Published in final edited form as:

Biometrics. 2017 March ; 73(1): 114–123. doi:10.1111/biom.12573.

Efficient Estimation for Accelerated Failure Time Model under Case-Cohort and Nested Case-Control Sampling

Suhyun Kang^{1,*}, Wenbin Lu^{1,**}, and Mengling Liu^{2,***}

¹Department of Statistics, North Carolina State University, Raleigh, North Carolina, U.S.A

²Department of Environmental Medicine, New York University School of Medicine, New York, New York, U.S.A

SUMMARY

Case-cohort (Prentice, 1986) and nested case-control (Thomas, 1977) designs have been widely used as a cost-effective alternative to the full-cohort design. In this article, we propose an efficient likelihood-based estimation method for the accelerated failure time model under case-cohort and nested case-control designs. An EM algorithm is developed to maximize the likelihood function and a kernel smoothing technique is adopted to facilitate the estimation in the M-step of the EM algorithm. We show that the proposed estimators for the regression coefficients are consistent and asymptotically normal. The asymptotic variance of the estimators can be consistently estimated using an EM-aided numerical differentiation method. Simulation studies are conducted to evaluate the finite-sample performance of the estimators and an application to a Wilms tumor data set is also given to illustrate the methodology.

Keywords

Accelerated failure time model; Case-cohort; Efficient estimation; Kernel smoothing; Nested case-control

1. Introduction

Case-cohort (CC) and nested case-control (NCC) designs are two popular sampling techniques used in epidemiological cohort studies when the primary endpoint of interest is a time-to-event. The former randomly samples a subcohort from the full cohort and ascertains the covariate information for all the cases from the whole cohort and only the controls in the subcohort. The latter randomly samples a fixed number of controls for each case from the risk set at the case's event time, excluding the case itself, and then ascertains covariate information for all the cases and the sampled controls.

* skang8@ncsu.edu. ** lu@stat.ncsu.edu. *** Mengling.Liu@nyumc.org.

Supplementary Materials

The Web Appendix referenced in Sections 4, 5 and Appendix is available with this paper at the Biometrics website on Wiley Online Library. In addition, the R codes for implementing the proposed estimation method are available in the Supplementary Materials.

Various models and approaches have been proposed for the analysis of CC and NCC data. One commonly used approach is based on the proportional hazards (PH) model with the inverse probability of selection weighted (IPSW) technique, such as Prentice (1986) and Self and Prentice (1988) for the CC design and Samuelsen (1997) for the NCC design. The IPSW estimation methods for CC and NCC data have been also studied under other survival models, for example, the additive hazards model (Kulich and Lin, 2000), the proportional odds model (Chen, 2001a), the linear transformation model (Chen, 2001b; Kong et al, 2004; Lu and Tsiatis, 2006; Lu and Liu, 2012), and the accelerated failure time model (Nan et al, 2006; Kong and Cai, 2009).

Another approach for analyzing CC and NCC data is the nonparametric maximum likelihood estimation (NPMLE). Scheike and Martinussen (2004) and Scheike and Juul (2004) studied the NPMLE for the PH model under CC and NCC samplings, respectively. Zeng and Lin (2014) studied the NPMLE for the transformation models for two-phase cohort studies, including CC and NCC designs. The NPMLE approach has two advantages compared with the IPSW methods. First, this likelihood-based approach does not need to model the sampling scheme in the likelihood function and hence can handle the estimation of CC and NCC data in a unified framework. Second, the NPMLE is semiparametric efficient.

In this article, we study the efficient estimation for the accelerated failure time (AFT) model under CC and NCC samplings. The AFT model is a useful alternative to the PH model (Reid, 1994) and has a natural log linear interpretation between the event time of interest and covariates. The estimation for the AFT model with the full cohort data has been extensively studied. In particular, Zeng and Lin (2007) developed an efficient estimation method for the AFT model based on a kernel-smoothed profile likelihood function. To the best of our knowledge, an efficient likelihood-based estimation method for the AFT model under CC and NCC samplings have not yet been studied. Different from the method of Zeng and Lin (2007) for full cohort data, an EM algorithm needs to be developed for handling missing covariates due to sampling schemes of the CC and NCC designs. In addition, in the M-step of the proposed EM algorithm, both the baseline hazard function and the distribution of covariates need to be estimated nonparametrically. We adopt a kernel smoothing technique as in Zeng and Lin (2007) to facilitate the estimation of the baseline hazard function. Note that the nonparametric estimation of two components (baseline hazard function and covariate distribution) using different techniques entails challenges in method development and theory derivation. The asymptotic properties of the resulting estimators will be investigated.

The rest of the article is organized as follows. Section 2 introduces model and data. Section 3 presents our estimation method and the asymptotic properties of the resulting estimators. Simulation studies are given in Section 4 to evaluate the performance of the proposed method. An application to a Wilms tumor data set is illustrated in Section 5, followed by a discussion section. All the technical proofs of asymptotic properties are given in the Appendix.

2. Model and Data

Let T denote the event time of interest. The AFT model assumed that

$$\log(T) = \beta' Z + \varepsilon, \quad (1)$$

where β and Z are, respectively, p -dimensional vectors of regression parameters and covariates, and ε is the error term with an unspecified continuous distribution independent of covariates. Let C denote the censoring time. It is assumed that T and C are conditionally

independent given Z . The full cohort data consist of $(\tilde{T}_i, \delta_i, Z_i)$, $i = 1, 2, \dots, N$, which are N independent replications of $\{\tilde{T}, \delta, Z\}$, where $\tilde{T}_i = \min\{T_i, C_i\}$ and $\delta_i = \mathbb{I}(T_i < C_i)$.

Under CC and NCC sampling designs, covariate information are ascertained for all cases and selected controls. Let $S = \{1, 2, \dots, N\}$ be the index set of all individuals in the full cohort, and let S_1 , S_0 , and S_2 denote the index sets of all cases, selected controls and

unselected controls, respectively. The observed data are then $\{\mathcal{O}_i = (\tilde{T}_i, \delta_i, Z_i) : i \in S_0 \cup S_1\}$ and $\{\mathcal{O}_i = (\tilde{T}_i, \delta_i) : i \in S_2\}$.

3. Estimation Method

The observed-data likelihood function can be written as

$$L_N^o(\beta, \lambda, F) = \prod_{i \in S_0 \cup S_1} \left[e^{-\beta' Z_i} \lambda \left\{ e^{R_i(\beta, Z_i)} \right\} \right]^{\delta_i} e^{-\Lambda \left\{ e^{R_i(\beta, Z_i)} \right\}} dF(Z_i) \prod_{i \in S_2} \int_z e^{-\Lambda \left\{ e^{R_i(\beta, z)} \right\}} dF(z), \quad (2)$$

where $R(\cdot)$ is the cumulative distribution function of Z , $\Lambda(\cdot)$ and $\lambda(\cdot)$ are, respectively, the cumulative hazard function and hazard function of e^Z , and $R_i(\beta, z) = \log(\tilde{T}_i) - \beta' z$. Note that both components of Λ and F are nonparametric and a direct maximization of the observed likelihood function in (2) over unrestricted (Λ, F) will be prohibitive. Here, we propose an EM algorithm to maximize (2) and develop a kernel smoothing technique to facilitate the estimation in the M-step of the EM algorithm.

Towards this end, following the technique of Wellner and Zhan (1997), we handle the distribution of Z nonparametrically by assuming that the nonparametric maximum likelihood estimator of the distribution of Z has point masses at its observed values. When F is assumed to be concentrated on $\{Z_i : i \in S_0 \cup S_1\}$, the existence of the maximum likelihood estimator can be shown as in Murphy and van der Vaart (2001). Similar ideas have also been used by Scheike and Juul (2004) and Zeng and Lin (2014) for deriving the NPMLE for CC and NCC data. More specifically, let $\{z_1, \dots, z_L\}$ denote the L distinct values among the observed

covariates of $\{Z_i, i \in S_0 \sim S_1\}$ and let $\mathbf{p} = \{p_1, \dots, p_L\}$ denote the corresponding point masses subject to $\sum_{l=1}^L p_l = 1$.

Then the complete-data log-likelihood function for the full cohort data is given by

$$\ell_N^c(\beta, \lambda, \mathbf{p}) = \frac{1}{N} \sum_{i=1}^N \left\{ \delta_i \left[-\beta' Z_i + \log \lambda \left\{ e^{R_i(\beta, Z_i)} \right\} \right] - \Lambda \left\{ e^{R_i(\beta, Z_i)} \right\} + \sum_{l=1}^L I(Z_i = z_l) \log p_l \right\}. \quad (3)$$

The E-step of the EM algorithm computes the conditional expectation of $\ell_N^c(\beta, \Lambda, \mathbf{p})$ given the observed data, $\{\mathcal{O}_1, \dots, \mathcal{O}_N\}$ and the current parameter estimates $\Omega^{[k]} \equiv (\beta^{[k]}, \lambda^{[k]}, \mathbf{p}^{[k]})$.

To do this, we need to calculate $P(Z_i = z_l | \mathcal{O}_i, \Omega^{[k]})$, $l = 1, \dots, L$, which are given by

$$\alpha_{il}^{[k]} \triangleq P(Z_i = z_l | \mathcal{O}_i, \Omega^{[k]}) = I(i \in S_0 \cup S_1) I(Z_i = z_l) + I(i \in S_2) \frac{\exp \left\{ -\Lambda^{[k]}(e^{R_i(\beta^{[k]}, z_l)}) \right\} p_l^{[k]}}{\sum_{l=1}^L \exp \left\{ -\Lambda^{[k]}(e^{R_i(\beta^{[k]}, z_l)}) \right\} p_l^{[k]}}. \quad (4)$$

Therefore, we have $E \left\{ \ell_N^c(\beta, \lambda, \mathbf{p}) | \mathcal{O}, \Omega^{[k]} \right\} = \tilde{\ell}_{N,1}(\beta, \lambda | \mathcal{O}, \Omega^{[k]}) + \tilde{\ell}_{N,2}(\mathbf{p} | \mathcal{O}, \Omega^{[k]})$, where

$$\tilde{\ell}_{N,1}(\beta, \lambda | \mathcal{O}, \Omega^{[k]}) = \frac{1}{N} \sum_{i=1}^N \left\{ \delta_i \left[-\beta' Z_i + \log \lambda \left\{ e^{R_i(\beta, Z_i)} \right\} \right] - \sum_{l=1}^L \alpha_{il}^{[k]} \Lambda \left\{ e^{R_i(\beta, z_l)} \right\} \right\},$$

and

$$\tilde{\ell}_{N,2}(\mathbf{p} | \mathcal{O}, \Omega^{[k]}) = \frac{1}{N} \sum_{i=1}^N \sum_{l=1}^L \alpha_{il}^{[k]} \log p_l.$$

Note that parameters of (β, λ) and $\mathbf{p}$ are now separated through the E-step and this separation greatly facilitates the maximization in the M-step.

In the M-step, the maximizer of $\tilde{\ell}_{N,2}(\mathbf{p} | \mathcal{O}, \beta^{[k]}, \Lambda^{[k]}, \mathbf{p}^{[k]})$ subject to $\sum_{l=1}^L p_l = 1$ can be easily obtained as

$$\hat{\mathbf{p}}^{[k+1]} \equiv \left(\hat{p}_1^{[k+1]}, \dots, \hat{p}_L^{[k+1]} \right)', \text{ with } \hat{p}_l^{[k+1]} = \frac{\sum_{i=1}^N \alpha_{il}^{[k]}}{N}.$$

However, the maximization of $\tilde{\ell}_{N,1}(\beta, \lambda | \mathcal{O}, \Omega^{[k]})$ over (β, Λ) , assuming $\Lambda(t)$ to be a step increasing function with the jumps only at the observed event times, may not always be

possible as discussed in Zeng and Lin (2007). The main reason is that this estimator for Λ is very unsmooth and thus the resulting profile likelihood function for β may not achieve its maximum for finite β . Following similar techniques in Zeng and Lin (2007), we derive a smoothed profile likelihood for β in the M-step. Specifically, we start with a piecewise constant hazard function,

$$\lambda(t) = \sum_{j=1}^{J_N} c_j I\{t \in [t_{j-1}, t_j)\}, \quad 0 \leq t \leq M$$

where $0 \equiv t_0 < t_1 < \dots < t_{J_N} \equiv M$ and M is the upper bound of

$\{e^{R_i(\beta, z_l)} : i=1, \dots, N, l=1, \dots, L\}$ for all possible β in a bounded set. The corresponding cumulative hazard function can be written as

$$\Lambda(t) = \sum_{j=1}^{J_N} c_j (t - t_j) I(t_{j-1} \leq t < t_j) + (M/J_N) \sum_{j=1}^{J_N} c_j I(t \geq t_j).$$

Plugging $\lambda(t)$ and $\Lambda(t)$ into $\tilde{\ell}_{N,1}(\beta, \lambda | \mathcal{O}, \Omega^{[k]})$ and maximizing it with respect to c_j for fixed β , we obtain the following maximizer

$$\hat{c}_j = \frac{\sum_{i=1}^N \delta_i I(t_{j-1} \leq e^{R_i(\beta, Z_i)} < t_j)}{\sum_{i=1}^N \sum_{l=1}^L \alpha_{il}^{[k]} \{(e^{R_i(\beta, z_l)} - t_j) I(t_{j-1} \leq e^{R_i(\beta, z_l)} < t_j) + (M/J_N) I(e^{R_i(\beta, z_l)} \geq t_j)\}}.$$

Plugging the estimators $\hat{c}_j$'s into $\tilde{\ell}_{N,1}(\beta, \lambda | \mathcal{O}, \Omega^{[k]})$, we can obtain a profile log-likelihood for β , which equals to

$$\begin{aligned} & -\frac{1}{N} \sum_{i=1}^N \delta_i \beta' Z_i + \sum_{j=1}^{J_N} \left\{ \frac{1}{N} \sum_{i=1}^N \delta_i I(e^{R_i(\beta, Z_i)} \in [t_{j-1}, t_j]) \right\} \left[\log \left\{ \frac{J_N}{NM} \sum_{i=1}^N \delta_i I(e^{R_i(\beta, Z_i)} \in [t_{j-1}, t_j]) \right\} \right. \\ & \left. - \log \left\{ \frac{J_N}{NM} \sum_{i=1}^N \sum_{l=1}^L \alpha_{il}^{[k]} (e^{R_i(\beta, z_l)} - t_{j-1}) I(e^{R_i(\beta, z_l)} \in [t_{j-1}, t_j]) + \frac{1}{N} \sum_{i=1}^N \sum_{l=1}^L \alpha_{il}^{[k]} I(e^{R_i(\beta, z_l)} \geq t_j) \right\} \right]. \end{aligned}$$

Following the similar techniques in Zeng and Lin (2007), as $N \rightarrow \infty$, $J_N \rightarrow \infty$ and $J_N/N \rightarrow 0$, the limit of the above profile log-likelihood for β can be approximated by the following kernel-smoothed function

$$\begin{aligned} \tilde{\ell}_{N,1}^S(\beta | \mathcal{O}, \Omega^{[k]}) &= \frac{1}{N} \sum_{i=1}^N \delta_i \log \tilde{T}_i + \frac{1}{N} \sum_{i=1}^N \delta_i \log \left\{ \frac{1}{N} \sum_{j=1}^N \delta_j K_h(R_j(\beta, Z_j) - R_i(\beta, Z_i)) \right\} \\ &\quad - \frac{1}{N} \sum_{i=1}^N \delta_i \log \left\{ \frac{1}{N} \sum_{j=1}^N \sum_{l=1}^L \alpha_{jl}^{[k]} \int_{-\infty}^{R_j(\beta, z_l) - R_i(\beta, z_i)} K_h(s) ds \right\}, \end{aligned} \quad (5)$$

where $K_h(\cdot)$ is a kernel function with the bandwidth h . Then, the smoothed profile log-likelihood function $\tilde{\ell}_{N,1}^S(\beta|\mathcal{O}, \Omega^{[k]})$ can be easily maximized using the Newton-Raphson or other gradient-based algorithms. Let $\beta^{[k+1]}$ denote the maximizer of $\tilde{\ell}_{N,1}^S(\beta|\mathcal{O}, \Omega^{[k]})$. Given $\beta^{[k+1]}$, we update $\lambda(t)$ and $\Lambda(t)$ by the following kernel-smoothed estimators

$$\lambda^{[k+1]}(t) = \frac{t^{-1} \sum_{i=1}^N \delta_i K_h \left\{ R_i(\beta^{[k+1]}, Z_i) - \log t \right\}}{\sum_{i=1}^N \sum_{l=1}^L \alpha_{il}^{[k]} \int_{-\infty}^{R_i(\beta^{[k+1]}, Z_i) - \log t} K_h(s) ds} \quad \text{and} \quad \Lambda^{[k+1]}(t) = \int_0^t \lambda(u) du.$$

The algorithm will be iterated until convergence. Let $\hat{\beta}_N$, $\hat{\lambda}_N(\cdot)$ and $\hat{\Lambda}_N(\cdot)$ denote the resulting estimators at convergence for β , $\lambda(\cdot)$ and $\Lambda(\cdot)$, respectively.

Next, we establish the asymptotic properties of $\hat{\beta}_N$ and $\hat{\lambda}_N(\cdot)$. Let β_0 , $\lambda_0(\cdot)$ and $\Lambda_0(\cdot)$ be the true values of β , $\lambda(\cdot)$ and $\Lambda(\cdot)$, respectively. For a function $a(t)$, let $\dot{a}(t)$ denote its first derivative with respect to t . We need the following regularity conditions.

- (C1) The covariates Z are bounded. If there exists a constant vector a such that $a'Z = 0$ almost surely, it implies $a = 0$.
- (C2) The hazard function $\lambda_0(t)$ is thrice-continuously differentiable on $[0, \tau]$ with bounded derivatives and $\Lambda_0(\tau) < \infty$, where τ is a fixed constant.
- (C3) The true parameter vector β_0 is an interior point of a known compact set $\mathcal{B}$ in $\mathbb{R}^p$.
- (C4) The probability $P(Ce^{-\beta'_0 Z} \geq \tau | Z)$ is bounded away from zero for all possible values of Z .
- (C5) The kernel function $K(\cdot)$ is thrice-continuously differentiable with bounded variations.
- (C6) The information matrix Σ defined in the Appendix is finite and positive definite.

Theorem 1

Assume that the regularity conditions (C1)–(C6) hold. As $N \rightarrow \infty$, $Nh^2 \rightarrow \infty$ and $Nh^4 \rightarrow 0$, we have: (i) $\sup_{t \in [0, \tau]} |\hat{\lambda}_N(t) - \lambda_0(t)| \xrightarrow{a.s.} 0$ and $\sup_{t \in [0, \tau]} |\dot{\hat{\lambda}}_N(t) - \dot{\lambda}_0(t)| \xrightarrow{a.s.} 0$; (ii) $N^{1/2}(\hat{\beta}_N - \beta_0)$ converges in distribution to a mean-zero normal random vector with the covariance matrix that achieves the semiparametric efficiency bound Σ^{-1} .

The proof of Theorem 1 is given in the Appendix. To estimate the asymptotic covariance matrix of $\hat{\beta}_N$, we adopt an EM-aided numerical differentiation method as in Lu (2010) and Liu et al (2013). Specifically, perturb the j th component of $\hat{\beta}_N$ by a small number d , denoted by $\hat{\beta}_j^{(1)} = \hat{\beta}_{N,j} + d$ and $\hat{\beta}_j^{(2)} = \hat{\beta}_{N,j} - d$, and fix other components of $\hat{\beta}_N$ unchanged. Let $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(2)}$ denote the resulting perturbed values of $\dots$. Applying the above EM algorithm with β

fixed at the perturbed values, say $\hat{\beta}^{(1)}$, we get the estimates of λ and $\mathbf{p}$, denoted by $\hat{\lambda}^{(1)}$ and $\hat{\mathbf{p}}^{(1)}$. Let $\hat{\Omega}^{(1)} = (\hat{\beta}^{(1)}, \hat{\lambda}^{(1)}, \hat{\mathbf{p}}^{(1)})$. Similarly, we can have $\hat{\Omega}^{(2)} = (\hat{\beta}^{(2)}, \hat{\lambda}^{(2)}, \hat{\mathbf{p}}^{(2)})$. Let $\tilde{\ell}_{N,1i}(\beta, \lambda | \mathcal{O}, \Omega)$ denote the i th summand of $\tilde{\ell}_{N,1}(\beta, \lambda | \mathcal{O}, \Omega)$. For $i = 1, \dots, N$ and $j = 1, \dots, p$, define

$$S_{ij} = \left\{ \tilde{\ell}_{N,1i}(\hat{\beta}^{(1)}, \hat{\lambda}^{(1)} | \mathcal{O}, \hat{\Omega}^{(1)}) - \tilde{\ell}_{N,1i}(\hat{\beta}^{(2)}, \hat{\lambda}^{(2)} | \mathcal{O}, \hat{\Omega}^{(2)}) \right\} / (2d),$$

and $S_i = (S_{i1}, \dots, S_{ip})'$. Then, the asymptotic covariance matrix of $\hat{\beta}_N$ can be estimated by $\sum_{i=1}^N S_i S_i'$, where $\sum_{i=1}^N S_i S_i' = \sum_{i=1}^N S_i S_i'$. In practice, we can choose $d = c/N$, where c is a positive constant, such as 1 or 2. In our numerical implementation, we chose $c = 2$ and it works well for all the considered settings.

4. Simulation Studies

We generate the full cohort data of size $N = 1000$ from the AFT model given in (1) and consider three error distributions: normal, logistic and extreme value distributions. For covariates, we consider three scenarios: (I) $Z = Z_1$ follows a Bernoulli distribution with success probability of 0.5; (II) $Z = Z_2$ follows a uniform distribution on $[-1.25, 1.25]$; (III) $Z = (Z_1, Z_2)'$. For the first two scenarios, we set $\beta = 1$; while for the third one, we set $\beta = (\beta_1, \beta_2)' = (1, 1)'$. The censoring time C is generated from a uniform distribution on $[0, c_0]$, where c_0 is chosen to achieve 90% censoring rate.

We consider both CC and NCC sampling from the full cohort. For the CC sampling, we study two cases. The first one is expected to have the same number of controls and cases (i.e. subjects with $T < C$) while the second one is expected to have twice as many controls as cases, denoted by CC1 and CC2, respectively. For the NCC sampling, we also study two cases. The first one randomly selects one control from the risk set at each of case event times while the second one selects two, denoted by NCC1 and NCC2, respectively. Since the censoring rate is 90%, the expected effective sample size (i.e. samples with complete covariate information) is approximately 200 for CC1, NCC1 and 300 for CC2, NCC2, respectively.

To select the bandwidth h in the kernel function used to smooth the profile log-likelihood function of β in the M-step, following Jones (1990) and Jones and Sheather (1991), we set $h = 4^{1/3} \min(\sigma, IQR/1.34) n^{-1/3}$, where σ and IQR are the standard deviation and interquartile range of $\left\{ \log(\tilde{T}_i) - \hat{\beta}'_{LS} Z_i \right\}$'s for all subjects in $S_0 \sim S_1$, and n is the size of $S_0 \sim S_1$. Here, $\hat{\beta}_{LS}$ is the least square estimate of regressing $\log(\tilde{T}_i)$'s on Z_i 's based on the data in $S_0 \sim S_1$. A similar choice was also used in Zeng and Lin (2007). In addition, in the EM algorithm, we chose $\hat{\beta}^{[0]} = \hat{\beta}_{LS}$ as the initial value for β ; set $\hat{\Lambda}^{[0]}(\cdot)$ as the Nelson estimator based on the data $\left\{ \log(\tilde{T}_i) - \hat{\beta}^{[0]'} Z_i, \delta_i : i \in S_0 \cup S_1 \right\}$; and set $\hat{\mathbf{p}}^{[0]} = (1/L, \dots, 1/L)'$. For each setting, we

conducted 500 simulations. The simulation results for Scenarios I and II are summarized in Table 1 while those for Scenario III are summarized in Table 2. For comparison, we also implement the inverse probability of selection weighted (IPSW) estimation method. Specifically, the IPSW estimators are obtained from the weighted Gehan-type log-rank estimating equations. For case-cohort sampling, selection probabilities of $p = 1/9$ and $p = 2/9$ are used for CC1 and CC2, respectively; while for nested case-control sampling, Samuelsen (1997)'s weight is used. They are similar to those in Lu and Tsiatis (2006) and Lu and Liu (2012) for the linear transformation model. Here, we focused on Scenario III, the cases with two covariates.

Based on the results, we observe that our proposed estimators of regression parameters are nearly unbiased under all the settings for both CC and NCC sampling; the estimated standard errors obtained using the proposed EM-aided numerical differentiation method are all close to the sample standard deviations of the estimators; the 95% Wald-type confidence intervals have proper empirical coverage probabilities; the estimators obtained under CC and NCC designs are less efficient than the corresponding full cohort estimators while the estimators obtained under the CC2 and NCC2 designs are more efficient than those for the CC1 and NCC1 designs as expected. In addition, the IPSW estimators are less efficient than the proposed estimators for almost all the considered simulation settings and the efficiency loss can be large, which show the advantages of the proposed estimators.

The relative efficiency comparing the CC and NCC estimators varies from case to case. From the perspective of statistical efficiency, the relative efficiency of the NCC or CC with respect to the full-cohort study depends on many things, such as numbers of controls per case or subcohort size, disease frequency, exposure distribution, relative risk, and whether more than one disease is to be studied (Ernster, 1994). For example, Langholz and Thomas (1990) showed that the asymptotic relative efficiencies of these two designs in an ideal intervention trial were close but slightly favorable to the CC design; however, they also showed using simulation studies that the CC design could be substantially worse than the NCC method. Therefore, the choice between these two designs should largely be determined by research question (e.g. the necessity of matching confounders) and resource availability (e.g. volume of stored biological samples).

We also consider the maximum likelihood estimator when the true error distribution is known, which serves as a benchmark. Based on the results, we observe that our proposed estimators are less efficient than the benchmark estimators as expected, but the efficiency loss of the proposed estimators is much less than that of the IPSW estimators for almost all the scenarios. In addition, we conduct additional simulations for cases with covariate-dependent censoring. Here, the distribution of censoring times is allowed to depend on the continuous covariate Z_2 . The proportional hazards model is used for generating censoring times. The regression coefficient in the PH model is chosen to be 0.25 and the constant baseline hazard function is chosen to obtain a 90% censoring rate. The simulation results show that the proposed method still works well when censoring is covariate dependent. For saving space, these simulation results are given in the online Supplementary Appendix.

5. Application to Wilms Tumor Data

We illustrate the proposed estimation method using a data set from the Wilms tumor studies (D'Angio et al, 1989; Green et al, 1998). The data consist of 3915 subjects who participate in the third (NWTSG-3) and fourth (NWTSG-4) Wilms tumor studies, which was treated as the full cohort. Among all subjects, about 17% of them have a cancer recurrence, which were referred to as cases. The primary interest is to evaluate the effect of NWTSG-4 (1 for NWTSG-4 and 0 for NWTSG-3), Stage (three binary variables), UH (unfavorable central history: 1 or 0), and tumor diameter on time to cancer recurrence. We conducted both CC and NCC sampling from the full cohort and considered CC1, CC2, NCC1 and NCC2 sampling schemes as in simulations. The bandwidth parameter h in the kernel function is chosen in the same way as in simulations. The results are given in Table 3. For comparison, we also included the estimates based on the PH model for the full cohort in Table 3.

Based on the results, we observe that the parameter estimates obtained by the proposed method are comparable to those obtained by the method of Zeng and Lin (2007) for the full cohort under all the considered sampling schemes, although the full cohort estimators have smaller standard errors compared with the proposed estimators. In addition, the proposed estimators for CC2 and NCC2 designs have smaller standard errors than those for CC1 and NCC1 designs as expected. For the full cohort estimators, all covariates are significant under the level of 0.05, while for the proposed estimators, except for Stage II and III indicators, all other covariates are significant under all sampling schemes.

We also conducted analyses using different choices of the bandwidth parameter h for estimating the coefficients and different choices of the c values for estimating the standard errors of the estimates. Specifically, (a) we tried bandwidths as in Section 4 but with the $n^{-1/3}$ factor replaced by $n^{-1/5}$, $n^{-1/7}$ and $n^{-1/9}$ as in Zeng and Lin (2007), and (b) c values 1.2, 1.6, 2.4, 2.8 were considered. Overall, the results are not affected much by choosing different values of h and c in our analysis. For saving space, these results are given in the online Supplementary Appendix.

In Figure 1, we plotted the estimated baseline cumulative hazard function of $\exp(\epsilon)$ under CC1, CC2, and full cohort. The estimated baseline cumulative hazard functions under NCC1 and NCC2 are very close to those under CC1 and CC2, respectively, and thus are omitted here. Overall, the estimated baseline cumulative hazard functions under different sampling designs based on the semiparametric AFT model agree with the corresponding full cohort estimate very well. We also fitted parametric AFT models with the log-normal, Weibull and log-logistic distributions for the full cohort data. The corresponding estimated baseline cumulative hazard function of $\exp(\epsilon)$ were also plotted in Figure 1. The estimated baseline cumulative hazard functions under parametric AFT models show big difference from the full cohort estimate under the semiparametric AFT model. This implies that the considered parametric AFT models do not fit the data well.

In addition, to check the goodness-of-fit of the semiparametric AFT model for the Wilms tumor data, we plotted the martingale residuals versus fitted value $\hat{\beta}'Z$ and the Nelson-Aalen estimator of the cumulative hazard rates of Cox-Snell residuals, which are given in Figure 2.

For the martingale residual plot, no systematic pattern is found. For the Cox-Snell residual plot, the estimated cumulative hazard rates are close to the diagonal line. Both indicate a proper fit of the AFT model to the data. In Lu et al (2014) for testing the goodness-of-fit of the PH model based on Wilms tumor data, the proportional hazard assumption is rejected for covariates NWTSG-4 and Diameter. This may partly explain why the two models (AFT and PH) yield different results for NWTSG-4 and Diameter in the full cohort analysis.

6. Discussions

In this paper, we propose an efficient nonparametric likelihood estimation approach for the inference of the AFT model in CC and NCC studies. The proposed method can be extended to incorporate general outcome-dependent sampling schemes. The sampling probability may depend on observed event times and covariates, such as counter matching and stratified matching. In practice, researchers often want to select controls matched with cases in some confounding variables. With such fine matching, the IPSW method becomes infeasible to implement. Our proposed method can be used to fairly evaluate the efficiency of various risk-set sampling schemes and identify an optimal scheme to complement the classical CC and NCC sampling designs. This warrants future research.

In cohort studies, it is common that extensive baseline or questionnaire data, say W , are available for the full cohort. When these covariates are correlated with the expensive covariates Z that are only measured in the CC and NCC samples, incorporating them in data analysis can improve the efficiency of the CC and NCC estimators. This will require the estimation of the conditional distribution of Z given W assuming a parametric model or using a nonparametric method. This however may be practically challenging when the dimension of W is high and some of them are continuous covariates.

Another common challenge in CC and NCC studies is to handle time-dependent exposure variables $Z(t)$. The proposed nonparametric estimation of Z 's distribution would be difficult to generalize to repeatedly measured covariates. We may need to assume some parametric models, such as a mixed-effects model, for the time-dependent covariate process and then incorporate it into the likelihood based on a joint model. Its associated nonparametric maximum likelihood estimation needs further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank an associate editor and two referees for their thoughtful and constructive comments, which help to improve an earlier version of the paper. This research was partly supported by NIH grants R01 CA140632 and R21 CA169739.

Appendix: Proof of Theorem 1

For saving the space, here we only prove the asymptotic distribution of $\hat{\beta}_N$ given in (ii) of Theorem 1. The proofs of the consistency results established in (i) of Theorem 1 are given in

the online Supplementary Appendix. We first derive the score functions based on the sub-models $\Lambda_\theta(t) = \int_0^t \{1 + \theta h_1(u)\} d\Lambda_0(u)$ and $dF_\theta(z) = \{1 + \theta h_2(z)\} dF_0(z)$, where $h_1 \in BV[0, \tau]$ and $h_2 \in BV(D_Z)$ with $E\{h_2(Z)\} = 0$. Here, $BV(A)$ denote the set of functions defined on A with bounded variation and D_Z is the support of Z . Let $S_\beta = \partial \ell_i^o(\beta, \lambda_0, F_0) / \partial \beta|_{\beta=\beta_0}$, $S_\Lambda(h_1) = \partial \ell_i^o(\beta_0, \lambda_\theta, F_0) / \partial \theta|_{\theta=0}$ and $S_F(h_2) = \partial \ell_i^o(\beta_0, \lambda_0, F_\theta) / \partial \theta|_{\theta=0}$. We have

$$\begin{aligned} & S_\beta(\beta_0, \dot{\lambda}_0, \Lambda_0, F_0) \\ &= \eta_i Z_i B_1(\tilde{T}_i, Z_i, \delta_i; \beta_0, \lambda_0, \dot{\lambda}_0) + (1 - \eta_i) \int_z z B_2(\tilde{T}_i, z; \Lambda_0, \lambda_0, \beta_0) B_3(\tilde{T}_i, z; \Lambda_0, \beta_0, F_0) dF_0(z), \\ & S_\Lambda(h_1)(\beta_0, \Lambda_0, F_0) = \eta_i \left\{ \delta_i h_1(e^{R_i(\beta_0, Z_i)}) - \int_0^{e^{R_i(\beta_0, Z_i)}} h_1(u) d\Lambda_0(u) \right\} \\ & \quad - (1 - \eta_i) \frac{\int_z e^{-\Lambda_0\{e^{R_i(\beta_0, z)}\}} \int_0^{e^{R_i(\beta_0, z)}} h_1(u) d\Lambda_0(u) dF_0(z)}{\int_z e^{-\Lambda_0\{e^{R_i(\beta_0, z)}\}} dF_0(z)}, \\ & S_F(h_2)(\beta_0, \Lambda_0, F_0) = \eta_i h_2(Z_i) + (1 - \eta_i) \frac{\int_z e^{-\Lambda_0\{e^{R_i(\beta_0, z)}\}} h_2(z) dF_0(z)}{\int_z e^{-\Lambda_0\{e^{R_i(\beta_0, z)}\}} dF_0(z)}, \end{aligned}$$

where

$$\begin{aligned} B_1(\tilde{T}_i, Z_i, \delta_i; \beta, \lambda, \dot{\lambda}) &= -\delta_i \left\{ 1 + \frac{\lambda(e^{R_i(\beta, Z_i)})e^{R_i(\beta, Z_i)}}{\lambda(e^{R_i(\beta, Z_i)})} \right\} + \lambda(e^{R_i(\beta, Z_i)})e^{R_i(\beta, Z_i)}, \\ B_2(\tilde{T}_i, z; \Lambda, \lambda, \beta) &= e^{-\Lambda(e^{R_i(\beta, z)})} \lambda(e^{R_i(\beta, z)}) e^{R_i(\beta, z)}, \text{ and } B_3(\tilde{T}_i, z; \Lambda, \beta, F) = \frac{1}{\int_z e^{-\Lambda(e^{R_i(\beta, z)})} dF(z)}. \end{aligned}$$

Define the information operator $\sigma(h_1, h_2) = (S_\beta, S_\Lambda, S_F)^*(S_\beta, S_\Lambda, S_F)$ that maps $R^p \times BV[0, \tau] \times BV(D_Z)$ to itself, where $(S_\beta, S_\Lambda, S_F)^*$ is the dual operator of $(S_\beta, S_\Lambda, S_F)$. Following Murphy et al (1997), we want to show that $\sigma(h_1, h_2)$ is continuously invertible on its range. To do this, we first show that $\sigma(h_1, h_2)$ is a one-to-one map, which is equivalent to show that the equation $\sigma(h_1, h_2) = 0$ has a unique solution $h_1 = h_2 = 0$. Note that $\sigma(h_1, h_2) = 0$ implies $S'_\beta u + S_\Lambda(h_1) + S_F(h_2) = 0$. Setting $\eta_i = 1$ gives, with probability one,

$$B_1(\tilde{T}_i, Z_i, \delta_i; \beta_0, \lambda_0, \dot{\lambda}_0) Z'_i u + \delta_i h_1(e^{R_i(\beta_0, Z_i)}) - \int_0^{e^{R_i(\beta_0, Z_i)}} h_1(u) d\Lambda_0(u) + h_2(Z_i) = 0.$$

Note that $h_2(Z_i)$ is orthogonal to $B_1(\tilde{T}_i, Z_i, \delta_i; \beta_0, \lambda_0, \dot{\lambda}_0)$ and

$\delta_i h_1(e^{R_i(\beta_0, Z_i)}) - \int_0^{e^{R_i(\beta_0, Z_i)}} h_1(u) d\Lambda_0(u)$. Multiplying both sides by $h_2(Z_i)$ and taking the expectation gives $h_2 = 0$. Letting $\delta_i = 1$ gives the equation

$B_1(\tilde{T}_i, Z_i, \delta_i; \beta_0, \lambda_0, \dot{\lambda}_0) Z'_i u + h_1(e^{R_i(\beta_0, Z_i)}) = \int_0^{e^{R_i(\beta_0, Z_i)}} h_1(u) d\Lambda_0(u)$, which is a homogeneous function of h_1 and has a unique solution $h_1 = 0$. It results in $u = 0$. Thus, the information operator is one-to-one. Following Kress (1989), after some tedious calculations, we can show that $\sigma(h_1, h_2)$ can be written as the sum of an invertible and a compact operator. It follows that the information operator is continuously invertible. Therefore, there exist $h_1 \in BV[0, \tau]$ and $h_2 \in BV(D_Z)$ such that $(S_\Lambda, S_F)^*(S_\Lambda, S_F)(h_1, h_2) = (S_\Lambda, S_F)^* S_\beta$. Here, h_1 and h_2 are the least favorable directions for β_0 .

Define $A_{N,1}(\beta, \dot{\lambda}, \lambda, \Lambda, F) = P_N S_\beta(\beta, \dot{\lambda}, \lambda, \Lambda, F)$, $A_{N,2}(\beta, \Lambda, F) = P_N S_\Lambda(h_1)(\beta, \Lambda, F)$, and $A_{N,3}(\beta, \Lambda, F) = P_N S_F(h_2)(\beta, \Lambda, F)$, where P_N denote the empirical measure of observed

data. We have $A_{N,1}(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) = A_{N,2}(\hat{\beta}_N, \hat{\Lambda}_N, \hat{F}_N) = A_{N,3}(\hat{\beta}_N, \hat{\Lambda}_N, \hat{F}_N) = 0$.

Define $A_N(\beta, \dot{\lambda}, \lambda, \Lambda, F) = A_{N,1}(\beta, \dot{\lambda}, \lambda, \Lambda, F) - A_{N,2}(\beta, \Lambda, F) - A_{N,3}(\beta, \Lambda, F)$ and $A(\beta, \dot{\lambda}, \lambda, \Lambda, F) = A_1(\beta, \dot{\lambda}, \lambda, \Lambda, F) - A_2(\beta, \Lambda, F) - A_3(\beta, \Lambda, F)$, where the function A_1 , A_2 and A_3 are the expectations of $A_{N,1}$, $A_{N,2}$ and $A_{N,3}$, respectively. We want to show that

$$N^{1/2} \left\{ A_N(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) - A(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) \right\} = -N^{1/2} A(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) \\ = N^{1/2} (P_n - P) \{ S_\beta - S_\Lambda(h_1) - S_F(h_2) \} + o_p(1).$$

(6)

To show this, we want to verify all the conditions in Theorem 2.11.22 of van der Vaart and Wellner (1996). Define the following classes of functions indexed by $(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N)$:

$$\mathcal{F}_{N,\beta} = \left\{ S_\beta(\beta, \dot{\lambda}, \lambda, \Lambda, F) : \|\beta - \beta_0\| + \|\Lambda - \Lambda_0\| + \|\lambda - \lambda_0\| + \|\dot{\lambda} - \dot{\lambda}_0\| + \|F - F_0\| \leq \varepsilon_0 \right\},$$

$$\mathcal{F}_{N,\Lambda} = \left\{ S_\Lambda(h_1)(\beta, \Lambda, F) : \|\beta - \beta_0\| + \|\Lambda - \Lambda_0\| + \|F - F_0\| \leq \varepsilon_0 \right\},$$

$$\mathcal{F}_{N,F} = \left\{ S_F(h_2)(\beta, \Lambda, F) : \|\beta - \beta_0\| + \|\Lambda - \Lambda_0\| + \|F - F_0\| \leq \varepsilon_0 \right\},$$

where $\|\Lambda - \Lambda_0\|$, $\|\lambda - \lambda_0\|$ and $\|\dot{\lambda} - \dot{\lambda}_0\|$ are the supreme norms on $[0, \tau]$, and $\|F - F_0\|$ is the supreme norm on D_Z . Based on the uniform consistency established in (i), we

can show that $S_\beta(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) \rightarrow S_\beta(\beta_0, \dot{\lambda}_0, \lambda_0, \Lambda_0, F_0)$, and $S_F(\hat{\beta}_N, \hat{\Lambda}_N, \hat{F}_N) \rightarrow S_F(\beta_0, \Lambda_0, F_0)$.

We first show that the classes of functions $\mathcal{F}_{N,\beta}$, $\mathcal{F}_{N,\Lambda}$ and $\mathcal{F}_{N,F}$ satisfy the uniform entropy condition in Theorem 2.11.22 of van der Vaart and Wellner (1996). Consider two functions f_1 and f_2 in $\mathcal{F}_{N,\beta}$, which are indexed by $(\beta_1, \dot{\lambda}_1, \lambda_1, \Lambda_1, F_1)$ and $(\beta_2, \dot{\lambda}_2, \lambda_2, \Lambda_2, F_2)$, respectively. After some algebra, we can show that

$$|f_1 - f_2| \leq M_1 \left\{ \|\beta_1 - \beta_2\| + |\lambda_1(e^{R_i(\beta_2, Z_i)}) - \lambda_2(e^{R_i(\beta_2, Z_i)})| + |\dot{\lambda}_1(e^{R_i(\beta_2, Z_i)}) - \dot{\lambda}_2(e^{R_i(\beta_2, Z_i)})| \right. \\ \left. + \int_z |F_1(z) - F_2(z)| dz + \int_z |\Lambda_1(e^{R_i(\beta_2, z)}) - \Lambda_2(e^{R_i(\beta_2, z)})| dF_2(z) \right. \\ \left. + \int_z |\lambda_1(e^{R_i(\beta_2, z)}) - \lambda_2(e^{R_i(\beta_2, z)})| dF_2(z) \right\},$$

where M_1 is a constant. Then, for any finite discrete probability measure Q , we have

$$\begin{aligned} \|f_1 - f_2\|_{Q,2} &\leq M_2[\|\beta_1 - \beta_2\| + \|\lambda_1(e^{R_i(\beta_2, Z_i)}) - \lambda_2(e^{R_i(\beta_2, Z_i)})\|_{Q,2} \\ &+ \|\dot{\lambda}_1(e^{R_i(\beta_2, Z_i)}) - \dot{\lambda}_2(e^{R_i(\beta_2, Z_i)})\|_{Q,2} + \left\{ \int_z \|\Lambda_1(e^{R_i(\beta_2, z)}) - \Lambda_2(e^{R_i(\beta_2, z)})\|_{Q,2}^2 dF_2(z) \right\}^{1/2} \\ &+ \left\{ \int_z \|\lambda_1(e^{R_i(\beta_2, z)}) - \lambda_2(e^{R_i(\beta_2, z)})\|_{Q,2}^2 dF_2(z) \right\}^{1/2} + \|F_1 - F_2\|_{D_z}] \end{aligned}$$

for some constant M_2 . Let $N(\varepsilon, \mathcal{F}_{N,\beta}, L_2(Q))$ denote the covering number of $\mathcal{F}_{N,\beta}$ with the $L_2(Q)$ norm. We have

$$\begin{aligned} N(\varepsilon, \mathcal{F}_{N,\beta}, L_2(Q)) &\leq \kappa N\{\varepsilon/5, (\beta: \|\beta - \beta_0\| < \varepsilon_0), \|\cdot\|\} \times N\{\varepsilon/5, (F: \|F - F_0\| < \varepsilon_0), \|\cdot\|_{D_z}\} \\ &\times N\{\varepsilon/5, (\Lambda: \|\Lambda - \Lambda_0\| < \varepsilon_0), L_2(Q)\} \times N\{\varepsilon/5, (\lambda: \|\lambda - \lambda_0\| < \varepsilon_0), L_2(Q)\} \\ &\times N\{\varepsilon/5, (\dot{\lambda}: \|\dot{\lambda} - \dot{\lambda}_0\| < \varepsilon_0), L_2(Q)\} \end{aligned}$$

for some constant κ . It is known that for any $\varepsilon > 0$, $N(\varepsilon, \mathcal{F}, \|\cdot\|) \leq N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$, where $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the bracketing number. On the right-hand side of the above inequality, the first covering number for β is $O(1/\varepsilon^p)$ and the other covering numbers are $O(\exp(1/\varepsilon))$ by Theorem 2.7.5 of van der Vaart and Wellner (1996). Thus, the uniform entropy condition holds for $\mathcal{F}_{N,\beta^*}$. The uniform entropy conditions for $\mathcal{F}_{N,\Lambda}$ and $\mathcal{F}_{N,F}$ can be proved in the same manner. Similarly, we can show that the functions in the classes $\mathcal{F}_{N,\beta^*}$, $\mathcal{F}_{N,\Lambda}$ and $\mathcal{F}_{N,F}$ are Lipschitz continuous. Therefore, by Theorem 2.11.22 of van der Vaart and Wellner (1996), equation (6) holds.

In addition, by Taylor expansion, we have

$$\begin{aligned} &-A_1(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) = -A_1(\beta_0, \dot{\lambda}_0, \lambda_0, \Lambda_0, F_0) \\ &-P\left[\eta_i Z_i \left\{ \frac{\partial}{\partial \beta'} B_1(\tilde{T}_i, \delta_i, Z_i; \beta^*, \lambda_0, \dot{\lambda}_0,) \right\} (\hat{\beta}_N - \beta_0)\right] \\ &+ P\left[\eta_i Z_i \int_0^{R_i(\beta_0, Z_i)} \left\{ 1 + \frac{t \dot{\lambda}_0(t)}{\lambda_0(t)} \right\} d\left\{ \hat{\Lambda}_N(t) \right\} - \Lambda_0(t) \right] \\ &-P\left[(1 - \eta_i) \int z \left\{ \frac{\partial}{\partial \beta'} B_2(\tilde{T}_i, z; \beta^*, \lambda_0, \Lambda_0) B_3(\tilde{T}_i, z; \beta^*, \Lambda_0, F_0) \right\} (\hat{\beta}_N - \beta_0) dF_0(z)\right] \\ &-P\left[(1 - \eta_i) \int z \left\{ \frac{\partial}{\partial \Lambda} B_2(\tilde{T}_i, z; \beta_0, \lambda^*, \Lambda^*) B_3(\tilde{T}_i, z; \beta_0, \Lambda^*, F_0) \right\} (\hat{\Lambda}_N - \Lambda_0) dF_0(z)\right] \\ &-P\left[(1 - \eta_i) \int z B_2(\tilde{T}_i, z; \beta_0, \lambda_0, \Lambda_0) \left\{ \frac{\partial}{\partial F} B_3(\tilde{T}_i, z; \beta_0, \Lambda_0, F^*) \right\} (\hat{F}_N - F_0) dF_0(z)\right] \\ &-P\left[(1 - \eta_i) \int z B_2(\tilde{T}_i, z; \beta_0, \lambda_0, \Lambda_0) B_3(\tilde{T}_i, z; \beta_0, \Lambda_0, F_0) \right] d\left\{ \hat{F}_N(z) - F_0(z) \right\} \end{aligned}$$

where $\dot{\lambda} / \Lambda$ and $\dot{\lambda} / F$ denote the path derivatives along the defined submodels, and $(\beta^*, \Lambda^*, F^*)$ lies between $(\hat{\beta}_N, \hat{\Lambda}_N, \hat{F}_N)$ and $(\beta_0, \Lambda_0, F_0)$. Similar expansions can be derived for $A_2(\hat{\beta}_N, \hat{\Lambda}_N, \hat{F}_N)$ and $A_3(\hat{\beta}_N, \hat{\Lambda}_N, \hat{F}_N)$. Therefore,

$$\begin{aligned}
& -N^{1/2}A(\hat{\beta}_N, \hat{\lambda}_N, \hat{\Lambda}_N, \hat{F}_N) = -N^{1/2}A(\beta_0, \lambda_0, \Lambda_0, F_0) \\
& -N^{1/2}(1+o_p(1))E\left\{S_{\beta\beta}(\hat{\beta}_N - \beta_0) + S_{\beta\Lambda}(\hat{\Lambda}_N - \Lambda_0) + S_{\beta F}(\hat{F}_N - F_0)\right. \\
& \quad - S_{\Lambda\beta}(h_1)(\hat{\beta}_N - \beta_0) - S_{\Lambda\Lambda}(h_1, \hat{\Lambda}_N - \Lambda_0) - S_{\Lambda F}(h_1, \hat{F}_N - F_0) \\
& \quad \left. - S_{F\beta}(h_2)(\hat{\beta}_N - \beta_0) - S_{F\Lambda}(h_2, \hat{\Lambda}_N - \Lambda_0) - S_{FF}(h_2, \hat{F}_N - F_0)\right\},
\end{aligned}$$

where $S_{\beta\beta}$ is the derivative of S_β with respect to β , $S_{\beta\Lambda}(h_1)$, and $S_{\beta F}(h_2)$ are the derivatives of S_β with respect to Λ and F along the directions h_1 and h_2 , respectively, and $S_{\Lambda\beta}(h_1)$, $S_{\Lambda\Lambda}(h_1, h_1)$, $S_{\Lambda F}(h_1, h_2)$, $S_{F\beta}(h_2)$, $S_{F\Lambda}(h_2, h_1)$, and $S_{FF}(h_2, h_2)$ are similarly defined.

By choosing h_1 and h_2 as the least favorable directions for β_0 , we have

$$E\left\{S_{\Lambda\Lambda}(h_1, \hat{\Lambda}_N - \Lambda_0) + S_{\Lambda F}(h_1, \hat{F}_N - F_0)\right\} = E\left\{S_{\beta\Lambda}(\hat{\Lambda}_N - \Lambda_0)\right\} \text{ and}$$

$$E\left\{S_{F\Lambda}(h_2, \hat{\Lambda}_N - \Lambda_0) + S_{FF}(h_2, \hat{F}_N - F_0)\right\} = E\left\{S_{\beta F}(\hat{F}_N - F_0)\right\}. \text{ In addition,}$$

$A(\beta_0, \lambda_0, \Lambda_0, F_0) = 0$. Thus, the right-hand side of the above equation can be written as

$\sum N^{1/2}(\hat{\beta}_N - \beta_0)$, where $\Sigma = -E\{S_{\beta\beta} - S_{\Lambda\beta}(h_1) - S_{F\beta}(h_2)\}$ is an invertible matrix due to the invertibility of the information operator. Based on (6), it follows that

$$N^{1/2} \sum (\hat{\beta}_N - \beta_0) = N^{1/2}(P_n - P)\{S_\beta - S_\Lambda(h_1) - S_F(h_2)\} + o_p(1).$$

Therefore, $N^{1/2}(\hat{\beta}_N - \beta_0)$ converges in distribution to a mean-zero normal random vector with the covariance matrix Σ^{-1} .

References

- Chen HY. Weighted semiparametric likelihood method for fitting a proportional odds regression model to data from the case-cohort design. *Journal of the American Statistical Association*. 2001a; 96:1446–1458.
- Chen HY. Fitting semiparametric transformation regression models to data from a modified case-cohort design. *Biometrika*. 2001b; 88:255–268.
- D'Angio GJ, Breslow N, Beckwith JB, Evans A, Baum H, de Lorimier A, Ferbach D, Hrabovsky E, Jones G, Kelalis P. Treatment of Wilms tumor. Results of the third national Wilms tumor study. *Cancer*. 1989; 64:349–360. [PubMed: 2544249]
- Ernster VL. Nested case-control studies. *Preventive Medicine*. 1994; 23:587–590. [PubMed: 7845919]
- Green DM, Breslow NE, Beckwith JB, Finklestein JZ, Grundy PG, Thomas PRM, Kim T, Shochat S, Haase GM, Ritchey ML, Kelalis PP, D'Angio GJ. Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms tumor: a report from the national Wilms tumor study group. *Journal of Clinical Oncology*. 1998; 16:237–245. [PubMed: 9440748]
- Jones MC. The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters*. 1990; 9:129–132.
- Jones MC, Sheather SJ. Using non-Stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics and Probability Letters*. 1991; 11:511–514.
- Kong L, Cai J, Sen PK. Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika*. 2004; 91:305–319.
- Kong L, Cai J. Case-cohort analysis with accelerated failure time model. *Biometrics*. 2009; 65:135–142. [PubMed: 18537948]

- Kress, R. Linear Integral Equations. New York: Springer-Verlag; 1989.
- Kulich M, Lin DY. Additive hazards regression for case-cohort studies. *Biometrika*. 2000; 87:73–87.
- Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *American Journal of Epidemiology*. 1990; 131:169–176. [PubMed: 2403467]
- Lu W. Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica*. 2010; 20:661–674. [PubMed: 20414460]
- Lu W, Liu M, Chen YH. Testing goodness-of-fit for the proportional hazards model based on nested case-control data. *Biometrics*. 2014; 70:845–851. [PubMed: 25298193]
- Lu W, Liu M. On estimation of linear transformation models with nested case-control sampling. *Lifetime Data Anal*. 2012; 18:80–93. [PubMed: 21912975]
- Lu W, Tsiatis AA. Semiparametric transformation models for the case-cohort Study. *Biometrika*. 2006; 93:207–214.
- Liu B, Lu W, Zhang J. Kernel smoothed profile likelihood estimation in the accelerated failure time frailty model for clustered survival data. *Biometrika*. 2013; 100:741–755. [PubMed: 24443587]
- Nan B, Yu M, Kalbfleisch JD. Censored linear regression for case-cohort studies. *Biometrika*. 2006; 93:747–762.
- Murphy SA, Rossini AJ, van der Vaart AW. Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*. 1997; 92:968–976.
- Murphy SA, van der Vaart AW. Semiparametric mixtures in case-control studies. *Journal of Multivariate Analysis*. 2001; 79:1–32.
- Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73:1–11.
- Reid N. A conversation with Sir David Cox. *Statistical Science*. 1994; 9:439–455.
- Samuelsen SO. A pseudo-likelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84:379–394.
- Scheike TH, Juul A. Maximum likelihood estimation for Cox's regression model under nested case-control sampling. *Biostatistics*. 2004; 5:193–206. [PubMed: 15054025]
- Scheike TH, Martinussen T. Maximum likelihood estimation for Cox's regression model under case-cohort sampling. *Scandinavian Journal of Statistics*. 2004; 31:283–293.
- Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort Studies. *The Annals of Statistics*. 1988; 16:64–81.
- Thomas, D. Addendum to: Methods of cohort analysis: appraisal by application to asbestos mining. In: Liddell, FDK, McDonald, JC., Thomas, DC., editors. *Journal of the Royal Statistical Society: Series A*. Vol. 140. 1977. p. 469–491.
- van der Vaart, AW., Wellner, JA. *Weak Convergence and Empirical Processes*. New York: Springer; 1996.
- Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*. 1997; 92:945–959.
- Zeng D, Lin DY. Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*. 2007; 102:1387–1396.
- Zeng D, Lin DY. Efficient estimation of semiparametric transformation models for two-phase cohort studies. *Journal of the American Statistical Association*. 2014; 109:371–383. [PubMed: 24659837]

Figure 1.
Estimated baseline cumulative hazard function for Wilms tumor data.

Figure 2.
Martingale and Cox-Snell residuals plots for Wilms tumor data.

Table 1

Simulation results for Scenarios I and II.

full cohort				case-cohort				nested-case control			
Parameters	Est	SD		Est	SD	SE	CP	Est	SD	SE	CP
Scenario I											
ε : Normal											
β	1.076	0.128	S1	1.013	0.173	0.178	0.952	1.015	0.172	0.179	0.952
			S2	1.020	0.168	0.176	0.960	1.021	0.160	0.178	0.966
ε : Logistic											
β	1.084	0.146	S1	1.015	0.185	0.176	0.956	1.001	0.174	0.179	0.958
			S2	1.021	0.152	0.175	0.962	1.014	0.172	0.177	0.956
ε : Extreme Value											
β	1.075	0.151	S1	1.014	0.170	0.166	0.936	1.009	0.165	0.167	0.950
			S2	1.021	0.153	0.165	0.966	1.014	0.159	0.166	0.956
Scenario II											
ε : Normal											
β	1.085	0.146	S1	0.992	0.173	0.199	0.966	0.960	0.164	0.195	0.946
			S2	1.029	0.166	0.189	0.960	1.002	0.162	0.186	0.966
ε : Logistic											
β	1.066	0.129	S1	1.000	0.172	0.187	0.954	0.964	0.187	0.185	0.948
			S2	1.024	0.166	0.179	0.954	1.002	0.174	0.176	0.960
ε : Extreme Value											
β	1.075	0.151	S1	0.972	0.181	0.177	0.940	0.974	0.194	0.186	0.934
			S2	1.002	0.171	0.168	0.944	0.978	0.182	0.178	0.940

[†] Est, mean of the estimates; SD, sample standard deviation of the estimates; SE, mean of the estimated standard errors; CP, empirical coverage probability of 95% confidence intervals; S1 stands for CC1 or NCC1 designs; S2 stands for CC2 or NCC2 designs.

Table 2

Simulation results for Scenario III.

full cohort				case-cohort				nested-case control			
Parameters	Est	SD		Est	SD	SE	CP	Est	SD	SE	CP
ε : Normal											
β_1	1.047	0.109	S1	0.957	0.233	0.220	0.970	0.945	0.186	0.203	0.954
β_2	1.056	0.108		0.968	0.195	0.235	0.952	0.976	0.196	0.235	0.952
β_1			S2	1.009	0.164	0.175	0.948	0.962	0.176	0.196	0.954
β_2				1.019	0.171	0.197	0.962	0.985	0.175	0.196	0.956
β_1	IPSW		S1	0.967	0.229			0.999	0.227		
β_2				0.974	0.314			1.019	0.311		
β_1	IPSW		S2	0.966	0.196			1.004	0.205		
β_2				0.975	0.274			1.010	0.293		
ε Logistic											
β_1	1.061	0.127	S1	1.000	0.183	0.199	0.958	0.964	0.181	0.198	0.964
β_2	1.061	0.129		1.005	0.197	0.207	0.958	0.985	0.204	0.209	0.964
β_1			S2	1.033	0.176	0.187	0.954	1.006	0.169	0.186	0.956
β_2				1.024	0.183	0.195	0.948	1.009	0.181	0.196	0.962
β_1	IPSW		S1	1.037	0.259			1.027	0.264		
β_2				1.054	0.349			1.013	0.334		
β_1	IPSW		S2	1.029	0.223			1.029	0.225		
β_2				1.020	0.313			1.019	0.326		
ε : Extreme Value											
β_1	1.029	0.123	S1	0.996	0.181	0.206	0.968	0.959	0.199	0.204	0.944
β_2	0.992	0.144		0.985	0.194	0.226	0.964	0.965	0.203	0.228	0.956

full cohort			case-cohort				nested-case control			
Parameters	Est	SD	Est	SD	SE	CP	Est	SD	SE	CP
β_1			1.031	0.169	0.176	0.944	1.005	0.174	0.174	0.942
β_2		S2	1.017	0.184	0.200	0.951	0.996	0.184	0.199	0.966
β_1	IPSW		1.022	0.261			1.026	0.256		
β_2		S1	1.012	0.351			1.000	0.370		
β_1	IPSW		1.021	0.229			1.021	0.241		
β_2		S2	1.015	0.316			1.008	0.328		

[†]Est, mean of the estimates; SD, sample standard deviation of the estimates; SE, mean of the estimated standard errors; CP, empirical coverage probability of 95% confidence intervals; S1 stands for CC1 or NCC1 designs; S2 stands for CC2 or NCC2 designs; IPSW, inverse probability of selection weighted estimator.

Table 3

The analysis results for Wilms tumor data.

	Full cohort		case-cohort			
	AFT		CC1		CC2	
	Est	SE	Est	SE	Est	SE
NWTSG-4	0.284	0.037	0.285	0.079	0.297	0.068
Stage II	-0.124	0.046	-0.158	0.114	-0.168	0.111
Stage III	-0.187	0.047	-0.184	0.113	-0.205	0.111
Stage IV	-0.428	0.053	-0.545	0.120	-0.514	0.119
UH	-0.460	0.036	-0.402	0.088	-0.379	0.084
Diameter	-0.017	0.005	-0.016	0.006	-0.017	0.006

	Full cohort		nested case control			
	PH		NCC1		NCC2	
	Est	SE	Est	SE	Est	SE
NWTSG-4	-0.138	0.116	0.290	0.079	0.300	0.078
Stage II	0.358	0.151	-0.152	0.116	-0.165	0.114
Stage III	0.636	0.156	-0.187	0.114	-0.207	0.112
Stage IV	0.991	0.178	-0.533	0.121	-0.506	0.108
UH	1.280	0.148	-0.383	0.092	-0.366	0.081
Diameter	1.062	0.557	-0.016	0.006	-0.017	0.006