Statistics and Its Interface Volume 9 (2016) 213–222

Semiparametric transformation models with length-biased and right-censored data under the case-cohort design

Huijuan Ma[*], Zhiping Qiu, and Yong Zhou

Case-cohort designs provide a cost effective way in large cohort studies. Semiparametric transformation models, which include the proportional hazards model and the proportional odds model as special cases, are considered here for length-biased right-censored data under case-cohort design. Weighted estimating equations, which can be used even when the censoring variables are dependent of the covariates, are proposed for simultaneous estimation of the regression parameters and the transformation function. The resulting regression estimators are shown to be asymptotically normal with a closed form of variance-covariance matrix and can be estimated by the plug-in method. Simulation studies show that the proposed approach performs well for practical use. An application to the Oscar data is also given to illustrate the methodology.

Keywords and phrases: Case-cohort design, Length-biased and right-censored data, Mean zero process, Transformation model, Weighted estimating equation.

1. INTRODUCTION

Prentice (1986) proposed a case-cohort design to reduce the cost in large cohort studies, where much of the covariate information on free subjects is largely redundant. Under this design, a random sample named subcohort is selected from the full cohort. Covariate information is collected only for the subjects in the subcohort and any cases who experience the event of interest.

After the creative work of Prentice (1986) under the proportional hazards model (Cox, 1972), Self and Prentice (1988) and Lin and Ying (1993) have made further developments. However, the proportion hazards assumption may not always be true in some applications. Or we may be interested in modeling the association between the failure time and the covariate from different aspects. Many authors studied other regression models such as additive hazards models (Kulich and Lin, 2000), proportional odds models (Chen, 2001b) and semiparametric transformation regression models (Chen, 2001a; Kong et al., 2004; Lu and Tsiatis, 2006). The semiparametric transformation model is specified by

$$H(T) = -\boldsymbol{\beta}'\mathbf{Z} + \varepsilon, \tag{1}$$

where H is an unknown monotone transformation function, ε is a random variable with a known distribution and is independent of $\mathbf{Z}$, $\boldsymbol{\beta}$ is an unknown p-dimensional regression parameter of interest. The proportional hazards model and the proportional odds model are special cases of (1) with ε following the extreme-value distribution and the standard logistic distribution, respectively. Equivalently, the models can be represented by

$$g\{S_z(t)\} = H(t) + \boldsymbol{\beta}'\mathbf{Z},$$

where $S_z(t)$ is the survival function for given the covariate vector $\mathbf{Z}$, g is a known transformation. For the Cox proportional hazards model, the link function is $g(\cdot) = \log\{-\log(\cdot)\}$, and for the proportional odds model, $g(\cdot) = -\text{logit}(\cdot)$.

To date, these studies have all involved right censored failure time data. In observational studies, such as studies of unemployment duration in the labor economy (Lancaster, 1992; de Una-Alvarez et al., 2003), cancer screening trials (Zelen and Feinleib, 1969; Simon, 1980), and the HIV prevalent cohort studies (Lagakos et al., 1988), one often encounters length-biased right-censored data. Length-biased sampling is a special case of left truncation, which assumes that the incidence of the disease onset follows a Poisson process (Zelen and Feinleib, 1969; Simon, 1980), and hence the probability of a survival time being sampled is proportional to its length. Recently, Tsai (2009), Qin and Shen (2010) and Huang and Qin (2012) have proposed semiparametric estimation under length-biased sampling of Cox model. Shen et al. (2009) studied the semiparametric transformation model and the accelerated failure time model. There have no study in the regression model for length-biased right-censored data under the case-cohort design.

In this article, we study the semiparametric analysis of transformation models with length-biased right-censored data under the case-cohort design. We propose mean zero estimating equations to estimate regression parameters and the monotone transformation function. Our approach is motivated by Lu and Tsiatis (2006), which make use of a mar-

[*] The corresponding author.

tingale integral representation and the inverse weighted selected probabilities. The main challenge here is that even though we can make use of a martingale integral representation by accommodating left-truncation, the resulting estimators are not fully efficient under length-biased sampling.

This article is organized as follows. In Section 2, we introduce some notation, the proposed estimating methods and the computational algorithms. The asymptotic distribution results are shown in Section 3, and the outline of their proofs is presented in the Appendix. Section 4 is devoted to simulation studies to examine the finite sample properties of the regression parameter estimators. In Section 5, the Oscar data is used to illustrate the estimating procedure. Section 6 contains a brief discussion.

2. METHODOLOGY

2.1 Notation

Let T^0 denote the time from the onset to the failure event of interest, and let A^0 the time between the onset and study enrollment. Here we assume that T^0 satisfies the transformation model which is specified through $H(T^0) = -\boldsymbol{\beta}'\mathbf{Z}+\varepsilon$, where the hazard function and cumulative hazard function of ε is denoted by $\lambda(t)$ and $\Lambda(t)$, respectively. In a length-biased sampling, a subject would be sampled only if the failure event does not occur before the sampling time, that is, T^0 is left truncated by A^0. Denote by T, A and $\mathbf{Z}$ the survival time, truncation time, and the covariates for individuals in the prevalent cohort. Then (T, A) has the same joint distributions as $(T^0, A^0)|T^0 \geq A^0$. Write $V = T - A$, thus V denotes the residual lifetime after enrollment. The observation of the survival time in the prevalent cohort is usually subject to right censoring due to study end or premature dropout. Instead of observing the actual value of T, we observe the censored survival time $Y = \min(T, A + C)$. Let $\tilde{V} = \min(T - A, C)$ be the observed residual lifetime, that is, $\tilde{V} = V$ for uncensored subjects and $\tilde{V} = C$ for censored subjects. We assume that the censoring time after enrollment C is independent of (T, A) given $\mathbf{Z}$, which is reasonable in many applications. However, that the survival time T and the total censoring time $A + C$ are dependent, as they share the same A.

To simplify the presentation, we will focus on the classical case-cohort design of Prentice (1986), in which the subcohort is a simple random sample from the full cohort. With slight modifications the methodology can also be applied to the stratified case-cohort design of Borgan et al. (2000), where the sampling of the subcohort can be stratified. Let n and $\tilde{n}$ denote the number of subjects in the full cohort and in the subcohort respectively. Under the classical case-cohort design, we observe $\{(Y_i, A_i, \delta_i), i = 1, 2, \cdots, n\}$ for all individuals in the full cohort, and $\mathbf{Z}_i$ only for subjects in the subcohort and all cases. Let ξ_i be the subcohort indicator, taking the value 1 or 0, whether the subject is included in the subcohort or not. Hence the data are summarised as $\{(Y_i, A_i, \delta_i, [\delta_i + (1-\delta_i)\xi_i]\mathbf{Z}_i,), i = 1, 2, \cdots, n\}$. Here ξ_i is independent of $(Y_i, A_i, \mathbf{Z}_i), i = 1, 2, \cdots, n$, while the $\{\xi_i, i = 1, 2, \cdots, n\}$ are dependent because of the sampling without replacement. Let p denote the probability of a subject being sampled into the subcohort. Given the failure status δ_i, the probability of subject i being selected into the case-cohort sample is $\delta_i + (1 - \delta_i)p$. We can define a weight $\rho_i = \delta_i + (1 - \delta_i)\xi_i/p$, for each individual in the full cohort by the inverse selection probabilities. In view of the results in Robins et al. (1994), we can replace p in the definition of ρ_i with its empirical estimate $\hat{p} = \tilde{n}/n$. Here $\pi_i = \delta_i + (1 - \delta_i)\xi_i/\hat{p}$ is denoted so as to exploit the information available in the full-cohort data.

2.2 Estimating methods

The method of Lu and Tsiatis (2006) can be modified to accommodate left-truncation. Let $N_i(t) = I(Y_i \leq t, \delta_i = 1)$ be the indicator of whether or not the ith individual failed before time t and $Y_i^a(t) = I(A_i \leq t \leq Y_i)$ be the indicator of whether or not the ith individual is at risk just before time t. As shown by Andersen et al. (1993), the counting process $N_i(\cdot)$ can be uniquely decomposed so that for every i and t,

$$N_i(t) \quad = \quad M_{1i}(t) + \int_0^t Y_i^a(u) d\Lambda_{T^0}(u|\mathbf{Z}_i). \tag{2}$$

where $M_{1i}(t)$ is a local square martingale, $\Lambda_{T^0}(u|\mathbf{Z})$ is the cumulative hazard function of T^0 given the covariate $\mathbf{Z}$.

Equation (2) leads to the following estimating equations:

$$\sum_{i=1}^n \int_0^\tau \pi_i \mathbf{Z}_i \left[dN_i(t) - Y_i^a(t) d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i\}\right] = 0, \tag{3}$$

$$\sum_{i=1}^n \pi_i \left[dN_i(t) - Y_i^a(t) d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i\}\right] = 0, \quad (t \geq 0) \tag{4}$$

where H is a nondecreasing function satisfying $H(0) = -\infty$ and τ is a prespecified constant such that $\Pr(\tilde{V} \geq \tau) > 0$. In fact, the estimating equations (3) and (4) can be used to analyze the general left-truncated and right-censored data for semiparametric transformation model under the case-cohort design. When using them to length-biased right-censored data, the resulting estimators are still consistent and have weak convergency properties. But they are inefficient because they do not exploit the special property of length-biased sampling.

The conditional density function and survival function of the survival time T^0 given $\mathbf{Z} = \mathbf{z}$ are denoted by $f(t|\mathbf{z})$ and $S(t|\mathbf{z})$. Let $G(t|\mathbf{z})$ denote the survival function of C given $\mathbf{Z} = \mathbf{z}$. Under length-biased sampling, the truncation variable A^0 follows a uniform distribution and the joint density function of (T, A) given $\mathbf{Z} = \mathbf{z}$ evaluated at (t, a) is

$$\frac{f(t|\mathbf{z})}{\mu(\mathbf{z})} I(t \geq a) \tag{5}$$

(Lancaster, 1992), where $\mu(\mathbf{z}) = \int u f(u|\mathbf{z}) du$ is the conditional mean of T^0 given $\mathbf{Z} = \mathbf{z}$. In the absence of censoring, it follows from (5) that the pair of random variables (A, V) has an exchangeable joint density function $f(a+v|\mathbf{z})/\mu(\mathbf{z})$ for $a \geq 0$ and $v \geq 0$, and the common marginal density function is $f_A(t|\mathbf{z}) = f_V(t|\mathbf{z}) = S(t|\mathbf{z})/\mu(\mathbf{z})$. In the presence of right censoring, the joint density function of $(A, \tilde{V})$ conditioned on $\delta = 1$ is

$$\begin{aligned} &(A = a, \tilde{V} = v)|(\delta = 1, \mathbf{Z} = \mathbf{z}) \\ \sim\ &\frac{f(a+v|\mathbf{z})}{\mu(\mathbf{z})} \frac{G(v|\mathbf{z})}{\Pr(\delta = 1|\mathbf{Z} = \mathbf{z})}, \quad a \geq 0,\ v \geq 0. \end{aligned}$$

That is, A and $\tilde{V}$ do not have an exchangeable joint distribution despite conditioning on $\delta = 1$. Interestingly, we have

$$\begin{aligned} &A = a|(\delta = 1, \tilde{V} = v, \mathbf{Z} = \mathbf{z}) \\ \sim\ &\frac{f(a+v|\mathbf{z})}{S(v|\mathbf{z})}, \quad a \geq 0,\ v \geq 0, \end{aligned}$$

which is identical to the conditional density function of $V = T - A$ given A in the prevalent population.

We then use this property to propose new mean-zero estimating equations to add efficiency. Based on this, we can use the mean-zero stochastic process $M_{2i}(t)$, where

$$M_{2i}(t) = N_i(t) - \int_0^t \delta_i I(\tilde{V}_i \leq t \leq Y_i) d\Lambda_{T^0}(u|\mathbf{Z}_i) du.$$

In order to express briefly, define $M_i(t) = [M_{1i}(t) + M_{2i}(t)]/2 = N_i(t) - \int_0^t R_i(u) d\Lambda_{T^0}(u|\mathbf{Z}_i) du$, where $R_i(u) = \frac{1}{2}\{I(A_i \leq u \leq Y_i) + \delta_i I(\tilde{V}_i \leq u \leq Y_i)\}$.

Lemma 1. *Under length-biased sampling, we have* $E[dM_1(t)] = 0$.

A proof of Lemma 1 is given in the Appendix. And it leads to the following two estimating equations:

$$\begin{aligned} &U(\boldsymbol{\beta}, H) = \\ (6)\quad &\sum_{i=1}^n \int_0^\tau \pi_i \mathbf{Z}_i \left[dN_i(t) - R_i(t) d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i\} \right] = 0, \\ (7)\quad &\sum_{i=1}^n \pi_i \left[dN_i(t) - R_i(t) d\Lambda\{H(t) + \boldsymbol{\beta}'\mathbf{Z}_i\} \right] = 0. \quad (t \geq 0). \end{aligned}$$

They are analogous to the estimating equations derived by Lu and Tsiatis (2006). The requirement $H(0) = -\infty$ ensures that $\Lambda(a + H(0)) = 0$ for any finite a. Let $\mathcal{H}$ be the collection of all nondecreasing step functions on $[0, \infty)$ with $H(0) = -\infty$ and with jumps only at the observed failure times $t_1, t_2, \cdots, t_k$. We denote by $(\hat{\boldsymbol{\beta}}, \hat{H})$ the solution of (6) – (7). It is then clear that $\hat{H} \in \mathcal{H}$.

There are some alternative versions of (7) that are simple for computational purposes (Chen et al., 2002). Note that (7) can be rewritten as:

$$\begin{pmatrix} 1 - \sum_{i=1}^n \pi_i R_i(t_1)[\Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_1)) - \Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_1-))] \\ 1 - \sum_{i=1}^n \pi_i R_i(t_2)[\Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_2)) - \Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_2-))] \\ \vdots \\ 1 - \sum_{i=1}^n \pi_i R_i(t_k)[\Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_k)) - \Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_k-))] \end{pmatrix}$$

$$(8) \qquad = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

with $H \in \mathcal{H}$, implying $H(t_1-) = -\infty$. Slightly differently from (8), one might also consider the following computational simpler estimating equations:

$$\begin{pmatrix} 1 - \sum_{i=1}^n \pi_i R_i(t_1)\Lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_1)) \\ 1 - \sum_{i=1}^n \pi_i R_i(t_2)\lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_2-))\Delta H(t_2) \\ \vdots \\ 1 - \sum_{i=1}^n \pi_i R_i(t_k)\lambda(\boldsymbol{\beta}'\mathbf{Z}_i + H(t_k-))\Delta H(t_k) \end{pmatrix}$$

$$(9) \qquad = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $H \in \mathcal{H}$ and $\Delta H(t) = H(t) - H(t-)$.

2.3 Computational algorithms

Following Chen et al. (2002), the equations (6) and (7) naturally suggest the following iterative algorithm for computing $(\hat{\boldsymbol{\beta}}, \hat{H})$.

Step 0: Choose an initial value of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}^{(0)}$.

Step 1: Obtain $H^{(0)}$ as follows. First obtain $H^{(0)}(t_1)$ by solving

$$\sum_{i=1}^n \pi_i R_i(t_1)\Lambda\{\boldsymbol{\beta}'\mathbf{Z}_i + H(t_1)\} = 1,$$

with $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$. Then, obtain $H^{(0)}(t_k), k = 2, 3, \cdots, K$, one-by-one by

$$\begin{aligned} &H^{(0)}(t_k) \\ &= H^{(0)}(t_{k-1}) + \frac{1}{\sum_{i=1}^n \pi_i R_i(t_k)\lambda\{\boldsymbol{\beta}'\mathbf{Z}_i + H^{(0)}(t_{k-1})\}} \end{aligned}$$

with $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$.

Step 2: Obtain new estimate of $\boldsymbol{\beta}$ by solving (6) with $H = H^{(0)}$.

Step 3: Set $\boldsymbol{\beta}^{(0)}$ to be the estimate obtained in Step 2 and repeat Steps 1 and 2 until prescribed convergence criteria are met.

3. THEORETICAL RESULTS

To identify the limiting distributions of the estimators, we define the following terms:

$$
\begin{aligned}
&B_1(t) = \mathsf{E}[\dot{\lambda}\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)],\\
&B_1^Z(t) = \mathsf{E}[\mathbf{Z}\dot{\lambda}\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)],\\
&B_2(t) = \mathsf{E}[\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)],\\
&B_2^Z(t) = \mathsf{E}[\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)],\\
&B(t,s) = \exp\left\{\int_s^t B_2(u)^{-1}B_1(u)dH_0(u)\right\},\\
&B_3(t,s) = \mathsf{E}[\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(s)\}I(s \geq t)B(t,s)],\\
&B_4(t,s) = \mathsf{E}[\delta\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(s)\}I(s \geq t)B(t,s)],\\
&\mu_z(t) = \frac{1}{2B_2(t)}\\
&\qquad \left[B_3(t,Y) - B_3(t,A) + B_4(t,Y) - B_4(t,\tilde{V})\right].
\end{aligned}
$$

Also define

$$
\begin{aligned}
A &= \int_0^\tau \mathsf{E}\left[\{\mathbf{Z} - \mu_z(t)\}\mathbf{Z}'\dot{\lambda}\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)\right]dH_0(t),\\
\Sigma &= \mathsf{E}\left\{\{\delta + (1-\delta)/p\}\left[\int_0^\tau \{\mathbf{Z} - \mu_z(t)\}dM(t)\right]^{\otimes 2}\right\}\\
&\quad -\frac{1-p}{p}\left\{\mathsf{E}\left[\delta\int_0^\tau \{\mathbf{Z} - \mu_z(t)\}dM(t)\right]\right\}^{\otimes 2}.
\end{aligned}
$$

We need to impose the following regularity conditions:

C1 For any finite K, $\lambda(x)$ is strictly positive and $\dot{\lambda}(x)$ is bounded and continuously differentiable on $(-\infty, K)$, where the superscript dot always denote derivatives.

C2 The covariate vector $\mathbf{Z}$ is bounded in the sense that $\Pr(\|\mathbf{Z}\| < M) = 1$ for some constant $M > 0$.

C3 The true transformation function H_0 has a continuous and positive derivative on $[0, \tau]$.

C4 The matrix A is nonsingular.

Remark: Condition C1 is a mild condition and is satisfied in commonly encountered transformation models. Condition C2 is imposed so that modern empirical process theory can be directly used. Condition C4 is necessary since otherwise the problem becomes singular.

Theorem 2. *Under regularity conditions C1–C4, $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_0$ in probability as $n \to \infty$. Furthermore, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(0, A^{-1}\Sigma(A^{-1})')$, and $n^{1/2}(\hat{H}(t) - H_0(t))$ converges weakly to a Gaussian process. Moreover, A and Σ can be consistently estimated by*

$$
\begin{aligned}
\hat{A} &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \pi_i\{\mathbf{Z}_i - \bar{\mathbf{Z}}(t)\}\mathbf{Z}_i'\dot{\lambda}\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(t)\}R_i(t)d\hat{H}(t),\\
\hat{\Sigma} &= \frac{1}{n}\sum_{i=1}^n \pi_i^2\left[\int_0^\tau \{\mathbf{Z}_i - \bar{\mathbf{Z}}(t)\}d\hat{M}_i(t)\right]^{\otimes 2}\\
&\quad -\frac{1-\hat{p}}{\hat{p}}\left[\frac{1}{n}\sum_{i=1}^n \int_0^\tau \delta_i\{\mathbf{Z}_i - \bar{\mathbf{Z}}(t)\}d\hat{M}_i(t)\right]^{\otimes 2},
\end{aligned}
$$

respectively, where

$$
\begin{aligned}
\bar{\mathbf{Z}}(t) = \frac{1}{2}\Bigg[&\frac{\sum_{i=1}^n \pi_i\mathbf{Z}_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(Y_i)\}I(Y_i \geq t)\hat{B}(t,Y_i)}{\sum_{i=1}^n \pi_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(t)\}R_i(t)}\\
&- \frac{\sum_{i=1}^n \pi_i\mathbf{Z}_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(A_i)\}I(A_i \geq t)\hat{B}(t,A_i)}{\sum_{i=1}^n \pi_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(t)\}R_i(t)}\\
&+ \frac{\sum_{i=1}^n \delta_i\mathbf{Z}_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(Y_i)\}I(Y_i \geq t)\hat{B}(t,Y_i)}{\sum_{i=1}^n \pi_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(t)\}R_i(t)}\\
&- \frac{\sum_{i=1}^n \delta_i\mathbf{Z}_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(\tilde{V}_i)\}I(\tilde{V}_i \geq t)\hat{B}(t,\tilde{V}_i)}{\sum_{i=1}^n \pi_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(t)\}R_i(t)}\Bigg],\\
\hat{B}(t,s) = \exp&\left(\int_s^t \frac{\sum_{i=1}^n \pi_i\dot{\lambda}\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(x)\}R_i(x)}{\sum_{i=1}^n \pi_i\lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(x)\}R_i(x)}d\hat{H}(x)\right),\\
\hat{M}_i(t) = N_i(t) &- \int_0^t R_i(s)d\Lambda\{\hat{\boldsymbol{\beta}}'\mathbf{Z}_i + \hat{H}(s)\}.
\end{aligned}
$$

for $t, s \in (0, \tau]$.

4. SIMULATION STUDIES

We conduct simulation studies to examine the finite sample performance of the proposed methodology. We present two simulation studies. In both examples, let the hazard function of ε be of the form $\lambda(t) = \exp(t)/\{1 + v\exp(t)\}$, with $v = 0, 1, 2$ (Dabrowska and Doksum, 1988). Note that the proportional hazards and the proportional odds models correspond to $v = 0$ and $v = 1$, respectively. The transformation function $H(t)$ is chosen as $\log(t)$ for $v = 0$, and $\log(\frac{1}{v}e^{vt} - 1)$ for $v \neq 0$. We set the sampling time ξ to be 100 and simulate the onset time of a stable disease, W^0, from a uniform distribution over $[0, 100]$. Two covariates of Z_1 and Z_2 are chosen to be independent of each other with Z_1 following the standard normal distribution and Z_2 taking values 0 or 1 with equal probability 0.5.

The first example focuses on covariate independent censoring. The regression parameter is $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (1, 1)'$ and the censoring times are generated from a $U(0, c)$ distribution, where c is chosen such that the expected proportion of censoring was 80%. Five hundred full cohorts with sample size 1, 000 are simulated and then two case-cohort samples are selected from each full cohort data set by simple random sampling without replacement. The first case-cohort study is designed to have the same number of controls and cases while the second case-cohort study is designed to have twice as many controls as cases. Simulation results are summarised in Table 1, where CCI and CCII are the first and second case-cohort design, FULL is the full cohort design. Bias and SD are the empirical bias and empirical standard deviation of 500 regression estimates, respectively. SE is the

Table 1. Simulation results when censoring is covariate independent

	$\boldsymbol{\beta}_1 = 1$			$\boldsymbol{\beta}_2 = 1$		
Study Design	CCI	CCII	FULL	CCI	CCII	FULL
Scenario I : $v = 0$						
BIAS	0.014	0.005	0.004	−0.041	−0.047	−0.049
SD	0.100	0.090	0.086	0.144	0.125	0.114
SE	0.119	0.105	0.094	0.156	0.163	0.126
CP	97.2	96.8	95.6	96.8	98.6	95.0
RE	0.74	0.91	1	0.63	0.83	1
Scenario II : $v = 1$						
BIAS	0.019	0.007	0.001	−0.047	−0.046	−0.051
SD	0.194	0.167	0.156	0.309	0.268	0.248
SE	0.169	0.149	0.135	0.310	0.274	0.251
CP	92.0	93.2	93.0	95.6	95.6	95.2
RE	0.65	0.87	1	0.64	0.86	1
Scenario III : $v = 2$						
BIAS	−0.001	−0.002	−0.008	−0.006	−0.039	−0.038
SD	0.293	0.259	0.247	0.551	0.494	0.458
SE	0.278	0.237	0.230	0.524	0.464	0.433
CP	93.2	94.2	93.4	94.2	94.4	94.0
RE	0.71	0.90	1	0.69	0.86	1

BIAS, the empirical bias; SD, the empirical standard deviation; SE, the mean of estimated standard error; CP, the empirical coverage probability of 95% confidence interval. RE, empirical relative efficiencies of our estimators, which is the ratio of sample variance of our estimators with the full cohort design being a reference.

averaged robust standard error estimator. RE is the empirical relative efficiency of our estimator, calculated by the ratio of sample variance of our estimator with the full cohort design being the reference. CP is the empirical coverage probability of 95% confidence interval. For all three hazard functions, our proposed estimators CCI and CCII have similar asymptotic results as the FULL estimator in terms of Bias, SE and CP. It is easy to see from Table 1 that the SEs of the proposed estimators are close to the corresponding SDs, and CP approximates the nominal 95% confidence interval. More importantly, CCI and CCII estimators do not lose too much efficiency in comparison to the FULL estimator according to the empirical relative efficiency RE values. This implies that our proposed estimators perform very well.

The second set of simulation studies are conducted for covariate dependent censoring. The settings are the same as the first one except that the censoring times are generated from $\beta_c Z_2 + U[0, c]$, β_c is selected as 0.5 and 0.2 respectively and c is chosen such that the expected proportion of censoring is 0.75. Here only the first case-cohort design is used. Simulation results are summarised in Table 2 which shows that the estimators from the proposed method still performs very well as the Table 1 when censoring is covariate dependent.

5. A REAL DATA EXAMPLE

In this Section, we analyze the Oscar Awards data using our proposed method. The Oscar Awards are the most prominent and most watched film awards ceremony in the world. They are presented annually by the Academy of Motion Pictures Arts and Sciences. The detailed description of the Oscar Awards data and the website where we can download this dataset could be found in Han et al. (2011), so we omit here. In the Oscar dataset, there are 766 people who have been nominated, and only 327 died before the study ended. This means that the censoring ratio is about 57.3%.

Several authors (Redelmeier and Singh, 2001; Sylvestre et al., 2006; Han et al., 2011) have studied whether winning an Oscar Award causes the actor's/actress's expected lifetime to increase. They used different statistical methods but all view this dataset as a right-censored survival dataset. Redelmeier and Singh (2001) fitted a Cox proportional hazards model, where whether a performer ever won an Oscar Award in his or her lifetime was treated as a time-independent covariate and survival was measured from the performer's date of birth. They stated that life expectancy was 3.9 years longer for Oscar Award winners than for other less recognized performers. Sylvestre et al. (2006) pointed out that this analysis suffers from immortal time bias. In other words, performers who live longer have more opportunities to win Oscar Awards. To eliminate immortal time bias, Sylvestre et al. (2006) fitted a Cox proportional hazards model with winning status treated as a time-dependent covariate and survival measured from a performer's date of first nomination. Sylvestre et al. (2006) estimated that winning an Oscar Award had a positive effect on lifetime, but the estimated effect was not significant. Although a valuable step forward, Han et al.(2011) pointed that Sylvestre et al.'s analysis still suffers from healthy performer survivor bias:

Table 2. Summary of simulation results when censoring is covariate dependent

		$\beta_1 = 1$				$\beta_2 = 1$			
Model	β_c	Bias	SD	SE	CP	Bias	SD	SE	CP
$v = 0$	0.5	0.014	0.088	0.098	97.0	−0.026	0.113	0.124	96.4
	0.2	0.012	0.087	0.096	97.0	−0.028	0.124	0.137	96.6
$v = 1$	0.5	−0.007	0.174	0.158	93.4	−0.037	0.311	0.290	93.0
	0.2	0.010	0.162	0.157	94.6	−0.024	0.284	0.281	94.6
$v = 2$	0.5	−0.020	0.260	0.245	94.2	−0.048	0.442	0.449	95.8
	0.2	0.021	0.309	0.262	95.6	−0.053	0.609	0.493	95.4
		$\beta_1 = 0$				$\beta_2 = 0$			
$v = 0$	0.5	0.001	0.055	0.054	95.0	−0.067	0.144	0.154	95.0
	0.2	−0.002	0.056	0.053	93.2	−0.020	0.108	0.112	95.8
$v = 1$	0.5	0.001	0.108	0.109	95.2	−0.055	0.232	0.227	94.2
	0.2	−0.004	0.129	0.124	93.4	−0.019	0.247	0.249	95.0
$v = 2$	0.5	0.002	0.201	0.202	95.0	−0.081	0.410	0.396	94.2
	0.2	−0.007	0.214	0.219	95.8	−0.033	0.455	0.428	93.4

Table 3. Estimated regression coefficients of the semiparametric transformation model for the Oscar data under the first case-cohort design

	$v = 1$		$v = 2$		$v = 0$	
Effect	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
SEX	0.981 (0.209)	<0.001	1.381 (0.286)	<0.001	0.558 (0.126)	<0.001
USA	0.362 (0.194)	0.062	0.570 (0.270)	0.035	0.228 (0.122)	0.062
NOFF	−0.076 (0.017)	<0.001	−0.105 (0.023)	<0.001	−0.045 (0.010)	<0.001
NOW	−0.314 (0.443)	0.479	−0.495 (0.689)	0.473	−0.074 (0.235)	0.754
NON	0.039 (0.073)	0.597	0.076 (0.105)	0.472	0.043 (0.042)	0.310
WIN	0.247 (0.519)	0.634	0.039 (0.805)	0.961	−0.055 (0.291)	0.849

SE, estimated standard error; SEX: male=1, female=0; USA: whether born in USA, yes=1, no=0; NOFF: number of four star films; NOW: number of times the person won an Oscar; NON: number of times the person was nominated for an Oscar; WIN: whether the person has won an Oscar, yes=1, no=0

Candidates who are healthier will be able to act in more films and have more chances to win Oscar Awards. Han et al. (2011) adapt Robins' rank preserving structural accelerated failure time model and g-estimation method, and there is no strong evidence that winning an Oscar increases life expectancy.

Interestingly, let T denote survival time, the time from birth to death. We denote by A the time between the performer's birth year and the first Oscar nomination year. Based on the formal test proposed by Addona and Wolfson (2006), we found that the Oscar data can be treated as the length-biased data. The p-value of this form test is about 0.3, which means that the data set satisfies the stationarity assumption and we can use the proposed method to analyze the data. The observation of survival time subject to right censoring due to study end. Then, the Oscar data can be treated as length-biased right-censored data.

The main interest here is also to assess the association between the performer's lifetime and the winning of an Oscar Award. The first case-cohort study is designed to have the same number of controls and cases. Other covariates such as sex (male=1, female=0), born in USA (yes=1, no=0), number of four star films, number of times the person won an Oscar and number of times the person was nominated for an Oscar are also included. Table 3 presents the estimates of the regression parameters, their estimated standard errors and the p-values based on Wald test. The three models give similar results, and the results shows that there is no evidence that winning an Oscar increases the performer's lifetime. The results are in agreement with those obtained by Sylvestre et al. (2006) and Han et al. (2011). Here only Sex and number of four star films are significant with $p < 0.05$.

6. REMARKS

In this paper, we use the time-independent weight to analyze the length-biased and right-censored data under the case-cohort design. In fact, referred to Kulich and Lin (2004) and Chen and Zucker (2009), we also can consider the time-dependent weighted estimator for p given by

$$\hat{p}(t) = \sum_{i=1}^{n} \xi_i (1 - \delta_i) m_i(t) \,/\, \sum_{i=1}^{n} (1 - \delta_i) m_i(t),$$

where $m_i(t)$ is a weight function. Various versions of the weight $m_i(t)$ have been suggested in Kulich and Lin (2004) for the Cox model. Thus, we can replace p in the definition of ρ_i, leading to a new weight $\pi_i(t) = \delta_i + (1-\delta_i)\xi_i/\hat{p}(t)$ to replace π_i in the estimating equations (6) and (7) to improve the estimation efficiency.

APPENDIX A

Proof of Lemma 1: Under length-biased sampling, we have

$$\begin{aligned} & E[dN_1(t) \mid \mathbf{Z}_1 = \mathbf{z}] \\ = & P(\delta_1 = 1, t-dt < Y_1 \le t \mid \mathbf{Z}_1 = \mathbf{z}) \\ = & \int_{t-dt}^{t}\int_0^s \frac{f(s \mid \mathbf{z})}{\mu(\mathbf{z})} G(a \mid \mathbf{z}) dads \\ = & \frac{f(t \mid \mathbf{z})}{\mu(\mathbf{z})}\omega_c(t \mid \mathbf{z})dt, \end{aligned}$$

where $\omega_c(t \mid \mathbf{z}) = \int_0^t G(u \mid \mathbf{z})du$.

Furthermore, since

$$\begin{aligned} & P(A_1 \le t \le Y_1 \mid \mathbf{Z}_1 = \mathbf{z}) \\ = & \int_t^{\infty}\int_0^t \frac{f(s \mid \mathbf{z})}{\mu(\mathbf{z})} G(t-a \mid \mathbf{z}) dads \\ = & \frac{S(t \mid \mathbf{z})}{\mu(\mathbf{z})}\omega_c(t \mid \mathbf{z}), \end{aligned}$$

and

$$\begin{aligned} & P(\delta_1 = 1, \tilde{V}_1 \le t \le Y_1 \mid \mathbf{Z}_1 = \mathbf{z}) \\ = & \int\int_{v \le t \le s} \frac{f(s \mid \mathbf{z})}{\mu(\mathbf{z})} G(v \mid \mathbf{z}) dvds \\ = & \frac{S(t \mid \mathbf{z})}{\mu(\mathbf{z})}\omega_c(t \mid \mathbf{z}). \end{aligned}$$

Therefore

$$\begin{aligned} & E[R_1(t) \mid \mathbf{Z}_1 = \mathbf{z}] \\ = & \frac{1}{2}\Big[P(A_1 \le t \le Y_1 \mid \mathbf{Z}_1 = \mathbf{z}) \\ & + P(\delta_1 = 1, \tilde{V}_1 \le t \le Y_1 \mid \mathbf{Z}_1 = \mathbf{z})\Big] \\ = & \frac{S(t \mid \mathbf{z})}{\mu(\mathbf{z})}\omega_c(t \mid \mathbf{z}). \end{aligned}$$

Hence $E[dM_1(t)] = 0$.

Proof of the Theorem 2: The details of the proof are followed by Kim et al. (2013) and Shen (2011). Here we only give the sketch of the proof. Following Chen et al. (2012) and Kim et al. (2013), we divide the sketch of proof into two steps.

Step 1: We show that $\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}} U\{\boldsymbol{\beta}, \hat{H}(\cdot, \boldsymbol{\beta})\} \mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ converges to $-A$. Let $a > 0$ and b be fixed finite numbers and define

$$\lambda^*\{H_0(t)\} = B(t,a), \qquad \Lambda^*(x) = \int_b^x \lambda^*(s)ds,$$

for $t > 0$ and $x \in (-\infty, \infty)$. We choose finite $a > 0$ and b to ensure that the integrals are finite. It is easy to see that

$$\begin{aligned} & B(t,s) = \lambda^*\{H_0(t)\}/\lambda^*\{H_0(s)\}, \\ & d\lambda^*\{H_0(t)\} = [\lambda^*\{H_0(t)\}/B_2(t)]B_1(t)dH_0(t). \end{aligned}$$

We have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \pi_i M_i(t) \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^t \pi_i R_i(s) \\ & \cdot \Big[d\Lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{H}(s, \boldsymbol{\beta}_0)\} - d\Lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}\Big] \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^t \pi_i R_i(s) \\ & \cdot d\left(\frac{\lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}}{\lambda^*\{H_0(s)\}}\Big[\Lambda^*\{\hat{H}(s, \boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(s)\}\Big]\right) \\ & + o_p(n^{-\frac{1}{2}}) \\ = & \int_0^t \frac{B_2(s)}{\lambda^*\{H_0(s)\}} d\Big[\Lambda^*\{\hat{H}(s, \boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(s)\}\Big] \\ & + o_p(n^{-\frac{1}{2}}). \end{aligned}$$

Therefore, for $t \in [0, \tau]$,

$$\begin{aligned} & \Lambda^*\{\hat{H}(t, \boldsymbol{\beta}_0)\} - \Lambda^*\{H_0(t)\} \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^t \pi_i \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dM_i(s) + o_p(n^{-\frac{1}{2}}). \end{aligned}$$

Differentiating (7) with respect to $\boldsymbol{\beta}$, we have that

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \int_0^t \pi_i R_i(s) \\ & \cdot d\left[\lambda\{\boldsymbol{\beta}'\mathbf{Z}_i + \hat{H}(s, \boldsymbol{\beta})\}\left\{\mathbf{Z}_i + \frac{\partial}{\partial \boldsymbol{\beta}}\hat{H}(s, \boldsymbol{\beta})\right\}\right] \\ = & \; 0. \end{aligned}$$

for every $t \in [0, \tau]$. Hence we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \int_0^t \pi_i R_i(s) \\ & \cdot d\left[\lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}\left\{\mathbf{Z}_i + \frac{\partial}{\partial \boldsymbol{\beta}}\hat{H}(s, \boldsymbol{\beta}) \mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right\}\right] \\ = & \; o_p(1). \end{aligned}$$

Then

$$\begin{aligned} & \int_0^t B_1^Z(s)dH_0(s) \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^t \pi_i R_i(s)\mathbf{Z}_i \dot{\lambda}\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}dH_0(s) + o_p(1) \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^t \pi_i R_i(s) \\
&\qquad \cdot d\left[\lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}\cdot\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(s,\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \\
&\quad +o_p(1) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^t \pi_i R_i(s) \\
&\quad \cdot d\left[\frac{\lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}}{\lambda^*\{H_0(s)\}}\cdot\lambda^*\{H_0(s)\}\cdot\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(s,\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \\
&\quad +o_p(1) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\pi_i R_i(s)\lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(s)\}}{\lambda^*\{H_0(s)\}} \\
&\qquad \cdot d\left[\lambda^*\{H_0(s)\}\cdot\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(s,\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \\
&\quad +o_p(1) \\
&= -\int_0^t \frac{B_2(s)}{\lambda^*\{H_0(s)\}} d\left[\lambda^*\{H_0(s)\}\cdot\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(s,\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \\
&\quad +o_p(1).
\end{aligned}$$

Hence

$$\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(t,\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= -\frac{1}{\lambda^*\{H_0(t)\}}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} \\
&\qquad \cdot \mathsf{E}[\pi\mathbf{Z}\dot{\lambda}\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(s)\}R(s)]dH_0(s) \\
&\quad +o_p(1) \\
&= -\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s) + o_p(1).
\end{aligned}$$

It follows from the law of large numbers that

$$\begin{aligned}
&\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}}U\{\boldsymbol{\beta},\hat{H}(\cdot,\boldsymbol{\beta})\}\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\pi_i\mathbf{Z}_iR_i(t) \\
&\quad \cdot d\left[\lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{H}(s,\boldsymbol{\beta}_0)\}\left\{\mathbf{Z}_i + \frac{\partial}{\partial\boldsymbol{\beta}}\hat{H}(s,\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right\}'\right] \\
&= -\int_0^{\tau}\mathsf{E}\left[\pi\{\mathbf{Z}-\mu_z(t)\}\mathbf{Z}'\dot{\lambda}\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)\right]dH_0(t) \\
&\quad +o_p(1).
\end{aligned}$$

Here

$$\begin{aligned}
\mu_z(t) &= \frac{1}{2}\left[\frac{\mathsf{E}[\pi\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(Y)\}I(Y\geq t)B(t,Y)]}{\mathsf{E}[\pi\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)]}\right. \\
&\quad -\frac{\mathsf{E}[\pi\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(A)\}I(A\geq t)B(t,A)]}{\mathsf{E}[\pi\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)]} \\
&\quad +\frac{\mathsf{E}[\Delta\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(Y)\}I(Y\geq t)B(t,Y)]}{\mathsf{E}[\pi\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)]} \\
&\quad \left.-\frac{\mathsf{E}[\Delta\mathbf{Z}\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(\tilde{V})\}I(\tilde{V}\geq t)B(t,\tilde{V})]}{\mathsf{E}[\pi\lambda\{\boldsymbol{\beta}_0'\mathbf{Z} + H_0(t)\}R(t)]}\right]
\end{aligned}$$

Step 2: Here we show the asymptotic normality of $U\{\boldsymbol{\beta}_0,\hat{H}(\cdot,\boldsymbol{\beta}_0)\}$. Using the results of step 1 and some empirical process approximation techniques, we can write

$$\begin{aligned}
&U\{\boldsymbol{\beta}_0,\hat{H}(\cdot,\boldsymbol{\beta}_0)\} \\
&= \sum_{i=1}^{n}\int_0^{\tau}\pi_i\mathbf{Z}_i\left[dN_i(t) - R_i(t)d\Lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{H}(t,\boldsymbol{\beta}_0)\}\right] \\
&= \sum_{i=1}^{n}\int_0^{\tau}\pi_i\mathbf{Z}_i dM_i(t) \\
&\quad -\sum_{i=1}^{n}\int_0^{\tau}\pi_i\mathbf{Z}_iR_i(t) \\
&\qquad \cdot\left[d\Lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + \hat{H}(t,\boldsymbol{\beta}_0)\} - d\Lambda\{\boldsymbol{\beta}_0'\mathbf{Z}_i + H_0(t)\}\right] \\
&= \sum_{i=1}^{n}\int_0^{\tau}\pi_i[\mathbf{Z}_i - \mu_z(t)]dM_i(t) + o_p(n^{-\frac{1}{2}}) \\
&= \sum_{i=1}^{n}\int_0^{\tau}[\mathbf{Z}_i - \mu_z(t)]dM_i(t) \\
&\quad +\sum_{i=1}^{n}\int_0^{\tau}(\pi_i - 1)[\mathbf{Z}_i - \mu_z(t)]dM_i(t) + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

Similarly to Kulich and Lin (2000) and Lu and Tsiatis (2006), it is easy to show that the first two terms on the right side of the above equation are uncorrelated. The first term is the sum of n independent zero-mean random vectors, which converges in distribution to a zero-mean normal random vector with covariance matrix

$$\Sigma_1 = \mathsf{E}\left[\int_0^{\tau}\{\mathbf{Z} - \mu_z(t)\}dM(t)\right]^{\otimes 2}.$$

By a slight extension of Hájek's (1960) central limit theorem for simple random sampling, the second term converges in distribution to a zero-mean normal random vector with covariance matrix

$$\begin{aligned}
\Sigma_2 &= \frac{1-p}{p}\mathsf{E}\left\{(1-\delta)\left[\int_0^{\tau}\{\mathbf{Z} - \mu_z(t)\}dM(t)\right]^{\otimes 2}\right\} \\
&\quad -\frac{1-p}{p}\left\{\mathsf{E}\left[(1-\delta)\int_0^{\tau}\{\mathbf{Z} - \mu_z(t)\}dM(t)\right]\right\}^{\otimes 2}.
\end{aligned}$$

Therefore, $n^{-1/2}U\{\boldsymbol{\beta}_0,\hat{H}(\cdot,\boldsymbol{\beta}_0)\}$ is asymptotically zero-mean normal with covariance matrix $\Sigma = \Sigma_1 + \Sigma_2$. By simple algebra, we show that

$$\Sigma = \mathsf{E}\left\{\{\delta + (1-\delta)/p\}\left[\int_0^{\tau}\{\mathbf{Z} - \mu_z(t)\}dM(t)\right]^{\otimes 2}\right\}$$

$$-\frac{1-p}{p}\left\{\mathsf{E}\left[\delta\int_0^\tau\{\mathbf{Z}-\mu_z(t)\}dM(t)\right]\right\}^{\otimes 2}.$$

By Taylor's expansion and some empirical process approximation techniques, $n^{1/2}(\hat{\boldsymbol\beta}-\boldsymbol\beta_0)$ is asymptotically zero-mean normal with covariance matrix $A^{-1}\Sigma A^{-1}$.

To show the weak convergence of $\sqrt{n}\{\hat{H}(t,\hat{\boldsymbol\beta})-H_0(t)\}$, we have

$$\begin{aligned}
&\sqrt{n}\{\hat{H}(t,\hat{\boldsymbol\beta})-H_0(t)\}\\
=\ &\sqrt{n}\{\hat{H}(t,\hat{\boldsymbol\beta})-\hat{H}(t,\boldsymbol\beta_0)\}+\sqrt{n}\{\hat{H}(t,\boldsymbol\beta_0)-H_0(t)\}\\
=\ &\sqrt{n}\frac{\partial}{\partial\boldsymbol\beta}\hat{H}(t,\boldsymbol\beta)|_{\boldsymbol\beta=\boldsymbol\beta_0}(\hat{\boldsymbol\beta}-\boldsymbol\beta_0)\\
&+\sqrt{n}\{\hat{H}(t,\boldsymbol\beta_0)-H_0(t)\}+o_p(1).
\end{aligned}$$

The first term on the right hand side equals

$$n^{-1/2}D(t)A^{-1}\sum_{i=1}^n\int_0^\tau\pi_i[\mathbf{Z}_i-\mu_z(t)]dM_i(t)+o_p(1),$$

where

$$D(t)=-\int_0^t\frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s).$$

To tackle the second term, observe that

$$\begin{aligned}
&\sum_{i=1}^n\pi_i dM_i(t)\\
=\ &\sum_{i=1}^n\pi_i R_i(t)\\
&\quad\cdot\left[d\Lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+\hat{H}(t,\boldsymbol\beta_0)\}-d\Lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}\right]\\
=\ &\sum_{i=1}^n\pi_i R_i(t)\\
&\quad\cdot d\left[\lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}\{\hat{H}(t,\boldsymbol\beta_0)-H_0(t)\}\right]\\
&+o_p(n^{-1/2})\\
=\ &\sum_{i=1}^n\pi_i R_i(t)\dot\lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}[\hat{H}(t,\boldsymbol\beta_0)-H_0(t)]dH_0(t)\\
&+\sum_{i=1}^n\pi_i R_i(t)\lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}d[\hat{H}(t,\boldsymbol\beta_0)-H_0(t)]\\
&+o_p(n^{-1/2}).
\end{aligned}$$

Let

$$J_n(t)=\frac{\sum_{i=1}^n\pi_i R_i(t)\dot\lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}}{\sum_{i=1}^n\pi_i R_i(t)\lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}},$$

and $J(t)=\lim_{n\to\infty}J_n(t)$. Then we have

$$\begin{aligned}
&\frac{\sum_{i=1}^n\pi_i dM_i(t)}{\sum_{i=1}^n\pi_i R_i(t)\lambda\{\boldsymbol\beta_0'\mathbf{Z}_i+H_0(t)\}}\\
=\ &J_n(t)[\hat{H}(t,\boldsymbol\beta_0)-H_0(t)]dH_0(t)\\
&+d[\hat{H}(t,\boldsymbol\beta_0)-H_0(t)]+o_p(n^{-1/2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\exp\left\{\int_0^t J(s)ds\right\}(\hat{H}(t,\boldsymbol\beta_0)-H_0(t))\\
=\ &\sum_{i=1}^n\int_0^t\frac{\exp\{\int_0^s J(u)du\}\pi_i dM_i(s)}{\sum_{j=1}^n\pi_j R_j(t)\lambda\{\boldsymbol\beta_0'\mathbf{Z}_j+H_0(t)\}}+o_p(n^{-1/2}).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\hat{H}(t,\boldsymbol\beta_0)-H_0(t)\\
=\ &\frac{1}{n}\sum_{i=1}^n\int_0^t\frac{B(s,t)}{B_2(s)}\pi_i dM_i(s)+o_p(n^{-1/2}).
\end{aligned}\tag{10}$$

Combining the above two equations, we have

$$\begin{aligned}
&\sqrt{n}\{\hat{H}(t,\hat{\boldsymbol\beta})-H_0(t)\}\\
=\ &\frac{1}{\sqrt{n}}D(t)A^{-1}\sum_{i=1}^n\int_0^\tau\pi_i[\mathbf{Z}_i-\mu_z(t)]dM_i(t)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{B(s,t)}{B_2(s)}\pi_i dM_i(s)+o_p(1).
\end{aligned}\tag{11}$$

By the Hájek's (1960) central limit theorem, $\sqrt{n}\{\hat{H}(t,\hat{\boldsymbol\beta})-H_0(t)\}$ converges in finite dimensional distribution to a mean-zero Gaussian process.

ACKNOWLEDGEMENTS

The authors are very grateful to the Co-Editor Ming-Hui Chen, the Associate Editor and two anonymous referees for their valuable comments and helpful suggestions. Ma's research is partially supported by National Institutes of Health grant R01 HL113548. Qiu's work was supported by the Natural Science Foundation of Fujian Province, China (Grant No.2015J01583). Zhou's work was supported by National Natural Science Foundation of China (NSFC) (71271128), the State Key Program of National Natural Science Foundation of China (71331006), NCMIS, Key Laboratory of RCSDS, CAS and Shanghai University of Finance and Economics through Project 211 Phase IV, Shanghai First-class Discipline A and IRTSHUFE, PCSIRT (IRT13077).

Received 1 April 2014

REFERENCES

[1] Addona, V. and Wolfson, D. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis* **12** 267–284. MR2328577

[2] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer. MR1198884

[3] Borgan, O., Langholz, B., Samuelsen, S. O., Goldstein, L., and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis* **6** 39–58. MR1767493

[4] Chen, H. Y. (2001a). Fitting semiparametric transformation regression models to data from a modified case-cohort design. *Biometrika* **88** 255–268. MR1841273

[5] Chen, H. Y. (2001b). Weighted semiparametric likelihood method for fitting a proportional odds regression model to data from the case-cohort design. *Journal of the American Statistical Association* **96** 1446–1457. MR1946589

[6] Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89** 659–668. MR1929170

[7] Chen, Y. and Zucker, D. M. (2009). Case-cohort analysis with semiparametric transformation models. *Journal of Statistical Planning and Inference* **139** 3706–3717. MR2549118

[8] Cox, D. R. (1972). Regression models and life–tables. *Journal of the Royal Statistical Society Series B* **34** 187–220. MR0341758

[9] Dabrowska, D. M and Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics* **15** 1–23. MR0967953

[10] de Una-Alvarez, J., Otero-GirALdez, M., Soledad de Una-Alvarez, J., Rodriguez-Casal, A., and Alvarez-Llorente, G. (2003). Estimation under length-bias and right-censoring: an application to unemployment duration analysis for married women. *Journal of Applied Statistics* **30** 283–291. MR1963203

[11] Hajek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematics Institute of the Hungarian Academy of Science* **5** 361–374. MR0125612

[12] Han, X., Small, D. S., Foster, D. P., and Patel, V. (2011). The effect of winning an oscar award on survival: Correcting for healthy performer survivor bias with a rank preserving structural accelerated failure time model. *The Annals of Applied Statistics* **5** 746–772. MR2840174

[13] 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. MR3010882

[14] Kim, J. P., Lu, W., Sit, T., and Ying, Z. (2013). A unified approach to semiparametric transformation models under general biased sampling schemes. *Journal of the American Statistical Association* **108** 217–227. MR3174614

[15] Kong, L., Cai, J., and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika* **91** 305–319. MR2081303

[16] Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika* **87** 73–87. MR1766829

[17] Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association* **99** 832–844. MR2090916

[18] Lagakos, S., Barraj, L., and Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75** 515–523. MR0967591

[19] Lancaster, T. (1992). *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press. MR1167199

[20] Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* **88** 1341–1349. MR1245368

[21] Lu, W. and Tsiatis, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* **93** 207–214. MR2277751

[22] Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73** 1–11.

[23] Qin, J. and Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* **66** 382–392. MR2758818

[24] Redelmeier, D. A. and Singh, S. M. (2001). Survival in academy award winning actors and actresses. *Annals of Internal Medicine* **134** 955–962.

[25] Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89** 846–866. MR1294730

[26] Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics* **16** 64–81. MR0924857

[27] Shen, P. S. (2011). Semiparametric analysis of transformation models with left-truncated and right-censored data. *Computational Statistics* **26** 521–537. MR2833146

[28] Shen, Y., Ning, J., and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association* **104** 1192–1202. MR2750244

[29] Simon, R. (1980). Length biased sampling in etiologic studies. *American Journal of Epidemiology* **111** 444–452.

[30] Sylvestre, M., Huszti, E., and Hanley, J. A. (2006). Do Oscar winners live longer than less successful peers? a reanalysis of the evidence. *Annals of Internal Medicine* **145** 361–363.

[31] Tsai, W. Y. (2009). Pseudo-partial likelihood for proportional hazards models with biased-sampling data. *Biometrika* **96** 601–615. MR2538760

[32] Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56** 601–614. MR0258224

Huijuan Ma
Department of Biostatistics and Bioinformatics
Emory University
Atlanta, GA, 30322
USA
E-mail address: hma30@emory.edu

Zhiping Qiu
School of Mathematical Sciences
Huaqiao University
Quanzhou, Fujian, 362021
China
E-mail address: qiuzhiping128@gmail.com

Yong Zhou
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing, 100190
China
and
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai, 200433
China
E-mail address: yzhou@amss.ac.cn