A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling

By CHIUNG-YU HUANG, JING QIN and DEAN A. FOLLMANN

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

huangchi@niaid.nih.gov   jingqin@niaid.nih.gov   dfollmann@niaid.nih.gov

Summary

This paper considers semiparametric estimation of the Cox proportional hazards model for right-censored and length-biased data arising from prevalent sampling. To exploit the special structure of length-biased sampling, we propose a maximum pseudo-profile likelihood estimator, which can handle time-dependent covariates and is consistent under covariate-dependent censoring. Simulation studies show that the proposed estimator is more efficient than its competitors. A data analysis illustrates the methods and theory.

Some key words: Approximate likelihood; Cross-sectional sampling; Product-limit estimator; Random truncation; Screening trials.

1. Introduction

When studying the natural history of a disease, the time from disease onset to an event or failure is usually the focus. An incident cohort approach, which studies initially disease-free subjects from disease onset to failure, can be very inefficient, especially if the disease is uncommon. A prevalent sampling design, which only includes diseased subjects who have not experienced the failure event at the time of recruitment, can be much more efficient. However, the observed survival time is subject to left truncation: those who have experienced the failure event before the recruitment time are not observable. Thus, individuals in the prevalent cohort tend to have slower progression of the disease than those in a typical incident study. As a result, statistical methods such as the Kaplan–Meier estimator that fail to account for left truncation can lead to substantial overestimation of the survival time.

In the case of stable disease, that is, the occurrence of disease onset follows a stationary Poisson process, the survival time in the prevalent cohort is a biased sample of that in the incident population, where the sampling weight is proportional to the length of the survival time. Similarly, the truncation time, from disease onset to recruitment, in the prevalent cohort is also a biased sample of the uniform truncation time in the incident population, and its distribution is related to the underlying survival distribution in a known fashion. We use the term length-biased sampling for left truncation under the assumption of stationary disease incidence. Examples of length-biased sampling include studies of cancer screening trials (Zelen & Feinleib, 1969; Zelen, 2004), HIV prevalent cohort studies (Lagakos et al., 1988) and unemployment duration (Lancaster, 1979; de Una-Alvarez et al., 2003).

This paper focuses on semiparametric estimation of the Cox proportional hazards model for right-censored survival data under length-biased sampling. Intuitively, efficient estimation can be

achieved by maximizing the full semiparametric likelihood with respect to the regression parameter and the baseline hazard function. The maximum likelihood approach, however, involves high-dimensional maximization, and hence may cause computational concerns for large sample sizes. Estimation of a finite-dimensional parameter in the presence of an infinite-dimensional nuisance parameter has been studied by a number of authors. In particular, Severini & Wong (1992) and Zucker (2005) generalized the profile likelihood method by replacing the nuisance parameters in the full likelihood or the partial likelihood with a consistent estimator that may depend on the parameter of interest. In this paper, we follow their idea to propose a semiparametric estimation procedure for the Cox model under length-biased sampling. Specifically, we replace the hazard function in the full likelihood with a Breslow-type estimator for the hazard function to obtain a pseudo-profile likelihood function. Thus, a consistent estimator of the regression parameters can be easily derived by maximizing the pseudo-profile likelihood. Unlike other bias-adjusted risk-set methods, including Ghosh (2008), Tsai (2009) and Qin & Shen (2010), the proposed estimation procedure does not involve estimation of the censoring distribution, so it is expected to be more stable when the censoring proportion is high.

2. MODEL AND ESTIMATION METHODS

2·1. *Data and model set-up*

For subjects in the target disease population, let T^0 denote the time from the disease incidence to the failure event of interest, W^0 denote the calendar time of the disease incidence and X^0 denote a $p \times 1$ vector of covariates. Assume that the sampling time, ξ, is independent of (W^0, T^0, X^0). An individual would be qualified to be sampled at time ξ only if $T^0 + W^0 \geqslant \xi \geqslant 0$. Denote by (W, T, X) the random variables from the prevalent population. The probability distribution of (W, T, X) is the same as the probability distribution of (W^0, T^0, X^0) conditional on $T^0 + W^0 \geqslant \xi \geqslant W^0$.

In practice, the observation of failure time T in the prevalent cohort is subject to right censoring due to the study ending or premature dropout. The censoring time measured from recruitment, C, is usually assumed to be independent of (T, A) given X. However, the total censoring time $A + C$ and the survival time T are correlated, as they share the same A. Let $Y = \min(T, A + C)$ denote the follow-up time until failure or censoring, and let $\Delta = I(T \geqslant A + C)$ be the indicator of failure. For subject $i \in \{1, \ldots, n\}$, denote by x_i the covariate vector, by y_i and a_i the observed survival time and truncation time, and by δ_i the indicator of an uncensored event time. The observed data $(y_i, a_i, \delta_i, x_i)$ for $i = 1, \ldots, n$ are assumed to be independent and identically distributed realizations of (Y, A, Δ, X).

Denote by $f(t \mid x)$ and $S(t \mid x)$ the conditional density function and survival function of $T^0 = t$ given $X^0 = x$, and let $\mu(x) = \int_0^\infty u f(u \mid x) \, du$ be the conditional mean of T^0 given $X^0 = x$. We impose the following conditions for incident population random variables.

Assumption 1. The variable (T^0, X^0) is independent of when the disease incidence occurs, W^0.

Assumption 2. Disease incidence occurs over calendar time at a constant rate.

Under Assumptions 1 and 2, the joint density function of (A, T) given $X = x$ evaluated at (a, t) is $f(t \mid x)\mu(x)^{-1}I(t > a > 0)$ (Lancaster, 1990, Ch. 3), and the survival time T given $X = x$ has a length-biased density function $t f(t \mid x)\mu(x)^{-1}$.

We assume that the survival time T^0 in the incident population follows the Cox (1972) proportional hazards model $\lambda(t \mid x) = \lambda(t) \exp(\beta' x)$, where $\lambda(t)$ is an unspecified, continuous baseline

hazard function and β is a vector of $p \times 1$ regression parameters. Let $\Lambda(t) = \int_0^t \lambda(u)\,du$ be the cumulative baseline hazard function. Under Assumptions 1 and 2 and the independence of C and (T, A) given X, the full likelihood function is proportional to

$$\mathcal{L}(\beta, \Lambda) = \prod_{i=1}^n \frac{f(y_i \mid x_i)^{\delta_i} S(y_i \mid x_i)^{1-\delta_i}}{\mu(x_i)} = \prod_{i=1}^n \frac{\{\lambda(y_i)\exp(\beta'x_i)\}^{\delta_i}\exp\{-\Lambda(y_i)\exp(\beta'x_i)\}}{\int_0^\infty \exp\{-\Lambda(u)\exp(\beta'x_i)\}\,du}.$$

(1)

2·2. *Brief review of existing methods*

The likelihood (1) can be re-expressed as the product of the truncation likelihood conditional on A and the marginal likelihood of A:

$$\mathcal{L}(\beta, \Lambda) = \mathcal{L}_{\mathrm{T}}(\beta, \Lambda) \times \mathcal{L}_{\mathrm{M}}(\beta, \Lambda) = \prod_{i=1}^n \left\{ \frac{f(y_i \mid x_i)^{\delta_i} S(y_i \mid x_i)^{1-\delta_i}}{S(a_i \mid x_i)} \right\} \times \prod_{i=1}^n \left\{ \frac{S(a_i \mid x_i)}{\mu(x_i)} \right\}.$$

Written in this way, we see that there is information about the regression parameter β in $\mathcal{L}_{\mathrm{M}}(\beta, \Lambda)$. The truncation likelihood $\mathcal{L}_{\mathrm{T}}$ can be further decomposed as the product of the partial likelihood (Kalbfleisch & Lawless, 1991)

$$\mathcal{L}_{\mathrm{P}}(\beta) = \prod_{i=1}^n \left\{ \frac{\exp(\beta'x_i)}{\sum_{j=1}^n \exp(\beta'x_j)I(a_j \leqslant y_i \leqslant y_j)} \right\}^{\delta_i},$$

and the residual likelihood $\mathcal{L}_{\mathrm{R}}(\beta, \Lambda)$. Wang et al. (1993) showed that $\mathcal{L}_{\mathrm{P}}$ is fully efficient with respect to $\mathcal{L}_{\mathrm{T}}$. However, under length-biased sampling the maximum partial likelihood estimator is expected to be inefficient, because it ignores information in $\mathcal{L}_{\mathrm{M}}(\beta, \Lambda)$.

Various methods that better exploit the special structure of length-biased survival data have been proposed in the literature. Let $G(t)$ be the survival function of the censoring time C, and let $\hat{G}(t)$ be the Kaplan–Meier estimator of $G(t)$ based on $\{(y_i - a_i, 1 - \delta_i) : i = 1, \dots, n\}$. Qin & Shen (2010) proposed to solve the weighted estimating equation

$$U_1(\beta) = \sum_{i=1}^n \delta_i \left[x_i - \frac{\sum_{j=1}^n \delta_j x_j \exp(\beta'x_j)\{y_j \hat{G}(y_j - a_j)\}^{-1} I(y_j \geqslant y_i)}{\sum_{j=1}^n \delta_j \exp(\beta'x_j)\{y_j \hat{G}(y_j - a_j)\}^{-1} I(y_j \geqslant y_i)} \right] = 0,$$

where the contribution of a subject in the risk set is inversely weighted by the probability of the subject being sampled and uncensored. This estimating method, however, might be unstable as the weight function $y_j^{-1}\hat{G}(y_j - a_j)^{-1}$ involves estimation of the tail probability of the censoring distribution. As an alternative, Qin & Shen (2010) considered solving the estimating equation

$$U_2(\beta) = \sum_{i=1}^n \delta_i \left[x_i - \frac{\sum_{j=1}^n \delta_j x_j \exp(\beta'x_j)\{\hat{w}_c(y_j)\}^{-1} I(y_j \geqslant y_i)}{\sum_{j=1}^n \delta_j \exp(\beta'x_j)\{\hat{w}_c(y_j)\}^{-1} I(y_j \geqslant y_i)} \right] = 0,$$

with $\hat{w}_c(y) = \int_0^y \hat{G}(u)\,du$. The weight function $\hat{w}_c(y_j)^{-1}$ is the integral of the censoring survival function, which is more stable than the weight function $y_j^{-1}\hat{G}(y_j - a_j)^{-1}$ in U_1. A major restriction of the two estimating equation-based methods is that the censoring time must not depend on the covariates. Moreover, the estimating equations only use covariate information from uncensored individuals, suggesting that there is still room for efficiency gains.

2·3. *Maximum pseudo-profile likelihood estimator*

The maximum likelihood estimator could be obtained by applying the semiparametric profile likelihood method (Murphy & van der Vaart, 2000) to deal with the nuisance parameter Λ. For length-biased sampling data, however, maximizing $\mathcal{L}$ with respect to Λ for fixed β is computationally difficult because $\mathcal{L}$ involves Λ in a complicated way. Instead of profiling out the nonparametric component Λ in $\mathcal{L}$, we propose to replace $\Lambda(t)$ with a simple estimate that is consistent and has a $n^{1/2}$-convergence rate. This approach has been used in various contexts under various names, including pseudo- and estimated-likelihood estimation (Gong & Samaniego, 1981; Pepe & Fleming, 1991; Severini & Wong, 1992; Zucker, 2005).

Our simple estimate is based on profiling the truncation likelihood $\mathcal{L}_{\mathrm{T}}(\beta, \Lambda)$. Specifically, for fixed β, the truncation likelihood $\mathcal{L}_{\mathrm{T}}(\beta, \Lambda)$ is maximized by the Breslow-type estimator

$$\hat{\Lambda}_\beta(t) = \int_0^t \frac{d\{\sum_{j=1}^n \delta_j I(y_j \leqslant u)\}}{\sum_{j=1}^n \exp(\beta' x_j) I(a_j \leqslant u \leqslant y_j)}$$

in the class of nondecreasing right-continuous functions which jump only at uncensored failure times. Note that $\hat{\Lambda}_\beta(t)$ can be generalized to handle time-varying covariates. Profiling out Λ from the truncation likelihood $\mathcal{L}_{\mathrm{T}}(\beta, \Lambda)$ yields the partial likelihood, that is, $\mathcal{L}_{\mathrm{T}}(\beta, \hat{\Lambda}_\beta) = \mathcal{L}_{\mathrm{P}}(\beta)$. Replacing Λ with $\hat{\Lambda}_\beta$ in the full likelihood $\mathcal{L}$, we obtain a pseudo-profile likelihood function,

$$\mathcal{L}(\beta, \hat{\Lambda}_\beta) = \mathcal{L}_{\mathrm{T}}(\beta, \hat{\Lambda}_\beta) \times \mathcal{L}_{\mathrm{M}}(\beta, \hat{\Lambda}_\beta) = \mathcal{L}_{\mathrm{P}}(\beta) \times \prod_{i=1}^n \frac{\exp\{-\hat{\Lambda}_\beta(a_i) \exp(\beta' x_i)\}}{\int_0^\infty \exp\{-\hat{\Lambda}_\beta(u) \exp(\beta' x_i)\} \, du}.$$

We propose to estimate the regression parameter β by maximizing the pseudo-profile likelihood.

Assume that T^0, and hence T, has a finite maximal support τ, where $\tau = \sup\{t : \mathrm{pr}(T^0 \leqslant t) < 1\} < \infty$. Then τ is also the maximal support for the truncation time random variable A, as A given T has a uniform distribution on $[0, T]$. We further assume that C is not degenerate at 0, that is, $\mathrm{pr}(C > 0) > 0$. Then it can be shown that $\max \Delta_i Y_i \to \tau$ as $n \to \infty$. Thus, $\Lambda(t)$ is estimable on the interval $[0, \tau]$; as a result, the conditional mean of T^0 given X is also estimable. Let $N_i(t) = \delta_i I(y_i \leqslant t)$ be the counting process of observed failure events for subject i, and denote $\bar{N}(t) = n^{-1} \sum_{i=1}^n N_i(t)$ and $F^u(t) = \mathrm{pr}(\Delta = 1, Y \leqslant t)$. Define the functions $S^{(k)}(u, \beta) = n^{-1} \sum_{i=1}^n x_i^{\otimes k} \exp(\beta' x_i) I(a_i \leqslant u \leqslant y_i)$ $(k = 0, 1, 2)$, and let $\mathcal{S}^{(k)}(u, \beta) = E\{X^{\otimes k} \exp(\beta' X) I(A \leqslant u \leqslant Y)\}$ be the expectations. Assume that X is bounded, that the two classes of functions $\{\Delta I(Y \leqslant t) : t \in [0, \tau]\}$ and $\{X^{\otimes k} \exp(\beta' X) I(A \leqslant t \leqslant Y) : t \in [0, \tau], \beta \in \Theta\}$ are both Glivenko–Cantelli, as the class of indicator functions and the class of bounded monotone functions are Glivenko–Cantelli (van der Vaart & Wellner, 1996, Theorems 2.4.1 and 2.7.5). Moreover, because $S^{(0)}(t, \beta)$ is bounded away from zero, we can show that $\sup_{t \in [0,\tau], \beta \in \Theta} | \hat{\Lambda}_\beta(t) - \Lambda_\beta(t) | \to 0$ almost surely as $n \to \infty$, where

$$\Lambda_\beta(t) = \int_0^t \frac{dF^u(u)}{\mathcal{S}^{(0)}(u, \beta)}. \tag{2}$$

The limit $\Lambda_\beta(t)$ of $\hat{\Lambda}_\beta(t)$ defines a smooth mapping in β, and it passes through the true baseline cumulative hazard function $\Lambda(t)$ when β equals the true parameter value. If we regard (2) as a known function of β, the function $\mathcal{L}(\beta, \Lambda_\beta)$ can be viewed as the full likelihood function derived under an induced parametric submodel $\lambda(t \mid x) = \lambda_\beta(t) \exp(\beta' X)$.

Replacing Λ with $\hat{\Lambda}_\beta$ in $\mathcal{L}(\beta, \Lambda)$, we obtain a log pseudo-profile likelihood function $\ell(\beta) = \ell_P(\beta) + \ell_M(\beta)$, where

$$\ell_P(\beta) = \sum_{i=1}^{n} \int_0^\tau [\beta' x_i - \log\{S^{(0)}(u, \beta)\}] \, dN_i(u)$$

is the log partial likelihood obtained by profiling out Λ from the truncation likelihood $\mathcal{L}_T$, and

$$\ell_M(\beta) = -\sum_{i=1}^{n} \{\hat{\Lambda}_\beta(a_i) \exp(\beta' x_i) + \log \hat{\mu}_\beta(x_i)\},$$

with $\hat{\mu}_\beta(x_i) = \int_0^\infty \exp\{-\hat{\Lambda}_\beta(u) \exp(\beta' x_i)\} \, du$. We show in the Appendix that, in a compact neighbourhood of the true regression parameter, $\ell(\beta)$ can be approximated by $\tilde{\ell}(\beta) = \tilde{\ell}_P(\beta) + \tilde{\ell}_M(\beta)$, where $\tilde{\ell}_P(\beta) = \sum_{i=1}^{n} \int_0^\tau [\beta' x_i - \log\{S^{(0)}(u, \beta)\}] \, dN_i(u)$, $\tilde{\ell}_M(\beta) = -\sum_{i=1}^{n} \{\Lambda_\beta(a_i) \exp(\beta' x_i) + \log \mu_\beta(x_i)\}$ and $\mu_\beta(x_i) = \int_0^\infty \exp\{-\Lambda_\beta(u) \exp(\beta' x_i)\} \, du$. Thus, ℓ and $\tilde{\ell}$ have similar local behaviour in the compact neighbourhood, and the asymptotic properties of the maximum pseudo-profile likelihood estimator can be investigated through $\tilde{\ell}$.

Define the limit function $\gamma(\beta) = \lim_{n\to\infty} n^{-1}\tilde{\ell}(\beta) = \lim_{n\to\infty} n^{-1}\{\tilde{\ell}_P(\beta) + \tilde{\ell}_M(\beta)\}$. We denote the true parameter values of the proportional hazards model by $\{\beta_0, \lambda_0(\cdot)\}$, and define $\Lambda_0(t) = \int_0^t \lambda_0(u) \, du$. Theorem 1 summarizes the consistency and asymptotic normality of $\hat{\beta}$ that maximizes the log pseudo-profile likelihood function $\ell(\beta)$, with proofs given in the Appendix.

THEOREM 1. *Assume the following conditions hold*: (a) β_0 *lies in the interior of a known compact set* Θ *in* $\mathbb{R}^p$; (b) X *is bounded*; (c) $\mathrm{pr}(Y \geqslant t)$ *is a continuous function for* $t \in [0, \tau]$ *and* (d) $\partial^2 \gamma(\beta_0)/\partial\beta'\partial\beta$ *is nonsingular. Then* $\hat{\beta} \to \beta_0$ *in probability as* $n \to \infty$. *Moreover*, $n^{1/2}(\hat{\beta} - \beta_0)$ *converges in distribution to a zero mean multivariate normal distribution with variance-covariance matrix* $\Sigma(\beta_0)$, *where* $\Sigma(\beta_0)$ *is specified in the Appendix*.

While the asymptotic variance $\Sigma(\beta_0)$ may be estimated by its empirical version, the computation is quite complicated. Since we have established the asymptotic normality, it is computationally more convenient to use the bootstrap method. The performance of the proposed estimator is evaluated in §3 via simulations.

2·4. *Efficiency considerations*

To investigate the potential efficiency gains in the proposed pseudo-profile likelihood estimator, we first consider the case that Λ is parameterized by a vector of $q \times 1$ parameters ν, that is, $\Lambda(t) = \Lambda(t, \nu)$. For model identifiability, we assume without loss of generality that $E(X) = 0$. Define the log truncation likelihood function $\ell_T = \log(\mathcal{L}_T)$. The proposed method is equivalent to solving the system of estimating equations $\partial\ell_T/\partial\beta + \partial\ell_M/\partial\beta = 0$ and $\partial\ell_T/\partial\nu = 0$. Let $\eta = (\partial\ell_T/\partial\beta, \partial\ell_T/\partial\nu, \partial\ell_M/\partial\beta)$ be a vector of score functions. Define

$$\begin{pmatrix} a_{11} & a_{12} \\ a'_{12} & a_{22} \end{pmatrix} = -E \begin{pmatrix} \partial^2\ell_T/\partial\beta'\partial\beta & \partial^2\ell_T/\partial\beta'\partial\nu \\ \partial^2\ell_T/\partial\beta\partial\nu' & \partial^2\ell_T/\partial\nu'\partial\nu \end{pmatrix},$$

and let $b_1 = -E(\partial^2\ell_M/\partial\beta'\partial\beta)$ and $b_2 = -E(\partial^2\ell_M/\partial\beta'\partial\nu)$. Denote $\nu^* = (\beta, \nu)$. Then the optimal linear combination of estimating functions is

$$E\left(\frac{\partial\eta}{\partial\nu^*}\right)' \mathrm{var}(\eta)^{-1}\eta = \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a'_{12} & a_{22} & b'_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & 0 \\ a'_{12} & a_{22} & 0 \\ 0 & 0 & b_1 \end{pmatrix}^{-1} \eta = \begin{pmatrix} I_p & 0 & I_p \\ 0 & I_q & b'_2 b_1^{-1} \end{pmatrix} \eta, \quad (3)$$

where, for convenience, 0 denotes a matrix of 0s of appropriate dimensions and I_p is a $p \times p$ identity matrix. It can be verified that, when evaluated at the true parameter values, $b_2 = n/2 \times E(X \exp(2\beta'X) \times \partial/\partial\nu'[\Lambda(A)^2 - E\{\Lambda(A) \mid X\}^2])$. Hence, if $b_2 = 0$, the system of estimating equations $\partial\ell_T/\partial\beta + \partial\ell_M/\partial\beta = 0$ and $\partial\ell_T/\partial\nu = 0$ is the optimal linear combination of estimating equations based on η. The partial likelihood method solves the system of estimating equations $\partial\ell_T/\partial\beta = 0$ and $\partial\ell_T/\partial\nu = 0$, which also belongs to the class of linear combinations of estimating equations based on η. Thus, the proposed method is more efficient than the partial likelihood method when $b_2 = 0$.

When the baseline hazard function λ is of infinite dimension, the proposed pseudo-profile likelihood method solves the system (van der Vaart, 1998, §25.12)

$$\mathbb{P}_n(\partial\ell_T/\partial\beta + \partial\ell_T/\partial\beta) = 0, \quad \mathbb{P}_n\Psi h - P\Psi h = 0, \quad (h \in H),$$

where Ψ is the score operator (Begun et al., 1983) for Λ based on the truncation likelihood $\ell_T = \log\mathcal{L}_T$ and H is a infinite-dimensional class of direction h from which paths of one-dimensional submodels for Λ may approach the true parameter. We use $\mathbb{P}_n$ to denote the empirical measure, and use P for the probability measures. Let $L^2(\mu)$ denote the Hilbert space that contains square integrable functions with the inner product $\langle g, h \rangle_\mu = \int g(u)h(u)\, d\mu(u)$ for $g, h \in L^2(\mu)$. It is easy to see that $\mathcal{H} \subset L^2(\Lambda)$. Applying a similar argument as in van der Vaart (1998, §25.12.1), we can show that the score operator $\Psi : L^2(\Lambda) \to L^2(P_{\beta,\Lambda})$ for Λ is given by $\Psi(h) = \int_0^\tau h(u)\, dM(u)$. Let $\bar{H}$ be a Hilbert space containing $\mathcal{H}$. The adjoint operator $\Psi^* : L^2(P_{\beta,\Lambda}) \to \bar{H}$ of Ψ, which satisfies $E\{\Psi(g)h\} = \int_0^\tau g(u)\Psi^*(h)(u)\, d\Lambda(u)$ for all $g \in \bar{H}$ and $h \in L^2(P_{\beta,\Lambda})$, can be shown to be $\Psi^*(g)(t) = E\{g\, dM(t)\}/d\Lambda(t)$. It can be further shown that $\Psi^*\Psi(h)(t) = E\{h(t)\exp(\beta'X)I(Y \geqslant t \geqslant A)\}$ and $\Psi^*(\partial\ell_T/\partial\beta)(t) = nE\{X\exp(\beta'X)I(Y \geqslant t \geqslant A)\}$ (Murphy & van der Vaart, 2000).

By a similar argument as above, we show in the Supplementary Material that the score operator $\Phi : L^2(\Lambda) \to L^2(P_{\beta,\Lambda})$ for Λ based on the marginal likelihood ℓ_M is $\Phi(h) = -\int_0^\tau [h(u)X\exp(\beta'X)\{I(A \geqslant u) - \mathrm{pr}(A \geqslant u \mid X)\}]\, d\Lambda(u)$. The adjoint operator $\Phi^* : L^2(P_{\beta,\Lambda}) \to \bar{H}$ of Φ can be shown to be $\Phi^*(g)(t) = -E[g\{I(A \geqslant t) - \mathrm{pr}(A \geqslant t \mid X)\}\exp(\beta'X)]$. Moreover, the adjoint operator Φ^* satisfies $\Phi^*\Phi(h)(t) = E[\int_0^\tau h(u)\{I(A \geqslant u) - \mathrm{pr}(A \geqslant u \mid X)\}\, d\Lambda(u)\{I(A \geqslant t) - \mathrm{pr}(A \geqslant t \mid X)\}\exp(2\beta'X)]$ and $\Phi^*(\partial\ell_M/\partial\beta)(t) = nE[\{I(A \geqslant t) - \mathrm{pr}(A \geqslant t \mid X)\}\{\Lambda(A) - E(\Lambda(A) \mid X)\}X\exp(2\beta'X)]$.

Analogous to (3) for parametric models, the optimal combination of estimating functions based on the score operators $\partial\ell_T/\partial\beta$, $\partial\ell_M/\partial\beta$ and Ψ is given by $\mathbb{P}_n(\partial\ell_T/\partial\beta + \partial\ell_T/\partial\beta) = 0$ and $(\mathbb{P}_n - P)\Psi h + (b_2^*)'b_1^{-1}\partial\ell_T/\partial\beta = 0$, $h \in H$, where $b_1 = -E(\partial^2\ell_M/\partial\beta'\partial\beta)$ and $b_2^* = \int_0^\tau \Phi^*(\partial\ell_M/\partial\beta)(t)\, d\Lambda(t) = nE([\Lambda(A)^2 - E\{\Lambda(A) \mid X\}^2]X\exp(2\beta'X))$. Hence, if $b_2^* = 0$, the proposed pseudo-partial likelihood method is the most efficient estimator in the class of linear combinations of estimating functions based on $\partial\ell_T/\partial\beta$, $\partial\ell_M/\partial\beta$ and Ψ. The weight $(b_2^*)^\mathrm{T}b_1^{-1}$ can be estimated by replacing Λ with $\hat{\Lambda}$ in the corresponding empirical estimators. In general, solving the optimal combination of estimation equations is computationally intensive, and hence is impractical. Moreover, there is no guarantee that it works better than the proposed method for small samples.

3. Simulations and data analysis

3·1. *Monte-Carlo simulations*

We conducted simulations to assess the performance of the proposed methods. In each simulation, 2000 studies were generated, each with $n = 400$. The sampling time ξ was set to be 100,

and the time of disease onset, W^0, was simulated from a uniform distribution over [0, 100] to mimic the incidence of a stable disease. For each subject, we generated X_1^0 from the Bernoulli distribution with $\mathrm{pr}(X_1^0 = 1) = \mathrm{pr}(X_1^0 = 0) = 0\cdot 5$ and generated X_2^0 from the standard normal distribution. The survival time T^0 was independently generated from one of the three models: (I) an exponential distribution with hazard function $2\exp(X_1^0 + X_2^0)$, (II) a Weibull distribution with hazard function $2t\exp(X_1^0 + X_2^0)$ or (III) a Weibull distribution with hazard function $0\cdot 5(t-2)^2 \exp(X_1^0 + X_2^0)$. Thus, we simulated failure time distributions with constant, increasing and U-shape hazards. To form a prevalent cohort of sample size n, realizations of (W^0, T^0, X_1^0, X_2^0) were generated repeatedly until there were n subjects satisfying the sampling constraint $W^0 + T^0 \geqslant \tau$. The time from enrolment ξ to loss to follow-up was generated from a uniform distribution so that the censoring rate was approximately 0, 30 and 50%.

We compared the finite-sample performance of the proposed pseudo-profile likelihood method with those of the weighted estimating equation methods studied in Qin & Shen (2010) and of the popular partial likelihood method for truncated survival time data. By applying these methods to estimate the Cox model $\lambda(t \mid X_1, X_2) = \lambda_0(t)\exp(\beta_1 X_1 + \beta_2 X_2)$, we evaluated the relative efficiency by comparing the bootstrap variance of the maximum partial likelihood estimator to that of the other methods. Table 1 summarizes the empirical bias, empirical standard error and the relative efficiency of these four estimation methods. All four estimators are close to their estimands. In the absence of censoring, the pseudo-profile likelihood method has a similar efficiency gain as the weighted estimating equation methods in Qin & Shen (2010). Overall, the relative efficiency of the proposed estimator increases with censoring rate. When the censoring proportion reaches 50%, the pseudo-profile likelihood estimator yields a significant improvement over the maximum partial likelihood estimator, with an efficiency gain greater than 50% in the exponential and Weibull cases, and an efficiency gain greater than 20% in the U-shape hazard function scenario. In the presence of censoring, the proposed pseudo-profile method always outperforms its competitors. In some scenarios, weighted estimating equation methods fail to show improvement, as these methods only use covariate information from uncensored subjects.

In addition to better efficiency, another advantage of the proposed pseudo-profile likelihood method is that it does not involve estimation of the censoring distribution. When this distribution depends on the covariate, the estimating equation methods may yield biased estimation. For demonstration, we simulated survival time data under Model (II). The censoring times for subjects with observed covariates $X_1 = 1$ and $X_2 < 0$ were generated from an exponential distribution with mean $5\exp(-X_2)$, while the censoring times for other subjects were generated from a uniform distribution. The overall censoring proportion was set at approximately 30 and 50%. As summarized in Table 2, the estimating equation-based methods yield biased estimators, while the bias of the pseudo-profile estimators remains small.

3·2. *Analysis of Canadian Study of Health and Aging*

In this section, we report the results of data analysis for a cohort of prevalent cases in one of the largest epidemiologic studies of dementia, the Canadian Study of Health and Aging. From February 1991 to May 1992, an extensive survey was carried out and a total of 1132 persons aged 65 and older with dementia were identified in this first phase of the study. For each study subject, a diagnosis of possible Alzheimer's disease, probable Alzheimer's disease, or vascular dementia was assigned, and the date of dementia onset was determined by interviewing caregivers. Information on mortality were collected between January 1996 and May 1997.

We considered a subset of the study data by excluding those with missing date of onset or missing dementia subtype classification. Moreover, as in Wolfson et al. (2001), those with

Table 1. *Summary statistics for the estimated regression parameters under independent censoring*

Proportion censored	Estimated coefficient	Partial		WEE-1		WEE-2		Profile		
		Bias	SE	Bias	SE	Bias	SE	Bias	SE	RE
Scenario I: $\lambda_0(t) = 2$										
0%	$\hat{\beta}_1$	6	133	−1	98	−1	98	−2	98	1·84
	$\hat{\beta}_2$	5	83	0	65	0	65	−2	65	1·60
30%	$\hat{\beta}_1$	6	151	−46	136	2	120	1	108	1·98
	$\hat{\beta}_2$	6	94	−57	90	2	82	4	77	1·50
50%	$\hat{\beta}_1$	10	171	−113	189	10	157	12	122	1·96
	$\hat{\beta}_2$	6	112	−125	115	2	98	10	90	1·53
Scenario II: $\lambda_0(t) = 2t$										
0%	$\hat{\beta}_1$	4	118	0	98	0	98	−1	97	1·45
	$\hat{\beta}_2$	2	74	0	63	0	63	−1	63	1·40
30%	$\hat{\beta}_1$	3	132	−21	137	9	121	3	105	1·58
	$\hat{\beta}_2$	5	89	−28	90	4	80	2	73	1·46
50%	$\hat{\beta}_1$	6	159	−101	206	10	154	5	118	1·82
	$\hat{\beta}_2$	8	100	−97	118	6	96	5	79	1·62
Scenario III: $\lambda_0(t) = 0{\cdot}5(t-2)^2$										
0%	$\hat{\beta}_1$	6	112	5	104	5	104	5	104	1·17
	$\hat{\beta}_2$	6	69	5	65	5	65	5	64	1·14
30%	$\hat{\beta}_1$	10	134	10	143	10	130	9	122	1·21
	$\hat{\beta}_2$	7	82	0	85	3	80	4	77	1·14
50%	$\hat{\beta}_1$	9	151	−34	216	7	154	6	134	1·27
	$\hat{\beta}_2$	7	97	−35	123	7	98	5	88	1·20

Partial, the maximum partial likelihood estimator; WEE-1 and WEE-2, estimators derived by solving $U_1(\beta) = 0$ and $U_2(\beta) = 0$; Profile, the maximum pseudo-profile likelihood estimator; Bias and ES, empirical bias ($\times 1000$) and empirical standard deviation ($\times 1000$) of 2000 regression parameter estimates; RE, the empirical variance of the maximum partial likelihood estimator divided by that of the maximum pseudo-profile likelihood estimator.

Table 2. *Empirical bias and standard error of estimators of estimated regression parameters under covariate dependent censoring*

Proportion censored	Estimated coefficient	Partial		WEE-1		WEE-2		Profile		
		Bias	SE	Bias	SE	Bias	SE	Bias	SE	RE
30%	$\hat{\beta}_1$	7	132	−388	135	−127	119	2	105	1·56
	$\hat{\beta}_2$	8	87	48	84	30	79	4	73	1·44
50%	$\hat{\beta}_1$	10	166	−819	175	−252	161	−2	128	1·69
	$\hat{\beta}_2$	8	103	82	107	51	97	6	80	1·64

See Table 1 for abbreviations.

observed survival times greater than or equal to 20 years were excluded because these subjects are considered unlikely to have Alzheimer's disease or vascular dementia. As a result, a total of 807 dementia patients were included in our analysis. Among them 388 had a diagnosis of probable Alzheimer's disease, 249 had possible Alzheimer's disease and 170 had vascular dementia. In the second phase of the study, a total of 627 deaths were recorded, among whom 302 had a diagnosis of probable Alzheimer's, 189 had possible Alzheimer's and 136 had vascular dementia.

The stationarity assumption that the incidence of dementia is constant over time was found to be reasonably met for this data using the method suggested in Wang (1991). To compare

Table 3. *Estimated regression coefficients for the Canadian study*

	β_1, probable Alzheimer's			β_2, vascular dementia		
Method	Estimate	SE	95% CI	Estimate	SE	95% CI
Partial	0·030	0·089	(−0·142, 0·203)	0·113	0·109	(−0·103, 0·323)
EE-1	0·130	0·095	(−0·058, 0·312)	0·278	0·107	(0·070, 0·497)
EE-2	0·157	0·088	(−0·022, 0·328)	0·257	0·121	(0·038, 0·519)
Profile	0·150	0·068	(0·016, 0·278)	0·241	0·088	(0·066, 0·419)

Partial, maximum partial likelihood estimator; EE-1 and EE-2, estimators derived by solving $U_1(\beta) = 0$ and $U_2(\beta) = 0$; Profile, the pseudo-profile likelihood estimator; SE, the empirical standard deviation of 2000 regression parameter.

the risk of death between different diagnoses, we fit a Cox proportional hazards model for the length-biased survival time data, with indicators of probable Alzheimer's and vascular dementia as covariates. We applied the pseudo-profile likelihood method, the two weighted estimating equation methods in Qin & Shen (2010), and the partial likelihood method. The estimated regression coefficients are summarized in Table 3. The proposed method yields similar estimates of the regression parameters as do the estimating equation methods, and the bootstrap standard errors of the proposed estimator are smaller than those of its competitors. The proposed pseudo-profile likelihood method estimates a significant higher risk of death in patients with probable Alzheimer's and those with vascular dementia. Specifically, as compared with patients with possible Alzheimer's, the risk of death increased by 16% among those with probable Alzheimer's and by 27% among those with vascular dementia. For β_1, the variance ratio for the competitors to the proposed method is always at least 1·67. This suggests that if a competitor method were used in lieu of the proposed method, the study would need to recruit at least 760 more subjects to achieve the same precision.

4. REMARK

The validity of the proposed method relies on the assumption of stable disease. When the stationarity assumption fails to hold, it is not uncommon that knowledge about the distribution of disease incidence can be obtained from other sources. If H denotes the distribution of the truncation time in the disease population, then the transformed survival time $H(T^0)$ is truncated by a uniformly distributed random variable. Thus, it follows from the fact that the Cox model is invariant under monotone transformation, that the regression coefficients in the Cox model can be consistently estimated by applying the proposed method to the transformed data $\{H(a_i), H(y_i), \delta_i\}$ $(i = 1, \ldots, n)$.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the derivation of the score operator Φ and the adjoint operator Φ^* based on the marginal likelihood function ℓ_M.

Appendix

Proofs

We begin by establishing the consistency of $\hat{\beta}$. In view of the proof of van der Vaart (1998, Theorem 5.7), it suffices to show that, as $n \to \infty$, $\sup_{\beta \in \Theta} |n^{-1}\ell(\beta) - \gamma(\beta)| \to 0$ almost surely and that β_0 is the unique maximizer of $\gamma(\beta)$ in a compact neighbourhood of β_0.

We first show that, for sufficiently large n, $\ell(\beta)$ has similar local behaviour to $\tilde{\ell}(\beta)$ in the compact neighbourhood Θ. Because $\{\exp(\beta'X)I(A \leqslant t \leqslant Y) : t \in [0, \tau], \beta \in \Theta\}$ is Glivenko–Cantelli and the logarithmic transformation is monotone, $\log\{S^{(0)}(t, \beta)\} - \log\{\mathcal{S}^{(0)}(t, \beta)\}$ converges to 0 uniformly over $\beta \in \Theta$ and $t \in [0, \tau]$. Hence, $\sup_{\beta \in \Theta} |n^{-1}\ell_P(\beta) - n^{-1}\tilde{\ell}_P(\beta)| \to 0$ almost surely. Following the result that $\hat{\Lambda}_\beta(t)$ converges to $\Lambda_\beta(t)$ uniformly over $\beta \in \Theta$ and $t \in [0, \tau]$, $\exp\{\hat{\Lambda}_\beta(t)\exp(\beta'X)\}$ converges to $\exp\{\Lambda_\beta(t)\exp(\beta'X)\}$ uniformly over $\beta \in \Theta$ and $t \in [0, \tau]$. Hence, $\hat{\mu}_\beta(x)$ converges to $\mu_\beta(x)$ uniformly over $\beta \in \Theta$. For a $\delta_n > 0$ with $\delta_n \to 0$ as $n \to \infty$, define the class $\mathcal{F} = [f(t) = \{g(t) - \Lambda_\beta(t)\}\exp(\beta'X)$, where $\beta \in \Theta$, g is nondecreasing and nonnegative and $\sup_{t \in [0, \tau]} |g(t) - \Lambda_\beta(t)| \leqslant \epsilon_n]$. Thus, by definition, $\sup_{f \in \mathcal{F}} |Pf| \leqslant \epsilon_n \times \sup_{\beta \in \Theta} |\exp(\beta'X)|$. Moreover, it follows from van der Vaart & Wellner (1996, Theorems 2.7.5 and 2.4.1) that $\mathcal{F}$ is Glivenko–Cantelli. Hence, $\sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \to 0$ almost surely. For a sufficiently large n, $|n^{-1}\sum_{i=1}^{n}\{\hat{\Lambda}_\beta(a_i) - \Lambda_\beta(a_i)\}\exp(\beta'X_i)| \leqslant \sup_{f \in \mathcal{F}}|Pf| + \sup_{f \in \mathcal{F}}|\mathbb{P}_n f - Pf|$. Thus, we show that $\sup_{\beta \in \Theta}|n^{-1}\sum_{i=1}^{n}\{\hat{\Lambda}_\beta(a_i) - \Lambda_\beta(a_i)\}\exp(\beta'X_i)| \to 0$ almost surely. By a similar argument, we can show that $\sup_{\beta \in \Theta}|n^{-1}\sum_{i=1}^{n}\{\hat{\mu}_\beta(x_i) - \mu_\beta(x_i)\}| \to 0$ almost surely, and hence $n^{-1}\ell_M(\beta) - n^{-1}\tilde{\ell}_M(\beta) \to 0$ uniformly over $\beta \in \Theta$. Thus, $\ell(\beta) = \ell_P(\beta) + \ell_M(\beta)$ and $\tilde{\ell}(\beta) = \ell_P(\beta) + \tilde{\ell}_M(\beta)$ have similar local behaviour in Θ.

Next, because Θ is compact and the function $m(\beta) = \int_0^\tau [\beta'X - \log\{\mathcal{S}^{(0)}(u, \beta)\}]\,dN(u) - \Lambda_\beta(A)\exp(\beta'X) - \log\mu_\beta(X)$ is continuous and dominated by an integrable function, the class of functions $\{m(\beta) : \beta \in \Theta\}$ is Glivenko–Cantelli (van der Vaart, 1998, Example 19.8). It follows from a uniform law of large numbers (Pollard, 1990) that $\sup_{\beta \in \Theta}|n^{-1}\tilde{\ell}(\beta) - \gamma(\beta)| \to 0$ almost surely. Thus, $\sup_{\beta \in \Theta}|n^{-1}\ell(\beta) - \gamma(\beta)| \to 0$ almost surely as $n \to \infty$.

Below we prove that β_0 is the unique maximizer in a neighbourhood of β_0 by showing that $\partial\gamma(\beta_0)/\partial\beta = 0$ and $\partial^2\gamma(\beta_0)/\partial\beta'\partial\beta$ is negative definite at $\beta = \beta_0$. Following the fact that the partial score function has expectation zero when evaluated at $\beta = \beta_0$ and that $E[\partial\{\Lambda_\beta(A)\exp(\beta'X)\}/\partial\beta|_{\beta=\beta_0}] = -E\{\mu_\beta(X)^{-1}\partial\mu_\beta(X)/\partial\beta|_{\beta=\beta_0}\}$ by double expectation, we can show that $\partial\gamma(\beta)/\partial\beta = 0$ when $\beta = \beta_0$. Write $S_\beta(u \mid x) = \exp\{-\Lambda_\beta(u)\exp(\beta'X)\}$. The second derivative of $\gamma(\beta)$ is

$$\frac{\partial^2\gamma(\beta)}{\partial\beta'\partial\beta} = E\left[\Delta\left\{\frac{\mathcal{S}^{(1)}(Y, \beta)^{\otimes 2}}{\mathcal{S}^{(0)}(Y, \beta)^2} - \frac{\mathcal{S}^{(2)}(Y, \beta)}{\mathcal{S}^{(0)}(Y, \beta)}\right\}\right] \tag{A1}$$

$$- E\left[\frac{\partial^2}{\partial\beta'\partial\beta}\{\Lambda_\beta(A)\exp(\beta'X)\}\right] \tag{A2}$$

$$+ E\left[\int_0^\tau \frac{S_\beta(u \mid X)}{\mu_\beta(X)}\frac{\partial^2}{\partial\beta'\partial\beta}\{\Lambda_\beta(u)\exp(\beta'X)\}\,du\right] \tag{A3}$$

$$- E\left(\int_0^\tau \frac{S_\beta(u \mid X)}{\mu_\beta(X)}\left[\frac{\partial}{\partial\beta}\{\Lambda_\beta(u)\exp(\beta'X)\}\right]^{\otimes 2}\,du\right) \tag{A4}$$

$$+ E\left(\left[\int_0^\tau \frac{S_\beta(u \mid X)}{\mu_\beta(X)}\frac{\partial}{\partial\beta}\{\Lambda_\beta(u)\exp(\beta'X)\}\,du\right]^{\otimes 2}\right). \tag{A5}$$

By applying the double expectation technique, it can be shown that (A2) + (A3) = 0 for $\beta = \beta_0$. Moreover, by the Cauchy–Schwarz inequality, both (A1) and (A4) + (A5) are negative semidefinite. Hence, it follows regularity condition (d) that $\partial^2\gamma(\beta)/\partial\beta'\partial\beta$ is negative definite at $\beta = \beta_0$. Because the function $\gamma(\beta)$ is continuous in β, there exists a compact neighbourhood Θ_0 of β_0 that β_0 is the unique maximizer $\gamma(\beta)$ in Θ_0. This completes the proof of consistency.

We now prove the asymptotic normality of the maximum pseudo-profile likelihood estimator. A Taylor series expansion yields $0 = \partial \ell(\beta)/d\beta \mid_{\beta=\hat{\beta}} = \partial \ell(\beta)/\partial \beta \mid_{\beta=\beta_0} + \partial^2 \ell(\beta)/\partial \beta' \partial \beta \mid_{\beta=\beta^*} (\hat{\beta} - \beta_0)$, where β^* lies between $\hat{\beta}$ and β_0. Thus, by consistency of $\hat{\beta}$, one has $\beta^* \to \beta_0$ in probability and

$$n^{1/2}(\hat{\beta} - \beta_0) = - \left\{ n^{-1} \frac{\partial^2 \ell(\beta)}{\partial \beta' \partial \beta} \bigg|_{\beta=\beta_0} \right\}^{-1} \left\{ n^{-1/2} \frac{\partial \ell(\beta)}{\partial \beta} \bigg|_{\beta=\beta_0} \right\} + o_p(1).$$

In what follows, we show that

$$n^{-1/2} \left\{ \frac{\partial \ell_M(\beta)}{\partial \beta} - \frac{\partial \tilde{\ell}_M(\beta)}{\partial \beta} \right\} = -n^{-1/2} \sum_{i=1}^n \frac{\partial}{\partial \beta} [\{\hat{\Lambda}_\beta(a_i) - \Lambda_\beta(a_i)\} \exp(\beta' x_i)] \tag{A6}$$

$$- n^{-1/2} \left\{ \sum_{i=1}^n \frac{1}{\hat{\mu}_\beta(x_i)} \frac{\partial \hat{\mu}_\beta(x_i)}{\partial \beta} - \frac{1}{\mu_\beta(x_i)} \frac{\partial \mu_\beta(x_i)}{\partial \beta} \right\} \tag{A7}$$

has an asymptotic independent and identically distributed representation. Let H be the joint probability measure of (A, X) and let $\hat{H}$ be the corresponding empirical measure for H. Then the right-hand side of (A6) can be expressed as

$$- n^{1/2} \frac{\partial}{\partial \beta} \int_{-\infty}^{\infty} \int_0^\tau \left[\left\{ \frac{d\bar{N}(u)}{S^{(0)}(u, \beta)} - \frac{dF^u(u)}{\mathcal{S}^{(0)}(u, \beta)} \right\} \exp(\beta' X) I(u \leqslant a \leqslant \tau) \right] d\hat{H}(a, x)$$

$$= -n^{1/2} \frac{\partial}{\partial \beta} \int_{-\infty}^{\infty} \int_0^\tau \left[\frac{d\bar{N}(u) - dF^u(u)}{\mathcal{S}^{(0)}(u, \beta)} - \frac{dF^u(u)}{\mathcal{S}^{(0)}(u, \beta)^2} \{S^{(0)}(u, \beta) - \mathcal{S}^{(0)}(u, \beta)\} \right]$$

$$\times \exp(\beta' X) I(u \leqslant a \leqslant \tau) \, d\hat{H}(a, x) + o_p(1)$$

$$= -n^{-1/2} \sum_{i=1}^n \frac{\partial}{\partial \beta} \int_{-\infty}^{\infty} \int_0^\tau \left[\left\{ \frac{dN_i(u)}{\mathcal{S}^{(0)}(u, \beta)} - \frac{\exp(\beta' x_i) I(a_i \leqslant u \leqslant y_i) \, dF^u(u)}{\mathcal{S}^{(0)}(u, \beta)^2} \right\} \right.$$

$$\times \exp(\beta' X) I(u \leqslant a \leqslant \tau) \Bigg] dH(a, x) = -n^{-1/2} \sum_{i=1}^n \phi_{1i}(\beta) + o_p(1).$$

Next, applying the functional delta method, we have $n^{1/2}\{\hat{\mu}_\beta(x) - \mu_\beta(x)\} = n^{-1/2} \sum_{i=1}^n \psi_i(\beta, x) + o_p(n^{-1/2})$, where

$$\psi_i(\beta, x)$$

$$= \sum_{i=1}^n \int_0^\tau \int_0^\tau \left[S_\beta(u \mid x) \exp(\beta' X) \left\{ \frac{dN_i(v)}{\mathcal{S}^{(0)}(v, \beta)} - \frac{dF^u(v)}{\mathcal{S}^{(0)}(v, \beta)^2} \exp(\beta' x_i) I(a_i \leqslant v \leqslant y_i) \right\} \right] du.$$

Thus, (A7) can be expressed as

$$n^{1/2} \int_{-\infty}^{\infty} \int_0^\tau \left[\frac{1}{\mu_\beta(x)} \left\{ \frac{\partial \hat{\mu}_\beta(x)}{\partial \beta} - \frac{\partial \mu_\beta(x)}{\partial \beta} \right\} - \frac{1}{\mu_\beta(x)^2} \frac{\partial \mu_\beta(x)}{\partial \beta} \{\hat{\mu}_\beta(x) - \mu_\beta(x)\} \right] d\hat{H}(a, x) + o_p(1)$$

$$= n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\infty} \int_0^\tau \left[\frac{1}{\mu_\beta(x)} \left\{ \frac{\partial \psi_i(\beta, x)}{\partial \beta} - \frac{\psi_i(\beta, x)}{\mu_\beta(x)} \frac{\partial \mu_\beta(x)}{\partial \beta} \right\} \right] dH(a, x) + o_p(1)$$

$$= n^{-1/2} \sum_{i=1}^n \phi_{2i}(\beta) + o_p(1).$$

Finally, applying the functional delta method, we can obtain the asymptotic representation for the partial score function: $n^{-1/2}\partial\ell_P(\beta_0)/\partial\beta = n^{-1/2}\sum_{i=1}^{n}\phi_{3i}(\beta_0) + o_p(1)$, where

$$\phi_{3i}(\beta_0) = \int_0^\tau \left\{ x_i - \frac{\mathcal{S}^{(1)}(u,\beta_0)}{\mathcal{S}^{(0)}(u,\beta_0)} \right\} \{dN_i(u) - \exp(\beta_0' x_i) I(a_i \leqslant u \leqslant y_i) \, d\Lambda_0(u)\}.$$

Let $\phi_i(\beta_0) = \phi_{1i}(\beta_0) + \phi_{2i}(\beta_0) + \phi_{3i}(\beta_0)$. We have $n^{-1/2}\partial\ell(\beta)/\partial\beta \mid_{\beta=\beta_0} = n^{-1/2}\sum_{i=1}^{n}\kappa_i(\beta_0) + o_p(1)$, where $\kappa_i(\beta_0) = \phi_i(\beta_0) - [\partial\{\Lambda_\beta(a_i)\exp(\beta' x_i)\}/\partial\beta + \mu_\beta(x_i)^{-1}\partial\mu_\beta(x_i)/\partial\beta] \mid_{\beta=\beta_0}$. Arguing as in the proof of consistency, we can show that, as $n \to \infty$, $n^{-1}\{\partial^2\ell(\beta)/\partial\beta'\partial\beta\} \mid_{\beta=\beta_0} \to \{\partial^2\gamma(\beta)/\partial\beta'\partial\beta\} \mid_{\beta=\beta_0}$ almost surely. Define $\Gamma(\beta_0) = E\{\kappa_i(\beta_0)^{\otimes 2}\}$. Hence, under the regularity conditions, as $n \to \infty$, $n^{1/2}(\hat{\beta} - \beta_0)$ converges weakly to a zero mean multivariate distribution with variance-covariance matrix $\Sigma(\beta_0) = \{\partial^2\gamma(\beta_0)/\partial\beta'\partial\beta\}^{-1}\Gamma(\beta_0)\{\partial^2\gamma(\beta_0)/\partial\beta'\partial\beta\}^{-1}$.

References

Begun, J. M., Hall, W. J., Huang, W.-M. & Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432–52.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–220.

de Una-Alvarez, J., Otero-Giraldez, M. & Alvarez-Llorente, G. (2003). Estimation under length-bias and right-censoring: an application to unemployment duration analysis for married women. *J. Appl. Statist.* **30**, 283–91.

Ghosh, D. (2008). Proportional hazards regression for cancer studies. *Biometrics* **64**, 141–8.

Gong, G. & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.* **9**, 861–9.

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

Lagakos, S., Barraj, L. & De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75**, 515–23.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* **47**, 939–56.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge, UK: Cambridge University Press.

Murphy, S. A. & van der Vaart, A. W. (2000). On profile likelihood. *J. Am. Statist. Assoc.* **95**, 449–65.

Pepe, M. S. & Fleming, T. R. (1991). A nonparametric method for dealing with mismeasured covariate data. *J. Am. Statist. Assoc.* **413**, 108–13.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Hayward, CA: Institute of Mathematical Statistics.

Qin, J. & Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* **66**, 382–92.

Severini, T. A. & Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **4**, 1768–802.

Tsai, W.-Y. (2009). Pseudo-partial likelihood for proportional hazards models with biased-sampling data. *Biometrika* **96**, 601–15.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

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

Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Am. Statist. Assoc.* **86**, 130–43.

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

Wolfson, C., Wolfson, D. B., Asgharian, M. & M'Lan, C. E. (2001). A reevaluation of the duration of survival after the onset of dementia. *New Engl. J. Med.* **344**, 1111–6.

Zelen, M. (2004). Forward and backward recurrence times and length biased sampling: age specific models. *Lifetime Data Anal.* **10**, 325–34.

Zelen, M. & Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56**, 601–14.

Zucker, D. M. (2005). A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *J. Am. Statist. Assoc.* **100**, 1264–77.