Biometrics xxx, 1–21 DOI: 0.0.0.0
Xxxx 201x

A Pairwise Likelihood Augmented Cox Estimator for Left-Truncated Data

Fan Wu[1], Sehee Kim[1], Jing Qin[2], Rajiv Saran[3] and Yi Li[1,*]

[1] Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

[2]Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

[3]Kidney Epidemiology and Cost Center, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

email: yili@umich.edu

Summary: Survival data collected from a prevalent cohort are subject to left truncation and the analysis is challenging. Conditional approaches for left-truncated data could be inefficient as they ignore the information in the marginal likelihood of the truncation times. Length-biased sampling methods may improve the estimation efficiency but only when the underlying truncation time is uniform; otherwise, they may generate biased estimates. We propose a semiparametric method for left-truncated data under the Cox model with no parametric distributional assumption about the truncation times. Our approach is to make inference based on the conditional likelihood augmented with a pairwise likelihood, which eliminates the truncation distribution, yet retains the information about the regression coefficients and the baseline hazard function in the marginal likelihood. An iterative algorithm is provided to solve for the regression coefficients and the baseline hazard function simultaneously. By empirical process and U-process theories, it has been shown that the proposed estimator is consistent and asymptotically normal with a closed-form consistent variance estimator. Simulation studies show substantial efficiency gain of our estimator in both the regression coefficients and the cumulative baseline hazard function over the conditional approach estimator. When the uniform truncation assumption holds, our estimator enjoys smaller biases and efficiency comparable to that of the full maximum likelihood estimator. An application to the analysis of a chronic kidney disease cohort study illustrates the utility of the method.

Key words: Chronic kidney disease; Composite likelihood; Empirical process; Self-consistency; U-Process.

1. Introduction

Survival data collected from a prevalent cohort that includes patients who already have the disease at the study enrollment are subject to left truncation. This is because those who died with the disease before the enrollment would have no chances to be selected, whereas the selected patients, having survived until the enrollment, are healthier on average. To avoid overestimating the survival, conventional approaches make inferences conditional on truncation times (Kalbfleisch and Lawless, 1991; Wang et al., 1993). These approaches disregard the information about the regression coefficients in the marginal likelihood of the truncation times, and hence loss of efficiency is expected when additional knowledge on the underlying truncation distribution is available (Huang and Qin, 2012).

If the underlying truncation time is uniformly distributed, left truncation reduces to length-biased sampling (Vardi, 1989), that is, the probability of selecting a subject is proportional to the length of his or her underlying failure time; see a comprehensive review by Shen et al. (2017). Among the newly developed regression methods for length-biased data, many show considerable improvement of efficiency in estimation compared with the conditional approach by incorporating information from the observed truncation times (Qin and Shen, 2010; Qin et al., 2011; Huang et al., 2012; Huang and Qin, 2012; Ning et al., 2014). Nevertheless, when the uniform truncation assumption is violated, these methods may yield inconsistent estimates (Huang and Qin, 2012).

The motivating study is a prevalent cohort study of patients with chronic kidney disease (CKD), sponsored by the Renal Research Institute (Perlman et al., 2003). Following the diagnosis, in general, CKD patients are referred to nephrologists to receive special care and treatments. The investigators were interested in whether the patient characteristics at referral were associated with the disease progression to end-stage renal disease (ESRD) or death. At the study recruitment from June 2000 to January 2006, subjects with glomerular filtration

rate (GFR) less than or equal to 50 ml/min/1.73 m^2 were invited to participate. The dataset is of a moderate sample size, so improving the estimation efficiency is important. However, statistical assessment in Section 4 indicated deviation of the motivating data from the uniform truncation assumption, which prompted us to seek an efficiency-improving method with consistent estimates.

Recently, Huang and Qin (2013) proposed a more efficient estimator for the additive hazards model under general left truncation. They used a pairwise likelihood of the truncation times to eliminate the unspecified truncation distribution (Liang and Qin, 2000). In practice, however, the Cox model is more commonly used than the additive hazards model, and its interpretation is more familiar to practitioners (Cox, 1972). Yet the challenge of applying the same approach to the Cox model lies in the complicated way that the pairwise likelihood involves the cumulative baseline hazard function, causing serious theoretical and computational difficulties.

In this paper, we propose to augment the conditional likelihood with a pairwise likelihood constructed from the marginal likelihood of the truncation times to improve the efficiency in estimation for the Cox model. We have achieved several important improvements. First, we design an nonparametric maximum likelihood estimating (NPMLE) procedure to estimate the cumulative baseline hazard function along with the regression coefficients. Second, with the asymptotic results proven by empirical process and U-process theories, we provide a closed-form consistent sandwich variance estimator. Finally, we provide an iterative algorithm that explores the self-consistency of the nonparametric estimator and guarantees a computationally efficient implementation. Our simulations show that efficiency of both the regression coefficients and the cumulative baseline hazard function, especially the former, can be improved using the proposed method. Moreover, even when the uniform truncation assumption holds, the proposed estimator of the regression coefficients has efficiency com-

parable to that of the full maximum likelihood estimator by Qin et al. (2011), and enjoys smaller biases. Thus, we believe the proposed estimator provides a promising alternative to improve the estimation efficiency for left-truncated survival data.

2. Proposed Method

2.1 *Preliminaries*

Suppose the time origin is the onset of disease. For a patient from the target population, let T^* denote the *underlying* failure time, i.e., the time to the event of interest, and A^* denote the *underlying* truncation time, i.e., the time to the study enrollment. We use f and S to denote the density and survival functions of T^*, and the distribution function of A^* is denoted as G. Let Z^* be a $p \times 1$ vector of covariates. We assume A^* and T^* are independent conditional on Z^*. A commonly used model that links the hazard function of T^* to the covariates Z^* is the Cox proportional hazards model (Cox, 1972):

$$\lambda(t \mid Z^*; \beta) = \lambda(t) \exp(\beta^{\mathrm{T}} Z^*),$$

where β is a $p \times 1$ vector of regression coefficients, and $\lambda(\cdot)$ is an unspecified baseline hazard function. The cumulative baseline hazard function is defined as $\Lambda(t) = \int_0^t \lambda(s)\, ds$. Data collected from a prevalent cohort only consist of patients with $A^* \leqslant T^*$. The same notations without asterisks, such as A, T, and Z, will be used to denote the *observed* random variables conditional on $A^* \leqslant T^*$, i.e., $(A, T, Z) \equiv (A^*, T^*, Z^*) \mid (A^* \leqslant T^*)$, throughout the paper.

Usually, the failure time is also subject to potential right censoring by C starting from the enrollment. Thus, what we observe are $X = \min(A + C, T)$ and $\Delta = I(T \leqslant A + C)$, where $I(\cdot)$ is the indicator function. Suppose we have independent and identically distributed observations $\{A_i, X_i, \Delta_i, Z_i;\, i = 1, \ldots, n\}$ on n individuals sampled from a prevalent cohort. The full likelihood of the observed data is proportional to

$$\prod_{i=1}^{n} \mathrm{pr}(A_i^*, T_i^*, C_i \mid A_i^* \leqslant T_i^*, Z_i^*) \propto \prod_{i=1}^{n} \frac{f(X_i \mid Z_i)^{\Delta_i} S(X_i \mid Z_i)^{1-\Delta_i} dG(A_i)}{\int_0^\infty S(a \mid Z_i) dG(a)} \equiv \mathcal{L}_n.$$

We assume C is independent of (A^*, T^*) given $A^* \leqslant T^*$ and Z^*, and that A^* does not depend on Z^*, i.e., the patient recruitment process does not depend on covariates. Note that the latter assumption does not imply independence between the *observed* A and Z, since the biased sampling scheme may induce correlations between them as well as between A and T given Z. The full likelihood can be further decomposed into two parts:

$$\mathcal{L}_n = \prod_{i=1}^{n} \frac{f(X_i \mid Z_i)^{\Delta_i} S(X_i \mid Z_i)^{1-\Delta_i}}{S(A_i \mid Z_i)} \times \prod_{i=1}^{n} \frac{S(A_i \mid Z_i) dG(A_i)}{\int_0^\infty S(a \mid Z_i) dG(a)} \equiv \mathcal{L}_n^C \times \mathcal{L}_n^M, \qquad (1)$$

where $\mathcal{L}_n^C$ is the conditional likelihood of (X_i, Δ_i) given (A_i, Z_i), $i = 1, \ldots, n$, and $\mathcal{L}_n^M$ is the marginal likelihood of A_i given Z_i, $i = 1, \ldots, n$.

2.2 *Pairwise Likelihood Augmented Cox (PLAC) Estimator*

In the presence of truncation, inference based on $\mathcal{L}_n^C$ only, using the Cox's partial likelihood (Cox, 1975) with the at-risk indicator $Y_i(t) = I(A_i \leqslant t \leqslant X_i)$, has been proposed by Kalbfleisch and Lawless (1991) and Wang et al. (1993). The conditional approaches yield consistent estimates, but they may be inefficient, since they completely ignore the information about the parameters contained in $\mathcal{L}_n^M$. Taking advantage of the fully specified uniform truncation distribution, regression methods for length-biased data generally result in more efficient estimators. Among them, the maximum likelihood estimator by Qin et al. (2011) is asymptotically efficient for the Cox model. More recently, Liu et al. (2016) extended the expectation-maximization algorithm in Qin et al. (2011) to general biased-sampling cases, where G is known up to some unspecified finite-dimensional parameters, and estimation efficiency of the Cox model parameters can be improved while jointly estimating the parameters in G.

Deviating from existing methods for left-truncated data, our method does not impose any parametric assumption on the underlying truncation distribution or on the baseline hazard function. Our approach to improving efficiency is to supplement $\mathcal{L}_n^C$ with major information in $\mathcal{L}_n^M$ that depends only on β and Λ. Specifically, we first apply the pairwise likelihood

method by Liang and Qin (2000) to $\mathcal{L}_n^M$ in order to eliminate the truncation distribution function, and then estimate β and Λ based on a composite likelihood consisting of $\mathcal{L}_n^C$ and $\mathcal{L}_n^P$, where the pairwise likelihood $\mathcal{L}_n^P$ is derived as follows.

Suppose a sample $\{(A_i, Z_i), (A_j, Z_j); 1 \leqslant i < j \leqslant n\}$ is available. A pseudo-likelihood of the pair (A_i, A_j), conditional on the order statistic of (A_i, A_j) and (Z_i, Z_j), is given by

$$\frac{\frac{S(A_i|Z_i)dG(A_i)}{\int_0^\infty S(a|Z_i)dG(a)} \times \frac{S(A_j|Z_j)dG(A_j)}{\int_0^\infty S(a|Z_j)dG(a)}}{\frac{S(A_i|Z_i)dG(A_i)}{\int_0^\infty S(a|Z_i)dG(a)} \times \frac{S(A_j|Z_j)dG(A_j)}{\int_0^\infty S(a|Z_j)dG(a)} + \frac{S(A_i|Z_j)dG(A_i)}{\int_0^\infty S(a|Z_j)dG(a)} \times \frac{S(A_j|Z_i)dG(A_j)}{\int_0^\infty S(a|Z_i)dG(a)}} = \frac{1}{1 + R_{ij}(\beta, \Lambda)},$$

where

$$R_{ij}(\beta, \Lambda) = \frac{S(A_i \mid Z_j)S(A_j \mid Z_i)}{S(A_i \mid Z_i)S(A_j \mid Z_j)} = \exp\left[(e^{\beta^{\mathrm{T}} Z_i} - e^{\beta^{\mathrm{T}} Z_j})\{\Lambda(A_i) - \Lambda(A_j)\}\right]$$

denotes the generalized odds ratio under the Cox model. The pairwise likelihood $\mathcal{L}_n^P$ of all pairs is then given by

$$\mathcal{L}_n^P = \prod_{i<j} \{1 + R_{ij}(\beta, \Lambda)\}^{-1}.$$

It is worth noting that, by canceling out the terms involving G, $\mathcal{L}_n^P$ is a function of (β, Λ) only, whereas $\mathcal{L}_n^M$ is a function of (β, Λ, G). Simulation studies (Qin and Liang, 1999; Liang and Qin, 2000) show that the pairwise likelihood can usually retain the majority of the information in the likelihood from which it is derived, and that the efficiency loss may not be substantial, depending on the model as well as the values of the parameters. Therefore, to estimate β and Λ, we propose using $\mathcal{L}_n^P$ as a reasonably good surrogate for $\mathcal{L}_n^M$ in the full likelihood approach. An analogous idea has been exploited in the additive hazards model by Huang and Qin (2013); however, the additive hazards model is less commonly used. Applying the pairwise likelihood augmentation method to the Cox model will greatly promote more practical use due to ease of interpretation to practitioners.

To account for the different magnitudes of $\log \mathcal{L}_n^C$ and $\log \mathcal{L}_n^P$ (there are n terms in $\log \mathcal{L}_n^C$

and $n(n-1)/2$ terms in $\log \mathcal{L}_n^P$), we maximize the following composite log-likelihood:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i\left\{\log\lambda(X_i)+\beta^{\mathrm{T}}Z_i\right\}-\exp(\beta^{\mathrm{T}}Z_i)\int_0^\infty Y_i(t)\lambda(t)dt\right]$$
$$-\frac{2}{n(n-1)}\sum_{i<j}\log\{1+R_{ij}(\beta,\Lambda)\},$$

over the domain of (β, Λ). Using the nonparametric maximum likelihood estimation, we treat $\Lambda(\cdot)$ as a nondecreasing step function such that $\Lambda(0) = 0$, and has jumps, denoted by $\Lambda\{\cdot\}$, only at the time points where events are observed (see Murphy et al., 1997; Zeng and Lin, 2006, among others). Let $w_1 < \cdots < w_m$ $(m \leqslant n)$ be the ordered distinct observed failure times, and $\lambda_1 \equiv \Lambda\{w_1\}$, ..., $\lambda_m \equiv \Lambda\{w_m\}$ be the corresponding positive jumps of Λ. We denote by $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_m)^{\mathrm{T}}$ the vector of all positive jumps. For $k = 0$, 1, 2, we define the following functions which appear in $\log \mathcal{L}_n^P$ and its derivatives:

$$Q_{ij}^{(k)}(t;\beta)=\left(Z_i^{\otimes k}e^{\beta^{\mathrm{T}}Z_i}-Z_j^{\otimes k}e^{\beta^{\mathrm{T}}Z_j}\right)\{I(t\leqslant A_i)-I(t\leqslant A_j)\},$$

where $Z^{\otimes 0} = 1$, $Z^{\otimes 1} = Z$, and $Z^{\otimes 2} = ZZ^{\mathrm{T}}$. Below we may suppress the dependence on model parameters, using R_{ij} and $Q_{ij}^{(k)}(t)$ to denote $R_{ij}(\beta, \Lambda)$ and $Q_{ij}^{(k)}(t;\beta)$ when the meanings of the notations are clear from the context. Replacing $\lambda(t)$ with $\Lambda\{t\}$, we modify the composite log-likelihood as a function of β and $\boldsymbol{\lambda}$:

$$\begin{aligned}\ell_n^c(\beta,\boldsymbol{\lambda}) \;=\; & \frac{1}{n}\sum_{i=1}^{n}\left\{\Delta_i\left(\log\Lambda\{X_i\}+\beta^{\mathrm{T}}Z_i\right)-\exp(\beta^{\mathrm{T}}Z_i)\sum_{k=1}^{m}\lambda_k Y_i(w_k)\right\}\\ & -\frac{2}{n(n-1)}\sum_{i<j}\log\{1+R_{ij}(\beta,\boldsymbol{\lambda})\},\end{aligned}\tag{2}$$

where $R_{ij}(\beta, \boldsymbol{\lambda}) = \exp\{\sum_{k=1}^m \lambda_k Q_{ij}^{(0)}(w_k)\}$. We refer to the maximizer $(\hat{\beta},\, \hat{\boldsymbol{\lambda}})$, or $(\hat{\beta},\, \hat{\Lambda})$, of the composite log-likelihood as the pairwise likelihood augmented Cox (PLAC) estimator, where Λ at a fixed time t is estimated by $\hat{\Lambda}(t) = \sum_{k=1}^m \hat{\lambda}_k\, I(w_k \leqslant t)$. Specifically, differentiating (2) with respect to $(\beta, \boldsymbol{\lambda})$ yields the composite score functions (the dependence on n is suppressed

in the notations):

$$U_\beta(\beta,\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n} Z_i\left\{\Delta_i - e^{\beta^{\mathrm{T}}Z_i}\sum_{k=1}^{m}\lambda_k Y_i(w_k)\right\} - \frac{1}{n(n-1)}\sum_{i\neq j}\frac{\sum_{k=1}^{m}\lambda_k Q_{ij}^{(1)}(w_k)}{1+R_{ij}^{-1}},$$

$$U_{\lambda_k}(\beta,\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n} I(X_i=w_k)\left\{\frac{\Delta_i}{\lambda_k} - Y_i(w_k)\,e^{\beta^{\mathrm{T}}Z_i}\right\} - \frac{1}{n(n-1)}\sum_{i\neq j}\frac{Q_{ij}^{(0)}(w_k)}{1+R_{ij}^{-1}}.$$

Let $U_{\boldsymbol{\lambda}}^{\mathrm{T}} = (U_{\lambda_1},\ldots,U_{\lambda_m})$, then the PLAC estimator $(\hat{\beta},\hat{\boldsymbol{\lambda}})$ is the solution to

$$U(\beta,\boldsymbol{\lambda}) = (U_\beta^{\mathrm{T}}, U_{\boldsymbol{\lambda}}^{\mathrm{T}})^{\mathrm{T}}(\beta,\boldsymbol{\lambda}) = 0, \tag{3}$$

which can be obtained numerically using the following algorithm, for example.

Unlike the conditional approach, directly solving the nonlinear system (3) is difficult due to the computational complexity brought by the pairwise structure. Therefore, we propose to solve for $\hat{\beta}$ and $\hat{\lambda}_k$ $(k=1,\ldots,m)$ iteratively:

Step 1. Start with initial values $\beta^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$.

Step 2. At the r-th iteration, update each $\lambda_k^{(r)}$ using

$$\lambda_k^{(r)} = \frac{n^{-1}\sum_{i=1}^{n}\Delta_i I(X_i=w_k)}{n^{-1}\sum_{i=1}^{n}Y_i(w_k)e^{Z_i^{\mathrm{T}}\beta^{(r-1)}} + \{n(n-1)\}^{-1}\sum_{i\neq j}\frac{Q_{ij}^{(0)}(w_k;\beta^{(r-1)})}{1+1/R_{ij}(\beta^{(r-1)},\boldsymbol{\lambda}^{(r-1)})}}. \tag{4}$$

Step 3. Update $\beta^{(r)}$ by one step of Newton-Raphson iteration:

$$\beta^{(r)} = \beta^{(r-1)} - \left\{\dot{U}_{\beta\beta}(\beta^{(r-1)},\boldsymbol{\lambda}^{(r)})\right\}^{-1}\left\{U_\beta(\beta^{(r-1)},\boldsymbol{\lambda}^{(r)})\right\},$$

where $\dot{U}_{\beta\beta}(\beta^{(r-1)},\boldsymbol{\lambda}^{(r)}) = \partial U_\beta(\beta,\boldsymbol{\lambda})/\partial\beta^{\mathrm{T}}|_{\beta=\beta^{(r-1)},\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(r)}}$.

Step 4. Repeat Steps 2 and 3 until convergence.

Initial values for the parameters in Step 1 can be set as $\beta^{(0)}=0$ and $\boldsymbol{\lambda}^{(0)}=(1/m,\ldots,1/m)$ or the estimates from the conditional approach. In our simulation studies, it is demonstrated that the algorithm is robust to the choice of initial values. In Step 2, updating λ_k using the self-consistent solution (4) is the crucial step which makes the computation of the PLAC estimator tractable. The above algorithm is implemented in the R package `plac`.

2.3 *Asymptotic Properties*

We establish consistency and asymptotic normality of the PLAC estimator $(\hat{\beta},\, \hat{\Lambda})$ under the regularity conditions in the Appendix, utilizing techniques from empirical process (van der Vaart and Wellner, 1996) and U-process theories (De la Peña and Giné, 1999). Detailed proofs are provided in the Web Supplementary, Section A.1. To ensure the NPMLE $\hat{\Lambda}$ exists, we prove its asymptotic properties on $[0, \tau]$, where $\tau > 0$ is a constant. In practice, τ is often chosen to be the maximum of $X = \min(A + C, T)$ (Huang and Qin, 2012). Denote the normalized score functions corresponding to $\mathcal{L}_n^C$ and $\mathcal{L}_n^P$ as $U^C(\beta, \boldsymbol{\lambda}) = n^{-1} \sum_{i=1}^n U_i^C(\beta, \Lambda)$ and $U^P(\beta, \boldsymbol{\lambda}) = 2\{n(n-1)\}^{-1} \sum_{i<j} U_{ij}^P(\beta, \boldsymbol{\lambda})$, respectively, where

$$U_i^C(\beta, \boldsymbol{\lambda}) \;=\; \begin{pmatrix} \Delta_i Z_i - Z_i e^{\beta^{\mathrm{T}} Z_i} \sum_{k=1}^m \lambda_k Y_i(w_k) \\ I(X_i = w_1)\{\Delta_i/\lambda_1 - Y_i(w_1)\, e^{\beta^{\mathrm{T}} Z_i}\} \\ \vdots \\ I(X_i = w_m)\{\Delta_i/\lambda_m - Y_i(w_m)\, e^{\beta^{\mathrm{T}} Z_i}\} \end{pmatrix}, \tag{5}$$

and

$$U_{ij}^P(\beta, \boldsymbol{\lambda}) \;=\; -\frac{1}{1 + R_{ij}^{-1}} \begin{pmatrix} \sum_{k=1}^m \lambda_k Q_{ij}^{(1)}(w_k) \\ Q_{ij}^{(0)}(w_1) \\ \vdots \\ Q_{ij}^{(0)}(w_m) \end{pmatrix}. \tag{6}$$

Theorem 1 (Consistency)**.** *Under Conditions (C1)-(C3),*

$$\hat{\beta} \to \beta_0 \quad \textit{and} \quad \left\| \hat{\Lambda} - \Lambda_0 \right\|_{L_\infty[0,\tau]} \to 0 \qquad \textit{almost surely as } n \to \infty,$$

where $(\beta_0,\, \Lambda_0)$ *are the true parameters, and* $\|\cdot\|_{L_\infty[0,\tau]}$ *is the supreme norm on* $[0, \tau]$*.*

The proof of Theorem 1 consists of three major steps. First, we show the parameters of interest are identifiable. Second, by the nature of the pairwise construction, $U_{ij}^P(\beta,\, \Lambda)$ is permutation symmetric in the observed data; thus, the pairwise score function $U^P(\beta,\, \Lambda)$ and its derivatives are U-processes of order two. We construct upper bounds for bracketing

numbers of the related function classes by combining the bracketing entropy results of uniformly bounded monotone functions with the preservation theorems for Lipschitz functions (see van der Vaart and Wellner, 1996, Chapter 2.7). The law of large numbers of these classes then follows from Corollary 3.2.5 of De la Peña and Giné (1999). In addition, we can show $\mathrm{E}\{U^P(\beta_0, \Lambda_0)\} = 0$ by the fact that $U_{ij}^P(\beta, \Lambda)$ is exactly the score function corresponding to the pseudo-likelihood of the pair (A_i, A_j), conditional on the order statistic of (A_i, A_j) and (Z_i, Z_j). In the last step, the strong consistency of the PLAC estimator can be proven through the likelihood equation argument similar to that given by Murphy et al. (1997), along with the composite Kullback-Leibler divergence (Varin and Vidoni, 2005) and the identifiability of the parameters.

For the weak convergence, we first establish the uniform $\sqrt{n}$-convergence rate and the asymptotic normality of the log-generalized odds ratio using the Hájek projection of U-processes (van der Vaart, 2000). The asymptotic normality of the PLAC estimator can be proved using Theorem 3.3.1 of van der Vaart and Wellner (1996). Noting that $\sqrt{n}\, U(\beta_0, \Lambda_0) = \sqrt{n}\, U^C(\beta_0, \Lambda_0) + \sqrt{n}\, U^P(\beta_0, \Lambda_0)$, the asymptotic normality of $\sqrt{n}\, U(\beta_0, \Lambda_0)$ is obtained by the separate contributions of $\sqrt{n}\, U^C(\beta_0, \Lambda_0)$ and $\sqrt{n}\, U^P(\beta_0, \Lambda_0)$, which are asymptotically independent (van der Vaart and Wellner, 1996, Example 1.4.6). The asymptotic normality of $\sqrt{n}\, U^C(\beta_0, \Lambda_0)$ follows from the martingale theory (Andersen and Gill, 1982; Wang et al., 1993), and our innovative contribution is to identify the limiting distribution of $\sqrt{n}\, U^P(\beta_0, \Lambda_0)$. The normality of the function classes involved in $U^P(\beta_0, \Lambda_0)$ and its derivative is shown through the results on the Vapnik-Chervonenkis subgraph classes, the normality of the log-generalized odds ratio, and the preservation theorems for Lipschitz functions (van der Vaart and Wellner, 1996, Chapter 2.10). Finally, the Fréchet differentiability of $\mathrm{E}\{U(\beta_0, \Lambda_0)\}$ and the invertibility of its derivative can be shown by Condition (C4) and the

Fredholm theory, following arguments similar to those in Zeng and Lin (2006). The weak convergence results are summarized in the following theorem.

Theorem 2 (Asymptotic normality)**.** *Under Conditions (C1)-(C4),* $\sqrt{n}\,(\hat{\beta}-\beta_0,\ \hat{\Lambda}(t)-\Lambda_0(t))$ *converges weakly to a mean-zero Gaussian process in* $\mathbb{R}^p \times \mathrm{BV}[0,\tau]$*, where* $\mathrm{BV}[0,\tau]$ *denotes the space of all functions with bounded total variation on* $[0,\tau]$*.*

One of the appealing features of our approach is that the covariance of the limiting process of the PLAC estimator can be consistently estimated by a closed-form sandwich estimator. Let

$$\begin{aligned}
\hat{V}^C &= \frac{1}{n}\sum_{i=1}^{n} U_i^C(\hat{\beta},\hat{\boldsymbol{\lambda}})^{\otimes 2},\\
\hat{J}^C &= -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial U_i^C(\beta,\boldsymbol{\lambda})}{\partial(\beta^{\mathrm{T}},\boldsymbol{\lambda}^{\mathrm{T}})}\bigg|_{\beta=\hat{\beta},\,\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}},\\
\hat{V}^P &= \frac{4}{n-1}\sum_{i=1}^{n}\left\{\frac{1}{n-1}\sum_{i\neq j} U_{ij}^P(\hat{\beta},\,\hat{\boldsymbol{\lambda}})\right\}^{\otimes 2},\\
\hat{J}^P &= -\frac{1}{n(n-1)}\sum_{i\neq j}\frac{\partial U_{ij}^P(\beta,\boldsymbol{\lambda})}{\partial(\beta^{\mathrm{T}},\boldsymbol{\lambda}^{\mathrm{T}})}\bigg|_{\beta=\hat{\beta},\,\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}},
\end{aligned}$$

where the exact expressions of $\partial U_i^C(\beta,\boldsymbol{\lambda})/\partial(\beta^{\mathrm{T}},\boldsymbol{\lambda}^{\mathrm{T}})$ and $\partial U_{ij}^P(\beta,\boldsymbol{\lambda})/\partial(\beta^{\mathrm{T}},\boldsymbol{\lambda}^{\mathrm{T}})$ are given in the Web Supplementary. To define the asymptotic covariance, consider a linear functional

$$\sqrt{n}\left[b_1^{\mathrm{T}}(\hat{\beta}-\beta_0)+\int_0^{\tau} h(t)\,d\left\{\hat{\Lambda}(t)-\Lambda_0(t)\right\}\right], \tag{7}$$

where $b_1 \in \mathbb{R}^p$, and $h(t)$ is an arbitrary function with bounded total variation on $[0,\tau]$. Let b_2 be the $m\times 1$ vector $(h(w_1),\ldots,h(w_m))^{\mathrm{T}}$, and $b=(b_1^{\mathrm{T}},b_2^{\mathrm{T}})^{\mathrm{T}}$. For example, when β_k $(k=1,\ldots,p)$ is the parameter of interest, we can set $b_1=e_k$ and b_2 as the $m\times 1$ vector of zeros, where e_k is a unit vector with 1 at the k-th element and 0 otherwise. In another example where $\Lambda(t)$ at a fixed time t is the parameter of interest as in our simulation, $b_1=0$ and $b_2=(I(w_1\leqslant t),\ldots,I(w_m\leqslant t))^{\mathrm{T}}$. As in Zeng and Lin (2006), we can treat β and $\boldsymbol{\lambda}$ in (2) as if they are finite-dimensional parameters since the PLAC estimator of Λ converges at a parametric rate. Then by the asymptotic properties of U-statistics (Sen, 1960) and the

composite likelihood theory, the linear functional (7) converges in distribution to a mean-zero Gaussian random variable with variance that can be consistently estimated by $b^{\mathrm{T}}\hat{\Sigma}b$, where

$$\hat{\Sigma} = (\hat{J}^C + \hat{J}^P)^{-1}(\hat{V}^C + \hat{V}^P)(\hat{J}^C + \hat{J}^P)^{-1} \tag{8}$$

is the observed inverse Godambe information (Varin et al., 2011). Naturally, the sandwich estimator (8) has the following partition:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{\beta\beta} & \hat{\Sigma}_{\beta\boldsymbol{\lambda}} \\ \hat{\Sigma}_{\boldsymbol{\lambda}\beta} & \hat{\Sigma}_{\boldsymbol{\lambda}\boldsymbol{\lambda}} \end{pmatrix},$$

where the sub-matrices are estimated asymptotic covariance matrices of the corresponding parameter estimates. The result of (8) is important, as we can directly apply the delta method to get the asymptotic variances of quantities of interest other than β and $\boldsymbol{\lambda}$. For instance, the asymptotic variance of $\hat{\Lambda}(t)$ at a fixed time t can be estimated by

$$\hat{\Sigma}_{\hat{\Lambda}(t)} = \sum_{k=1}^{m}\sum_{l=1}^{m} I(w_k \leqslant t, w_l \leqslant t)\hat{\sigma}_{kl}^{(\boldsymbol{\lambda}\boldsymbol{\lambda})},$$

where $\hat{\sigma}_{kl}^{(\boldsymbol{\lambda}\boldsymbol{\lambda})}$ is the covariance (variance) estimate corresponding to $\hat{\lambda}_k$ and $\hat{\lambda}_l$ in $\hat{\Sigma}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}$.

3. Simulation

We conducted extensive simulation studies to evaluate the finite-sample performance of the proposed PLAC estimator, and compared it with those of the conditional approach estimator (Kalbfleisch and Lawless, 1991; Wang et al., 1993) and the maximum likelihood estimator for length-biased data (LBML) proposed by Qin et al. (2011). The *underlying* failure time T^* was associated with two independent covariates in the following Cox model:

$$\lambda(t \mid Z_1, Z_2) = \lambda(t)\exp(\beta_1 Z_1 + \beta_2 Z_2), \tag{9}$$

where $Z_1 \sim$ Bernoulli(0.5), $Z_2 \sim$ uniform$(-1, 1)$, and the true values $\beta_1 = \beta_2 = 1$. The baseline hazard function $\lambda(t) = 2t$. For the *underlying* truncation time, we considered two cases: (1) length-biased data, and (2) non-length-biased data with $A^* \sim$ exponential(1). To generate samples in Case 1, we used the property that the truncation distribution is uniform,

which led to the *observed* failure time T having the density function $tf(t|Z)/\int_0^\infty tf(t|Z)dt$, where $f(t|Z)$ is the density corresponds to the Cox model (9). That is, we first generated *observed* failure time T from its density and then drew the corresponding *observed* truncation time A from uniform$(0, T)$, as suggested by (Mandel and Betensky, 2007). In Case 2, the *underlying* failure times T^* were generated from (9), and the *underlying* truncation times A^* were independently generated from exponential(1); yet only the pairs satisfying $A^* \leqslant T^*$ were kept. Repeat the data generation process until the desired sample size was reached. The censoring times C were generated independently from uniform$(0, C_{\max})$, where $C_{\max}$ was chosen to designate censoring rates of approximately 50% and 80%. The event indicator for subject i was obtained by calculating $\Delta_i = I(T_i \leqslant A_i + C_i)$. Sample sizes of 200, 400 and 800 were considered, and we generated 1000 datasets under each scenario.

For each dataset, we estimated β_1, β_2, and $\Lambda(t)$ at two fixed times $t = (\tau_{30}, \tau_{60})$, where τ_{30} and τ_{60} were the 30% and 60% percentiles of the observed survival times, $X = \min(T, A+C)$, under each scenario. Summary statistics for datasets with sample sizes of 400 and 800, including the average of the estimates minus the true value, the empirical standard error of the estimates, the average of the standard error estimates, the 95% coverage probability, and the relative efficiency to the conditional estimator (as the ratio between the mean squared errors), are provided in Table 1.

[Table 1 about here.]

The empirical biases of the PLAC estimates, like the conditional approach estimates, are close to zero under all scenarios. When data are length-biased (Case 1), the LBML estimates also have biases that are small yet larger than those of the other two estimators. The moderate biases in LBML estimates have been consistently observed by Liu et al. (2016). In contrast, the LBML estimates in Case 2 are severely biased, and the biases remain at similar magnitudes even when the sample size increases to 800.

The proposed method yields considerable efficiency gains compared with the conditional approach estimator under different sample sizes and censoring rates. The efficiency gains in $\hat{\beta}_1$ and $\hat{\beta}_2$ range from 49% to 173% in Case 1 and 36% to 97% in Case 2. The efficiency gains in $\hat{\Lambda}_{\tau_{30}}$ and $\hat{\Lambda}_{\tau_{60}}$ are not as large, but improvement over the conditional approach has been clearly shown, i.e., all relative efficiencies are greater than one. For the length-biased data (Case 1), although the proposed estimator of the regression coefficients has larger standard errors than the LBML estimator, the difference between the two is smaller than the improvement of PLAC estimator achieves over the conditional estimator. Due to smaller biases, the mean squared errors of the PLAC estimator are comparable to those of the LBML estimator. The relative efficiency gains of our estimator increase as the censoring rate increases, because the augmenting pairwise likelihood is not subject to censoring. These higher gains when censoring rate increases are also observed in the LBML estimates, but they are undermined by the simultaneously inflated biases. We also performed additional simulations with the baseline hazard function $\lambda(t) = 1$. The results were as good as those in Table 1 or even better with slightly increased efficiency gains and thus are omitted. Taking the biases and the variances altogether, the mean squared errors of our estimator are either the smallest or comparable to the best performer.

Comparing the empirical and estimated standard errors of the proposed estimator, we demonstrate that the variance of the proposed estimator is consistently estimated by the sandwich variance estimator (8). We notice that the standard errors for the PLAC estimates under $n = 200$ (Web Supplementary, Table B.1) are approximately twice of those under $n = 800$, which confirms the $\sqrt{n}$-convergence rate as proven in Section 2.3. In the scenarios with $n = 400$ and 80% censoring, the 95% coverage probabilities for the proposed estimator are close to the nominal level, except for $\hat{\Lambda}_{\tau_{30}}$ and $\hat{\Lambda}_{\tau_{60}}$. This is because of the small numbers of observed events which attenuate the normal approximation not only in our approach, but

also in the competitors. For example, the corresponding coverage probabilities for $\hat{\Lambda}_{\tau_{30}}$ and $\hat{\Lambda}_{\tau_{60}}$ using the conditional approach are 92% and 92%, both of which are also below the nominal level. When the sample size increases to 800, all coverage probabilities of the PLAC estimator get closer to the nominal level.

In summary, the proposed estimator performs well under finite sample sizes. It has small empirical biases, and enjoys substantial gains in efficiency in both the regression coefficients and the cumulative baseline hazard function. The performance of our estimator is robust to violation of the uniform truncation assumption as well as high censoring rates. The proposed sandwich estimator results in good variance estimates for all parameters, and yields reasonable confidence intervals with close-to-nominal coverage.

4. Data Application

We applied the proposed method to the RRI-CKD study. Investigators were interested in the risk of ESRD progression associated with the patient's characteristics at referral. In this study, the failure time was measured from the referral to the composite renal outcome defined as either death, long-term dialysis or kidney transplantation, whichever came first. The truncation time was measured from the referral to the study enrollment. The failure time was also subject to right censoring by non-participating physicians, consent withdraw, lost to follow up, protocol deviation, or the end of study. The baseline patient characteristics included age group (45 to 65 and older than 65), gender, race (white and non-white), the presence of diabetes, the presence of hypertension, and advanced-stage CKD (defined by estimated GFR less than 30 ml/min/1.73 m^2). Patients without referral information or important covariates were excluded. A total of 545 patients were included in our analysis, of which 256 experienced the composite renal outcome during the study follow-up. The censoring rate was 53%.

We first assessed the uniform truncation assumption. When it holds, the observed truncation time A follows the same distribution as the residual survival time, $V = X - A$ (Mandel and Betensky, 2007). We conducted a paired log-rank test for A and V (Jung, 1999), and the null hypothesis of the same distribution was rejected ($p < 0.001$). Moreover, the estimated survival functions for A and V deviated from each other with non-overlapping point-wise confidence intervals (Web Supplementary, Figure C.1). Therefore, the uniform truncation assumption did not hold in the RRI-CKD data, and hence regression methods for length-biased data might yield invalid inference. The violation of the uniform truncation assumption may be explained by the absence of general guidelines for when to refer to a nephrologist in practice; patients can be referred at either early or late stages of the disease.

Table 2 gives the regression coefficient estimates and their standard errors from the RRI-CKD data using the conditional approach and the proposed PLAC estimator. Comparing to the conditional approach, we observed consistently smaller standard error estimates and narrower confidence intervals (Web Supplementary, Figure C.2) for all regression coefficients in the analysis. The variance ratio of the conditional approach estimate to the corresponding PLAC estimate is 1.30 or greater. This implies that the conditional approach requires at least 30% more CKD patients to achieve the same estimating precision as the PLAC estimator. It is worth noting that the estimated coefficient for Non-white using the proposed estimator indicates a significant survival difference between the white and the non-white (estimated hazard ratio is 1.30, $p = 0.045$), whereas the conditional approach estimate does not suggest such a significant difference (estimated hazard ratio is 1.24, $p = 0.185$).

[Table 2 about here.]

To illustrate the use of the closed-form variance estimator (8), we estimated the survival curves of the patients with and without diabetes at referral, and constructed the correspond-

ing 95% point-wise confidence intervals (Figure 1). The estimated median survival times and the corresponding confidence intervals are also displayed in Figure 1.

[Figure 1 about here.]

5. Discussion

We have proposed a semiparametric estimation method for the Cox model with the issue of general left truncation. By constructing a pairwise likelihood from the marginal likelihood of the truncation times, we have eliminated the unknown truncation distribution from the full likelihood. Based on our simulation studies, the proposed estimator has been shown to be robust to heavy censoring and violation of the uniform truncation assumption, where the robustness means consistency and efficiency gain over the conditional approach estimator across all scenarios considered. On the contrary, efficiency improvement of length-biased sampling methods relies on the uniform truncation assumption that is required for consistent estimates.

We have utilized a nonparametric maximum likelihood approach to estimate the cumulative baseline hazard function along with the regression coefficients. Under regularity conditions, the consistency and asymptotic normality of $(\hat{\beta}, \hat{\Lambda})$ have been rigorously proved, which results in a closed-form consistent sandwich variance estimator. We avoid estimating the truncation distribution G, deemed as a nuisance parameter in our application, because eliminating it in the likelihood may simplify inference. The convenience, however, may come at the expense of some efficiency loss. Alternatively, one can estimate G directly and plug the estimate into the full likelihood. However, several drawbacks may present. First, if G is estimated nonparametrically, the numerical instability might undermine the estimation of β and Λ. Second, the inference with plug-in type estimators is challenging. Often the variance estimator is so complicated that a resampling-based method should be used (Huang et al.,

2012). Nevertheless, a full likelihood approach that incorporates additional information in G may yield a more efficient estimator, and further research is warranted.

Even though we relax the uniform truncation assumption, our proposed method, as well as all existing regression methods for length-biased data, still requires independence assumption between A^* and Z^*. Our sensitivity analysis (not shown) indicated the proposed method would yield biased estimates under covariate-dependent truncation. However, under weak dependent cases, we still observed smaller mean squared errors (with 12% biases) compared with the conditional approach estimator. To apply the PLAC estimator, a rigorous model checking tool for the independence assumption between A^* and Z^* is worth pursuing in the future. For our RRI-CKD example, a graphical inspection tool has been illustrated in Web Supplementary, Figure C.3, where we plot and compare the estimated $\hat{G}$ for each level of the covariate of interest. There is no apparent deviation between the estimated curves for the demographics and hypertension status. As for the CKD stage and diabetes status, the estimated $\hat{G}$'s have overlapping confidence intervals. Thus, we conclude that there was no obvious violation of the independence assumption in the RRI-CKD data, which is further supported by the similar point estimates as shown in Table 2 and Figure C.2.

The gain in efficiency is the greatest advantage of the proposed method. For length-biased data, the PLAC estimator is less efficient than the full maximum likelihood estimator of Qin et al. (2011), because the latter is based on the correctly specified uniform truncation distribution. However, the loss of efficiency is not substantial, and the proposed estimator has smaller biases. For non-length-biased data, if the truncation distribution is known, we can apply the monotone transformation $G(\cdot)$ to both A and X and apply the regression methods for length-biased data to the transformed data as suggested by Huang and Qin (2012). In additional simulations with $A^* \sim$ exponential(1), the transformation approach only outperformed PLAC when the censoring rate was small to moderate (Web Supplementary, Table

B.2). When the censoring rate increased, the transformation approach suffered from large biases and its mean squared errors would be larger than those of the PLAC estimator.

When we combined $\mathcal{L}_n^C$ and $\mathcal{L}_n^P$ in the composite log-likelihood, we used the weights proportional to the reciprocals of their magnitudes (number of terms), which may not be optimal, and further investigation is needed. In the context of additive hazards model, Huang and Qin (2013) studied the optimal weights with which the resulting estimator would have the smallest variance. Their simulation showed that the estimator using the optimal weights was less efficient compared to the estimator using the reciprocals of the magnitudes as weights. They discussed it was because that the optimal weights involves estimation of the variances of the scores, which requires larger sample sizes to obtain the benefit.

Lastly, while the proposed estimator focuses on handling time-independent covariates, the extension to time-dependent covariates is promising based on our preliminary work. We expect to derive asymptotic properties and devote more effort to reducing computation time, which is magnified by the need of expanding the dataset with the time-dependent covariates.

6. Supplementary Materials

Web Appendices, Tables and Figures referenced in Sections 2–5, and an R package to implement the PLAC method are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgments

The authors thank the editor, the associate editor, and two anonymous referees for their constructive comments that improved the presentation of the paper.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10,** 1100–1120.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **34,** 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

De la Peña, V. and Giné, E. (1999). *Decoupling: from dependence to independence*. Springer, New York, NY.

Huang, C.-Y. and Qin, J. (2012). Composite partial likelihood estimation under length-biased sampling, with application to a prevalent cohort study of dementia. *Journal of the American Statistical Association* **107,** 946–957.

Huang, C.-Y. and Qin, J. (2013). Semiparametric estimation for the additive hazards model with left-truncated and right-censored data. *Biometrika* **100,** 877–888.

Huang, C.-Y., Qin, J., and Follmann, D. A. (2012). A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. *Biometrika* **99,** 199–210.

Jung, S.-H. (1999). Rank tests for matched survival data. *Lifetime Data Analysis* **5,** 67–79.

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

Liang, K.-Y. and Qin, J. (2000). Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 773–786.

Liu, H., Ning, J., Qin, J., and Shen, Y. (2016). Semiparametric maximum likelihood inference for truncated or biased-sampling data. *Statistica Sinica* **26,** 1087–1115.

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

Murphy, S., Rossini, A., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association* **92,** 968–976.

Ning, J., Qin, J., and Shen, Y. (2014). Semiparametric accelerated failure time model for length-biased data with application to dementia study. *Statistica Sinica* **24,** 313–333.

Perlman, R. L., Kiser, M., Finkelstein, F., Eisele, G., Roys, E., Liu, L., Burrows-Hudson, S., Port, F., Messana, J. M., Bailie, G., et al. (2003). Renal Research Institute symposium: The longitudinal chronic kidney disease study: A prospective cohort study of predialysis renal failure. *Seminars in Dialysis* **16,** 418–423.

Qin, J. and Liang, K. (1999). Generalized odds ratio model and pairwise conditional likelihood. Technical report, Johns Hopkins University, Baltimore, MD.

Qin, J., Ning, J., Liu, H., and Shen, Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association* **106,** 1434–1449.

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

Sen, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statist. Assoc. Bull* **10,** 18.

Shen, Y., Ning, J., and Qin, J. (2017). Nonparametric and semiparametric regression estimation for length-biased survival data. *Lifetime data analysis* **23,** 3–24.

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York, NY.

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

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21,** 5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection.

Biometrika **92,** 519–528.

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

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93,** 627–640.

Received October 2016. *Revised June* 2017. *Accepted June* 2017.

Appendix

The asymptotic properties of the PLAC estimator are proved under the following regularity conditions.

(C1) The true regression coefficient vector β_0 lies in the interior of a compact set $B \subset \mathbb{R}^p$. The true cumulative baseline hazard function $\Lambda_0(t)$ is continuously differentiable and strictly increasing on $[0, \tau]$, and satisfies $\Lambda_0(0) = 0$.

(C2) The vector Z is bounded almost surely. If there exist a deterministic function $\gamma_0(t)$ and $\gamma \in \mathbb{R}^p$, such that $\gamma_0(t) + \gamma^{\mathrm{T}} Z = 0$ with probability one, then $\gamma_0(t) = 0$ and $\gamma = 0$.

(C3) With probability one, there exists a constant $\delta_1 > 0$ such that $\mathrm{pr}(A^* \leqslant T^* \mid Z^*) > \delta_1$, $\mathrm{pr}(A + C \geqslant \tau \mid Z) > \delta_1$, and $\mathrm{pr}(\bar{Y}(\tau) = 1 \mid Z) > \delta_1$, where $\bar{Y}(\tau) = 1$ implies $Y(t) = 1$ for all $t \in [A, \tau]$.

(C4) Let $b \in \mathbb{R}^p$, and h be a function with bounded total variation on $[0, \tau]$, then the information operator corresponding to the conditional likelihood evaluated at (β_0, Λ_0), $J_0^C(b, h) = \left(\lim_{n\to\infty} \partial U^C(\beta, \Lambda)/\partial(\beta, \Lambda)\big|_{\beta=\beta_0, \Lambda=\Lambda_0} \right)(b, h)$, is invertible.

If $\mathrm{pr}(T^* < A^* \mid Z^*) > \delta_2 > 0$, then $\mathcal{L}_n^P$ is non-degenerate, so that we can attain efficiency gain beyond the conditional approach. When this condition does not hold, $\mathcal{L}_n^P$ is zero, and thus the PLAC estimator becomes identical to the conditional approach estimator.

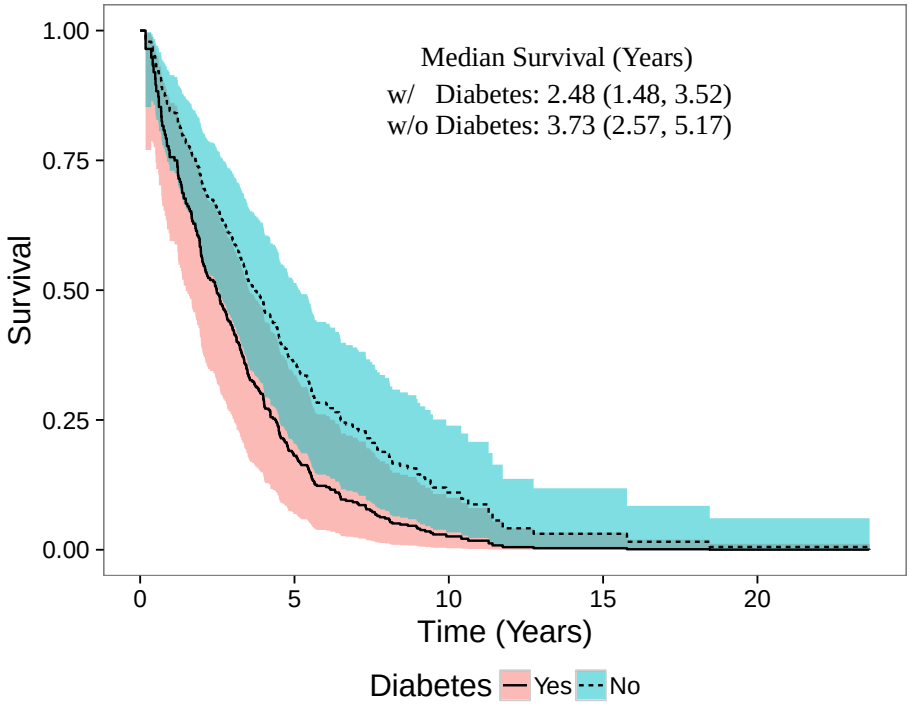

Figure 1. Estimated survival curves of patients with diabetes (solid) or without diabetes (dashed) at referral using the proposed method (PLAC). 95% point-wise confidence intervals are shown as shaded areas. The estimated median survival times for both groups are displayed with the corresponding 95% confidence intervals. The other covariates are set to their reference levels. (This figure appears in color in the electronic version of this article.)

Table 1

Summary from 1000 simulated datasets. PC: censoring percentage; True: true values; Bias, SE, SEE and CP: empirical bias ($\times 10^3$), standard error ($\times 10^3$), standard error estimate ($\times 10^3$) and 95% coverage probability; RE: relative efficiency with respect to the conditional approach estimator (ratio of the mean squared errors). The estimate of $\hat{\Lambda}(t)$ is evaluated at the 30% and 60% percentiles (τ_{30} and τ_{60}) of the observed survival times.

n	PC		Conditional			LBML			PLAC				
			True	Bias	SE	Bias	SE	RE	Bias	SE	SEE	CP	RE
						Case 1: length-biased sampling							
400	50	$\hat{\beta}_1$	1	5	169	−46	115	1.85	10	129	125	94	1.71
		$\hat{\beta}_2$	1	5	150	−49	109	1.60	9	118	113	94	1.60
		$\hat{\Lambda}_{\tau_{30}}$	0.212	1	45	18	40	1.07	−1	41	39	92	1.18
		$\hat{\Lambda}_{\tau_{60}}$	0.546	2	91	42	78	1.04	−1	84	79	93	1.17
	80	$\hat{\beta}_1$	1	24	265	−61	141	2.99	30	169	166	95	2.39
		$\hat{\beta}_2$	1	23	241	−67	133	2.63	30	162	152	94	2.15
		$\hat{\Lambda}_{\tau_{30}}$	0.103	0	37	30	35	0.62	−2	32	30	91	1.31
		$\hat{\Lambda}_{\tau_{60}}$	0.329	−2	92	49	71	1.14	−7	79	76	92	1.37
800	50	$\hat{\beta}_1$	1	8	116	-35	82	1.71	6	89	88	95	1.71
		$\hat{\beta}_2$	1	0	99	-38	75	1.41	4	81	80	94	1.49
		$\hat{\Lambda}_{\tau_{30}}$	0.212	1	30	18	30	0.75	0	28	28	95	1.18
		$\hat{\Lambda}_{\tau_{60}}$	0.546	1	62	36	54	0.90	1	56	56	94	1.22
	80	$\hat{\beta}_1$	1	10	194	-49	101	3.00	14	121	116	94	2.54
		$\hat{\beta}_2$	1	12	203	-48	101	3.31	16	122	116	94	2.73
		$\hat{\Lambda}_{\tau_{30}}$	0.103	−1	31	43	39	0.28	−1	30	30	93	1.08
		$\hat{\Lambda}_{\tau_{60}}$	0.329	−4	62	51	62	0.60	−4	58	58	94	1.15
						Case 2: non-length-biased sampling							
400	50	$\hat{\beta}_1$	1	3	150	−243	103	0.32	3	128	129	94	1.38
		$\hat{\beta}_2$	1	13	157	−232	105	0.38	18	134	129	94	1.36
		$\hat{\Lambda}_{\tau_{30}}$	0.207	−2	40	105	51	0.11	−2	39	38	94	1.02
		$\hat{\Lambda}_{\tau_{60}}$	0.538	−2	65	233	77	0.07	−3	64	64	94	1.01
	80	$\hat{\beta}_1$	1	11	262	−359	117	0.48	27	185	181	95	1.97
		$\hat{\beta}_2$	1	11	260	−364	122	0.46	19	194	181	93	1.78
		$\hat{\Lambda}_{\tau_{30}}$	0.099	−2	34	106	56	0.08	−3	33	31	91	1.04
		$\hat{\Lambda}_{\tau_{60}}$	0.270	−4	61	221	85	0.07	−5	59	58	93	1.05
800	50	$\hat{\beta}_1$	1	−1	107	−227	72	0.20	2	90	91	95	1.44
		$\hat{\beta}_2$	1	2	107	−227	73	0.20	3	92	91	96	1.36
		$\hat{\Lambda}_{\tau_{30}}$	0.207	−1	28	110	40	0.06	−1	27	27	96	1.04
		$\hat{\Lambda}_{\tau_{60}}$	0.538	−1	46	230	56	0.04	−1	45	45	95	1.04
	80	$\hat{\beta}_1$	1	7	176	−347	89	0.24	15	136	130	93	1.66
		$\hat{\beta}_2$	1	−1	179	−348	88	0.25	13	135	129	93	1.75
		$\hat{\Lambda}_{\tau_{30}}$	0.099	−1	25	112	55	0.04	−1	24	23	93	1.02
		$\hat{\Lambda}_{\tau_{60}}$	0.270	−1	44	233	71	0.03	−3	43	42	93	1.02

Table 2
Coefficient estimates from the RRI-CKD data using the conditional approach and the proposed method. LHR: log hazards ratio (β); SE: standard error; p: p-value.

	Conditional			PLAC		
	LHR	SE	p	LHR	SE	p
Older than 65	0.093	0.129	0.473	0.113	0.111	0.311
Male	0.517	0.131	<.001	0.422	0.113	<.001
Non-white	0.213	0.161	0.185	0.262	0.130	0.045
Diabetes	0.424	0.130	0.001	0.507	0.110	<.001
Hypertension	0.168	0.225	0.455	0.075	0.189	0.693
Late-Stage	0.950	0.146	<.001	1.020	0.128	<.001