length-biased interval-censored data. Direct maximization of likelihood function is intractable, thus we develop a computationally simple and stable expectation-maximization algorithm through introducing two layers of data augmentation. We establish the strong consistency, asymptotic normality and efficiency of the proposed estimator and provide an inferential procedure through profile likelihood. We assess the performance of the proposed methods through extensive simulations and apply the proposed methods to the Massachusetts Health Care Panel Study.

Keywords

left truncation; nonparametric maximum likelihood estimation; proportional hazards model; semiparametric efficiency

1 | INTRODUCTION

Interval-censored data arise when a failure time is not recorded precisely but is rather known to lie within a time interval. Such data are encountered in prospective follow-up studies, where the ascertainment of the event of interest is made over a series of examination times. Regression analysis of unbiased interval-censored survival data has been extensively studied. In particular, nonparametric maximum likelihood estimation for the proportional hazards and transformation models have been studied by Huang (1996) and Zeng et al. (2016), respectively. Due to intractable likelihood, sieve estimation is also proposed for the proportional hazards model by Huang and Rossini (1997) and Cai and Betensky (2003), among others. A comprehensive review is given in Sun (2007).

Correspondence Fei Gao, Kwun Chuen Gary Chan, Department of Biostatistics, University of Washington, Seattle, WA, feigao@uw.edu (FG); kcgchan@u.washington.edu (KCGC).

SUPPORTING INFORMATION
The Web Appendices and Figure, referenced in Sections 3 and 4 and Appendix A, and the computation codes for the simulation studies are available with this paper at the Biometrics website on Wiley Online Library.

Although sampling incident cases in a follow-up study is common, it may require a long follow-up period to observe enough failure events for meaningful analysis. Alternatively, a prevalent cohort design samples individuals who have experienced an initial event but not the failure event at enrollment, and is often considered as a more focused and economical design (Brookmeyer and Gail, 1987). However, subjects with a longer survival time are preferentially sampled in a prevalent cohort. When the incidence of the initial event is stationary over time, a prevalent cohort collects length-biased data (Wang, 1991; Shen et al., 2009), where the probability of observing the failure time is proportional to its value.

Prospective follow-up of a prevalent cohort can be subject to interval censoring. An example is the Massachusetts Health Care Panel Study (Chappell, 1991), where the time to loss of active life for elderly individuals were assessed approximately 1.25, 6, and 10 years after study recruitment. Since only functionally independent individuals were enrolled, subjects with a longer time to loss of active life were more likely to be sampled. Although a prevalent cohort is a biased sample that requires special methods for analysis, it may provide information that is otherwise unavailable in an incident cohort. For example, in an incident sampling design, the right tail of survival distribution may not be identified because of limited study duration. Using prevalent sampling design, the identifiable region for the survival distribution and marked variables that are observed only at the event occurrence could be enlarged (Chan and Wang, 2010). In the Massachusetts Health Care Panel Study data, even though the last monitoring time is 10 years after study recruitment, we can identify a survival distribution that ranges over 30 years (see Section 3.2), because individuals are event-free for a period before enrollment. An added advantage for interval-censored data is that, even when the monitoring time has a discrete distribution, we can identify the continuous survival distribution of the failure event because the event-free period before enrollment is typically continuous. For incident sampling with discrete monitoring time, in contrast, we can only identify a discrete survival distribution.

Statistical methodology for regression modeling of length-biased data are mostly proposed for uncensored and right-censored data. Wang (1996) and Chen (2010) considered uncensored data. For right-censored data, Qin and Shen (2010) proposed inverse weighted estimating equation and Huang and Qin (2012) proposed a composite likelihood approach for the proportional hazards model. Qin et al. (2011) considered the nonparametric maximum likelihood estimator and derived an expectation-maximization (EM) algorithm for computation, and showed that the estimator is efficient.

Even though there has been limited literature on length-biased interval-censored data, several methods were proposed for left-truncated interval-censored data without the length-biased assumption for the truncation time. In particular, Pan and Chappell (1998) considered the proportional hazards model and applied a gradient projection-based method for non-parametric maximum likelihood estimation, where the baseline survival function may be underestimated. Pan and Chappell (2002) considered the same model and suggested a marginal likelihood approach that avoids estimating the baseline hazards function and a monotone maximum likelihood approach assuming that the baseline distribution has a nondecreasing hazard function. Kim (2003) studied the special case of left-truncated current status data, where there is only one examination time, and established the asymptotic properties of the nonparametric maximum likelihood estimators. Recently, Wang et al. (2015) studied the additive hazard model with left-truncated interval-censored data and proposed a sieve estimation method.

In this paper, we study the nonparametric maximum likelihood estimation for the proportional hazards model with length-biased interval-censored data. Through introducing pseudo-truncated data and latent Poisson random variables, we develop a simple and computational stable EM algorithm. We establish the strong consistency and asymptotic normality of the proposed estimators and provide inference through a profile likelihood approach. We assess the performance of the proposed estimator and inferential procedures through extensive simulations and apply the proposed methods to the Massachusetts Health Care Panel Study data.

2 | THE PROPOSED METHODOLOGY

2.1 | Model and data

For individuals in the target population, let $\widetilde{T}$ be the time to a failure event and $\widetilde{\boldsymbol{Z}}$ be a p-vector of covariates. We assume that $\widetilde{T}$ follows a proportional hazards model with a cumulative hazard function

$$\Lambda(t \mid \widetilde{\boldsymbol{Z}}) = \Lambda(t)\exp\left(\beta^{\mathrm{T}}\widetilde{\boldsymbol{Z}}\right),$$

where $\boldsymbol{\beta}$ is a p-vector of unknown regression parameters, and $\Lambda(\cdot)$ is an arbitrary increasing function with $\Lambda(0) = 0$.

For length-biased sampling, it is common to assume that the incidence rate of the initial event is constant over calendar time and $\widetilde{A}$, the truncation time, is uniformly distributed in [0, $\boldsymbol{\tau}$], where $\boldsymbol{\tau}$ is the maximum support of $\widetilde{T}$ (Wang, 1991; Qin et al., 2011). In a prevalent cohort study, a subject is included only if the failure time does not occur before the truncation time, that is, $\widetilde{T} \ge \widetilde{A}$. We let T, A, and $\boldsymbol{Z}$ be the failure time, truncation time and covariates, respectively, in the prevalent cohort. Then, $(T, A, \boldsymbol{Z})$ has the same joint distribution as $(\widetilde{T}, \widetilde{A}, \widetilde{\boldsymbol{Z}})$ conditional on $\widetilde{T} \ge \widetilde{A}$. Suppose that the occurrence of the failure is not exactly observed but only determined at a sequence of examination times, denoted as $A < U_1 \cdots < U_M \le \tau$. The failure time is then known to lie in the interval (L, R), where $L = \max\left\{U_m : U_m < T, m = 0, \ldots, M\right\}$, $R = \min\left\{U_m : U_m \ge T, m = 1, \ldots, M + 1\right\}$, $U_0 = A$, and $U_{M+1} = \infty$. In particular, if the failure occurs before the first examination time, then $(L, R) = (A, U_1)$; if the failure has not occurred at the last examination time, then $(L, R) = \left(U_M, \infty\right)$. Let $V_m = U_m - A$ for $m = 0, \ldots, M$, so that $V_0 = 0$.

We assume the following non-informative sampling time condition, that M and $\left\{V_m : m = 1, \ldots, M\right\}$ are independent of (T, A) conditional on $\boldsymbol{Z}$. For a length-biased sample of n subjects, the observed data are $\left\{\mathcal{O}_i : i = 1, \ldots, n\right\}$, where $\mathcal{O}_i = \left\{L_i, R_i, A_i, \boldsymbol{Z}_i\right\}$. The observed-data likelihood is then given by

$$L_n(\boldsymbol{\beta}, \Lambda) = \prod_{i=1}^{n} \frac{\left[\exp\left\{-\Lambda\left(L_i\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\} - I\left(R_i < \infty\right)\exp\left\{-\Lambda\left(R_i\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\}\right]}{\int_0^{\tau}\exp\left\{-\Lambda(a)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\}da/\tau}. \tag{1}$$

The likelihood function $L_n(\boldsymbol{\beta}, \Lambda)$ involves $\boldsymbol{\tau}$, which is a constant related to study design but not a parameter of interest. In the E-step of the proposed EM algorithm, however, it is required to redistribute mass on [0, $\boldsymbol{\tau}$], and an approximation of $\boldsymbol{\tau}$ is required. Let $0 = t_0 < t_1 < \cdots < t_k < \infty$ be the ordered sequence of all L_i and $R_i I\left(R_i < \infty\right)$. Following Qin et al. (2011), we approximate $\boldsymbol{\tau}$ by t_k, which converges to $\boldsymbol{\tau}$ at a rate faster than $n^{1/2}$, and therefore does not alter subsequent results.

2.2 | Nonparametric maximum likelihood estimation

We adopt the nonparametric maximum likelihood estimation approach, where the estimator for Λ is a step function with nonnegative finite jumps at the ends of the intervals that bracket the failure times. Specifically, we let λ_0, $\lambda 1$,…, λ_k be the respective jump sizes at t_0, t_1, …,t_k, where $\lambda_0 = 0$. Write $\lambda = \left(\lambda_1, \ldots, \lambda_k\right)$. We maximize the objective function

$$\begin{aligned} l_n(\boldsymbol{\beta}, \lambda) \equiv \sum_{i=1}^{n} \Bigg(& \log\Bigg[\exp\Bigg\{-\sum_{t_j \le L_i} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\Bigg\} \\ & -I\left(R_i < \infty\right)\exp\Bigg\{-\sum_{t_j \le R_i} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\Bigg\}\Bigg] \\ & -\log\int_0^{\tau}\frac{1}{\tau}\exp\Bigg\{-\sum_{t_j \le a} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\Bigg\}da\Bigg). \end{aligned}$$

Direct maximization of $l_n(\boldsymbol{\beta}, \lambda)$ is difficult due to a lack of analytical expressions. We introduce two layers of data augmentation and propose an EM algorithm to facilitate computation. First, to handle left truncation, we introduce pseudo-truncated data, which is also referred to as "ghost data" (Turnbull, 1976). In particular, let $\mathcal{O}_i^* \equiv \left\{\left(T_{im}^*, A_{im}^*, \boldsymbol{Z}_i\right): T_{im}^* < A_{im}^*, m = 1, \ldots, n_i\right\}$ denote the pseudo-truncated samples corresponding to subject i, where $\left(T_{i1}^*, A_{i1}^*\right), \ldots, \left(T_{i,n_i}^*, A_{i,n_i}^*\right)$ are independent and identically distributed given $\boldsymbol{Z}_i$. Since the estimator for Λ only takes jump at $t_j(j = 1, \ldots, n)$, the failure time T_{im}^* can only take values from $\left\{t_1, \ldots, t_k\right\}$. The number of truncated samples n_i follows a negative binomial distribution with parameter

$$\begin{aligned} \pi_i = P\left(T_{im}^* < A_{im}^* \mid \boldsymbol{Z}_i\right) &= \sum_{j=0}^{k} P\left(T_{im}^* = t_j, A_{im}^* > t_j \mid \boldsymbol{Z}_i\right) \\ &= \sum_{j=1}^{k}\left(1 - t_j/\tau\right)\lambda_j \exp\left(\beta^{\mathrm{T}}\boldsymbol{Z}_i\right)\exp\left\{-\sum_{l=1}^{j}\lambda_l \exp\left(\beta^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\}, \end{aligned}$$

such that $E\left(n_i \mid \mathcal{O}_i\right) = \pi_i / \left(1 - \pi_i\right)$. Let $n_{ij} = \sum_{m=1}^{n_i} I\left(T_{im}^* = t_j\right)$. Given the total number n_i, $\left(n_{i1}, \ldots, n_{ik}\right)$ follows a multinomial distribution with probabilities $\left(p_{i1}, \ldots, p_{ik}\right)$, where

$$\begin{aligned} p_{ij} &= P\left(T_{im}^* = t_j \mid T_{im}^* < A_{im}^*, \mathbf{Z}_i\right) \\ &= \frac{\left(1 - t_j/\tau\right)\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\exp\left\{-\sum_{l=1}^{j}\lambda_l \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\right\}}{\pi_i}. \end{aligned} \tag{2}$$

By the missing information principle (Lai and Ying, 1994), the maximization of $l_n(\boldsymbol{\beta}, \lambda)$ is equivalent to maximizing the conditional expectation of the log-likelihood function of the “complete-data” $\left\{\left(\mathcal{O}_i, \mathcal{O}_i^*\right) : i = 1, \ldots, n\right\}$ given the observed data. The “complete-data” log-likelihood function is given by

$$\begin{aligned} \tilde{l}_n^C(\boldsymbol{\beta}, \lambda) \equiv \sum_{i=1}^{n} \Bigg(& \log\Bigg[\exp\Bigg\{-\sum_{t_j \le L_i} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\Bigg\} \\ & - I\left(R_i < \infty\right)\exp\Bigg\{-\sum_{t_j \le R_i} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\Bigg\}\Bigg]\Bigg] \\ & + \sum_{j=1}^{k} n_{ij} \log\Big[\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right) \\ & \times \exp\Bigg\{-\sum_{l=1}^{j} \lambda_l \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\Bigg\}\Bigg]\Bigg]\Bigg). \end{aligned}$$

While we may propose an EM algorithm based on $\tilde{l}_n^C(\boldsymbol{\beta}, \lambda)$, its maximization step is still difficult to obtain, since $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ cannot be separated in the complete-data log-likelihood function due to the interval censoring structure.

Therefore, we further introduce data augmentation based on independent Poisson random variables $W_{ij}\left(i = 1, \ldots, n; j = 1, \ldots, k, t_j \le R_i^*\right)$ with means $\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)$, where $R_i^* = L_i I\left(R_i = \infty\right) + R_i I\left(R_i < \infty\right)$. The joint density function for $W_{ij}\left(j = 1, \ldots, k, t_j \le R_i^*\right)$ is given by

$$\prod_{j=1, t_j \le R_i^*}^{k} \frac{\left\{\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\right\}^{W_{ij}}}{W_{ij}!} \exp\left\{-\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\right\}.$$

Let $N_{i1} = \sum_{t_j \le L_i} W_{ij}$ and $N_{i2} = I\left(R_i < \infty\right)\sum_{L_i < t_j \le R_i} W_{ij}$. Suppose that we observe $N_{i1} = 0$ and $N_{i2} > 0$. The observed-data likelihood for $\widetilde{\mathcal{O}}_i \equiv \left\{N_{i1} = 0, N_{i2} > 0\right\}$ is equal to

$$
\begin{aligned}
&\Pr\left(N_{i1} = 0, N_{i2} > 0\right) \\
&= \Pr\left(\sum_{t_j \le L_i} W_{ij} = 0\right)\left\{1 - I\left(R_i < \infty\right)\Pr\left(\sum_{L_i < t_j \le R_i} W_{ij} = 0\right)\right\} \\
&= \exp\left\{-\sum_{t_j \le L_i} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\right\} \\
&\quad - I\left(R_i < \infty\right)\exp\left\{-\sum_{t_j \le R_i} \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\right\}.
\end{aligned}
$$

Therefore, $\tilde{l}_n^C(\boldsymbol{\beta}, \lambda)$ can be viewed as the observed log-likelihood function for $\widetilde{\mathcal{O}} \equiv \left\{\left(\widetilde{\mathcal{O}}_i, \mathcal{O}_i^*\right); i = 1, \ldots, n\right\}$ with $W_{ij}\left(i = 1, \ldots, n; j = 1, \ldots, k, j \le R_i^*\right)$ and $n_{ij}(i = 1, \ldots, n; j = 1, \ldots, k)$ as latent variables. In particular, the complete-data log-likelihood function based on $\left(W_{ij}, n_{ij}\right)$ is given by

$$
\begin{aligned}
&\sum_{i=1}^{n} \sum_{j=1}^{k} I\left(t_j \le R_i^*\right)\left\{-\log\left(W_{ij}!\right)\right. \\
&\quad \left. + W_{ij}\left(\log\lambda_j + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right) - \lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\right\} \\
&\quad + \sum_{i=1}^{n} \sum_{j=1}^{k} n_{ij}\left\{\log\left(1 - t_j/\tau\right)\right. \\
&\quad \left. + \log\lambda_j + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i - \sum_{l=1}^{j} \lambda_l \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\right\}.
\end{aligned}
$$

Based on this formulation, we propose the following EM algorithm. In the E-step, we evaluate the conditional expectations of W_{ij} and n_{ij} given the observed data. In particular, we have

$$
\begin{aligned}
\widehat{E}\left(W_{ij}\right) &= I\left(L_i < t_j \le R_i, R_i < \infty\right) \\
&\quad \times \frac{\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)}{1 - \exp\left\{-\sum_{L_i < t_l \le R_i} \lambda_l \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\right\}},
\end{aligned}
$$

and

$$
\widehat{E}\left(n_{ij}\right) = \frac{\left(1 - t_j/\tau\right)\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\exp\left\{-\sum_{l=1}^{j} \lambda_l \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i\right)\right\}}{1 - \pi_i},
$$

where $\widehat{E}(\cdot)$ denote the conditional expectation with respect to the observed data $\widetilde{\mathcal{O}}$. In the M-step, we maximize the expected complete-data log-likelihood function. We update λ_j by

$$\lambda_j = \frac{\sum_{i=1}^{n}\left\{I\left(t_j \le R_i^*\right)\widehat{E}\left(W_{ij}\right) + \widehat{E}\left(n_{ij}\right)\right\}}{\sum_{i=1}^{n}\left\{I\left(t_j \le R_i^*\right) + \sum_{l=j}^{k}\widehat{E}\left(n_{il}\right)\right\}\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)}$$

and update $\boldsymbol{\beta}$ by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\left\{I\left(t_j \le R_i^*\right)\widehat{E}\left(W_{ij}\right) + \widehat{E}\left(n_{ij}\right)\right\} \times \left[\mathbf{Z}_i - \frac{\sum_{i'=1}^{n}\left\{I\left(t_j \le R_{i'}^*\right) + \sum_{l=j}^{k}\widehat{E}\left(n_{i',l}\right)\right\}\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{i'}\right)\mathbf{Z}_{i'}}{\sum_{i'=1}^{n}\left\{I\left(t_j \le R_{i'}^*\right) + \sum_{l=j}^{k}\widehat{E}\left(n_{i',l}\right)\right\}\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{i'}\right)}\right] = \mathbf{0}.$$

We iterate between the E-step and M-step until convergence. We denote the final estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ as $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\lambda}}$.

In summary, through introducing two layers of latent random variables, we proposed a stable computing algorithm to obtain the estimators that maximize the nonparametric likelihood function. The latent truncated "ghost data" were introduced to deal with the complications that arise from left truncation and the latent Poisson random variables were introduced to deal with the incomplete data caused by interval-censoring.

2.3 | Asymptotic properties

In this section, we establish the strong consistency and asymptotic normality of the proposed estimators. We assume the following regularity conditions.

Condition 1. The true value of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}_0$, belongs to the interior of a known compact set $\mathscr{B} \subset \mathbb{R}^p$.

Condition 2. The true value $\Lambda_0(\cdot)$ of $\Lambda(\cdot)$ is strictly increasing and continuously differentiable in $[0, \boldsymbol{\tau}]$ with $\Lambda_0(0) = 0$.

Condition 3. The covariate $\boldsymbol{Z}$ has bounded support and is not concentrated on any proper subspace of $\mathbb{R}^p$.

Condition 4. The examination times have finite support $\mathscr{V}$ with the least upper bound $\boldsymbol{\tau}$. The number of potential examination times M is positive with $E(M) < \infty$. There exists a positive constant $\boldsymbol{\eta}$ such that $\Pr\left(U_{m+1} - U_m \ge \eta \mid M, Z\right) = 1$. In addition, there exists a probability measure μ in $\mathscr{V}$ such that the bivariate distribution function of (U_m, U_{m+1}) conditional on $(\boldsymbol{M}, \boldsymbol{Z})$ is dominated by $\mu \times \mu$ and its Radon-Nikodym derivative, denoted by $\tilde{f}_m(u, v; M, \boldsymbol{Z})$, can be expanded to a positive and twice-continuously differentiable function in the set $\{(u, v): 0 \le u \le \tau, 0 \le v \le \tau, v - u \ge \eta\}$.

Conditions 1, 2, and 3 are standard conditions for failure time regression. Condition 4 pertains to the joint distribution of examination times. It requires that two adjacent examination times are separated by at least η; otherwise, the data may contain exact observations, which require a different theoretical treatment. The dominating measure μ is chosen as the Lebesgue measure if the examination times are continuous random variables and as the counting measure if the examinations occur only at a finite number of time points. The number of potential examination times $\boldsymbol{M}$ can be fixed or random, is possibly different among study subjects, and is allowed to depend on covariates.

We state the strong consistency of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ and the weak convergence of $\hat{\boldsymbol{\beta}}$ in two theorems.

Theorem 1.—*Under Conditions 1–4,* $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \to_{a.s.} 0$, *and* $\|\widehat{\Lambda} - \Lambda_0\|_{l^\infty(\mathscr{V})} \to_{a.s.} 0$, *where* $\| \cdot \|_{l\infty(\mathscr{V})}$ *denotes the supremum norm on* $\mathscr{V}$, *and* $\widehat{\Lambda}(t) = \sum_{t_j \le t} \hat{\lambda}_j$.

Theorem 2.—*Under Conditions 1–4,* $n^{1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ *converges weakly to a p-dimensional zero-mean normal random vector with a covariance matrix that attains the semiparametric efficiency bound.*

The proofs of both theorems are provided in Appendix A.

2.4 | Variance estimation

We estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ by a profile likelihood approach. Let $\widehat{\Lambda}_{\boldsymbol{\beta}} = \operatorname{argmax}\log_{\Lambda \in \mathscr{C}} L_n(\boldsymbol{\beta}, \Lambda)$, where $\mathscr{C}$ is the set of bounded step functions with non-negative jumps at t_l ($l = 1,\ldots, m$). The maximizer $\widehat{\Lambda}_{\boldsymbol{\beta}}$ can be obtained using the EM algorithm of Section 2.2 if we fix $\boldsymbol{\beta}$ and only update $\boldsymbol{\lambda}$ in the M-step. The profile log-likelihood function is defined as

$$pl_n(\boldsymbol{\beta}) = \max_{\Lambda \in \mathscr{C}} \log L_n(\boldsymbol{\beta}, \Lambda) = \log L_n\left(\boldsymbol{\beta}, \widehat{\Lambda}_{\boldsymbol{\beta}}\right).$$

Let $\widetilde{p}l_i(\boldsymbol{\beta})$ be the ith subject's contribution to $pl_n(\boldsymbol{\beta})$. We estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ by the inverse of

$$\sum_{i=1}^{n} \begin{pmatrix} \dfrac{\widetilde{p}l_i\left(\hat{\boldsymbol{\beta}} + h_n e_1\right) - \widetilde{p}l_i(\hat{\boldsymbol{\beta}})}{h_n} \\ \vdots \\ \dfrac{\widetilde{p}l_i\left(\hat{\boldsymbol{\beta}} + h_n e_p\right) - \widetilde{p}l_i(\hat{\boldsymbol{\beta}})}{h_n} \end{pmatrix}^{\otimes 2},$$

where e_j is the jth canonical vector in $\mathbb{R}^p$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$, and h_n is a constant of order $n^{-1/2}$. In the numerical studies, we used $h_n = 5n^{-1/2}$ as suggested by Zeng et al. (2016).

The above profile likelihood approach is different from that of Murphy and Vaart (2000). They estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ by the negative inverse of the Hessian matrix of $pl_n(\boldsymbol{\beta})$ at $\hat{\boldsymbol{\beta}}$, which is obtained by second order numerical differences. The estimated matrix may not be positive semidefinite, especially in small samples. Here, we estimate the covariance matrix by the inverse of the empirical covariance matrix of the gradient of $\widetilde{p}l_i(\boldsymbol{\beta})$ using the first-order numerical differences, similar to Zeng et al. (2017). The calculation is quicker than the approach requiring second-order numerical differences, and the estimated covariance matrix is guaranteed to be positive semidefinite.

3 | NUMERICAL STUDIES

3.1 | Simulation

We conducted simulation studies to evaluate the performance of the proposed methods. We considered two covariates Z_1 ~ Bernoulli(0.5) and z_2 ~ Uniform(−0.5, 0.5). We set $\boldsymbol{\beta} = (0.5, 1)^{\mathrm{T}}$, $\Lambda(t) = 0.3t$, and $\boldsymbol{\tau} = 15$. We generated the truncation time $\tilde{A}$ from Uniform(0, $\boldsymbol{\tau}$) and generated the sequence of potential examination times $U_m \sim U_{m-1} + 0.1 +$ Uniform(0, 2) with $U_0 = \tilde{A}$. We set n = 100, 200, or 400 and examined 10000 replicates for each sample size. We compared the proposed nonparametric maximum likelihood method with the maximum conditional likelihood method of Pan and Chappell (1998), which is applicable to left-truncated interval-censored data. For coherent comparisons, we compute both estimators by EM algorithms, where the algorithm for the conditional likelihood estimator is an adaption of the proposed algorithm and is given in Appendix B. We set the initial value of $\boldsymbol{\beta}$ to $\mathbf{0}$ and the initial value of λ_l to $1/k$.

Table 1 summarizes the simulation results on the estimation of $\boldsymbol{\beta}$ using the proposed and conditional likelihood approaches. The biases for the proposed estimators are small and decrease as sample size increases. The biases for the conditional likelihood estimators are larger than those for the proposed estimators for all sample sizes, but decrease as sample size increases. The variance estimators for $\hat{\boldsymbol{\beta}}$ using both approaches are accurate and the confidence intervals have proper coverage probabilities. As expected, the proposed estimator shows substantial efficiency gain compared to the conditional likelihood estimator. Web Figure S1 in the Supplementary Materials gives the estimated baseline survival functions. The nonparametric maximum likelihood estimation gives unbiased estimates, while the condition likelihood estimators tend to underestimate the true values, as indicated in Pan and Chappell (1998).

We further assess the robustness of the nonparametric maximum likelihood estimator when the uniform assumption for the truncation time does not hold. In particular, we considered the same simulation setting but generated the truncation time $\tilde{A}$ from $Exp(0.1)$ such that the stationary incidence assumption is violated. Table 2 shows the simulation results. The nonparametric maximum likelihood estimators are slightly biased, while the coverages of the 95% confidence intervals are acceptable. Even though the bias of the proposed estimator is larger than the conditional likelihood estimator when sample size is large (n = 400) and length-biased sampling is violated, the mean squared error of the proposed estimator is still smaller.

3.2 | Massachusetts health care panel study

We apply the proposed methods to the Massachusetts Health Care Panel Study, which has been described and analyzed previously (Chappell, 1991; Pan and Chappell, 1998; Hudgens, 2005). The study aimed at assessing the risk at which elderly individuals lose active life, which is defined as a continued ability to perform various activities of daily living such as dressing and bathing. The study was first conducted in 1975 taking a baseline survey of Massachusetts residents over the age of 65. Since only subjects who were active at baseline were included, the time to loss of active life was subject to left truncation. Three follow-up waves were then taken at 1.25, 6, and 10 years after baseline to determine if subjects are still living actively, so the time to loss of active life was also interval-censored.

The data set includes 1286 subjects with enrollment age ranges from 65 to 97.3. Since the study population were defined to be over age 65, we consider the failure time as age at loss of active life minus 65. Since the subjects are active at the enrollment, the truncation time is the age at enrollment minus 65. We applied the proposed methods to study the association between loss of active life and gender.

Table 3 shows the estimation results for the regression parameter in the Massachusetts Health Care Panel Study. The point estimates from the proposed approach and the conditional likelihood approach are both positive, indicating that male subjects are associated with a higher risk of losing active life than females. The standard error estimate of the proposed nonparametric maximum likelihood estimator is smaller than that of the conditional likelihood estimator, so that at a 5% significance level, the null hypothesis of no association between gender and loss of active life is only rejected by the nonparametric maximum likelihood approach.

Figure 1 shows the estimated survival probabilities for male and female subjects using the two approaches. Even though they give similar estimates for the survival probabilities, the nonparametric maximum likelihood approach gives an estimate with finer jumps, resulting from the additional assumption on the truncation time. Using both approaches, the female subjects have a higher survival probability than the male subjects.

If the length-biased sampling assumption does not hold, the regression coefficients of the two methods would converge to different values. Therefore, we estimated the difference of the estimators and construct a 95% confidence interval by bootstrapping with 1000 replications, to see if the stationary assumption for the truncation time holds. The differences of the two estimators was 0.011 with a 95% confidence interval (−0.092, 0.107), indicating that the length-biased assumption is possibly valid.

4 | DISCUSSION

In this paper, we adopt the nonparametric maximum likelihood estimation where the estimator for Λ is a step function that is right-continuous which is usually considered in the literature. As mentioned by a reviewer, if Λ is only restricted to be nondecreasing, then the true maximizer of the likelihood should involve a left-continuous Λ, that is, $\Lambda(t) = \Lambda(t_{j+1})$ on $(t_j, t_{j+1}]$ for $j = 0, \ldots, m - 1$. The two versions are asymptotically equivalent since any two adjacent step points get closer as sample size increases. In Web Appendix A of the Supplementary Materials, we implemented the version with left-continuous Λ and demonstrated that the numerical difference between the two versions is ignorable.

The iterative convex minorant (ICM) algorithm (Pan, 1999) is an alternative algorithm for the EM algorithm adopted in the paper to obtain the nonparametric maximum likelihood estimators for interval-censored data. Even though it is generally faster than the EM algorithm considered in the paper, it may become unstable for large datasets because it attempts to update a large number of parameters simultaneously using a quasi-Newton method (Zeng et al., 2016). Wang et al. (2016) also advocated the use of an EM algorithm by comparing it with the R package intcox (Henschel and Mansmann, 2013) that adopts the algorithm of Pan (1999). They found that ICM algorithm often exhibits larger biases, indicating that it may not converge to the true maximizer of the likelihood function.

In this paper, we studied the nonparametric maximum likelihood estimation of the proportional hazards model for length-biased interval-censored data. Although length-biasedness requires a stationary incidence distribution for the initial event, the proposed methods can be extended to situations when the incidence distribution follows a parametric model (Huang et al., 2015). In that case, the denominator of the individual components in $L_n(\boldsymbol{\beta}, \Lambda)$ need to be modified corresponding to the distribution of the truncation time, and the proposed EM algorithm can be adjusted accordingly.

The efficiency gain of the proposed nonparametric maximum likelihood estimators over the conditional likelihood estimators mainly comes from the information of the (uniform) distributional assumption on the truncation time. Relatively, the conditional likelihood estimators are more robust against the assumption. In practice, one need to carefully ascertain the assumption to apply the proposed approach. For the right-censored left-truncated data, graphical methods (Wang, 1991; Asgharian et al., 2006) and a goodness-of-fit test (Mandel and Betensky, 2007) have been proposed to test the length-biasedness assumption. In the numerical examples, we used a bootstrapped method to the difference of the nonparametric maximum likelihood and conditional likelihood approaches as an indirect test of length-biasedness. The diagnostic methods for right-censored data cannot be directly extended to the interval-censoring case, since the estimator for the survival function converges in a different, $n^{1/3}$, rate. Formal tests for the length-biased assumption with interval-censored data will be developed in the future.

The individuals in the MHCPS data may also be subject to the risk of a competing cause, for example, death, such that the subjects who died before loss of active life were right-censored. In addition, there may be selection effect such that only alive subjects were included in the study. Therefore, the assumptions of conditional independent censoring and truncation times may be questionable. However, the information on the cause of right censoring is not available in MHCPS data, so we are not able to access the validity of the assumptions. The regression analysis of competing risks interval-censored data has been studied (Mao et al., 2017), however, no existing methods considered the scenario when the competing risks interval-censored data are also subject to left truncation. Methods incorporating such complications are important future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Prof Rick Chappell for sharing the data from Massachusetts Health Care Panel Study. We also thank the associate editor and two referees for constructive feedbacks that improve the paper. The authors were partially supported by the National Heart, Lung, and Blood Institute of the U.S. National Institutes of Health.

Appendix A

Proof of asymptotic results

We use $\mathbb{P}_n$ to denote the empirical measure from n independent subjects and $\mathbb{P}$ to denote the true probability measure. Write $\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - \mathbb{P})$. Let $L(\boldsymbol{\beta}, \Lambda)$ be the observed-data likelihood for a single subject

$$L(\boldsymbol{\beta}, \Lambda) = \frac{\sum_{m=0}^{M} \Delta_m \left[\exp\left\{-\Lambda\left(U_m\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\} - \exp\left\{-\Lambda\left(U_{m+1}\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\}\right]}{\int_0^{\tau} \exp\left\{-\Lambda(a)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\} da/\tau},$$

where $\Delta_m = I\left(U_m < T \leq U_{m+1}\right)$.

Write $l(\boldsymbol{\beta}, \Lambda) = \log L(\boldsymbol{\beta}, \Lambda)$. Let $\widetilde{\Lambda}$ be a step function that takes jumps only at $t_1, \ldots, t_k$ with $\widetilde{\Lambda}\left(t_j\right) = \Lambda_0\left(t_j\right)$ for $j = 1, \ldots, k$. Let

$$m(\boldsymbol{\beta}, \Lambda) = \log\left\{\frac{L(\boldsymbol{\beta}, \Lambda) + L\left(\boldsymbol{\beta}_0, \widetilde{\Lambda}\right)}{2}\right\}$$

and

$$\mathscr{M} = \left\{m(\boldsymbol{\beta}, \Lambda) : \boldsymbol{\beta} \in \mathscr{B}, \Lambda \in D_M\right\},$$

where $D_M = \{\Lambda : \Lambda$ is increasing with $\Lambda(0) = 0$, $\Lambda(\boldsymbol{\tau}) \leq M\}$, and $M < \infty$. The proofs make use of two lemmas, whose proofs are given in Web Appendix B.

Lemma 1.

Under Conditions 1–4, the classes of functions $\mathscr{M}$ is $\mathbb{P}$-Donsker.

Lemma 2.

Under Conditions 1–4,

$$E\left[\sum_{m=0}^{M}\left\{\widehat{\Lambda}\left(U_m\right)-\Lambda_0\left(U_m\right)\right\}^2\right]=O_P\left(n^{-2/3}\right)+O\left(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\right).$$

Proof of Theorem 1

The jump points $\{t_1,..., t_k\}$ depend on sample size n and for any $\epsilon > 0$, $\cup_j B_\epsilon(t_j)$ covers the support $\mathscr{V}$ as $n \to \infty$, where $B_r(t)$ is the open ball around t_j with radius r. By the continuity of Λ_0, $\widetilde{\Lambda}(t)$ converges uniformly to $\Lambda_0(t)$. It follows from Lemma 1 that the class $\mathscr{M}$ is Donsker. By the concavity of the log function,

$$\begin{aligned}\mathbb{P}_n m(\hat{\boldsymbol{\beta}},\widehat{\Lambda}) &\geq \frac{1}{2}\left\{\mathbb{P}_n l(\hat{\boldsymbol{\beta}},\widehat{\Lambda})+\mathbb{P}_n l\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right)\right\}\\ &\geq \mathbb{P}_n l\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right)=\mathbb{P}_n m\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right).\end{aligned}$$

Since $\mathscr{B}$ is bounded, for any subsequence of $\hat{\boldsymbol{\beta}}$, we can find a further subsequence converging to $\boldsymbol{\beta}_* \in \mathscr{B}$.. In addition, by Helly's selection lemma, for any subsequence of $\widehat{\Lambda}$, there exists a further subsequence that converges to some increasing function Λ_*. We choose the converging subsequences of $\hat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}$ such that we can obtain without loss of generality that $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_*$ and $\widehat{\Lambda} \to \Lambda_*$ pointwise on any interior set of $\mathscr{V}$. Therefore,

$$\begin{aligned}0 &\leq \mathbb{P}_n m(\hat{\boldsymbol{\beta}},\widehat{\Lambda})-\mathbb{P}_n m\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right)\\ &=\mathbb{P}\log\frac{L(\hat{\boldsymbol{\beta}},\widehat{\Lambda})+L\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right)}{2L\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right)}+o_P(1)\\ &=\mathbb{P}\log\left\{\frac{1}{2}+\frac{L\left(\boldsymbol{\beta}_*,\Lambda_*\right)}{2L\left(\boldsymbol{\beta}_0,\widetilde{\Lambda}\right)}\right\}+o_P(1),\end{aligned}$$

such that the negative Kullback-Leibler information is positive. Therefore,

$$\begin{aligned}&\frac{\sum_{m=0}^{M}\Delta_m\left[\exp\left\{-\Lambda_*\left(U_m\right)\exp\left(\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{Z}\right)\right\}-I(R<\infty)\exp\left\{-\Lambda_*\left(U_{m+1}\right)\exp\left(\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{Z}\right)\right\}\right]}{\int_0^\tau\exp\left\{-\Lambda_*(a)\exp\left(\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da}\\ &=\frac{\sum_{m=0}^{M}\Delta_m\left[\exp\left\{-\Lambda_0\left(U_m\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}-I(R<\infty)\exp\left\{-\Lambda_0\left(U_{m+1}\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}\right]}{\int_0^\tau\exp\left\{-\Lambda_0(a)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da}\end{aligned}$$

with probability 1. For any $m \in \{0,..., M\}$, we set $\Delta_{m'} = 1$ in the above equation $m' = m,...,$ M and take the sum of the resulting equations to obtain

$$\frac{\exp\left\{-\Lambda_*\left(U_m\right)\exp\left(\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{Z}\right)\right\}}{\int_0^\tau \exp\left\{-\Lambda_*(a)\exp\left(\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da} = \frac{\exp\left\{-\Lambda_0\left(U_m\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}}{\int_0^\tau \exp\left\{-\Lambda_0(a)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da}.$$

Because m is arbitrary, we can replace U_m in the above equation by any $t \in \mathscr{V}$. We take the logarithm and differentiate both sides with respect to t to find

$$\Lambda_*'(t)\exp\left(\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{Z}\right) = \Lambda_0'(t)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right),$$

such that $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$ and $\Lambda_*'(t) = \Lambda_0'(t)$ for $t \in \mathscr{V}$. Hence, $\Lambda_*(t) = \Lambda_0(t)$ for $t \in \mathscr{V}$. We conclude that $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \to 0$ and $\left|\widehat{\Lambda}(t) - \Lambda_0(t)\right| \to 0$ for any $t \in \mathscr{V}$. Because Λ_0 is continuous, $\widehat{\Lambda}$ converges uniformly to Λ_0 on $\mathscr{V}$.

Proof of Theorem 2.

Let

$$Q_1(t, u, v; \boldsymbol{\beta}, \Lambda) = \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\frac{I(t \le v)\exp\left\{-\Lambda(v)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\} - I(t \le u)\exp\left\{-\Lambda(u)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\}}{\exp\left\{-\Lambda(u)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\} - \exp\left\{-\Lambda(v)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\}},$$

$$Q_2(t; \boldsymbol{\beta}, \Lambda) = \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right) \times \frac{\int_0^\tau I(t \le a)\exp\left\{-\Lambda(a)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da}{\int_0^\tau \exp\left\{-\Lambda(a)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da},$$

and $Q(t; \boldsymbol{\beta}, \Lambda) = \sum_{m=0}^{M} \Delta_m Q_1\left(t, U_m, U_{m+1}; \boldsymbol{\beta}, \Lambda\right) + Q_2(t; \boldsymbol{\beta}, \Lambda)$. The score equations for $\boldsymbol{\beta}$ is given by

$$\boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \Lambda) = Z\int Q(t; \boldsymbol{\beta}, \Lambda)d\Lambda(t).$$

The score operator for Λ along the submodel $d\Lambda_{\epsilon,h} = (1 + \epsilon h)d\Lambda$ for $h \in L_2(\mu)$ is given by

$$l_\Lambda(\boldsymbol{\beta}, \Lambda)(h) = \int Q(t; \boldsymbol{\beta}, \Lambda)h(t)d\Lambda(t).$$

Clearly,

$$\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}, \widehat{\Lambda})\right\} = -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}, \widehat{\Lambda}) - \boldsymbol{l}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0, \Lambda_0\right)\right\},$$

and

$$\mathbb{G}_n\left\{l_\Lambda(\widehat{\boldsymbol{\beta}}, \widehat{\Lambda})(h)\right\} = -\sqrt{n}\mathbb{P}\left\{l_\Lambda(\widehat{\boldsymbol{\beta}}, \widehat{\Lambda})(h) - l_\Lambda\left(\boldsymbol{\beta}_0, \Lambda_0\right)(h)\right\}.$$

We apply the Taylor series expansions at $(\boldsymbol{\beta}_0, \Lambda_0)$ to the right sides of the above two equations. In light of Lemma 2, the second-order terms are bounded by $O_P\left(n^{-1/6} + \sqrt{}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right)$. Therefore,

$$\begin{aligned}\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}, \widehat{\Lambda})\right\} = &-\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \boldsymbol{l}_{\boldsymbol{\beta}\Lambda}\left(\widehat{\Lambda} - \Lambda_0\right)\right\} \\ &+ O_P\left(n^{-1/6} + \sqrt{}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right),\end{aligned}$$

and

$$\begin{aligned}\mathbb{G}_n\left\{l_\Lambda(\widehat{\beta}, \widehat{\Lambda})(h)\right\} = &-\sqrt{n}\mathbb{P}\left\{l_{\Lambda\beta}(h)\left(\widehat{\beta} - \beta_0\right) + l_{\Lambda\Lambda}\left(h, \widehat{\Lambda} - \Lambda_0\right)\right\} \\ &+ O_P\left(n^{-1/6} + \sqrt{}\|\widehat{\beta} - \beta_0\|^2\right),\end{aligned}$$

$d\Lambda_{\epsilon,h}$, $l_{\Lambda\boldsymbol{\beta}}(h)$ is the derivative of $l_\Lambda(h)$ with respect to $\boldsymbol{\beta}$, and $l_{\Lambda\Lambda}(h, \widehat{\Lambda} - \Lambda_0)$ is the derivative of $l_\Lambda(h)$ along the submodel $d\Lambda_0 + \epsilon d(\widehat{\Lambda} - \Lambda_0)$. All derivatives are evaluated at $(\boldsymbol{\beta}_0, \Lambda_0)$.

If the least favorable direction exists, we denote it as $\boldsymbol{h}^* \in L_2(\mu)^p$. We first show the existence of $\boldsymbol{h}^*$, which is the solution of $l_\Lambda^* l_\Lambda\left(\boldsymbol{h}^*\right) = l_\Lambda^* \boldsymbol{l}_{\boldsymbol{\beta}}$, where l_Λ^* is the adjoint operator of l_Λ. We equip $L_2(\mu)$ with an inner product defined as

$$< h^{(1)}, h^{(2)} > = \int h^{(1)}(t) h^{(2)}(t) d\mu(t).$$

On the same space, we define

$$\begin{aligned}\| h \| &= \mathbb{P}\left\{l_\Lambda\left(\boldsymbol{\beta}_0, \Lambda_0\right)(h)^2\right\}^{1/2} \\ &= \mathbb{P}\left[\left\{\int Q\left(t; \boldsymbol{\beta}_0, \Lambda_0\right) h(t) d\Lambda(t)\right\}^2\right]^{1/2}.\end{aligned}$$

It is easy to show that $\|\cdot\|$ is a seminorm on $L_2(\mu)$. Furthermore, if $\|h\| = 0$, then $\mathbb{P}\left\{l_\Lambda\left(\boldsymbol{\beta}_0, \Lambda_0\right)(h)^2\right\} = 0$. Thus, with probability 1, $l_\Lambda\left(\boldsymbol{\beta}_0, \Lambda_0\right)(h) = 0$. By the arguments in the Lemma 2, $h(t) = 0$ for $t \in \mathscr{V}$. Clearly, $\|h\| \leq c < h, h >^{1/2}$ for some constant c by the Cauchy-Schwarz inequality. According to the bounded inverse theorem in Banach spaces, we have

$\langle h, h \rangle^{1/2} \le \tilde{c} \parallel h \parallel$ for another constant $\tilde{c}$. By the Lax-Milgram theorem (Zeidler, 1995), there exists $\boldsymbol{h}^* \in L_2(\mu)^p$ that satisfies

$$\int_0^\tau \mathbb{P}\left\{Q\left(t; \boldsymbol{\beta}_0, \Lambda_0\right)Q\left(s; \boldsymbol{\beta}_0, \Lambda_0\right)\right\}\boldsymbol{h}^*(s)d\Lambda_0(s) = \int_0^\tau \mathbb{P}\left\{\boldsymbol{Z}Q\left(t; \boldsymbol{\beta}_0, \Lambda_0\right)Q\left(s; \boldsymbol{\beta}_0, \Lambda_0\right)\right\}d\Lambda_0(s)$$

for $t \in \mathscr{V}$. Differentiation of the integral equations with respect to t yields

$$q_1(t)\boldsymbol{h}^*(t) + \int_t^\tau q_2(s,t)\boldsymbol{h}^*(s)ds + \int_0^t q_3(s,t)\boldsymbol{h}^*(s)ds = \boldsymbol{q}_4(t),$$

where $q_1(t) > 0$, and q_j ($j = 1, 2, 3$) and $\boldsymbol{q}_4$ are continuously differentiable functions. Thus, $\boldsymbol{h}^*$ can be expanded to be a continuously differentiable function in $\mathscr{V}$ with bounded total variations. It follows that

$$\begin{aligned}
&\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}, \widehat{\Lambda})\right\} - \mathbb{G}_n\left\{l_\Lambda(\hat{\boldsymbol{\beta}}, \widehat{\Lambda})\left(h^*\right)\right\} \\
&= -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \boldsymbol{l}_{\boldsymbol{\beta}\Lambda}\left(\widehat{\Lambda} - \Lambda_0\right)\right\} \\
&\quad + \sqrt{n}\mathbb{P}\left\{l_{\Lambda\boldsymbol{\beta}}\left(\boldsymbol{h}^*\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + l_{\Lambda\Lambda}\left(\boldsymbol{h}^*, \widehat{\Lambda} - \Lambda_0\right)\right\} \\
&\quad + O_P\left(n^{-1/6} + \sqrt{n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right) \\
&= \sqrt{n}\mathbb{P}\left[\left\{l_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0, \Lambda_0\right) - \boldsymbol{l}_\Lambda\left(\boldsymbol{\beta}_0, \Lambda_0\right)\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \\
&\quad + O_P\left(n^{-1/6} + \sqrt{n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right).
\end{aligned}$$

Using the arguments in the proof of Lemma 1, we can show that $\boldsymbol{l}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0, \Lambda_0\right) - l_\Lambda\left(\boldsymbol{\beta}_0, \Lambda_0\right)\left(\boldsymbol{h}^*\right)$ belongs to a Donsker class. Next, we show that $\mathbb{P}\left[\left\{\boldsymbol{l}_\beta - l_\Lambda\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]$ is invertible. If the matrix is singular, then there exists an vector $\boldsymbol{v} \in \mathbb{R}^p$ such that $\boldsymbol{v}^{\mathrm{T}}\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\beta}} - l_\Lambda\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]\boldsymbol{v} = 0$. It follows that, with probability 1, the score function along the submodel $\left\{\boldsymbol{\beta}_0 + \epsilon\boldsymbol{v}, \Lambda_{\epsilon, -\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}^*}\right\}$ is zero. That is,

$$\begin{aligned}
&\sum_{m=0}^{M} \Delta_m \int \frac{I\left(t \le U_{m+1}\right)\exp\left\{-\Lambda_0\left(U_{m+1}\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\} - I\left(t \le U_m\right)\exp\left\{-\Lambda_0\left(U_m\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}}{\exp\left\{-\Lambda_0\left(U_m\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\} - \exp\left\{-\Lambda_0\left(U_{m+1}\right)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}} \\
&\times \boldsymbol{v}^{\mathrm{T}}\left\{\boldsymbol{Z} - \boldsymbol{h}^*(t)\right\}d\Lambda_0(t) + \int \frac{\int_0^\tau I(t \le a)\exp\left\{-\Lambda_0(a)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da}{\int_0^\tau \exp\left\{-\Lambda_0(a)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\}da}\boldsymbol{v}^{\mathrm{T}}\left\{\boldsymbol{Z} - \boldsymbol{h}^*(t)\right\}d\Lambda_0(t) = 0.
\end{aligned}$$

For any $m \in \{0, \ldots, M\}$, we sum over all possible $\Delta_{m'}$ with $m' = m, \ldots, M$ to obtain

$$
\begin{aligned}
&-\int_0^{U_m} \boldsymbol{v}^{\mathrm{T}}\left\{\boldsymbol{Z}-\boldsymbol{h}^*(t)\right\} d\Lambda_0(t) \\
&+\int \frac{\int_0^a \exp\left\{-\Lambda_0(a)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\} da}{\int_0^\tau \exp\left\{-\Lambda_0(a)\exp\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}\right)\right\} da}\boldsymbol{v}^{\mathrm{T}} \\
&\times\left\{\boldsymbol{Z}-\boldsymbol{h}^*(t)\right\} d\Lambda_0(t)=0.
\end{aligned}
$$

Because m is arbitrary, we can replace U_m in the above equation by $t \in \mathscr{V}$. We differentiate both sides with respect to t to obtain

$$
\boldsymbol{v}^{\mathrm{T}}\left\{\boldsymbol{Z}-\boldsymbol{h}^*(t)\right\}\Lambda_0'(t)=0
$$

for any $t \in \mathscr{V}$. It then follows that $\boldsymbol{\upsilon} = \mathbf{0}$. Hence, the matrix $\mathbb{P}\left[\left\{\boldsymbol{l}_\beta - l_\Lambda\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]$ is invertible.

Then, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_P\left(n^{-1/2}\right)$, and

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=&\left(\mathbb{P}\left[\left\{l_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0,\Lambda_0\right)-l_\Lambda\left(\boldsymbol{\beta}_0,\Lambda_0\right)\left(h^*\right)\right\}^{\otimes 2}\right]\right)^{-1} \\
&\times\mathbb{G}_n\left\{l_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}},\widehat{\Lambda})-l_\Lambda(\widehat{\boldsymbol{\beta}},\widehat{\Lambda})\left(h^*\right)\right\}+o_P(1)\,.
\end{aligned}
$$

The influence function for $\widehat{\boldsymbol{\beta}}$ is the efficient influence function, such that $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)$ converges weakly to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

Appendix B

An EM algorithm for maximum conditional likelihood estimation

A nonparametric maximum conditional likelihood estimator is considered in Pan and Chappell (1998) for the proportional hazards model with left-truncated interval-censored data. A slight variation of the proposed EM algorithm can be used to compute their estimator, and is used in the numerical comparisons.

The observed-data conditional likelihood given the truncation time is given by

$$
\prod_{i=1}^n \frac{\exp\left\{-\Lambda\left(L_i\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\}-I\left(R_i<\infty\right)\exp\left\{-\Lambda\left(R_i\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\}}{\exp\left\{-\Lambda\left(A_i\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i\right)\right\}}.
$$

We estimate Λ nonparametrically such that the estimator for Λ is a step function that jumps only at $t_1, \ldots, t_k$, which are the ordered sequence of all L_i, R_i, and A_i. We maximize the objective function

$$\sum_{i=1}^{n}\left(\log\left[\exp\left\{-\sum_{A_i \le t_j \le L_i}\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\right\}\right.\right.$$
$$\left.\left. - I\left(R_i < \infty\right)\exp\left\{-\sum_{A_i \le t_j \le R_i}\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\right\}\right]\right).$$

We introduce a sequence of independent Poisson random variables $W_{ij}\left(i = 1, \ldots, n; j = 1, \ldots, k, A_i \le t_j \le R_i^*\right)$ with means $\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)$. Let $N_{i1} = \sum_{A_i \le t_j \le L_i} W_{ij}$, and $N_{i2} = I\left(R_i < \infty\right)\sum_{L_i < t_j \le R_i} W_{ij}$. The objective function can be viewed as the observed-data log-likelihood for $\{N_{i1} = 0, N_{i2} > 0 : i = 1, \ldots, n\}$ with $W_{ij}\left(j = 1, \ldots, k, A_i \le t_j \le R_i^*\right)$ as latent variables. We propose an EM algorithm. In the E-step, we evaluate

$$\widehat{E}\left(W_{ij}\right) = I\left(L_i < t_j \le R_i, R_i < \infty\right)$$
$$\times \frac{\lambda_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)}{1 - \exp\left\{-\sum_{L_i < t_l \le R_i}\lambda_l \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)\right\}}.$$

In the M step, we update λ_j by

$$\lambda_j = \frac{\sum_{i=1}^{n} I\left(A_i \le t_j \le R_i^*\right)\widehat{E}\left(W_{ij}\right)}{\sum_{i=1}^{n} I\left(A_i \le t_j \le R_i^*\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\right)}$$

and update β by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{k} I\left(A_i \le t_j \le R_i^*\right)\widehat{E}\left(W_{ij}\right)$$
$$\times\left\{Z_i - \frac{\sum_{i'=1}^{n} I\left(A_{i'} \le t_j \le R_{i'}^*\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}Z_{i'}\right)Z_{i'}}{\sum_{i'=1}^{n} I\left(A_{i'} \le t_j \le R_{i'}^*\right)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}Z_{i'}\right)}\right\} = \mathbf{0}.$$

We iterate between the E-step and M-step until convergence.

REFERENCES

Asgharian M, Wolfson DB, and Zhang X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. Stat Med 25, 1751–1767. [PubMed: 16220462]

Brookmeyer R. and Gail MH (1987). Biases in prevalent cohorts. Biometrics 43, 739–749. [PubMed: 3427161]

Cai T. and Betensky RA (2003). Hazard regression for interval-censored data with penalized spline. Biometrics 59, 570–579. [PubMed: 14601758]

Chan KCG And Wang M-C (2010). Backward estimation of stochastic processes with failure events as time origins. Ann Appl Stat 4, 1602–1620. [PubMed: 21359167]

Chappell R. (1991). Sampling design of multiwave studies with an application to the Massachusetts health care panel study. Stat Med 10, 1945–1958. [PubMed: 1805320]

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

Henschel V. and Mansmann U. (2013). intcox: Iterated convex minorant algorithm for interval-censored event data. R package version 0.9.3.

Huang J. (1996). Efficient estimation for the proportional hazards model with interval censoring. Ann Stat 24, 540–568.

Huang C-Y and Qin J. (2012). Composite partial likelihood estimation under length-biased sampling, with application to a prevalent cohort study of dementia. J Am Stat Assoc 107, 946–957. [PubMed: 24000265]

Huang J. and Rossini AJ (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. J Am Stat Assoc 92, 960–967.

Huang C-Y, Ning J, and Qin J. (2015). Semiparametric likelihood inference for left-truncated and right-censored data. Biostatistics 16, 785–798. [PubMed: 25796430]

Hudgens MG (2005). On nonparametric maximum likelihood estimation with interval censoring and left truncation. J R Stat Soc B 67, 573–587.

Kim JS (2003). Efficient estimation for the proportional hazards model with left-truncated and “Case 1” interval-censored data. Stat Sin 13, 519–537.

Lai TL and Ying Z. (1994). A missing information principle and M-estimators in regression analysis with censored and truncated data. Ann Stat 22, 1222–1255.

Mandel M. and Betensky RA (2007). Testing goodness of fit of a uniform truncation model. Biometrics 63, 405–412. [PubMed: 17688493]

Mao L, Lin D-Y, and Zeng D. (2017). Semiparametric regression analysis of interval-censored competing risks data. Biometrics 73, 857–865. [PubMed: 28211951]

Murphy SA and Vaart AW (2000). On profile likelihood. J Am Stat Assoc 95, 449–465.

Pan W. (1999). Extending the iterative convex minorant algorithm to the Cox model for interval-censored data. J Comput Graph Stat 8, 109–120.

Pan W. and Chappell R. (1998). Computation of the NPMLE of distribution functions for interval censored and truncated data with applications to the Cox model. Comput Stat Data Anal 28, 33–50.

Pan W. and Chappell R. (2002). Estimation in the Cox proportional hazards model with left-truncated and interval-censored data. Biometrics 58, 64–70. [PubMed: 11890328]

Qin J. and Shen Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. Biometrics 66, 382–392. [PubMed: 19522872]

Qin J, Ning J, Liu H, and Shen Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. J Am Stat Assoc 106, 1434–1449. [PubMed: 22323840]

Shen Y, Ning J, and Qin J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. J Am Stat Assoc 104, 1192–1202. [PubMed: 21057599]

Sun J. (2007). The Statistical Analysis of Interval-Censored Failure Time Data. New York: Springer.

Turnbull BW (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc B 38, 290–295.

Wang M-C (1991). Nonparametric estimation from cross-sectional survival data J Am Stat Assoc 86, 130–143.

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

Wang P, Tong X, Zhao S, and Sun J. (2015). Regression analysis of left-truncated and case i interval-censored data with the additive hazards model. Commun Stat Theory Methods 44, 1537–1551.

Wang L, McMahan CS, Hudgens MG, and Qureshi Z. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. Biometrics 72, 222–231. [PubMed: 26393917]

Zeidler E. (1995). Applied Functional Analysis—Applications to Mathematical Physics. New York: Springer.

Zeng D, Mao L, and Lin DY (2016). Maximum likelihood estimation for semiparametric transformation models with interval-scensored data. Biometrika 103, 253–271. [PubMed: 27279656]

Zeng D, Gao F, and Lin DY (2017). Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. Biometrika 104, 505–525. [PubMed: 29391606]

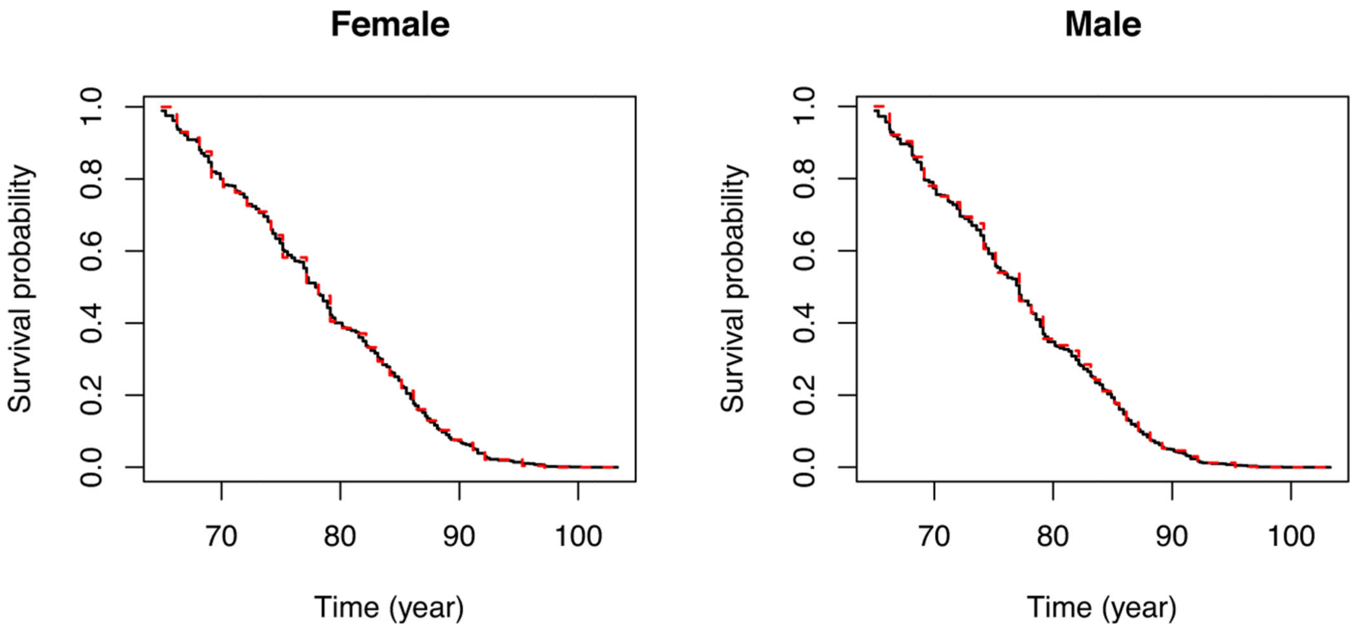

FIGURE 1.

Estimated survival probabilities for subgroups of subjects in the Massachusetts Health Care Panel Study. The solid and dashed curves pertain to the nonparametric maximum likelihood and conditional likelihood estimation approaches, respectively.

TABLE 1

Summary statistics for the simulation studies with length-biased assumption

		NPMLE					CLE				
		Bias	**SE**	**SEE**	**RMSE**	**CP**	**Bias**	**SE**	**SEE**	**RMSE**	**CP**
n=100	β_1	0.008	0.169	0.171	0.169	0.956	0.049	0.253	0.235	0.258	0.928
	β_2	0.015	0.295	0.317	0.295	0.965	0.089	0.444	0.436	0.453	0.942
n=200	β_1	0.004	0.117	0.116	0.117	0.950	0.025	0.170	0.159	0.172	0.933
	β_2	0.005	0.206	0.212	0.206	0.957	0.044	0.296	0.290	0.299	0.943
n=400	β_1	0.002	0.082	0.081	0.082	0.944	0.014	0.115	0.111	0.116	0.937
	β_2	0.002	0.144	0.146	0.144	0.951	0.024	0.200	0.199	0.202	0.949

Note: NPMLE and CLE are the nonparametric maximum likelihood and conditional likelihood estimators. SE, SEE, RMSE, and CP are the empirical standard error, mean standard error estimator, root mean squared error, and empirical coverage probability of the 95% confidence interval, respectively.

TABLE 2

Summary statistics for the simulation studies without length-biased assumption

		NPMLE					CLE				
		Bias	**SE**	**SEE**	**RMSE**	**CP**	**Bias**	**SE**	**SEE**	**RMSE**	**CP**
n=100	β_1	−0.039	0.160	0.167	0.165	0.953	0.045	0.244	0.231	0.248	0.933
	β_2	−0.077	0.285	0.307	0.295	0.951	0.083	0.442	0.428	0.450	0.941
n=200	β_1	−0.044	0.111	0.113	0.119	0.933	0.023	0.164	0.157	0.165	0.935
	β_2	−0.088	0.198	0.205	0.217	0.930	0.045	0.293	0.285	0.296	0.943
n=400	β_1	−0.047	0.078	0.078	0.091	0.899	0.012	0.113	0.109	0.113	0.938
	β_2	−0.095	0.137	0.140	0.167	0.894	0.021	0.198	0.196	0.199	0.947

Note: NPMLE and CLE are the nonparametric maximum likelihood and conditional likelihood estimators. SE, SEE, RMSE, and CP are the empirical standard error, mean standard error estimator, root mean squared error, and empirical coverage probability of the 95% confidence interval, respectively.

TABLE 3

Estimation results for the regression parameter in the Massachusetts Health Care Panel Study

	NPMLE			**CLE**		
Covariate	**Estimate**	**Std Err**	***p*-value**	**Estimate**	**Std Err**	***p*-value**
Male	0.144	0.059	0.014	0.133	0.076	0.081

Note: NPMLE and CLE are the nonparametric maximum likelihood and conditional likelihood estimators.