Statistics and Its Interface Volume 11 (2018) 657–668

Case-cohort design for accelerated hazards model

Jieli Ding[†], Xiaolong Chen, Huaying Fang, and Yanyan Liu[*,‡]

Case-cohort design is widely used in biomedical studies of rare diseases as an efficient way to reduce cost. Relevant covariate histories, which are costly or difficult to obtain, are observed only on cases and a random subcohort in such studies. It is often that a lag period exists before the treatment or other covariates is fully effective. This phenomenon may be described well by an accelerated hazards model. Existing methods for the accelerated hazards model do not handle case-cohort data. This paper proposes a semiparametric inference method for the accelerated hazards model with data from a case-cohort design. The proposed estimators are shown to be consistent and asymptotically normally distributed. The finite sample properties of proposed case-cohort estimator and its relative efficiency to full cohort estimator are assessed via simulation studies. An application to a burn study demonstrates the utility of the proposed method in practice.

AMS 2000 subject classifications: Primary 62Nxx, 62D05; secondary 62N01, 62N02.
Keywords and phrases: Case-cohort design, Accelerated hazards model, Estimating equation.

1. INTRODUCTION

Epidemiologic cohort studies often involve following of a large number of subjects for a long period of time. Obtaining expensive covariate information on all members in such studies might not be feasible due to a limited budget. Case-cohort design (Prentice, 1986) provides an efficient solution to reduce cost for large cohort studies involving rare diseases (cases). Under the case-cohort design, the complete information of covariates is only assembled for a random sample of the entire cohort (subcohort) and all the subjects who experience the event of interest. In addition to being cost efficient, the case-cohort design can be used when survival times to different diseases are of interest (Self and Prentice, 1988).

Extensive inference methodologies for case-cohort designs are based on the proportional hazards (PH) model (e.g. Self and Prentice, 1988; Lin and Ying, 1993; Chen and Lo, 1999; Borgan et al., 2000; Chen, 2001; Cai and Zeng, 2004; Kulich and Lin, 2004; Qi et al., 2005; Breslow and Wellner, 2007), the additive hazards (AdH) model (e.g. Kang and Cai, 2013), the accelerated failure time (AFT) model (e.g. Kong and Cai, 2009; Kang et al., 2016 etc) and the semiparametric transformation models (Chen, 2001; Kong, Cai and Sen, 2004). All the aforementioned semiparametric models assume an immediate treatment effect at the start time of clinical trials. However, in many randomized clinical trials with the goal to compare a treatment with a placebo, it is reasonable to assume that the risks of failure for both the treatment and placebo groups are the same at the start time and change as the trial proceeds. To accommodate this phenomenon, Chen and Wang (2000) proposed a so-called accelerated hazards (AcH) model. In practical studies, a crossing in survivor curves may occur when one group receives oral medication, while the other receives a riskier treatment which involves surgery. The AcH model can also characterize exhibition of crossings in either or both of survivor and hazard functions for the treatment and control groups, a feature that the PH, AdH and AFT models cannot accommodate.

For the AcH model, the statistical inference has been developed when data are complete (Chen and Wang, 2000). In this paper, we consider a more generalized AcH model than that of Chen and Wang (2000) by including other confounding covariates effects into the model. We refer to the proposed model still as the AcH model in the following. In Section 2, we propose an estimating procedure for the regression parameters. We study the asymptotic properties of the proposed estimators and develop an easy resampling approach to estimate the asymptotic covariance in Section 3. In Section 4, simulation studies are conducted to investigate the performance of the case-cohort estimator under practical sample sizes, as well as its efficiency relative to the full-cohort estimator. A real data from a burn study is used to illustrate the proposed method in Section 5. A discussion concludes in Section 6. The outline of the proofs is provided in Appendix A.

2. ESTIMATION PROCEDURES

2.1 Model and procedures

Let T be the failure time and C be the censoring time. The observed time is $U = \min(T, C)$ and $\Delta = I(T \leq C)$ denotes the right-censoring indicator, where $I(\cdot)$ is the indicator function. Let $\boldsymbol{Z} = (Z_1, Z_2)$ be a $(p+1)$-dimensional covariate vector where Z_1 is a binary treatment effect with

*Corresponding author.
†This research is supported by the National Science Foundation of China 11671310 to J.D.
‡This research is supported by the National Science Foundation of China 11571263, 11371299 to Y.L.

values 0 and 1 and Z_2 is a p-vector of potential confounding variables or risk factors. Assume that T and C are conditionally independent given $\boldsymbol{Z}$. Suppose that the underlying cohort consists of N independent subjects and $(U_i, \Delta_i, \boldsymbol{Z}_i)$, $i = 1, \cdots, N$, are N independent copies of $(U, \Delta, \boldsymbol{Z})$. The realization of $(U_i, \Delta_i, \boldsymbol{Z}_i)$ is denoted by $(u_i, \delta_i, \boldsymbol{z}_i)$.

The most widely used model for survival data is the Cox proportional hazard model (Cox, 1972) with the hazard function

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp(\beta'\boldsymbol{Z}), \tag{1}$$

where β is the relative risk parameter to be estimated, and $\lambda_0(t)$ is an unspecified baseline hazard function. This model assumes that the hazard curves are proportional for individuals with different covariates and the unknown parameters can be estimated by solving the following partial likelihood score equation:

$$\sum_{i=1}^{N}\delta_i\left[\boldsymbol{z}_i - \frac{\sum_{j=1}^{N} I(u_j \geq u_i)\exp(\beta'\boldsymbol{z}_i)\boldsymbol{z}_i}{\sum_{j=1}^{N} I(u_j \geq u_i)\exp(\beta'\boldsymbol{z}_j)}\right] = 0.$$

In practical studies, it is often happens that the hazards of treatment group and non-treatment group are identical at the starting time ($t = 0$). To accommodate this nonproportionality phenomenon, Chen and Wang (2000) proposed the so-called accelerated hazards (AcH) model as follows:

$$\lambda(t|Z_1, Z_2) = \lambda_0(\beta_1^{Z_1}t)\exp(\beta_2'Z_2), \tag{2}$$

where $\lambda_0(\cdot)$ is the unspecified baseline hazard function and $\boldsymbol{\beta} = (\beta_1, \beta_2')' \in \mathcal{B}_1 \times \mathcal{B}_2 \subset \mathbb{R}^+ \times \mathbb{R}^p$ is the unknown parameter. It can be seen that the above AcH model assumes that the binary treatment Z_1 accelerates or decelerates the baseline hazard progression while the other covariates Z_2 influence the baseline hazard function proportionally.

Note that the only difference between $\lambda(t|Z_1 = 0, Z_2)$ and $\lambda(t|Z_1 = 1, Z_2)$ is a time scale change. Therefore, the following proposition can be directly obtained as a simple generalization of Property 1 in Chen and Wang (2000).

Proposition 2.1. *If β_a is an arbitrary scale in $\mathcal{B}_1 \subset \mathbb{R}^+$ and the failure time T has the hazard function $\lambda(t|\boldsymbol{Z})$ in (2), then the hazard function for the transformed time $T_a = \beta_a^{Z_1}T$ is*

$$\lambda_0(\{\beta_1/\beta_a\}^{Z_1}t)\exp\{Z_2'\beta_2 - Z_1\log\beta_a\}. \tag{3}$$

Obviously, when $\beta_a = \beta_1$, (3) equals

$$\lambda_0(t)\exp\{Z_2'\beta_2 - Z_1\log\beta_1\}.$$

Proposition 2.1 means that T_a recovers the proportionality between the hazard functions with a ratio of $\exp\{Z_2'\beta_2 - Z_1\log\beta_1\}$ when β_1 is used to transform $(T, \boldsymbol{Z})$ to $(T_a, \boldsymbol{Z})$. Inspired by this fact, for the full-cohort design, we extend the approach by Chen and Wang (2000) to estimate the parameter $\boldsymbol{\beta} = (\beta_1, \beta_2')'$ by solving the following equations jointly:

$$\begin{cases}\sum_{i=1}^{N}\delta_i\left[\boldsymbol{z}_i - \frac{\sum_{j=1}^{N} I(\beta_a^{z_{1j}}u_j \geq \beta_a^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}\boldsymbol{z}_j}{\sum_{j=1}^{N} I(\beta_a^{z_{1j}}u_j \geq \beta_a^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}}\right] = 0,\\ \beta_1 = \beta_a.\end{cases} \tag{4}$$

It is easy to see that the solution to (4) can be equivalently obtained by solving the following equation:

$$\sum_{i=1}^{N}\delta_i\left[\boldsymbol{z}_i - \frac{\sum_{j=1}^{N} I(\beta_1^{z_{1j}}u_j \geq \beta_1^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}\boldsymbol{z}_j}{\sum_{j=1}^{N} I(\beta_1^{z_{1j}}u_j \geq \beta_1^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}}\right] = 0. \tag{5}$$

Chen and Wang (2000) proposed a grid search algorithm to realize the calculation.

In case-cohort studies, covariate observation is not completely available for each subject. Suppose a subcohort of size $\widetilde{n}$ is randomly selected without replacement from the full cohort. The subjects from the subcohort and the additional failures outside the subcohort constitute the case-cohort sample, which is supposed to have the size of n. Let $\widetilde{\mathcal{C}}$ and $\mathcal{C}$ be the index set of the subcohort and the case-cohort sample, respectively. Since covariate measurements are collected only for the case-cohort sample, equation (5) cannot be calculated directly. Following the idea of Prentice (1986), we proposed the following case-cohort estimating equation:

$$U(\boldsymbol{\beta}) = \sum_{i\in\mathcal{C}}\delta_i\left[\boldsymbol{z}_i - \frac{\sum_{j\in\mathcal{R}(u_i)} I(\beta_1^{z_{1j}}u_j \geq \beta_1^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}\boldsymbol{z}_j}{\sum_{j\in\mathcal{R}(u_i)} I(\beta_1^{z_{1j}}u_j \geq \beta_1^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}}\right] = 0, \tag{6}$$

where $\mathcal{R}(t) = \widetilde{\mathcal{C}} \cup \mathcal{D}(t)$ and $\mathcal{D}(t) = \{i = 1, \cdots, N \,|\, N_i(t) \neq N_i(t^-)\}$, which is empty unless a failure occurs at time t. Note that the proposed estimating equation is a modification of equation (5) that weights the contributions of the failures and subcohort differently.

Denote $N_i(t) = \Delta_i I(U_i \leq t)$ and $Y_i(t) = I(U_i \geq t)$ to be the counting process and the at-risk process, respectively. Equation (6) can be rewritten as:

$$U(\boldsymbol{\beta}) = \sum_{i\in\mathcal{C}}\int_0^{\tau}\left[\boldsymbol{z}_i - \frac{\sum_{j\in\mathcal{R}(t)} Y_j\left(t/{\beta_1}^{z_{1j}}\right)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}\boldsymbol{z}_j}{\sum_{j\in\mathcal{R}(t)} Y_j\left(t/{\beta_1}^{z_{1j}}\right)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}}\right] \times dN_i\left(t/{\beta_1}^{z_{1i}}\right) = 0,$$

where τ denotes the end time of the study. The proposed estimator of $\boldsymbol{\beta}$ is the solution to the above estimating equation, denoted by $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1,\ \widehat{\beta}_2')'$.

Furthermore, the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ can be estimated by a Breslow-Aalen type estimator as:

$$\widehat{\Lambda}_0(t) = \frac{\widetilde{n}}{n}\int_0^t \frac{\sum_{i\in\mathcal{C}} dN_i(s/\widehat{\beta}_1^{z_{1i}})}{\sum_{j\in\widetilde{\mathcal{C}}} Y_j(s/\widehat{\beta}_1^{z_{1j}})\ e^{z_{2j}'\widehat{\beta}_2 - z_{1j}\log\widehat{\beta}_1}}. \tag{7}$$

The asymptotic properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(t)$ are presented in Section 3.

2.2 Implementation

The computation of the proposed estimator $\widehat{\boldsymbol{\beta}}$ in practice is quite complicated because equation (6) is not smooth with respect to β_1. We establish a greed search algorithm for the implementation of $\widehat{\boldsymbol{\beta}}$ by modifying the approach developed by Chen and Wang (2000).

Algorithm.

Step 1: Choose an arbitrary positive real number β_a and transform all observed time u_i's to $\beta_a^{z_{1i}}u_i$, $i\in\mathcal{C}$.
Step 2: Consider the working model as follows:

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{Z_2'\beta_2 - Z_1\log\beta_1\}, \tag{8}$$

and solve the estimating equation for the fixed β_a,

$$\begin{aligned} &U(\boldsymbol{\beta};\beta_a) \\ &= \sum_{i\in\mathcal{C}}\delta_i\left[\boldsymbol{z}_i - \frac{\sum_{j\in\mathcal{R}(u_i)} I(\beta_a^{z_{1j}}u_j \ge \beta_a^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}\boldsymbol{z}_j}{\sum_{j\in\mathcal{R}(u_i)} I(\beta_a^{z_{1j}}u_j \ge \beta_a^{z_{1i}}u_i)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}}\right] \\ &= 0, \end{aligned} \tag{9}$$

to obtain the solution as $\widehat{\boldsymbol{\beta}}(\beta_a) = (\widehat{\beta}_1(\beta_a),\ \widehat{\beta}_2(\beta_a)')'$.
Step 3: Repeat Step 1 and Step 2 until $\widehat{\boldsymbol{\beta}}_n$ is found to satisfy

$$\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_n) = \widehat{\boldsymbol{\beta}}_n \text{ or } \left[\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_n+) - \widehat{\boldsymbol{\beta}}_n\right]\left[\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_n-) - \widehat{\boldsymbol{\beta}}_n\right] < 0.$$

To realize the above algorithm, we adopt a direct grid search technique initially proposed by Chen and Wang (2000) for independent and identically distributed observations. Specifically, we set β_a from a starting point a to an ending point b with a constant increment η, and calculate the solution, $\widehat{\boldsymbol{\beta}}(\beta_a^{(k)}) = (\widehat{\beta}_1(\beta_a^{(k)}), \widehat{\beta}_2(\beta_a^{(k)})')'$, to equation (9) for each $\beta_a^{(k)} = a + k\eta$. Locate the solution by searching the crossing of $y = \widehat{\beta}_1(\beta_a)$ and $y = \beta_a$. As far as the choice of a and b is concerned, we first estimate the regression parameter based on the subcohort subjects under the working model (8) which takes a proportional hazards form as

$$\lambda(t|\boldsymbol{W}) = \lambda_0(t)\exp(\boldsymbol{W}'\boldsymbol{\gamma}),$$

where $\boldsymbol{W} = (Z_2',\ Z_1)'$ and $\boldsymbol{\gamma} = ({\beta_2}',\ -\log\beta_1)'$. The resulting estimator of $\boldsymbol{\gamma}$ is denoted by $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1',\ \widehat{\gamma}_2)'$. We then set $a = \exp\{-\widehat{\gamma}_2\}(1 - 3\widehat{\sigma}_{\gamma_2})$ and $b = \exp\{-\widehat{\gamma}_2\}(1 + 3\widehat{\sigma}_{\gamma_2})$, where $\widehat{\sigma}_{\gamma_2}$ is the estimated standard error of $\widehat{\gamma}_2$.

3. ASYMPTOTIC PROPERTIES

3.1 Asymptotic properties for $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\Lambda}_0(t)$

Let $\boldsymbol{\beta}_0$ denote the true value of $\boldsymbol{\beta}$. We first introduce some notations. For $d = 0,\ 1,\ 2$, define

$$\begin{aligned} S^{(d)}(\boldsymbol{\beta}, t) =& \frac{1}{n}\sum_{i\in\mathcal{C}} Y_i\left(t/{\beta_1}^{z_{1i}}\right)e^{z_{2i}'\beta_2 - z_{1i}\log\beta_1}{\boldsymbol{z}_i}^{\otimes d}, \\ \widetilde{S}^{(d)}(\boldsymbol{\beta}, t) =& \frac{1}{\widetilde{n}}\sum_{i\in\widetilde{\mathcal{C}}} Y_i\left(t/{\beta_1}^{z_{1i}}\right)e^{z_{2i}'\beta_2 - z_{1i}\log\beta_1}{\boldsymbol{z}_i}^{\otimes d}, \\ Q^{(d)}(\boldsymbol{\beta}, t, w) =& \frac{1}{n}\sum_{i\in\mathcal{C}} Y_i\left(t/{\beta_1}^{z_{1i}}\right)Y_i\left(w/{\beta_1}^{z_{1i}}\right) \\ &\times e^{2\left\{z_{2i}'\beta_2 - z_{1i}\log\beta_1\right\}}{\boldsymbol{z}_i}^{\otimes d}, \\ \widetilde{Q}^{(d)}(\boldsymbol{\beta}, t, w) =& \frac{1}{\widetilde{n}}\sum_{i\in\widetilde{\mathcal{C}}} Y_i\left(t/{\beta_1}^{z_{1i}}\right)Y_i\left(w/{\beta_1}^{z_{1i}}\right) \\ &\times e^{2\left\{z_{2i}'\beta_2 - z_{1i}\log\beta_1\right\}}{\boldsymbol{z}_i}^{\otimes d}, \end{aligned}$$

where $\nu^{\otimes 0} = 1$, $\nu^{\otimes 1} = \nu$ and $\nu^{\otimes 2} = \nu\nu'$ for a vector ν. Define

$$\bar{Z}(\boldsymbol{\beta}, t) = \frac{S^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)}, \quad \widetilde{Z}(\boldsymbol{\beta}, t) = \frac{\widetilde{S}^{(1)}(\boldsymbol{\beta}, t)}{\widetilde{S}^{(0)}(\boldsymbol{\beta}, t)}.$$

The risk set of the entire cohort involved in the proposed equation (6) can be approximated by their subcohort counterparts, and the efficiency loss of the resulting estimators relative to the estimator obtained from (6) is ignorable especially for large cohorts with infrequent failure occurrence (Self and Prentice, 1988). In this spirit, we approximate $U(\boldsymbol{\beta})$ by using the index set $\widetilde{\mathcal{C}}$ instead of $\mathcal{R}(t)$, and construct the estimating function:

$$\begin{aligned} \widetilde{H}_n(\boldsymbol{\beta}) =& \sum_{i=1}^n\int_0^\tau\left[\boldsymbol{z}_i - \frac{\sum_{j\in\widetilde{\mathcal{C}}} Y_j\left(t/{\beta_1}^{z_{1j}}\right)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}\boldsymbol{z}_j}{\sum_{j\in\widetilde{\mathcal{C}}} Y_j\left(t/{\beta_1}^{z_{1j}}\right)e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}}\right] \\ &\times dN_i\left(t/{\beta_1}^{z_{1i}}\right). \end{aligned} \tag{10}$$

Suppose that $\widetilde{\boldsymbol{\beta}}$ is the solution to $\widetilde{H}_n(\boldsymbol{\beta}) = 0$, and $\widetilde{\Lambda}_0(t)$ is the corresponding estimator of $\Lambda_0(t)$ by replacing $\widehat{\boldsymbol{\beta}}$ in (7) with $\widetilde{\boldsymbol{\beta}}$. It can be proved that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(t)$ have the same asymptotic properties as $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\Lambda}_0(t)$ by similar arguments of Self and Prentice (1988). Therefore, it is sufficient to prove the asymptotic properties of $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\Lambda}_0(t)$.

We present the main results and outline the conditions and proofs in Appendix A.

Theorem 3.1. *Under Conditions A) to G) in Appendix A, $\widetilde{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$, and*

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0,\ \Sigma^{-1}\left\{\Omega + \Omega_G\right\}(\Sigma')^{-1}),$$

where

$$\Omega = \int_0^\tau \left[s^{(2)}(\boldsymbol{\beta}_0, t) - \frac{s^{(1)}(\boldsymbol{\beta}_0, t)^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] \lambda_0(t)dt,$$

$$\Omega_G = \int_0^\tau\!\!\int_0^\tau G(\boldsymbol{\beta}_0, t, w) s^{(0)}(\boldsymbol{\beta}_0, t) s^{(0)}(\boldsymbol{\beta}_0, w)\lambda_0(t)\lambda_0(w)dtdw,$$

and

$$\begin{aligned} G(\boldsymbol{\beta}_0, t, w) =& (1-\rho)\rho^{-1}\left\{ s^{(0)}(\boldsymbol{\beta}_0, t) s^{(0)}(\boldsymbol{\beta}_0, w)\right\}^{-1} \\ &\times \Bigg\{ q^{(2)}(\boldsymbol{\beta}_0, t, w) - q^{(1)}(\boldsymbol{\beta}_0, w, t)\left[\frac{s^{(1)}(\boldsymbol{\beta}_0, w)}{s^{(0)}(\boldsymbol{\beta}_0, w)}\right]' \\ &- \left[\frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)}\right] q^{(1)}(\boldsymbol{\beta}_0, t, w)' \\ &+ q^{(0)}(\boldsymbol{\beta}_0, t, w)\left[\frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)}\right]\left[\frac{s^{(1)}(\boldsymbol{\beta}_0, w)}{s^{(0)}(\boldsymbol{\beta}_0, w)}\right]' \Bigg\}, \end{aligned}$$

where the definition of $s^{(d)}(\boldsymbol{\beta}, t)$ and $q^{(d)}(\boldsymbol{\beta}, t, w)$ ($d = 0, 1, 2$) are given in Appendix A.

Let $s^{(1)}(\boldsymbol{\beta}, t) = (s_1^{(1)}(\boldsymbol{\beta}, t),\ s_2^{(1)}(\boldsymbol{\beta}, t))$ be the decomposition of the vector $s^{(1)}(\boldsymbol{\beta}, t)$ according to the dimensions of β_1 and β_2. The following theorem gives the asymptotic property of $\widetilde{\Lambda}_0(t)$.

Theorem 3.2. *Under Conditions A) to G) in Appendix A, $\widetilde{\Lambda}_0(t)$ converges uniformly in $t \in [0, \tau/\delta]$ to zero in probability, and $\sqrt{n}\left\{\widetilde{\Lambda}_0(t) - \Lambda_0(t)\right\}$ converges weakly to a zero-mean Gaussian process with the covariance function at (t, w) equals to*

$$\Psi(t)\Sigma^{-1}\left\{\Omega + \Omega_G\right\}(\Sigma')^{-1}\Psi(w)' + \int_0^{\min(t,w)} \frac{\lambda_0(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} du,$$

where

$$\Psi(t) = \left(\int_0^t \frac{s_1^{(1)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} d\left\{\lambda_0(u)u\right\},\ \int_0^t \frac{s_2^{(1)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} \lambda_0(u)du \right).$$

Remark. When the data is complete, i.e. $\rho = 1$, the asymptotic variances are $\Sigma^{-1}\Omega(\Sigma')^{-1}$ for $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\Psi(t)\Sigma^{-1}\Omega(\Sigma')^{-1}\Psi(w)' + \int_0^{\min(t,w)} \frac{\lambda_0(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} du$ for $\sqrt{n}\left\{\widetilde{\Lambda}_0(t) - \Lambda_0(t)\right\}$ at (t, w), which are coincident with the results of Chen and Wang (2000).

3.2 Estimation of the asymptotic covariance of $\widetilde{\boldsymbol{\beta}}$

Matrice Ω and Ω_G can be estimated straightforwardly by replacing the population quantities with their corresponding sample quantities. To obtain the estimator of Σ, we need to estimate the unknown baseline hazard function $\lambda_0(t)$ and its derivative. There are some possible approaches, such as nonparametric kernel density estimation (Tsiatis, 1990), computer-intensive resampling algorithm (Parzen et al., 1994), and recursive bisection method (Chen and Jewell, 2001), etc. Here we adopt a least square (LS) approach proposed by Zeng and Lin (2008), which was originally designed for nonsmooth estimating functions. One advantage of LS approach is that it is easy to implement.

Define $\bar{N}(t) = \sum_{i \in \mathcal{C}} N_i(t)$. Write $S^{(1)}(\boldsymbol{\beta}, t) = (S_1^{(1)}(\boldsymbol{\beta}, t),\ S_2^{(1)}(\boldsymbol{\beta}, t))$, and $S^{(2)}(\boldsymbol{\beta}, t) = (S_1^{(2)}(\boldsymbol{\beta}, t),\ S_2^{(2)}(\boldsymbol{\beta}, t))$, $\Sigma_n(\boldsymbol{\beta}) = (\Sigma_{1n}(\boldsymbol{\beta}),\ \Sigma_{2n}(\boldsymbol{\beta}))$, according to the dimensions of β_1 and β_2 in the following, where

$$\Sigma_{1n}(\boldsymbol{\beta}) = \int_0^\tau \left[S_1^{(2)}(\boldsymbol{\beta}, t) - S^{(1)}(\boldsymbol{\beta}, t)\frac{S_1^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \right] d\left\{\lambda_0(t)t\right\},$$

$$\begin{aligned} \Sigma_{2n}(\boldsymbol{\beta}) = \int_0^\tau & \left[\frac{S_2^{(2)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - S^{(1)}(\boldsymbol{\beta}, t)\frac{S_2^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)^2} \right] \\ & \times d\left\{ \frac{\bar{N}(te^{-z_{1i}\log\beta_1})}{n} \right\}. \end{aligned}$$

By the proof of Theorem 3.1 given in Appendix A, we obtain that, in the neighborhood of $\boldsymbol{\beta}_0$,

$$\begin{aligned} n^{-1/2}\widetilde{H}_n(\boldsymbol{\beta}) =& n^{-1/2}\widetilde{H}_n(\boldsymbol{\beta}_0) - \widetilde{\Sigma}_n(\boldsymbol{\beta}_0)\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &+ o_P(1 + \sqrt{n}\,\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|), \end{aligned} \tag{11}$$

where $\widetilde{\Sigma}_n(\boldsymbol{\beta})$ is defined by replacing $S^{(d)}$ in $\Sigma_n(\boldsymbol{\beta})$ with $\widetilde{S}^{(d)}$ for $d = 0,\ 1,\ 2$. Following the approach of Zeng and Lin (2008), we first generate $\widetilde{\boldsymbol{\beta}}_w$ by $\widetilde{\boldsymbol{\beta}}_w = \widetilde{\boldsymbol{\beta}} + n^{-1/2}W$, where W is a zero-mean random vector independent of the data. It follows from (11) and Theorem 3.1 that

$$n^{-1/2}\widetilde{H}_n(\widetilde{\boldsymbol{\beta}}_w) - n^{-1/2}\widetilde{H}_n(\widetilde{\boldsymbol{\beta}}) = -\widetilde{\Sigma}_n(\boldsymbol{\beta}_0)\sqrt{n}(\widetilde{\boldsymbol{\beta}}_w - \widetilde{\boldsymbol{\beta}}) + o_p(1).$$

Since $\widetilde{H}_n(\widetilde{\boldsymbol{\beta}}) = 0$ and $\widetilde{\boldsymbol{\beta}}_w - \widetilde{\boldsymbol{\beta}} = n^{-1/2}W$, we have

$$-n^{-1/2}\widetilde{H}_n(\widetilde{\boldsymbol{\beta}}_w) = \widetilde{\Sigma}_n(\boldsymbol{\beta}_0)W + o_p(1).$$

Here $\widetilde{\Sigma}_n(\boldsymbol{\beta}_0)$ could be viewed as parameter of a linear regression model and be estimated via LS approach. We summarize the algorithm for the calculation of the covariance estimator as follows.

Step 1: Calculate the estimators of Ω and Ω_G, denoted by $\widehat{\Omega}$ and $\widehat{\Omega}_G$.

Step 2: Generate B realizations of W, denoted by $W_1, \cdots, W_B$. For $b = 1, \cdots, B$ calculate $-n^{-1/2}\widetilde{H}_n(\widetilde{\boldsymbol{\beta}} + n^{-1/2}W_b)$.

Step 3: For $b = 1, \cdots, B$, calculate the jth LS estimator based on $-n^{-1/2}h_j(\widetilde{\boldsymbol{\beta}} + n^{-1/2}W_b)$, where h_j denotes the jth component of $\widetilde{H}_n$. Let $\widehat{\Sigma}$ be the matrix whose jth row is the jth LS estimator.

Step 4: Estimate the covariance matrix of $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ by $\widehat{\Sigma}^{-1}\left\{\widehat{\Omega} + \widehat{\Omega}_G\right\}(\widehat{\Sigma}')^{-1}$.

4. NUMERICAL STUDIES

Simulation studies were conducted to access the finite-sample performance of the proposed method. We compared the proposed estimator $\widehat{\beta}_P$ with two estimators, $\widehat{\beta}_F$ which is the estimator based on the full cohort, and $\widehat{\beta}_N$ which is the estimator based on a simple random sample of the same size as the case-cohort sample. The goal of this comparison is to address the following related issues: First, what is the efficiency loss of $\widehat{\beta}_P$ relative to $\widehat{\beta}_F$? Second, how much can the efficiency be increased by using a case-cohort design instead of a random sampling design?

We considered the following accelerated hazards model:

$$\lambda(t|Z_1, Z_2) = \lambda_0(te^{\beta_1 Z_1}) \exp\{\beta_2 Z_2\}. \tag{12}$$

The covariate Z_1 was generated from a Bernoulli distribution with success probability of 0.5, and Z_2 was generated from a standard normal distribution. We set $\beta_1 = 0.693,\ 0,$ or -0.693 and $\beta_2 = 0.5,\ 0,$ or -0.5. The baseline hazard function $\lambda_0(t)$ was set to follow the log-normal distribution $LN(0, 1)$ or the log-logistic distribution with the shape parameter 2 and the scale parameter 1. The censoring time C was generated from a uniform distribution $U[0, c]$ with c being chosen to obtain an approximately 90% censoring rate.

The subcohort of size $\widetilde{n} = 440$ was randomly selected from the full cohort of size 4000 without replacement. The subcohort and additional failures outside the subcohort constitute the case-cohort sample of size around 800. We calculated the proposed estimator $\widehat{\beta}_P$ based on the algorithm in Section 2.2 with $\eta = 0.005$, and applied the approach in Section 3.2 to estimate the asymptotic variance based on 50 bootstrap samples. The results based on 1000 replications are summarized in Tables 1 and 2.

The three estimators are all practically unbiased for all the cases considered here. The means of the estimated standard errors ($\widehat{SE}$s) are close to the sample standard errors of the estimates (SEs). The empirical 95% confidence intervals have reasonable coverage probabilities (CPs). The proposed estimator $\widehat{\beta}_P$ is more efficient than $\widehat{\beta}_N$. The estimated efficiencies relative to $\widehat{\beta}_F$ (REs) show that the proposed estimator of β_1 reaches about 40% of the efficiency of the full-cohort estimator when only about 20% subjects of the entire cohort are included in the case-cohort design. Meanwhile, the proposed estimator of β_2 reaches over 30% of the efficiency in most of the considered cases by using 20% of the subjects.

Additionally, we considered the following model:

$$\lambda(t|Z_1, Z_2, Z_3) = \lambda_0(te^{\beta_1 Z_1}) \exp\{\beta_2 Z_2 + \beta_3 Z_3\}, \tag{13}$$

which assumes two exposure variables in addition to the treatment factor. Z_1 was generated from a Bernoulli distribution with success probability of 0.5. $(Z_2, Z_3)'$ was generated from a two-dimensional normal distribution with mean 0 and variance matrix $\Sigma = (\sigma_{ij})_{2\times 2}$ where $\sigma_{ij} = 0.5^{|i-j|}$. We set $\beta_1 = 0.693$, $\beta_2 = 0.5$ and $\beta_3 = 0.5$. The baseline hazard function $\lambda_0(t)$ was set to follow the log-normal distribution $LN(0, 1)$. The censoring time C was generated from a uniform distribution $U[0, c]$ with c being chosen to obtain the desired censoring rate $\rho = 80\%,\ 85\%$ or 90%.

For the case-cohort design, we set the ratio between the sample sizes of case and control cohorts to be $r = 1$ or 2 to assess the impact of sample sizes on the performance of the proposed method. For example, in the case that the size of the full cohort was 2000 and the censoring rate ρ was 80%, we randomly selected a subcohort of size $\widetilde{n} = 500$ to obtain a case-cohort sample of size 800 which consisted of 400 cases and 400 controls, i.e., $r = 1$. The results based on 1000 replications are summarized in Table 3.

The three estimators are all practically unbiased, $\widehat{SE}$s are close to SEs and CPs are around 95%. The proposed estimator $\widehat{\beta}_P$ is more efficient than $\widehat{\beta}_N$ for all the settings considered here. REs relative to $\widehat{\beta}_F$ show that $\widehat{\beta}_P$ reaches around 2 times of the efficiency of $\widehat{\beta}_N$ under the same sample size. Both $\widehat{\beta}_P$ and $\widehat{\beta}_N$ are more efficient in the cases that $r = 2$ because the sizes of case-cohort sample are larger. Taking the sample size reduction into consideration, $\widehat{\beta}_P$ is more efficient when the censoring rate is higher.

5. REAL EXAMPLE

We illustrated the proposed method with a data set from a burn study (Ichida et al, 1993; Klein and Moeschberger, 2003). Infection of a burn wound is a common complication resulting in extended hospital stays and in the death of severely burned patients. Control of infection remains a prominent component of burn management. Medical records of 154 patients treated during the 18-month study period were reviewed, and information on their burn wound infections and other medical information were provided. The time to excision was recorded in days along with an indicator variable on whether or not the patient's wound had been excised. Wound was excised in 64.3% of the individuals.

The purpose of this study is to compare two treatment methods. One treatment is a routine bathing care method ($Z_1 = 0$) using initial surface decontamination with 10% povidone-iodine followed with regular bathing with Dial

Table 1. Results based on model $\lambda(t|Z_1, Z_2) = \lambda_0(te^{\beta_1 Z_1})\exp\{\beta_2 Z_2\}$, with the baseline hazard following the log-normal distribution $LN(0,1)$

		$\widehat{\beta}_1$					$\widehat{\beta}_2$				
$(\beta_1,\ \beta_2)$	*Method*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*
(0.693, 0.5)	$\widehat{\beta}_F$	0.698	0.153	0.156	94.6	1.00	0.497	0.049	0.050	93.9	1.00
	$\widehat{\beta}_N$	0.692	0.352	0.360	92.1	0.19	0.504	0.114	0.116	93.9	0.18
	$\widehat{\beta}_P$	0.686	0.236	0.236	93.6	0.42	0.504	0.092	0.088	92.6	0.28
(0.693, 0)	$\widehat{\beta}_F$	0.696	0.161	0.165	94.5	1.00	−0.001	0.050	0.050	93.3	1.00
	$\widehat{\beta}_N$	0.685	0.365	0.378	91.9	0.19	−0.001	0.112	0.113	94.0	0.20
	$\widehat{\beta}_P$	0.684	0.235	0.236	92.7	0.47	0.001	0.082	0.080	93.9	0.37
(0.693, −0.5)	$\widehat{\beta}_F$	0.696	0.153	0.153	93.9	1.00	−0.499	0.052	0.050	92.7	1.00
	$\widehat{\beta}_N$	0.677	0.345	0.359	91.9	0.20	−0.502	0.113	0.115	93.4	0.21
	$\widehat{\beta}_P$	0.684	0.235	0.235	93.4	0.42	−0.504	0.093	0.089	91.9	0.31
(0, 0.5)	$\widehat{\beta}_F$	0.003	0.133	0.132	94.4	1.00	0.497	0.050	0.050	94.4	1.00
	$\widehat{\beta}_N$	−0.001	0.295	0.299	93.1	0.20	0.505	0.117	0.114	94.4	0.18
	$\widehat{\beta}_P$	−0.002	0.217	0.206	91.5	0.38	0.505	0.092	0.090	93.2	0.30
(0, 0)	$\widehat{\beta}_F$	0.001	0.144	0.139	93.4	1.00	−0.001	0.050	0.049	94.0	1.00
	$\widehat{\beta}_N$	0.001	0.305	0.312	93.0	0.22	−0.001	0.115	0.112	92.8	0.19
	$\widehat{\beta}_P$	−0.002	0.217	0.205	91.3	0.44	0.001	0.083	0.079	93.0	0.36
(0, −0.5)	$\widehat{\beta}_F$	0.003	0.137	0.131	92.1	1.00	−0.498	0.053	0.050	92.2	1.00
	$\widehat{\beta}_N$	0.000	0.295	0.296	91.1	0.22	−0.503	0.115	0.115	93.1	0.21
	$\widehat{\beta}_P$	−0.001	0.217	0.205	91.6	0.40	−0.504	0.095	0.089	92.0	0.31
(−0.693, 0.5)	$\widehat{\beta}_F$	−0.694	0.121	0.118	93.1	1.00	0.496	0.050	0.050	94.3	1.00
	$\widehat{\beta}_N$	−0.702	0.275	0.278	93.5	0.19	0.505	0.117	0.114	94.2	0.18
	$\widehat{\beta}_P$	−0.691	0.187	0.177	92.5	0.42	0.503	0.095	0.092	93.1	0.28
(−0.693, 0)	$\widehat{\beta}_F$	−0.694	0.127	0.122	92.5	1.00	−0.001	0.050	0.049	93.7	1.00
	$\widehat{\beta}_N$	−0.702	0.273	0.291	93.6	0.22	0.001	0.111	0.112	94.0	0.20
	$\widehat{\beta}_P$	−0.691	0.183	0.175	92.0	0.48	0.001	0.083	0.081	92.5	0.36
(−0.693, −0.5)	$\widehat{\beta}_F$	−0.693	0.123	0.118	91.7	1.00	−0.499	0.051	0.050	92.1	1.00
	$\widehat{\beta}_N$	−0.700	0.265	0.272	92.5	0.22	−0.504	0.111	0.113	94.5	0.21
	$\widehat{\beta}_P$	−0.691	0.186	0.179	91.5	0.44	−0.505	0.096	0.091	92.6	0.28

Note: $\widehat{\beta}_F$, the estimator based on the full cohort; $\widehat{\beta}_N$, the estimator based on a simple random sample of the same size as the case-cohort sample; $\widehat{\beta}_P$, the proposed case-cohort estimator. Results are based on 1000 simulations.

soap. The other is a body cleansing method ($Z_1 = 1$) using 4% chlorhexidine gluconate. Among the 154 patients, 84 patients received the new bathing solution and 70 patients received the routine bathing care. We estimated the treatment effects with the proposed AcH model while adjusting for the additional 7 covariates such as sex and burn site. Among the cohort subjects, 22.08% of patients are female ($Z_2 = 1$). 45.45% of patients are with burn site on head ($Z_3 = 1$), 22.73% on buttock ($Z_4 = 1$), 84.42% on trunk ($Z_5 = 1$), 40.91% on upper leg ($Z_6 = 1$), 30.52% on lower leg ($Z_7 = 1$), and 29.22% in respiratory tract ($Z_8 = 1$). Table 4 provides a summary of the covariates considered.

To motivate the possible use of the proposed AcH model, we plotted the smoothed hazard functions with a bandwidth of 4 or 7 days in Figure 1 for both treatment groups, and chose two different study periods of the first 15 days and the whole 30 days to draw these plots. Figure 1 suggests that the assumption of constant proportionality may not be satisfied because both treatment groups share a similar hazard at the starting time of the trial but differ gradually as the time goes by. In addition, the body cleansing treatment has a tendency to increase the hazard risk between 2 and 20 days and after around 28 days, while it decreases the hazard risk between 20 and 28 days. The apparent crossovers of the hazard curves in Figure 1(c) and 1(d) may indicate further violation of the constant proportionality. These features suggest that the proposed AcH model may be more appropriate for this data set.

Since all the 8 considered covariates were measured in the data set, we started by fitting the AcH model (2) which included the binary treatment as Z_1 and other 7 covariates as $(Z_2, \cdots, Z_8)$ to the full cohort. To evaluate the

Table 2. Results based on model $\lambda(t|Z_1, Z_2) = \lambda_0(te^{\beta_1 Z_1})\exp\{\beta_2 Z_2\}$, with baseline hazard following the log-logistic distribution with shape parameter 2 and scale parameter 1

		$\widehat{\beta}_1$					$\widehat{\beta}_2$				
$(\beta_1,\ \beta_2)$	*Method*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*
(0.693, 0.5)	$\widehat{\beta}_F$	0.702	0.137	0.140	94.2	1.00	0.497	0.048	0.050	94.7	1.00
	$\widehat{\beta}_N$	0.714	0.325	0.365	93.9	0.18	0.507	0.113	0.115	93.8	0.18
	$\widehat{\beta}_P$	0.696	0.206	0.211	94.2	0.44	0.503	0.092	0.091	93.1	0.27
(0.693, 0)	$\widehat{\beta}_F$	0.701	0.139	0.144	94.5	1.00	−0.001	0.050	0.049	93.5	1.00
	$\widehat{\beta}_N$	0.706	0.326	0.375	93.7	0.18	−0.002	0.113	0.111	94.8	0.20
	$\widehat{\beta}_P$	0.695	0.198	0.204	94.3	0.49	0.001	0.083	0.080	92.8	0.36
(0.693, −0.5)	$\widehat{\beta}_F$	0.699	0.139	0.141	93.6	1.00	−0.499	0.053	0.050	92.3	1.00
	$\widehat{\beta}_N$	0.701	0.317	0.359	93.8	0.19	−0.501	0.112	0.113	92.8	0.22
	$\widehat{\beta}_P$	0.693	0.206	0.208	94.2	0.46	−0.505	0.094	0.090	92.2	0.32
(0, 0.5)	$\widehat{\beta}_F$	0.003	0.119	0.120	94.7	1.00	0.498	0.050	0.050	94.0	1.00
	$\widehat{\beta}_N$	−0.003	0.281	0.306	95.4	0.18	0.505	0.116	0.115	94.3	0.19
	$\widehat{\beta}_P$	−0.001	0.199	0.195	93.9	0.36	0.505	0.093	0.090	93.0	0.29
(0, 0)	$\widehat{\beta}_F$	0.002	0.126	0.124	94.3	1.00	−0.001	0.050	0.050	93.1	1.00
	$\widehat{\beta}_N$	0.000	0.281	0.307	94.8	0.20	0.000	0.115	0.112	93.9	0.19
	$\widehat{\beta}_P$	−0.001	0.193	0.186	92.6	0.43	0.001	0.082	0.079	92.9	0.37
(0, −0.5)	$\widehat{\beta}_F$	0.002	0.123	0.119	93.2	1.00	−0.498	0.052	0.050	92.3	1.00
	$\widehat{\beta}_N$	−0.003	0.281	0.299	94.2	0.19	−0.503	0.114	0.115	93.5	0.21
	$\widehat{\beta}_P$	−0.002	0.200	0.193	93.3	0.38	−0.504	0.095	0.090	93.0	0.30
(−0.693, 0.5)	$\widehat{\beta}_F$	−0.699	0.135	0.132	93.7	1.00	0.496	0.049	0.050	93.2	1.00
	$\widehat{\beta}_N$	−0.730	0.324	0.353	93.9	0.17	0.504	0.116	0.115	93.5	0.18
	$\widehat{\beta}_P$	−0.698	0.202	0.195	92.9	0.45	0.503	0.094	0.091	93.4	0.27
(−0.693, 0)	$\widehat{\beta}_F$	−0.697	0.139	0.135	93.3	1.00	−0.001	0.050	0.050	93.9	1.00
	$\widehat{\beta}_N$	−0.733	0.321	0.356	93.5	0.19	0.001	0.113	0.113	95.1	0.20
	$\widehat{\beta}_P$	−0.698	0.196	0.187	93.0	0.50	0.001	0.084	0.081	93.7	0.35
(−0.693, −0.5)	$\widehat{\beta}_F$	−0.699	0.139	0.132	91.4	1.00	−0.499	0.051	0.050	93.5	1.00
	$\widehat{\beta}_N$	−0.723	0.312	0.340	94.7	0.20	−0.505	0.112	0.115	94.7	0.21
	$\widehat{\beta}_P$	−0.698	0.203	0.197	91.8	0.47	−0.505	0.095	0.091	92.6	0.29

Note: $\widehat{\beta}_F$, the estimator based on the full cohort; $\widehat{\beta}_N$, the estimator based on a simple random sample of the same size as the case-cohort sample; $\widehat{\beta}_P$, the proposed case-cohort estimator. Results are based on 1000 simulations.

case-cohort design, we randomly sampled 80 patients as the subcohort, and implemented our proposed method under the AcH model. To compare with the proposed method, we additionally applied the method developed by Prentice (1986) for the PH model $\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{Z})$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_8)'$ and $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_8)'$. For the case-cohort design, the models were fitted for the first 15-day period and the whole 30-day period, respectively. Table 5 summaries the results, where columns under AcHM-FULL are the results from the full-cohort analysis under the AcH model. The columns under AcHM-CC and PHM-CC are the results from the case-cohort analysis under the AcH and PH models, respectively.

For the 15-day period study, the results from both the PH and AcH models are similar and suggest that the treatment effect is not significant, which means that there exists a lag period before the treatment begins to be effective. Based on the 30-day period, all the case-cohort estimators are close to the full-cohort estimators. The treatment effect is not significant in the PH model analysis, but it is significant in the AcH model analysis. For example, the body cleansing method would significantly accelerate the time of hazard progression by $e^{1.16} = 3.2$ times (AcHM-FULL) or $e^{1.015} = 2.8$ times (AcHM-CC). This means to reach the same level of risk, the patients received the body cleansing would only need around 0.31 times (AcHM-FULL) or 0.36 times (AcHM-CC) of the time needed by the patients in the routine bathing care. The results reported in Table 5 also suggest that female patients tend to have a higher risk.

The acceleration of the new bathing solution is statistically significant in the analysis of the AcH model while nonsignificant in the analysis of the PH model. This sug-

Table 3. Results based on model $\lambda(t|Z_1, Z_2, Z_3) = \lambda_0(te^{\beta_1 Z_1})\exp\{\beta_2 Z_2 + \beta_3 Z_3\})$, with the baseline hazard following the log-normal distribution $LN(0,1)$

			$\beta_1 = 0.693$					$\beta_2 = 0.5$					$\beta_3 = 0.5$				
ρ	r	*Method*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*	*Mean*	*SE*	$\widehat{SE}$	*CP*	*RE*
0.80	1	$\widehat{\beta}_F$	0.680	0.210	0.207	93.5	1.00	0.497	0.059	0.059	93.7	1.00	0.500	0.056	0.059	93.3	1.00
		$\widehat{\beta}_N$	0.678	0.468	0.455	91.4	0.20	0.507	0.140	0.137	92.1	0.18	0.508	0.136	0.137	93.8	0.17
		$\widehat{\beta}_P$	0.678	0.316	0.309	92.5	0.44	0.504	0.099	0.097	93.2	0.36	0.509	0.103	0.097	92.1	0.30
	2	$\widehat{\beta}_F$	0.680	0.210	0.207	93.5	1.00	0.497	0.059	0.059	93.7	1.00	0.500	0.056	0.059	93.3	1.00
		$\widehat{\beta}_N$	0.700	0.296	0.314	94.0	0.50	0.504	0.089	0.087	93.2	0.44	0.498	0.090	0.089	92.8	0.38
		$\widehat{\beta}_P$	0.680	0.243	0.246	93.7	0.75	0.500	0.074	0.072	94.5	0.64	0.505	0.071	0.073	92.2	0.64
0.85	1	$\widehat{\beta}_F$	0.693	0.205	0.202	93.8	1.00	0.501	0.068	0.066	92.6	1.00	0.501	0.066	0.068	93.8	1.00
		$\widehat{\beta}_N$	0.674	0.460	0.458	91.8	0.20	0.518	0.155	0.158	94.4	0.19	0.504	0.157	0.161	93.7	0.18
		$\widehat{\beta}_P$	0.678	0.304	0.311	92.5	0.46	0.516	0.129	0.121	92.9	0.28	0.509	0.124	0.121	92.7	0.29
	2	$\widehat{\beta}_F$	0.693	0.205	0.202	93.8	1.00	0.501	0.068	0.066	92.6	1.00	0.501	0.066	0.068	93.8	1.00
		$\widehat{\beta}_N$	0.688	0.394	0.385	92.7	0.27	0.509	0.133	0.133	94.6	0.26	0.506	0.134	0.134	94.1	0.25
		$\widehat{\beta}_P$	0.690	0.260	0.250	93.2	0.62	0.506	0.097	0.093	92.8	0.50	0.505	0.099	0.093	92.4	0.45
0.90	1	$\widehat{\beta}_F$	0.694	0.193	0.194	92.1	1.00	0.496	0.084	0.083	92.2	1.00	0.504	0.084	0.083	93.6	1.00
		$\widehat{\beta}_N$	0.686	0.465	0.462	91.3	0.17	0.504	0.200	0.195	94.9	0.18	0.521	0.197	0.197	92.3	0.18
		$\widehat{\beta}_P$	0.690	0.336	0.336	93.5	0.33	0.509	0.168	0.164	93.2	0.25	0.523	0.174	0.164	92.4	0.23
	2	$\widehat{\beta}_F$	0.694	0.193	0.194	92.1	1.00	0.496	0.084	0.083	92.2	1.00	0.504	0.084	0.083	93.6	1.00
		$\widehat{\beta}_N$	0.681	0.383	0.376	93.4	0.25	0.503	0.160	0.153	92.4	0.27	0.503	0.160	0.156	92.4	0.27
		$\widehat{\beta}_P$	0.687	0.263	0.271	94.3	0.54	0.511	0.122	0.125	93.1	0.47	0.508	0.123	0.125	94.5	0.46

Note: r denotes the ratio between the sample sizes of case and control cohorts. $\widehat{\beta}_F$, the estimator based on the full cohort; $\widehat{\beta}_N$, the estimator based on a simple random sample of the same size as the case-cohort sample; $\widehat{\beta}_P$, the proposed case-cohort estimator. Results are based on 1000 simulations.

Table 4. Demographics and Characteristics of the Burn Study Data

Treatment (Z_1)	%
0 = Routine bathing	45.45 (70/154)
1 = Body cleansing	54.55 (84/154)
Gender (Z_2)	
0 = Male	77.92 (120/154)
1 = Female	22.08 (34/154)
Burn site indicator: Head (Z_3)	
0 = No	54.55 (84/154)
1 = Yes	45.45 (70/154)
Burn site indicator: Buttock (Z_4)	
0 = No	77.27 (119/154)
1 = Yes	22.73 (35/154)
Burn site indicator: Trunk (Z_5)	
0 = No	15.58 (24/154)
1 = Yes	84.42 (130/154)
Burn site indicator: Upper leg (Z_6)	
0 = No	59.09 (91/154)
1 = Yes	40.91 (63/154)
Burn site indicator: Lower leg (Z_7)	
0 = No	69.48 (107/154)
1 = Yes	30.52 (47/154)
Burn site indicator: Respiratory tract (Z_8)	
0 = No	70.78 (109/154)
1 = Yes	29.22 (45/154)

gests that the AcH model helps to detect some significant effects which may be missed by the PH model analysis when the hazard curves of different groups have crossovers. Note that the proposed estimator, using a much smaller size of sample, performs very closely to the full-cohort estimator. Therefore, the case-cohort design can be a cost-effective alternative to the simple random sampling design in cohort studies.

6. CONCLUDING REMARKS

It is usual that a lag period may exist in medical studies before the treatment is fully effective. The AcH model is motivated by the need to accommodate this phenomenon. Since the case-cohort design is widely used as a cost-effective sampling method in large cohort studies, it is desirable to develop corresponding methodologies for the AcH model. We developed how to fit the AcH model to case-cohort data and showed that the proposed estimators are consistent and asymptotically normally distributed. Simulation studies suggest that the case-cohort estimator has a nice performance under the finite sample size. The numerical results indicate that the efficiency loss of the proposed estimator relative to the full-cohort estimator remains acceptable, compared to the sample size reduction. We proposed to use

Table 5. Results for Analysis of the Burn Study Data

	AcHM-FULL			the first 15 days study						the whole 30 days study					
				AcHM-CC			PHM-CC			AcHM-CC			PHM-CC		
	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
Treatment	1.160	0.316	$< 0.001^*$	0.261	0.407	0.522	0.317	0.379	0.403	1.015	0.427	0.017^*	0.582	0.307	0.058
Gender	0.696	0.270	0.010^*	0.125	0.446	0.779	0.084	0.377	0.824	0.772	0.382	0.043^*	0.771	0.334	0.021^*
Head	−0.296	0.266	0.267	−0.579	0.467	0.215	−0.519	0.393	0.186	−0.523	0.389	0.179	−0.493	0.336	0.143
Buttock	−0.410	0.321	0.201	−0.401	0.496	0.418	−0.374	0.428	0.383	−0.150	0.439	0.733	−0.120	0.376	0.749
Trunk	−0.262	0.361	0.468	−0.592	0.612	0.333	−0.659	0.542	0.223	−0.001	0.529	0.998	0.051	0.465	0.913
Upperleg	0.410	0.249	0.099	0.135	0.440	0.759	0.177	0.383	0.645	−0.009	0.391	0.982	0.024	0.326	0.940
Lowerleg	−0.476	0.276	0.085	−0.469	0.524	0.371	−0.503	0.399	0.208	−0.510	0.367	0.165	−0.591	0.331	0.074
Tract	0.267	0.274	0.330	0.030	0.501	0.952	−0.057	0.397	0.886	0.042	0.382	0.912	0.051	0.325	0.876

NOTE: * indicates significant effect at 5% level.

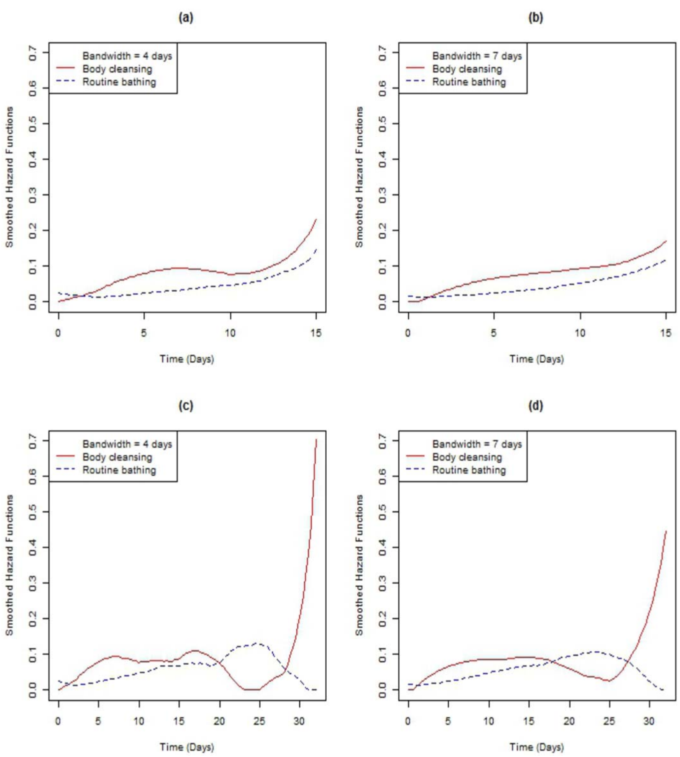

Figure 1. Smoothed hazard functions for the two treatments (——, Body cleansing; - - -, Routine bathing) in the burn study. (a) 4-day smoothing bandwidth for the first 15 days; (b) 7-day smoothing bandwidth for the first 15 days; (c) 4-day smoothing bandwidth for the full cohort; (d) 7-day smoothing bandwidth for the full cohort.

a least square approach, which is easy to implement in practice, to estimate the covariance for the proposed estimator.

Although we have presented an exploratory visual inspection of the adequacy of the proposed model, we have not fully discussed statistical model checking of the AcH model. The development of model-checking procedures for the AcH model in case-cohort studies will be an interesting and also challenging issue in the future. Several possible ideas could be used for such an issue. For example, we can extend the idea of Kolmogorov-Smirnov test by Lin et al. (1993) and Chen (2001), and the idea of Gill-Schumacher test by Gill and Schumacher (1987) and Chen (2001).

Here we assumed that the covariates Z_2 are time-independent. The proposed method can be easily extended to the case that the covariates Z_2 depend on time. Some easily measured covariates are generally available for each cohort member, the covariate information of controls outside the subcohort does not be incorporated in our estimation procedure. The future study on an estimating function, which takes all the information on these covariates into account, will be more efficient than the current one. The case-cohort design has been proved to be a cost-effective sampling method for rare event. When the censoring rate is medium or low, a generalized case-cohort design (Cai and Zeng, 2007) or a survival-data outcome-dependent sampling design (ODS) (Ding et al, 2014; Yu et al, 2015) were proposed to improve the efficiency. Future studies of how to fit the AcH model to data from a generalized case-cohort design or a survival-data ODS design are guaranteed.

APPENDIX A. PROOFS OF THEOREMS

We impose the following regularity conditions throughout the paper. $\|\cdot\|$ denotes the Euclidean norm and the convergence is referred to $n \to \infty$.

A) The parameter space $\mathcal{B}$ and the covariate space $\mathcal{Z}$ are compact.
B) $\widetilde{n}/n \to \rho$ for some $\rho \in (0, 1)$.
C) There exist some $\varepsilon_0 > 0$ and $\delta > 0$, such that $\mathrm{P}(T > \tau/\delta) > \varepsilon_0$.
D) $\lambda_0(t)$ is uniformly bounded and has the second derivative in $[0, \tau/\delta]$.
E) The sequence of the distributions of $\sqrt{n}(\bar{Z}(\boldsymbol{\beta}_0, t) - \widetilde{Z}(\boldsymbol{\beta}_0, t))$ is tight on the product space of left-continuous functions with right limits equipped with the product Skorohod topology.

F) For $d = 0, 1, 2$, there exist functions $s^{(d)}(\boldsymbol{\beta}, t)$ and $q^{(d)}(\boldsymbol{\beta}, t, w)$, defined on $\mathcal{B} \times [0, \tau/\delta]$ and $\mathcal{B} \times [0, \tau/\delta]^2$, respectively, such that

F1) $s^{(d)}(\boldsymbol{\beta}, t)$ is continuous of $\boldsymbol{\beta} \in \mathcal{B}$ uniformly in $t \in [0, \tau/\delta]$, satisfying

$$\sup_{\substack{\boldsymbol{\beta} \in \mathcal{B},\\ t \in [0, \tau/\delta]}} \|S^{(d)}(\boldsymbol{\beta}, t) - s^{(d)}(\boldsymbol{\beta}, t)\| \xrightarrow{\mathrm{P}} 0,$$

and $s^{(0)}(\boldsymbol{\beta}, t)$ is bounded away from zero;

F2) $q^{(d)}(\boldsymbol{\beta}, t, w)$ is continuous and bounded of $\boldsymbol{\beta} \in \mathcal{B}$ uniformly in $(t, w) \in [0, \tau/\delta]^2$, satisfying

$$\sup_{\substack{\boldsymbol{\beta} \in \mathcal{B},\\ (t,w) \in [0, \tau/\delta]^2}} \|Q^{(d)}(\boldsymbol{\beta}, t, w) - q^{(d)}(\boldsymbol{\beta}, t, w)\| \xrightarrow{\mathrm{P}} 0;$$

F3) Suppose that

$$\sup_{\substack{\boldsymbol{\beta} \in \mathcal{B},\\ t \in [0, \tau/\delta]}} \|\widetilde{S}^{(d)}(\boldsymbol{\beta}, t) - s^{(d)}(\boldsymbol{\beta}, t)\| \xrightarrow{\mathrm{P}} 0,$$

$$\sup_{\substack{\boldsymbol{\beta} \in \mathcal{B},\\ (t,w) \in [0, \tau/\delta]^2}} \|\widetilde{Q}^{(d)}(\boldsymbol{\beta}, t, w) - q^{(d)}(\boldsymbol{\beta}, t, w)\| \xrightarrow{\mathrm{P}} 0.$$

G) Decompose $e(\boldsymbol{\beta}, t) = s^{(2)}(\boldsymbol{\beta}, t) - \left\{s^{(1)}(\boldsymbol{\beta}, t)\right\}^{\otimes 2}$ according to the dimensions of β_1 and β_2 as $e(\boldsymbol{\beta}, t) = (e_1(\boldsymbol{\beta}, t), e_2(\boldsymbol{\beta}, t))$. The matrix

$$\Sigma = \left(\int_0^\tau e_1(\boldsymbol{\beta}_0, t) d\left\{\lambda_0(t)t\right\},\ \int_0^\tau e_2(\boldsymbol{\beta}_0, t)\lambda_0(t)dt \right),$$

is finite positive definite.

Conditions A)–G) are regularity conditions similar to those of Self and Prentice (1988), Tsiatis (1990), and Chen and Wang (2000). We first define the estimating equation

$$H_n(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \left[\boldsymbol{z}_i - \frac{\sum_{j=1}^n Y_j\left(t/{\beta_1}^{z_{1j}}\right) e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1} \boldsymbol{z}_j}{\sum_{j=1}^n Y_j\left(t/{\beta_1}^{z_{1j}}\right) e^{z_{2j}'\beta_2 - z_{1j}\log\beta_1}} \right] \times dN_i\left(t/{\beta_1}^{z_{1i}}\right) = 0.$$

We then construct a linear approximation of $H_n(\boldsymbol{\beta})$ as

$$H_n^*(\boldsymbol{\beta}) = H_n(\boldsymbol{\beta}_0) - n\Sigma_n(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0).$$

Therefore, the solution to $H_n^*(\boldsymbol{\beta}) = 0$ takes the form as:

$$\widehat{\boldsymbol{\beta}}^* = \boldsymbol{\beta}_0 + n^{-1}\Sigma_n^{-1}(\boldsymbol{\beta}_0)H_n(\boldsymbol{\beta}_0). \tag{14}$$

To prove the main results, we first prove the following three lemmas.

Lemma A.1.

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0) \xrightarrow{d} N(0,\ \Sigma^{-1}\Omega(\Sigma')^{-1}).$$

Proof of Lemma A.1. By equation (14), we have that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0) = \Sigma_n^{-1}(\boldsymbol{\beta}_0)\left\{n^{-1/2}H_n(\boldsymbol{\beta}_0)\right\},$$

and based on the discussion in Andersen and Gill (1982), we have

$$n^{-1/2}H_n(\boldsymbol{\beta}_0) \xrightarrow{d} N(0,\ \Omega).$$

Due to Conditions A)–D), Lemma A.1 holds since $\Sigma_n(\boldsymbol{\beta}_0) \xrightarrow{\mathrm{P}} \Sigma$. □

Lemma A.2.

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^*) \xrightarrow{P} 0.$$

Proof of Lemma A.2. Since $\widehat{\boldsymbol{\beta}}$ is the solution to $H_n(\boldsymbol{\beta}) = 0$, it is sufficient to prove the following Lemma A.3 to ensure Lemma A.2 (Jurecokova, 1971; Tsiatis, 1990). □

Lemma A.3. *For any positive constant K,*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| \le Kn^{-1/2}} n^{-1/2}\|H_n(\boldsymbol{\beta}) - H_n^*(\boldsymbol{\beta})\| \xrightarrow{P} 0.$$

Proof of Lemma A.3. By the finite covering theorem (i.e., the Heine-Borel Theorem), it suffices to show that

(a) $$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| = dn^{-1/2}} n^{-1/2}\|H_n(\boldsymbol{\beta}) - H_n^*(\boldsymbol{\beta})\| \xrightarrow{\mathrm{P}} 0,$$

and

(b) $$\sup_{\substack{\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = dn^{-1/2},\\ \|\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}}\| \le Kn^{-1/2}}} n^{-1/2}\left\|H_n(\boldsymbol{\beta}^*) - H_n^*(\widetilde{\boldsymbol{\beta}})\right\| \xrightarrow{\mathrm{P}} 0,$$

hold for any positive constants d and K.

By Conditions A)–D), we can obtain that $\Sigma_{1n}(\boldsymbol{\beta}_0) \xrightarrow{\mathrm{P}} \Sigma_1(\boldsymbol{\beta}_0)$, and $\Sigma_{2n}(\boldsymbol{\beta}_0) \xrightarrow{\mathrm{P}} \Sigma_2(\boldsymbol{\beta}_0)$. Due to

$$\begin{aligned} H_n(\boldsymbol{\beta}) - H_n^*(\boldsymbol{\beta}) =& H_n(\boldsymbol{\beta}) - H_n(\boldsymbol{\beta}_0) \\ &+ n\Sigma(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ =& H_n(\beta_1, \beta_2) - H_n(\beta_{10}, \beta_2) \\ &+ n\Sigma_{1n}(\boldsymbol{\beta}_0)(\beta_1 - \beta_{10}) \\ &+ H_n(\beta_{10}, \beta_2) - H_n(\beta_{10}, \beta_{20}) \\ &+ n\Sigma_{2n}(\beta_0)(\beta_2 - \beta_{20}), \\ \equiv& J_1 + J_2, \end{aligned} \tag{15}$$

we have, by Taylor expansion,

$$\begin{aligned} \Lambda_0(te^{\{\log\beta_{10} - \log\beta_1\}z_{1i}}) - \Lambda_0(t) &= \{\log\beta_{10} - \log\beta_1\} z_{1i}\lambda_0(t)t \\ &+ \mathrm{O_P}(1)\left|\log\beta_{10} - \log\beta_1\right|^2 \lambda_0(t)t. \end{aligned}$$

Then, for the two terms on the right-hand side of equation (15), we have

$$J_1 = n\left\{\Sigma_{1n}(\boldsymbol{\beta}_0) - \Sigma_{1n}(\boldsymbol{\beta})\right\}\left\{\log\beta_{10} - \log\beta_1\right\} + \left\{\sqrt{n}\,|\log\beta_{10} - \log\beta_1|\right\}^2 \mathrm{O_P}(1),$$

and

$$J_2 = \left\{\sqrt{n}\,\|\beta_{20} - \beta_2\|\right\}^2 \mathrm{O_P}(1).$$

It follows that

$$n^{-1/2}\left\{H_n(\boldsymbol{\beta}) - H_n^*(\boldsymbol{\beta})\right\} = \left\{\Sigma_{1n}(\boldsymbol{\beta}_0) - \Sigma_{1n}(\boldsymbol{\beta})\right\} \times \sqrt{n}\left\{\log\beta_{10} - \log\beta_1\right\} + \left\{\sqrt{n}\,\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2\right\}\mathrm{O_P}(1).$$

Based on the above fact, we have

$$n^{-1/2}\left\{H_n(\boldsymbol{\beta}) - H_n^*(\boldsymbol{\beta})\right\} = \mathrm{o_P}(1),$$

and this completes the proof of (a).

For (b), similarly, we have

$$H_n(\boldsymbol{\beta}^*) - H_n(\widetilde{\boldsymbol{\beta}}) = \left\{H_n(\beta_1^*, \beta_2^*) - H_n(\widetilde{\beta}_1, \beta_2^*)\right\} + \left\{H_n(\widetilde{\beta}_1, \beta_2^*) - H_n(\widetilde{\beta}_1, \widetilde{\beta}_2)\right\}.$$

By the similar arguments of Chen and Jewell (2001), we have

(16)
$$\sup_{\substack{||\widetilde{\beta}_1 - \beta_1^*|| \le Kn^{-1/2}, \\ ||\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| \le (K+d)n^{-1/2}}} n^{-1/2}\left\|H_n(\beta_1^*, \beta_2^*) - H_n(\widetilde{\beta}_1, \beta_2^*)\right\| \xrightarrow{\mathrm{P}} 0.$$

By Condition F) and Taylor expansion, we have

(17)
$$\sup_{\substack{||\beta_2^* - \widetilde{\beta}_2|| \le \delta n^{-1/2}, \\ ||\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| = dn^{-1/2}}} n^{-1/2}\left\|H_n(\widetilde{\beta}_1, \beta_2^*) - H_n(\widetilde{\beta}_1, \widetilde{\beta}_2)\right\| \xrightarrow{\mathrm{P}} 0.$$

Due to equations (16) and (17), and Cauchy-Schwarz inequality, we complete the proof of (b). □

Proof of Theorem 3.1. We first construct a linear approximation of $\widetilde{H}_n(\boldsymbol{\beta})$ as

$$\widetilde{H}_n^*(\boldsymbol{\beta}) = \widetilde{H}_n(\boldsymbol{\beta}_0) - n\widetilde{\Sigma}_n(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

where $\widetilde{\Sigma}_n(\boldsymbol{\beta})$ is defined by replacing $S^{(d)}$ in $\Sigma_n(\boldsymbol{\beta})$ with $\widetilde{S}^{(d)}$, $d = 0, 1, 2$. Then, the solution to $\widetilde{H}_n^*(\boldsymbol{\beta}) = 0$ takes the form as:

$$\widetilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta}_0 + n^{-1}\widetilde{\Sigma}_n^{-1}(\boldsymbol{\beta}_0)\widetilde{H}_n(\boldsymbol{\beta}_0).$$

Define

$$dM_i(te^{-z_{1i}\log\beta_{10}}) = dN_i(te^{-z_{1i}\log\beta_{10}}) - Y_i(te^{-z_{1i}\log\beta_{10}})d\Lambda_i(te^{-z_{1i}\log\beta_{10}}),$$

and it can be proved that $M_i(te^{-z_{1i}\log\beta_{10}})$ are martingale processes with respect to the filtration defined by

$$\mathscr{F}^{(n)}(t, \boldsymbol{\beta}_0) = \sigma\left\{I(\beta_{10}^{Z_{1j}} \le t),\ \Delta_j,\ Z_{1j},\ Z_{2j};\ j = 1, \cdots, n\right\}.$$

It is obvious that

(18)
$$n^{-1/2}\widetilde{H}_n(\boldsymbol{\beta}_0) = n^{-1/2}H_n(\boldsymbol{\beta}_0) + n^{-1/2}\int_0^\tau \left\{\bar{Z}(\boldsymbol{\beta}_0, t) - \widetilde{Z}(\boldsymbol{\beta}_0, t)\right\} \times d\overline{\Lambda}(te^{-z_{1i}\log\beta_{10}}) + n^{-1/2}\int_0^\tau \left\{\bar{Z}(\boldsymbol{\beta}_0, t) - \widetilde{Z}(\boldsymbol{\beta}_0, t)\right\} \times d\overline{M}(te^{-z_{1i}\log\beta_{10}}),$$

where $\overline{\Lambda}(t) = \sum_{i\in\mathcal{C}}\Lambda_i(t)$ and $\overline{M}(t) = \sum_{i\in\mathcal{C}}M_i(t)$. By Lemma A.1, the first term on the right-hand side of equation (18) converges in distribution to $N(0,\,\Omega)$. Similar arguments as Self and Prentice (1988), we can prove that the second term converges in distribution to $N(0,\,\Omega_G)$ and is independent of the first term. The third term is a martingale and converges in probability to 0. Due to $\widetilde{\Sigma}_n^{-1}(\boldsymbol{\beta}_0) \xrightarrow{\mathrm{P}} \Sigma^{-1}$, we complete Theorem 3.1 by Slutsky's theorem. □

Proof of Theorem 3.2. For the consistency of $\widetilde{\Lambda}_0(t)$, we have

$$\widetilde{\Lambda}_0(t) - \Lambda_0(t) = \left\{\widetilde{\Lambda}_0(t) - \widehat{\Lambda}_0(t)\right\} + \left\{\widehat{\Lambda}_0(t) - \Lambda_0(t)\right\}.$$

$\widetilde{\Lambda}_0(t) - \widehat{\Lambda}_0(t)$ converges in probability to zero uniformly in $t \in [0, \tau/\delta]$. $\widehat{\Lambda}_0(t) - \Lambda_0(t)$ converges in probability to zero uniformly in $t \in [0, \tau/\delta]$ by the standard decomposition techniques of Andersen and Gill (1982) and the results of Self and Prentice (1988). Furthermore, we have

$$\sqrt{n}\left\{\widehat{\Lambda}_0(t) - \Lambda_0(t)\right\} = \left(\int_0^t \frac{S_1^{(1)}(\boldsymbol{\beta}_0, u)}{S^{(0)}(\boldsymbol{\beta}_0, u)} d\left\{\lambda_0(u)u\right\},\ \int_0^t \frac{S_2^{(1)}(\boldsymbol{\beta}_0, u)}{S^{(0)}(\boldsymbol{\beta}_0, u)^2} d\left\{\frac{\bar{N}(ue^{-z_{1i}\log\beta_{10}})}{n}\right\}\right) \times \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sqrt{n}\int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta}_0, u)} d\bar{M}(ue^{-z_{1i}\log\beta_{10}}) + \mathrm{o_P}(1).$$

Following the similar techniques of Self and Prentice (1988), we can obtain that $\sqrt{n}\left\{\widetilde{\Lambda}_0(t) - \Lambda_0(t)\right\}$ converges weakly to a zero-mean Gaussian process with the covariance function at (s,t) equals to

$$\Psi(s)\Sigma^{-1}\left\{\Omega+\Omega_G\right\}(\Sigma')^{-1}\Psi(t)'+\int_0^{\min(s,t)}\frac{\lambda_0(u)}{s^{(0)}(\boldsymbol{\beta}_0,u)}du.$$

□

ACKNOWLEDGEMENTS

The authors thank the Editor, an Associate Editor and two reviewers for their insightful comments and suggestions that greatly improved the article. This research is supported in part by the National Science Foundation of China (11671310 to J.D., 11571263, 11371299 to Y.L.).

Received 19 May 2017

REFERENCES

[1] Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, **10**, 1100–1120. MR0673646

[2] 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

[3] Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics*, **34**, 86–102. MR2325244

[4] Cai, J. and Zeng, D. (2004). Sample size/power calculation for case-cohort studies. *Biometrics*, **60**, 1015–1024. MR2133554

[5] Cai, J. and Zeng, D. (2007). Power calculation for case-cohort studies with nonrare events. *Biometrics*, **63**, 1288–1295. MR2414609

[6] Chen, K. (2001). Generalized case-cohort sampling. *Journal of the Royal Statistical Society, Series B*, **63**, 791–809 MR1872067

[7] Chen, K. and Lo, S. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika*, **86**, 755–764. MR1741975

[8] Chen, Y. Q. (2001). Accelerated hazards regression model and its adequacy for censored survival data. *Biometrics*, **57**, 853–860. MR1859820

[9] Chen, Y. Q. and Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, **88**, 687–702. MR1859402

[10] Chen, Y. Q. and Wang, M. C. (2000). Analysis of accelerated hazards model. *Journal of the American Statistical Association*, **95**, 608–618. MR1803176

[11] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187–202. MR0341758

[12] Ding, J., Zhou, H., Liu, L., Cai, J., and Longnecker, M. P. (2014). Estimating effect of environmental contaminants on women's subfecundity for the MoBa study data with an outcome-dependent sampling scheme. *Biostatistics*, **15**, 636–650.

[13] Gill, R. D. and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika*, **74**, 289–300.

[14] Ichida, J. M., Wassell, J. T., Keller, M. D., and Ayers, L. W. (1993). Evaluation of protocol change in burn-care management using the cox proportional hazards model with time-dependent covariates. *Statistics in Medicine*, **12**, 301–310.

[15] Jurecokova, J. (1971). Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*, **42**, 1328–1338. MR0295487

[16] Kang, S., Lu, W. and Liu, M. (2017). Efficient estimation for accelerated failure time model under case-cohort and nested case-control sampling. *Biometrics*, **73**, 114–123. MR3632357

[17] Klein, J. P. and Moeschberger, M. L. (2003). Survival analysis techniques for censored and truncated data. Springer-Verlag New York.

[18] Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics*, **65**, 135–142. MR2665853

[19] 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

[20] Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, **80**, 557–572. MR1248021

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

[22] Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, **81**, 341–350. MR1294895

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

[24] Qi, L., Wang, C. Y., and Prentice, R. L. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, **100**, 1250–1263. MR2236439

[25] 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

[26] Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, **18**, 354–372. MR1041397

[27] Yu, J., Liu, Y., Sandler, D. P. and Zhou, H. (2015). Statistical inference for the additive hazards model under outcome-dependent sampling. *The Canadian Journal of Statistics*, **43**, 436–453. MR3388326

[28] Zeng, D. L. and Lin, D. Y. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics*, **9**, 355–363.

Jieli Ding
School of Mathematics and Statistics
Wuhan University
Wuhan, Hubei 430072
China
E-mail address: jlding.math@whu.edu.com

Xiaolong Chen
School of Mathematics and Statistics
Wuhan University
Wuhan, Hubei 430072
China
E-mail address: chenxiaolog@gmail.com

Huaying Fang
School of Mathematical Sciences
and Center for Quantitative Biology
Peking University
Beijing 100871
China
E-mail address: hyfang@pku.edu.cn

Yanyan Liu
School of Mathematics and Statistics
Wuhan University
Wuhan, Hubei 430072
China
E-mail address: liuyy@whu.edu.cn