Information and asymptotic efficiency of the case-cohort sampling design in Cox's regression model

By: Haimeng Zhang, Larry Goldstein

Abstract:

Efficiencies of the maximum pseudolikelihood estimator and a number of related estimators for the case-cohort sampling design in the proportional hazards regression model are studied. The asymptotic information and lower bound for estimating the parametric regression parameter are calculated based on the effective score, which is obtained by determining the component of the parametric score orthogonal to the space generated by the infinite-dimensional nuisance parameter. The asymptotic distributions of the maximum pseudolikelihood and related estimators in an i.i.d. setting show that these estimators do not achieve the computed asymptotic lower bound. Simple guidelines are provided to determine in which instances such estimators are close enough to efficient for practical purposes.

Keywords: Efficiency | Information bound | Semiparametric model | Epidemiology | Case-cohort design | Counting process | Cox regression model

Article:

1. Introduction

Cox's proportional hazards regression model [7] is often used to quantify the effects of prognostic factors on survival. One common choice of hazard function used in the Cox model, for an individual having covariate $Z(t)$ at time t, is

equation(1)

$$\lambda(t \mid Z) = \lambda_0(t) \exp\{\theta_0 Z(t)\}, \quad t \geq 0$$

where $\theta_0 \in R$ is an unknown parameter of interest to be estimated, and $\lambda_0(t)$ is a baseline hazard function, common to all subjects in the cohort. More modeling flexibility is obtained by leaving the baseline hazard function $\lambda_0(t)$ unspecified, as we may only know that the hazard function is, say, monotone, but be otherwise unaware of it having any particular functional form. In this form, the Cox model is semi-parametric, as it is determined by the real valued parameter θ_0, and the function λ_0.

We consider a cohort of n individuals, $\mathscr{R} = \{1, 2, \ldots, n\}$, with $Z_i(t)$ denoting the value of the covariate of individual i at time t. Suppose that $t_1 < t_2 < \cdots$ are the ordered failure times, and that ij is the index of the failure at time t_j. Let $\mathscr{R}_j$ be the risk set at time t_j, that is, the set consisting of all individuals who are still at risk at time t_j, and the failed individual ij. Estimates of θ_0 are often based on first choosing a sample $\mathscr{S}_j \subset \mathscr{R}_j$ according to some rule, and then maximizing the function

equation(2)

$$L(\theta) = \prod_{t_j} \left\{ \frac{\exp\{\theta Z_{i_j}(t_j)\}}{\sum_{l \in \mathscr{S}_j} \exp\{\theta Z_l(t_j)\}} \right\}.$$

One advantage of an estimator of this form is that it can be computed without making assumptions on the baseline hazard function λ_0, the censoring mechanism, or the distribution of the covariates.

When information is available on the entire cohort, the choice $\mathscr{S}_j = \mathscr{R}_j$ is possible and yields the maximum partial likelihood estimator, or MPLE, which we denote by $\hat{\theta}$. When the cohort is large or the collection of covariate information is costly or difficult, sampling schemes for which covariate information need to be collected on only a small subset $\mathscr{S}_j$ of $\mathscr{R}_j$ are clearly desirable. In this paper we address the question of the efficiency of certain estimators for the case-cohort sampling design.

In the case-cohort design, following Self and Prentice [17], the sampled risk set $\mathscr{S}_j$ at failure time t_j is chosen to be $\tilde{\mathscr{R}}_j$, consisting of all individuals included in a simple random sample $\tilde{\mathscr{R}}$ at time $t=0$ who are still at risk at time t_j; that is, $\tilde{\mathscr{R}}_j = \tilde{\mathscr{R}} \cap \mathscr{R}_j$. We term the estimator obtained by maximizing (2) with $\mathscr{S}_j = \tilde{\mathscr{R}}_j$ the SP88 estimator, and denote it $\tilde{\theta}$. In this paper, we consider a slight variation on the model in [17] and take $\tilde{\mathscr{R}}$ to be a random sample of $\mathscr{R}$ selected by i.i.d. inclusion indicators. We mention in Section 3, that by the same technique

as that used in [17], under mild moment conditions in our i.i.d. setting, the estimator $\tilde{\theta}$ is asymptotically equivalent to the maximum pseudolikelihood estimator specified in [16], where $\mathscr{S}_j = \mathscr{R}_j \cup \{i_j\}$. Therefore we henceforth consider $\tilde{\theta}$ and the maximum pseudolikelihood estimator of Prentice [16] asymptotically interchangeable.

Chen and Lo [5] proposed to improve the SP88 estimator by incorporating information from all cases rather than only those cases included in the random sample. If covariates are time dependent then use of such an estimator may require additional data collection, but when the covariates are time fixed then the inclusion of case information at all times previous to the failure of the case carries no burden. Chen and Lo [5] showed that these estimators, referred to here as the CL99 estimators, generally perform somewhat better than the SP88 estimator (see also Table 1). In this paper, we consider the CL99 estimators under the independent sampling model described above.

In Theorem 1 of Section 1, we present a formula which shows, under the null $\theta_0=0$, how close asymptotically the SP88 estimator and the CL99 estimators are to efficiency, as compared to an information bound over a set of 'reasonable' estimators based on the same data. In particular, in Section 5we show that for a simple model with exponential failure time and uniform censoring over the time interval [0,1], the efficiencies of the SP88 and CL99 estimators are

$$e^{SP} = \left[1 + \frac{2(1-p)}{p} J_1(d)\right][1 + (1-p)J_2(d)], e^{CL} = \left[1 - \frac{(1-p)}{p} J_2(d)\right][1 + (1-p)J_2(d)],$$

where $J_1(d)$ and $J_2(d)$ are given by

equation(3)

$$J_1(d) = 2 - d\log(1-d)/(d + \log(1-d)),$$

equation(4)

$$J_2(d) = 2 - (2d\log(1-d) + (1-d)\log^2(1-d))/(d + \log(1-d));$$

here p is the sampling fraction, and d is the probability of disease before time 1. In particular, in the case of small disease probability d, the formulas show that the SP88 and the CL99 estimators are close to efficient when the sampling fraction is at least 10 or 30 percent. In these cases, even an 'optimal' estimator could not improve these estimators significantly. In other cases, for example, when p is very small, the formulas show that both SP88 estimator and CL99 estimators are not efficient, and hence there may exist other, perhaps more complicated estimators, that may perform better given the same data.

The question of the efficiency of estimators that use sampled data is real one, as the need for sampling arises often in practice. For example, in a study to explore the relationship between

particulate exposures and esophagus cancer in a certain aircraft manufacturing firm with 14,067 employees (see [9]), computing the MPLE using the full cohort where $\mathscr{S}_j = \mathscr{R}_j$ would require the collection of a great deal of information. For each individual, such information could include the date and age at entry into and exit from the cohort, mortality status, the date and cause of death if dead, and exposure status. Such detailed exposure and job histories would be expensive or impossible to collect for the entire cohort. Furthermore, if the disease is rare, there is only a diminishing amount of information obtained by adding more controls to the risk set.

There are a number of sampling designs employed in epidemiological cohort studies, including case-cohort [16], nested case-control [3], [4], [14], [18] and [19], counter-matching [12] and its derivatives such as counter-matching with additional randomly sampled controls. All these schemes involves selecting, according to some rule, a sampled risk set $\mathscr{S}_j$. These schemes offer a substantial reduction of the work and expense of data collection as compared to what is required when working under the full cohort model.

Naturally, there is some information loss inherent in any sampling scheme, the extent of which can be determined by computing the asymptotic relative efficiency of the estimator under sampling to that under full cohort information. But additionally, under any sampling scheme, the question arises as to whether estimators obtained by maximizing (2) are using the available data in the most efficient manner. In a regular parametric model, the Cramer‑Rao lower bound provides the answer to such questions in terms of a variance lower bound for estimators of the unknown parameter. Under regularity it is well known that the maximum likelihood estimator achieves this lower bound and so is asymptotically efficient. But the partial likelihood $L(\theta)$ is not a likelihood in the usual sense since, for instance, the terms in the product (2) are not independent, and any information over intervals between failures is neglected. However, tools for calculating theoretical lower bounds for semi-parametric models, as developed by LeCam [13], and Hájek[11], may be applied. In the case of full cohort information, it was shown that the maximum partial likelihood estimator achieves a theoretical asymptotic variance lower bound (cf. [1] and [10]).

In this paper, we provide an analysis to determine the efficiencies, defined in reference to a theoretical lower bound, of the SP88 estimator and the CL99 estimators for the i.i.d. case-cohort design. Although we do not know if there exists any estimators which achieve this lower bound, we believe our analysis will aid in the development of better, or perhaps even efficient, estimators in the future.

Our paper is organized as follows. In Section 2, after presenting the i.i.d. case-cohort design model formally, we derive the information and variance lower bounds in the null case, $\theta_0=0$, when there is no relation between exposure and disease. These results give a bound on the performance of any reasonable estimator based on the same data as that available to these estimators.

In 3 and 4, we derive the asymptotic distributions of the SP88 estimator and the CL99 estimators using a counting process and martingale theory approach. All these analysis are based on the techniques in [17].

For the purposes of comparing the computed lower bounds to the actual variance achieved by the SP88 and the CL99 estimators, models satisfying the conditions in 2, 3 and 4 are considered, and such a comparison is carried out in Section 5 explicitly under the null $\theta_0=0$, assuming exponential failure times and uniform censoring on [0,1], independent of covariates. We show that the SP88 estimator $\tilde{\theta}$ (or equivalently, the maximum pseudolikelihood estimator of Prentice [16]) and the CL99 estimators do not achieve the asymptotic lower bound. The asymptotic variances of these estimators along with the asymptotic lower bounds are tabulated for certain subcohort sampling fractions and disease probabilities in Table 1. It turns out that for small disease probabilities, the maximum pseudolikelihood estimator and the CL99 estimators generally perform well if the sampled risk set is of an appropriate size. Some concluding remarks are given in Section 6.

2. Information and asymptotic variance lower bounds

The case-cohort sampling design as originally proposed [16] requires the collection of covariate histories on the subjects who develop the disease of interest, and on a control set selected by a simple random sample of the entire cohort at the start of the study. We will consider the related model where the control set is selected using independent Bernoulli random variables. We obtain the lower bound on the information for this setup by closely following the treatment of Begun et al. [1], referred to as BHHW in what follows.

We now specify the model of this section more formally. The variable Z denotes covariate value, Y censoring time, X^0 the failure time and B the indicator of inclusion into the sampled risk set.

Condition 2.1.

The covariate Z is time independent and has density $h(z)$ with respect to Lebesgue measure ν in R.

Each individual is observed up to the time when either the individual fails, or is censored; the distribution of the failure time may depend on the covariate Z.

Condition 2.2 Independent censoring.

Given $Z=z$, the failure time X^0 has density function $g(t|z)$ with respect to Lebesgue measure ν. Moreover, the censoring time Y has density function $c(t)$ with respect to Lebesgue measure ν, independent of both the covariate Z and the failure time X^0. Neither $h(z)$ nor $c(t)$ involves θ or λ.

Some independence between the failure and censoring times is necessary since it would clearly be impossible to obtain meaningful survival data if, for example, individuals were withdrawn from the study when they appeared to be at high risk for failure.

With failure occurring according to intensity (1), we let $G(t|z)$ denote the cumulative distribution function of X^0 when $Z=z$, and let $\overline{G}(t|z) = 1 - G(t|z)$; note that the dependence of these quantities on θ_0 and λ_0 is suppressed. The distribution function $G(t|z)$ and density function $g(t|z)$ are connected to the hazard function $\lambda(t|z)$ given in (1) through the relation $\lambda(t|z) = g(t|z) / \overline{G}(t|z)$; therefore, $\overline{G}(t|z) = (\overline{G}(t))^{\exp(\theta_0 z)}$. Note that for $\theta_0 = 0$ we have $\lambda(t|z) = \lambda_0(t) = g(t) / \overline{G}(t)$. Further, let $C(t)$ denote the cumulative distribution function of the censoring time Y and let $\overline{C}(t) = 1 - C(t)$.

In what follows, $\theta \in R^1$ is a real-valued parameter and g is an element of $\mathscr{G}$, a fixed subset of the set of all densities with respect to Lebesgue measure ν on $R^+ = [0, \infty)$.

We also assume

Condition 2.3.

$E\{Z^2 e^{\theta Z}\}$ is bounded uniformly in a neighborhood of the true value θ_0.

For each individual i there is an associated time $T_i = \min(X_i^0, Y_i)$ of withdrawal from the study, and the indicator $\Delta_i = 1(T_i = X_i^0)$ that the withdrawal use due to failure. To build the case cohort sampling mechanism into our model, we introduce the sampling indicator B that specifies whether an individual is included in the sample taken at time 0, where

Condition 2.4.

The indicator B is a Bernoulli random variable with success probability p.

Finally, we operate under an i.i.d. cohort model:

Condition 2.5.

The variables Z_i, Y_i, X_i^0, B_i over individuals in the cohort are i.i.d. copies of Z, Y, X^0 and B.

In the case cohort framework, information is only available on failed individuals, i.e., those with $\Delta_i = 1$, and those selected to be in the sampled risk set, i.e., those with $B_i = 1$. We summarize the data for each member of the cohort by the i.i.d. vectors $X_i = (T_i, \Delta_i, B_i, Z_i)$.

With ν Lebesgue measure on R, and τ counting measure on $\{0,1\}$, the vectors $X_i = (T_i, \Delta_i, B_i, Z_i)$, which take values in the space

equation(5)

$$\mathcal{X} = \mathbf{R}^+ \times \{0,1\} \times \{0,1\} \times \mathbf{R}$$

have density $f(x) = f(x;\theta, g)$, with respect to the product measure $\mu = \nu \times \tau \times \tau \times \nu$, given by

equation(6)

$$f(x) = \begin{cases} g(t \mid z)\,\overline{C}(t)\,h(z), & \Delta = 1, \\ pc(t)\,\overline{G}(t \mid z)\,h(z), & \Delta = 0, B = 1, \\ (1-p)\left(\int_{-\infty}^{\infty} \int_0^{\infty} c(t)\overline{G}(t \mid z)h(z)\,d\nu(t)\,d\nu(z)\right)\varphi(x), & \Delta = 0, B = 0, \end{cases}$$

where $x = (t, \Delta, B, z)$, and $\varphi(x)$ is an arbitrary density function. The density $\varphi(x)$ does not depend on either θ or λ since the subjects who do not fail and are not included in the sampled cohort do not provide any information for θ or λ. The density g is the 'nuisance parameter' which prevents the parametric estimation of θ.

To interpret the density function $f(x)$ above, consider for example $f(t,0,1,z)$, the 'probability' that there is an event at time t for an individual with $\Delta=0$, sampling indicator $B=1$ and covariate $Z=z$. Since $\Delta=0$ the individual is censored. The covariate value z occurs with density $h(z)$, and given this covariate value, censoring occurs at time t with density $c(t)$. In addition, being censored at time t means that the failure time is greater than t, an event of probability $\overline{G}(t \mid z)$. Lastly, such an individual is included in the sample with probability p. Multiplying these factors gives the density for such individuals. The other factors can be understood similarly.

We closely follow BHHW to develop our asymptotic lower bound.
Let $L^2(\mu) = L^2(\mathcal{X}, \mu)$ and $L^2(\nu)=L^2(R^+, \nu)$ denote the usual L^2-spaces of square integrable functions and let $\langle \cdot, \cdot \rangle_\mu$ ($\|\cdot\|_\mu$) and $\langle \cdot, \cdot \rangle_\nu$ ($\|\cdot\|_\nu$) denote the usual inner products (and norms) in $L^2(\mu)$ and $L^2(\nu)$, respectively. To compute the effective information for θ in the presence of the unknown function g, we need to parametrize $\mathcal{G}$ locally by a subspace $\mathcal{B}$ of $L^2(\nu)$, where each $\beta \in \mathcal{B}$ is a possible "direction" in which to approach g. Explicitly, for $g \in \mathcal{G}$ and $\beta \in L^2(\nu)$, let $\mathscr{C}(g,\beta)$ denote the collection of all sequences of densities $\{g_n\} \subset \mathcal{G}$ such that

equation(7)

$$\|n^{\frac{1}{2}}(g_n^{\frac{1}{2}} - g^{\frac{1}{2}}) - \beta\|_\nu \to 0$$

as $n \to \infty$. Note that (7) implies that β is orthogonal to $g^{1/2}$ since $\|g_n^{1/2}\|_\nu = \|g^{1/2}\|_\nu = 1$ for all $n \geqslant 1$.

As mentioned in BHHW, for the stability of the model we need to restrict attention to those sequences $\{g_n\} \in \mathscr{C}(g,\beta)$ in which each *gn* is absolutely continuous with respect to *g*. Doing so implies that the support of the associated β is contained in that of *g*. Therefore, for every $g \in \mathscr{G}$, let

equation(8)

$$\mathscr{B} \equiv \left\{ \beta \in L^2(\nu) : \|n^{\frac{1}{2}}(g_n^{\frac{1}{2}} - g^{\frac{1}{2}}) - \beta\|_\nu \to 0 \text{ as } n \to \infty \text{ for some } \{g_n\} \subset \mathscr{G} \text{ and } g_n \text{ absolutely continuous with respect to } g \right\}$$

and

$$\mathscr{C}_0(g,\beta) \equiv \{ g_n \in \mathscr{G}; \text{ (7) holds with } g_n \text{ absolutely continuous with respect to } g\},$$

i.e. $\mathscr{C}_0(g,\beta)$ is $\mathscr{C}(g,\beta)$ if the support of β is contained in the support of *g*. Furthermore, we let $\mathscr{C}_0(g)$ be the union of all $\mathscr{C}_0(g,\beta)$ over $\beta \in \mathscr{B}$.

Similarly, let $\Theta(\theta,h)$ denote all sequences $\{\theta_n\}_{n \geqslant 1}$ such that

equation(9)

$$\left|n^{\frac{1}{2}}(\theta_n - \theta) - h\right| \to 0, \quad \text{as } n \to \infty,$$

and $\Theta(\theta) = \bigcup_{h \in R^1} \Theta(\theta,h)$. Given $(\theta_n, g_n)_{n \geqslant 1} \in \Theta(\theta) \times \mathscr{C}_0(g)$ let $f_n \equiv f(\cdot\ ; \theta_n, g_n)$ denote the corresponding sequence of densities.

In order to apply the results of BHHW, we require the following result.

Proposition 1.

The set $\mathscr{B}$ is a subspace of $L^2(\nu)$.

Proof.

Definition (8) implies that

equation(10)

$$\mathscr{B} = \{\beta \in L^2; \beta \perp g^{1/2}, \text{ support}(\beta) \subset \text{support}(g)\},$$

and the proposition follows. $\square$

Before we introduce our main result, we need the following definition.

Definition 2.1.

We say that an estimator $\widehat{\theta}_n$ of θ is *regular* at (θ,g) if for every sequence $\{f_n\}_{n \geqslant 1}$, $f_n \equiv f(\cdot ; \theta_n, g_n)$ with $(\theta_n, g_n)_{n \geqslant 1} \in \Theta(\theta) \times \mathscr{C}_0(g)$ the distribution of $n^{1/2}(\widehat{\theta}_n - \theta_n)$ (under *fn*) converges weakly to a law which depends on f (and hence θ and g) but not on the particular sequence ($\theta n, gn$).

This is a type of stability property on an estimator and it is implied by uniform weak convergence of $n^{1/2}(\widehat{\theta}_n - \theta_n)$ (under *fn*) to a law which might depend on f in neighborhoods of g and θ; for more details see BHHW.

Now we present the main results for this section, the proofs of the theorems are deferred to the end of this section.

Theorem 1.

Consider a cohort $\mathscr{R}$ withnindividuals and assume Conditions 2.1-2.5 are satisfied. Suppose that $\widehat{\theta}_n$ is any regular estimator of θ based on the case-cohort design with i.i.d. sampling such that, under $\theta_0=0$, its limit law is $\mathscr{L} = \mathscr{L}(f)$. Then $\mathscr{L}$ may be represented as the convolution of a $N(0,1/I_*)$ distribution with $\mathscr{L}_1 = \mathscr{L}_1(f)$, a distribution depending only on $f=f(\cdot ; \theta_0, g)$, where

equation(11)

$$I_* = \text{var}(Z) \int_0^{\infty} (1 + 2(1-p)\log \overline{G}(t) + (1-p)\log^2 \overline{G}(t))\overline{C}(t)\,dG(t)$$

for $\overline{C}(t) = 1 - C(t)$ *and* $\overline{G}(t) = 1 - G(t)$.

To present our asymptotic minimax result, we make the following

Definition 2.2.

We say $l: R^1 \to R^+$ is a *loss function* if it is subconvex, that is, if $\{x: l(x) \leqslant y\}$ is closed, convex, and symmetric for every $y \geqslant 0$, and satisfies

equation(12)

$$\int_{-\infty}^{\infty} l(z)\phi(sz)\,dz < \infty$$

for all $s>0$, where ϕ denotes the standard normal density function.

Theorem 2.

Consider a cohort $\mathscr{R}$ with n individuals and assume Conditions 2.1-2.5 are satisfied. Let $l(x)$ be a loss function and $\mathscr{F}(f)$ be given by (16) below. Then under $\theta_0=0$ and with

$$B_n(c) \equiv \{ f_n \in \mathscr{F}(f) : n^{1/2} \| f_n^{1/2} - f^{1/2} \|_\mu \leqslant c \},$$

$$\lim_{c \to \infty} \lim_{n \to \infty} \inf_{\hat{\theta}_n} \sup_{f_n \in B_n(c)} E_{f_n} l(n^{1/2}(\hat{\theta}_n - \theta_n)) \geqslant E l(Z_*),$$

where $Z_* \sim N(0, 1/I_*)$ for I_* given by (11) in Theorem 1. If $l(x)=x^2$, then we say that $1/I_*$ is the asymptotic lower bound for the variance of any regular estimator when $\theta_0=0$.

Here the infimum over estimators $\hat{\theta}_n$ is taken over the class of "generalized procedures," the class of randomized (Markov kernel) procedures, as in BHHW.

The following proposition is required for the computation of the asymptotic information for regular estimators of θ, and hence for the proofs of Theorem 1 and Theorem 2.

Proposition 2.

Suppose $(\theta, g) \in R^1 \times \mathscr{G}$. If $\{(\theta_n, g_n)\}_{n \geqslant 1} \in \Theta(\theta, h) \times \mathscr{C}_0(g, \beta)$ for $h \in R^1$, $\beta \in L^2(\nu)$, and $f_n \equiv f(\cdot \ ; \theta_n, g_n)$ and $f \equiv f(\cdot \ ; \theta, g)$, then under $\theta_0=0$, we have

equation(13)

$$\| n^{\frac{1}{2}}(f_n^{\frac{1}{2}} - f^{\frac{1}{2}}) - \alpha \|_\mu \to 0 \quad \text{as } n \to \infty$$

with $\alpha \in L^2(\mu)$ given by

equation(14)

α=hρ+Aβ,

and $\rho \in L^2(\mu)$ and $A : L^2(\nu) \to L^2(\mu)$ are given by

$$\rho(t,1,B,z) = \tfrac{1}{2} z(1 + \log \overline{G}(t)) f^{\frac{1}{2}}(t,1,B,z), A\beta(t,1,B,z) = \left(R\beta(t) + \frac{\int_t^\infty \beta g^{\frac{1}{2}} \, d\nu}{\overline{G}(t)} \right) f^{\frac{1}{2}}(t,1,B,z), \rho(t,0,1,z)$$

$$= \tfrac{1}{2} z \log \overline{G}(t) f^{\frac{1}{2}}(t,0,1,z), A\beta(t,0,1,z) = \left(\frac{\int_t^\infty \beta g^{\frac{1}{2}} \, d\nu}{\overline{G}(t)} \right) f^{\frac{1}{2}}(t,0,1,z),$$

for $B=0$ or 1, and

equation(15)

$$R_\beta(t) = \beta(t)g^{-\frac{1}{2}}(t) - \frac{\int_t^\infty \beta g^{\frac{1}{2}}\, d\nu}{\overline{G}(t)}.$$

Proof.

For $\theta_0=0$, the verification of (13) and the determination of α,ρ and *A* parallel computations in Section 6 of BHHW for the full cohort case, and Lemma 1 of Begun and Wellner [2] for the two-sample case without censoring. □

From now on we will focus on the case $\theta_0=0$.

Let $H \equiv \{\alpha \in L^2(\mu): \alpha = h\rho + A\beta \text{ for some } h \in R^1, \beta \in \mathscr{B}\}$. Note that by Proposition 1, *H* is a subspace of $L^2(\mu)$ since it is the image of a subspace (of $R^1 \times L^2(\nu)$) under a bounded linear transformation. For $\alpha \in H$, we let $\mathscr{F}(f,\alpha)$ denote the collection of all sequences {*fn*} such that (13) holds for the given α and let

equation(16)

$$\mathscr{F}(f) \equiv \bigcup_{\alpha \in H} \mathscr{F}(f,\alpha).$$

To obtain the effective information for θ in the presence of the unknown function *g*, we orthogonally project ρ onto the nuisance space $\{A\beta : \beta \in \mathscr{B}\}$ to yield the "effective score" for θ,ρ−Aβ*, whereAβ*, the orthogonal projection, is such that β* satisfies the "normal equation"

equation(17)

A*Aβ*=A*ρ,

where A* is the adjoint operator of *A*. The effective asymptotic information then equals

equation(18)

$$I_*(\theta) = 4\|\rho - A\beta^*\|_\mu^2.$$

We are now ready for the proofs of the theorems.

Proof of Theorem 1 **and** Theorem 2**.**

Our proofs parallel those of Theorems 3.1 and 3.2 of BHHW. We have verified the subspace condition of BHHW in Proposition 1, and the conclusion of Proposition 2.1 of BHHW in Proposition 2. Therefore, it remains only to compute I_*.

First, following the notations in Proposition 2, we compute $\beta^*(t)$, the solution of the "normal equation"(17), and so the orthogonal projection $A\beta^*(x)$. This is a technically challenging part of the computation. Note that with classical functional analysis theory (cf. [15]), we have

equation(19)

$$A^* A\beta(t) = p\left[R\beta(t)\frac{M_0(t)}{\overline{G}(t)} - \int_0^t R\beta(s)\frac{M_0(s)}{\overline{G}(s)}\frac{dG}{\overline{G}}\right]g^{1/2}(t) + \left[(1-p)\beta(t)g^{-1/2}(t)\frac{M_0(t)}{\overline{G}(t)}\right]$$

equation(20)

$$A^*\rho(t) = \frac{1}{2}g^{\frac{1}{2}}(t)\left[\frac{M_1(t)}{\overline{G}(t)} + (1-p)\frac{\log \overline{G}(t)}{\overline{G}(t)}M_1(t) - p\int_0^t \frac{M_1(s)}{\overline{G}(s)}\frac{dG}{\overline{G}}\right],$$

where

equation(21)

$$M_i(t) = E\{Z^i 1(T > t)\} = \int_{-\infty}^{\infty} z^i \overline{C}(t)\overline{G}(t|z)h(z)\,d\nu(z).$$

Now notice that, with $\theta_0 = 0$ and the independence of Y and Z, we have

equation(22)

$$M_1(t) = M_0(t)E(Z), \quad M_0(t) = \overline{G}(t)\overline{C}(t).$$

Therefore, the normal equation (17) is simplified as

equation(23)

$$p\left[R\beta(t)\overline{C}(t) - \int_0^t R\beta(s)\overline{C}(s)\frac{dG}{\overline{G}}\right] + (1-p)\beta(t)g^{-1/2}(t)\overline{C}(t)$$

$$= \frac{E(Z)}{2}\left[p\left(\overline{C}(t) - \int_0^t \overline{C}(s)\frac{dG}{\overline{G}}\right) + (1-p)(1+\log \overline{G}(t))\overline{C}(t)\right].$$

Let $\beta^*(t) = \frac{E(Z)}{2}(1+\log \overline{G}(t))g^{\frac{1}{2}}(t)$. It is not too hard to see that $(\rho - A\beta^*)\perp A\beta$ for any $\beta \in \mathcal{B}$ after simple and straightforward calculations. Therefore, the orthogonal projection $A\beta^*(x)$ of $\rho(x)$ onto the closed space $\{A\beta : \beta \in \mathcal{B}\}$ of $L^2(\nu)$ with $\theta_0 = 0$ is given by

equation(24)

$$A\beta^*(t,1,B,z) = \frac{E(Z)}{2}(1 + \log \overline{G}(t))f^{\frac{1}{2}}(t,1,B,z),$$

equation(25)

$$A\beta^*(t,0,1,z) = \frac{E(Z)}{2}\log \overline{G}(t)f^{\frac{1}{2}}(t,0,1,z),$$

for B=0 or 1. Therefore, with , and above, it is easy to get (11) after tedious but straightforward computations. Theorem 2 now follows by a direct application of Theorem 3.2 of BHHW with I_* given by (11). $\square$

3. The SP88 estimator under independent sampling

To evaluate the properties of the SP88 estimator for the case-cohort design, we introduce a counting process and martingale framework. This framework and the subsequent analysis in this section parallel the treatment of Self and Prentice [17]. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\{\mathcal{F}_t\}_{t \in [0,1]}$ a right continuous, nondecreasing family of sub-σ-algebras of $\mathcal{F}$ with $\mathcal{F}_0$ containing all P null subsets of $\mathcal{F}$. We suppose that $\{\mathcal{F}_t\}$ includes failure time and covariate histories up to time t, and censoring histories to t^+ for all subjects in a cohort $\mathcal{R} = \{1,2,\ldots,n\}$. To the ith individual, $i \in \mathcal{R}$, we associate the triple $(Ni(t),Yi(t),Zi(t))$, which are independent replicates of $(N(t),Y(t),Z(t))$, where $N_i(t) = \sum_{j \geq 1} 1(t_j \leq t, i_j = i)$ is the counting process counting the number of times individual i is observed to fail in $(0,t]$, $Yi(t)$ is the censoring process so that $Yi(t)$=1 if the ith subject is observed at time t, and $Yi(t)$=0 otherwise, and $Zi(t)$ is the (possibly) time-dependent covariate process corresponding to the ith subject. We also assume $Ni(1)<\infty$ a.s. for every i. Note that Ni can only jump when $Yi(t)$=1.

Corresponding to each counting process $Ni(t)$, define the intensity process

equation(26)

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\theta_0 Z_i(t))$$

determining the rate at which individual i is observed to fail at time t, given the cohort history $\mathcal{F}_{t-}$ up to just before time t.

Recalling the notation introduced in Section 1, the maximum partial likelihood estimator $\widehat{\theta}$ and the SP88 estimator $\widetilde{\theta}$ are obtained by maximizing L of (2) for $\mathcal{S}_j = \mathcal{R}_j$ and $\widetilde{\mathcal{S}}_j = \widetilde{\mathcal{R}}_j$, respectively. Equivalently, in the counting process setting, $\widehat{\theta}$ and $\widetilde{\theta}$ are the respective solutions of the estimating equations

$$U_l(\theta,1) = 0, \quad l = 0,1,$$

where

equation(27)

$$U_1(\theta, t) \equiv \sum_{i=1}^{n} \int_0^t [Z_i(w) - E_1(\theta, w)] \, dN_i(w), \quad l = 0,1, \quad E_0(\theta, w) \equiv \frac{S^{(1)}(\theta, w)}{S^{(0)}(\theta, w)}, \quad E_1(\theta, w) \equiv \frac{\widetilde{S}^{(1)}(\theta, w)}{\widetilde{S}^{(0)}(\theta, w)},$$

and for $j=0,1$,

equation(28)

$$S^{(j)}(\theta, w) \equiv (1/n) \sum_{l \in \mathscr{R}} Y_1(w) Z_1^j(w) e^{\theta Z_1(w)}, \quad \widetilde{S}^{(j)}(\theta, w) \equiv (1/\tilde{n}) \sum_{l \in \tilde{\mathscr{R}}} Y_1(w) Z_1^j(w) e^{\theta Z_1(w)}.$$

To study the asymptotic properties of these estimators we assume the following conditions.

Condition 3.1 Finite interval condition.

$\int_0^1 \lambda_0(t) \, dt < \infty$;

Condition 3.2.

There exists a neighborhood $\mathscr{N}_0$ of the true value θ_0 such that

$$E \left\{ \sup_{t \in [0,1], \theta \in \mathscr{N}_0} Y(t) |Z(t)|^2 \exp\{\theta Z(t)\} \right\} < \infty.$$

Condition 3.3.

$P\{ Y(t) = 1, \forall t \in [0,1] \} > 0$.

Condition 3.4.

$\Sigma = \int_0^1 v(\theta_0, t) s^{(0)}(\theta_0, t) \lambda_0(t) \, dt > 0$, where $s^{(0)}, s^{(1)}$ and $s^{(2)}$ are defined
by $s^{(j)}(\theta, t) \equiv E\{ Y(t) Z^j(t) e^{\theta Z(t)} \}$, and $v(\theta, t) \equiv s^{(2)}(\theta, t)/s^{(0)}(\theta, t) - e^2(\theta, t)$, where $e(\theta, t) \equiv s^{(1)}(\theta, t)/s^{(0)}(\theta, t)$.

Condition 3.5 Stability of subcohort averages.

The sequence of distributions of $n^{1/2}\{E_1(\theta_0, t) - E_0(\theta_0, t)\}$ is tight on the space $D=D[0,1]$ of left-continuous functions with right-hand limits equipped with Skorohod topology, where E_0 and E_1 are defined by (27).

Theorem 3.

Under Conditions 3.1-3.5, as $n \to \infty$,

1. (Consistency of $\tilde{\theta}$*)* $\tilde{\theta} \xrightarrow{P} \theta_0$*, the true value of* θ*, and*

2. (Asymptotic normality of $\tilde{\theta}$*)*

equation(29)

$$n^{1/2}(\tilde{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{-1}\left(1 + \frac{1-p}{p}\Sigma^{-1}\Gamma\right)\right),$$

where

equation(30)

$$\Gamma = 2\int_0^1 \int_0^t E\{Y(u)\,Y(w)(Z(u) - e(u))(Z(w) - e(w))e^{\theta(Z(w) + Z(u))}\}\,d\Lambda_0(u)\,d\Lambda_0(w),$$

equation(31)

$$\Sigma = \int_0^1 E\{Y(t)(Z(t) - e(t))^2 e^{\theta\,Z(t)}\}\,d\Lambda_0(t),$$

for $\Lambda_0(t) = \int_0^t \lambda_0(w)\,dw$ *the cumulative hazard function, and* $e(t) = e(\theta_0, t) = s^{(1)}(\theta_0, t)/s^{(0)}(\theta_0, t)$.

The proof of the theorem requires some preliminary lemmas.

Lemma 1.

Let $I_1, I_2, \cdots, I_n$ *be i.i.d. Bernoulli random variables with success probability* $0 < p < 1$ *and* $\tilde{n} = \sum_{i=1}^n I_i$. *From a population of* n *items labeled with deterministic values* $f_1, f_2, \cdots, f_n$*, let* $\bar{Y}$ *denote the sample average, i.e.* $\bar{Y} = \tilde{n}^{-1}\sum_{i=1}^n I_i f_i$*, and let* $\bar{f} = n^{-1}\sum_{i=1}^n f_i$ *be the population average. If*

equation(32)

$$n^{-1}\sum_{i=1}^n (f_i - \bar{f})^2 \to \sigma_f^2 > 0 \quad \text{and} \quad \frac{f_n - \bar{f}}{\sqrt{n}} \to 0$$

as $n \to \infty$*, then*

$$n^{1/2}(\bar{Y} - \bar{f}) \xrightarrow{d} \mathcal{N}(0, \sigma_f^2(1-p)/p).$$

Proof.

Let $Z_i = I_i(f_i - \bar{f})$ and let $S_n = \sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} I_i(f_i - \bar{f})$, then $E Z_i = p(f_i - \bar{f})$, $E S_n = 0$ and $s_n^2 = \mathrm{var}(S_n) = p(1-p) \sum_{i=1}^{n} (f_i - \bar{f})^2$. We first demonstrate that

equation(33)

$$n^{1/2}(\bar{Y} - \bar{f}) = \frac{n^{-1/2} \sum_{i=1}^{n} I_i(f_i - \bar{f})}{n^{-1} \sum_{i=1}^{n} I_i} = \frac{n^{-1/2} S_n}{n^{-1} \sum_{i=1}^{n} I_i}$$

Using the Central Limit Theorem for independent but nonidentically distributed random variables (cf. [8, Chapter 2]), and then Chow and Teicher [6, p. 314] to replace εs_n by εs_i in the Lindeberg condition, it suffices to show

equation(34)

$$\frac{1}{s_n^2} \sum_{i=1}^{n} E\{|Z_i - E(Z_i)|^2 \mathbf{1}(|Z_i - E(Z_i)| > \epsilon s_i)\} \to 0.$$

From (32) we know that there exists $n_0 > 0$ such that $|f_i - \bar{f}| < \epsilon s_i$ for all $i > n_0$. In addition, $|Z_i - E(Z_i)| = |(I_i - p)(f_i - \bar{f})| < |f_i - \bar{f}|$. Therefore, the summation in (34) consists of at most n_0 terms, and division by $sn \to \infty$ yields the desired limit. Hence, (33) holds and therefore $n^{-1/2} S_n \xrightarrow{d} \mathcal{N}(0, p(1-p)\sigma_f^2)$. Lastly, note that

$$n^{1/2}(\bar{Y} - \bar{f}) = \frac{n^{-1/2} \sum_{i=1}^{n} I_i(f_i - \bar{f})}{n^{-1} \sum_{i=1}^{n} I_i} = \frac{n^{-1/2} S_n}{n^{-1} \sum_{i=1}^{n} I_i}$$

and $\tilde{n} n^{-1} = n^{-1} \sum_{i=1}^{n} I_i \xrightarrow{P} p \in (0,1)$. The lemma follows by Slustky's theorem. $\square$

Proposition 3.

Let $X_n = (X_{1n}, X_{2n}, \ldots, X_{nn})$ *and* $I_n = (I_{1n}, I_{2n}, \ldots, I_{nn})$ *be independent random sequences such that*

1. *$I_1 n, I_2 n, \cdots, I_{nn}$ are i.i.d. Bernoulli random variables with success probability $p \in (0,1)$, and $\tilde{n} = I_{1n} + \cdots + I_{nn}$.*

2. *For some scalar functions* $f_{in}(X_n)$ *of* X_n*, and for* $\sigma f > 0$*, with*

$$\bar{f}_n(X_n) = n^{-1} \sum_{i=1}^{n} f_{in}(X_n) \quad \text{and} \quad S_{fn}^2 = n^{-1} \sum_{i=1}^{n} [f_{in}(X_n) - \bar{f}_n(X_n)]^2,$$

we have

equation(35)

$$S_{fn}^2 \xrightarrow{P} \sigma_f^2 > 0 \quad \text{and} \quad \frac{f_{nn}(X_n) - \overline{f}_n(X_n)}{\sqrt{n}} \xrightarrow{P} 0.$$

3. *The scalar functions* $g_n(X_n)$ *of* X_n *in distribution to a Gaussian random variable with mean zero and variance* σg^2.

Then for $h_n(X_n, I_n) = n^{\frac{1}{2}} [\tilde{n}^{-1} \sum_{i=1}^{n} I_{in} f_{in}(X_n) - \overline{f}_n(X_n)]$, *we have that* $(g_n(X_n), h_n(X_n, I_n))$ *converges in distribution to a bivariate normal random variable with mean zero and covariance matrix given by*

$$\begin{pmatrix} \sigma_g^2 & 0 \\ 0 & \frac{1-p}{p} \sigma_f^2 \end{pmatrix}.$$

Note the proposition above is almost the same as Proposition 1 in [17] except the second convergence in (35). However, the independent sampling assumption makes the proof much simpler.

Proof of Proposition 3.

Let

$$a_n(X_n) = |S_{fn}^2 - \sigma_f^2| + \left| \frac{f_{nn}(X_n) - \overline{f}_n(X_n)}{\sqrt{n}} \right|,$$

and (*gn,hn*) denote $(g_n(X_n), h_n(X_n, I_n))$. The sequence *an* converges to zero in probability by hypotheses. The vector (*gn,hn*) converges in distribution to (*g,h*) if and only if every subsequence *nk* has a further subsequence *nkj* such that convergence in distribution to (*g,h*) occurs along the further subsequence. Let *nk* be any subsequence. Clearly *ank* converges to zero in probability. But since every sequence converging in probability has a further subsequence converging almost surely, there exists a further subsequence of *nk*, say *nkj*, such that *an* evaluated along *nkj* converges to zero a.s. Hence, it suffices to show that (*gn,hn*) evaluated along *nkj* has the claimed limiting distribution. For notational simplicity, we relabel (*gn,hn*) evaluated along *nkj* as (*gn,hn*), and hence, under the relabeling, $P(A) = 1$ where $A = \{\lim_{n \to \infty} a_n = 0\}$.

Let

$$\Phi_g(w) = \Phi(w / \sigma_g) \quad \text{and} \quad \Phi_f(v) = \Phi(v / (\sqrt{(1-p)/p}\sigma_f)).$$

Write

$$P\{g_n \leqslant w, h_n \leqslant v\} = E\{1(A)1(g_n \leqslant w, h_n \leqslant v)\} = E\{1(g_n \leqslant w)1(A)P[h_n \leqslant v|X_n]\}.$$

Lemma 1 gives that $1(A)P[h_n \leqslant v|X_n] \to 1(A)\Phi_f(v)$ a.s. Hence

$$|P\{g_n \leqslant w, h_n \leqslant v\} - \Phi_g(w)\Phi_f(v)| \leqslant |E\{1(g_n \leqslant w)(1(A)P[h_n \leqslant v|X_n] - 1(A)\Phi_f(v))\}| + \Phi_f(v)|E\{1(g_n \leqslant w) - \Phi_g(w)\}|$$

$$\leqslant E|1(A)P[h_n \leqslant v|X_n] - 1(A)\Phi_f(v)| + |P\{g_n \leqslant w\} - \Phi_g(w)| \to 0. \qquad \square$$

Proof of Theorem 3.

The consistency of $\tilde{\theta}$ follows easily by consistency arguments as in Lemma 3.1 of [17]. To derive the asymptotic distribution of $\tilde{\theta}$, we reason the same as in the proofs of Theorems 3.1 and 3.2 of Self and Prentice [17] but with the independent sampling scheme. Therefore, it is enough to verify (35) in Proposition 3, where *Xin* represents $\{Yi(u),Ni(u),Zi(u);0 \leqslant u \leqslant 1\}$, and $f_{in}(X_n)$ represents a linear combination of $Yi(t)e^{\theta_0 Zi(t)}$ and $Yi(t)Zi(t)e^{\theta_0 Zi(t)}$. Notice that for the fixed time *t*, $f_{in}(X_n), i = 1,2,\ldots,n$ are actually independent replicates of $f(X)$, a linear combination of the processes $Y(t)e^{\theta_0 Z(t)}$ and $Y(t)Z(t)e^{\theta_0 Z(t)}$, evaluated at the same time point *t*.

Because of the i.i.d. cohort assumption,
$n^{-1/2}\overline{f}_n(X_n) = n^{-1/2}n^{-1}\sum_{i=1}^{n} f_{in}(X_n)$ converges to 0 a.s. by Condition 3.2 and the law of large numbers. In addition, for $\epsilon > 0$,

$$P\left(\left|\frac{f_{nn}(X_n)}{\sqrt{n}}\right| > \epsilon\right) = P(|f(X)| > n^{1/2}\epsilon) \leqslant \frac{E|f(X)|}{n^{1/2}\epsilon}.$$

Therefore, $f_{nn}(X_n)/\sqrt{n}$ converges to 0 in probability, again, by Condition 3.2. Lastly, the first part of (35) is trivial because of the i.i.d. assumption of the full cohort and the independence of the sampling. Hence our result follows with the easy simplification from Theorem 3.2 in [17] due to the independent sampling and exponential risk function. $\square$

Finally, we mention the asymptotic equivalence of the maximum pseudolikelihood estimator in [16] and the SP88 estimator $\tilde{\theta}$ under our i.i.d. sampling model. Using the same proof as in [17, Section 4], under the condition

Condition 3.6.

$$E\{\sup_{t \in [0,1]} Y(t)|Z(t)|^2 \exp\{2\theta_0 Z(t)\}\} < \infty,$$

we have

Proposition 4.

Under Conditions 3.1‑3.5 and 3.6 above, $\tilde{\theta}$ and the maximum pseudolikelihood estimator under the i.i.d. sampling are asymptotically equivalent, i.e. both have the same asymptotic distribution.

4. The CL99 estimators under independent sampling

In this section, we apply the same techniques as those used in Section 3 to derive the asymptotic properties of the CL99 estimators under independent sampling. Following the notation from previous sections and Chen and Lo [5], we define

equation(36)

$$W = \int_0^1 (Z(t) - e(t))e^{\theta Z(t)} 1(T \geqslant t)\, d\Lambda_0(t), \quad \alpha = P(\Delta = 1), V_1 = \mathrm{var}(W|\Delta = 1), \quad V_0 = \mathrm{var}(W|\Delta = 0), \quad K_0 = E(W|\Delta = 0).$$

It is seen that $\Gamma = \mathrm{var}(W) = \alpha V_1 + (1 - \alpha) V_0 + (1 - \alpha)/\alpha\, K_0^2$, where Γ is given by (30) of Theorem 3 in Section 3. Recall that $\mathscr{R}$ and $\widetilde{\mathscr{R}}$ denote the set of all individuals in the cohort with size of n and subcohort with size of $\tilde{n}$, respectively. We let $n_1(\tilde{n}_1)$ and $n_0(\tilde{n}_0)$ be the respective numbers of the cases and controls in the cohort (subcohort). We further let $\mathscr{R}^1$, $\widetilde{\mathscr{R}}^1$ and $\mathscr{R}^0$, $\widetilde{\mathscr{R}}^0$ to denote, respectively, the index sets of all cases and all controls in $\mathscr{R}$, $\widetilde{\mathscr{R}}$.

From Chen and Lo [5], the CL99 estimators, $\widetilde{\theta}_1$, $l=2,3$, respectively, are the solutions of the estimating equation

equation(37)

$$U_1(\theta,1) = 0, \quad l = 2,3,$$

where

$$U_1(\theta, t) \equiv \sum_{i=1}^{n} \int_0^t [Z_i(w) - E_1(\theta, w)]\, dN_i(w), \quad l = 2,3,$$

$$E_2(\theta, w) \equiv \frac{\{\tilde{n}_1/(\tilde{n}\, n_1)\} \sum_{j \in \mathscr{R}^1} Y_j(w) Z_j(w)e^{\theta Z_j(w)} + (1/\tilde{n}) \sum_{j \in \widetilde{\mathscr{R}}^0} Y_j(w) Z_j(w)e^{\theta Z_j(w)}}{\{\tilde{n}_1/(\tilde{n}\, n_1)\} \sum_{j \in \mathscr{R}^1} Y_j(w)e^{\theta Z_j(w)} + (1/\tilde{n}) \sum_{j \in \widetilde{\mathscr{R}}^0} Y_j(w)e^{\theta Z_j(w)}}$$

and

$$E_3(\theta, w) \equiv \frac{(1/n) \sum_{j \in \mathscr{R}^1} Y_j(w) Z_j(w)e^{\theta Z_j(w)} + \{n_0/(n\, \tilde{n}_0)\} \sum_{j \in \widetilde{\mathscr{R}}^0} Y_j(w) Z_j(w)e^{\theta Z_j(w)}}{(1/n) \sum_{j \in \mathscr{R}^1} Y_j(w)e^{\theta Z_j(w)} + \{n_0/(n\, \tilde{n}_0)\} \sum_{j \in \widetilde{\mathscr{R}}^0} Y_j(w)e^{\theta Z_j(w)}}.$$

We consider $\widetilde{\theta}_2$ and $\widetilde{\theta}_3$, the two most practically useful estimators discussed in [5]. Since it is claimed in that paper that $\widetilde{\theta}_3$ is better than $\widetilde{\theta}_2$ we will focus on the asymptotic properties of $\widetilde{\theta}_3$ and only briefly comment on $\widetilde{\theta}_2$.

First, similar to Self and Prentice [17], the score process $n^{-1/2}U_3(\theta_0,t)$ can be simplified as

$$n^{-1/2} U_3(\theta_0, t) = \sum_{i=1}^{n} \int_0^t n^{-1/2} [Z_i(w) - E_0(\theta_0, w)] \, dM_i(w) - \int_0^t n^{-1/2} \{ E_3(\theta_0, w) - E_0(\theta_0, w) \} \, d\overline{\Lambda}(w),$$

where $\overline{\Lambda} = \Lambda_1 + \Lambda_2 + \cdots + \Lambda_n$ and E_0 is given by (27) in Section 3. Note that under the independence and Conditions 3.1‑3.4 in Section 3, the first term above, i.e., the score process of the full-cohort case, converges to a continuous Gaussian process with limiting variance function $\int_0^t E\{ Y(u)(Z(u) - e(u))^2 e^{\theta_0 Z(u)} \} \, d\Lambda_0(u)$, which equals Σ at $t=1$.

Regarding the second term above, we first assume the following tightness condition similar to Condition 3.5 in Section 3.

Condition 4.6.

The sequence of distributions of $n^{1/2}\{E_3(\theta_0, t) - E_0(\theta_0, t)\}$ is tight on the space $D=D[0,1]$ of left-continuous functions with right-hand limits equipped with the Skorohod topology.

Under the independent sampling and Condition 4.6 above, one can show that the second term converges to a Gaussian process independent of the first term, using techniques similar to [17] and in Section 3 of this paper. Hence, it will be sufficient to derive the limiting covariance process of $\int_0^t n^{-1/2}\{ E_3(\theta_0, t) - E_0(\theta_0, t) \} \, d\overline{\Lambda}(w)$.

The following theorem provides the asymptotic properties of $\tilde{\theta}_3$.

Theorem 4.

Assume that Conditions 3.1‑3.4 and 4.6 hold. With $\tilde{\theta}_3$ the solution of the estimating equation (37) for l=3, under independent Bernoulli sampling, $\tilde{\theta}_3$ is consistent and asymptotically normal with asymptotic variance

equation(38)

$$\sigma^2_{3,\text{Ber}} = \Sigma^{-1} + \frac{1-p}{p} \Sigma^{-1}(1-\alpha) V_0 \Sigma^{-1},$$

where Σ and Γ are given by (31) and (30) in Section 3, respectively, and V_0 is defined in (36). Here the subscript Ber *stands for the independent Bernoulli sampling.*

Proof.

The proof of the consistency of $\tilde{\theta}_3$ is similar to that for SP88 estimator in Section 3 and so is omitted. To prove the normality, we apply the usual Taylor series expansion of the score function $U_3(\theta,1)$ about θ_0 evaluated at $\tilde{\theta}_3$, similar to the expansion in Theorem 3.2 of Self and

Prentice [17]. In addition, $n^{-1} \partial U_3(\theta,1)/\partial \theta$ converges in probability to Σ of (31) in a small neighborhood of the true value θ_0. Therefore, it is sufficient to derive the limiting distribution of the score statistic $n^{-1/2} U_3(\theta_0,1)$. Hence, as we mentioned earlier, we only need to focus on the covariance function of the limiting Gaussian process

of $\int_0^t n^{-1/2}\{E_3(\theta_0,t) - E_0(\theta_0,t)\}\, d\bar{A}(w)$ evaluated at time $t=1$.

We introduce a convenient representation of $E_3(\theta,t)$. For each $j=0,1$, we define

equation(39)

$$\check{S}^{(j)}(\theta,t) \equiv (n_1/n)\check{S}_1^{(j)}(\theta,t) + (n_0/n)\check{S}_0^{(j)}(\theta,t)$$

with

equation(40)

Then

$$E_3(\theta,t) = \check{S}^{(1)}(\theta,t) / \check{S}^{(0)}(\theta,t).$$

In addition, we write $S^{(j)}(\theta,t), j=0,1$ of (28) in Section 3

equation(41)

$$S^{(j)}(\theta,t) = (n_1/n)S_1^{(j)}(\theta,t) + (n_0/n)S_0^{(j)}(\theta,t),$$

where

equation(42)

By the law of large numbers and Condition 3.2, one can notice that $E_0(\theta_0,t)$ and $\check{S}^{(0)}(\theta_0,t)$ converge to $e(t)=e(\theta_0,t)$ and $s^{(0)}(\theta_0,t)$ in probability. Therefore, from and , and applying the same calculation as in [17], we have

$n^{1/2}(E_3(\theta_0,t) - E_0(\theta_0,t)) = n^{1/2}(\check{S}^{(1)}(\theta_0,t) / \check{S}^{(0)}(\theta_0,t) - S^{(1)}(\theta_0,t) / S^{(0)}(\theta_0,t)) =$

$-n^{1/2}\{(\check{S}^{(0)}(\theta_0,t) - S^{(0)}(\theta_0,t))e(t) - (\check{S}^{(1)}(\theta_0,t) - S^{(1)}(\theta_0,t))\}s^{(0)}(\theta_0,t)^{-1} + o_P(1), =$

$-n^{1/2}(n_0/n)\{(\check{S}_0^{(0)}(\theta_0,t) - S_0^{(0)}(\theta_0,t))e(t) - (\check{S}_0^{(1)}(\theta_0,t) - S_0^{(1)}(\theta_0,t))\}s^{(0)}(\theta_0,t)^{-1} + o_P(1) =$

$-(n_0^{1/2}/n^{1/2})n_0^{1/2}\{(\check{S}_0^{(0)}(\theta_0,t) - S_0^{(0)}(\theta_0,t))e(t) - (\check{S}_0^{(1)}(\theta_0,t) - S_0^{(1)}(\theta_0,t))\}s^{(0)}(\theta_0,t)^{-1} + o_P(1) \equiv \eta^* + o_P(1),$

say. It is easily seen from the i.i.d. and independent sampling assumptions that $\tilde{n}/n \to P$ and $n_1/n \to \alpha$, in probability as $n \to \infty$. Furthermore, from Lemma

1 and Proposition 3 in Section 3 and the idea similar to that in [17], one can show that, after a straightforward but tedious computation, the covariance function of η^* is given by

$$G(\theta_0, x, w) = \frac{1-p}{p}(1-\alpha) \times s^{(0)}(\theta_0, x)^{-1}s^{(0)}(\theta_0, w)^{-1}$$

$$\times [h^{(1)}(\theta_0, x, w) + h^{(0)}(\theta_0, x, w)\, e(x)\, e(w) - h^{(2)}(\theta_0, x, w)\, e(x) - h^{(2)}(\theta_0, w, x)\, e(w)],$$

where

$$h^{(0)}(\theta_0, x, w) = E\{1(T \geqslant x)1(T \geqslant w)e^{\theta_0 Z(x)}e^{\theta_0 Z(w)}|\Delta = 0\} - E\{1(T \geqslant x)e^{\theta_0 Z(x)}|\Delta = 0\}E\{1(T \geqslant w)e^{\theta_0 Z(w)}|\Delta = 0\},$$

$$h^{(1)}(\theta_0, x, w) = E\{1(T \geqslant x)1(T \geqslant w) Z(x) Z(w)e^{\theta_0 Z(x)}e^{\theta_0 Z(w)}|\Delta = 0\}$$

$$- E\{1(T \geqslant x) Z(x)e^{\theta_0 Z(x)}|\Delta = 0\}E\{1(T \geqslant w) Z(w)e^{\theta_0 Z(w)}|\Delta = 0\},$$

$$h^{(2)}(\theta_0, x, w) = E\{1(T \geqslant x)1(T \geqslant w) Z(w)e^{\theta_0 Z(x)}e^{\theta_0 Z(w)}|\Delta = 0\} - E\{1(T \geqslant x)e^{\theta_0 Z(x)}|\Delta = 0\}E\{1(T \geqslant w) Z(w)e^{\theta_0 Z(w)}|\Delta = 0\}.$$

The weak convergence of the process $n^{1/2}$ $(E_3(\theta_0, t) - E_0(\theta_0, t))$ is implied by the finite-dimensional convergence of the process similar to that in [17] and the tightness Condition 4.6. Therefore, $\int_0^t n^{-1/2}\{E_3(\theta_0, t) - E_0(\theta_0, t)\}\, d\bar{\Lambda}(w)$ converges to a Gaussian process with limiting covariance function at time $t=1$ given by

$$\frac{1-p}{p}(1-\alpha)\int_0^1\int_0^1 G(\theta_0, x, w)s^{(0)}(\theta_0, x)s^{(0)}(\theta_0, w)\lambda_0(x)\lambda_0(w)\, dx\, dw = \frac{1-p}{p}(1-\alpha)V_0. \qquad \square$$

The same idea can be applied to derive the asymptotic normality of $\tilde{\theta}_2$, the solution of estimating equation (37) with $l=2$.

Proposition 5.

Under independent (Bernoulli) sampling, Conditions 3.1-3.4 and 4.6 with the replacement of $E_3(\theta_0, t)$ by $E_2(\theta_0, t)$, $\tilde{\theta}_2$ is consistent and asymptotically normal with asymptotic variance

equation(43)

$$\sigma^2_{2,Ber} = \Sigma^{-1} + \frac{1-p}{p}\Sigma^{-1}(1-\alpha)V_0\Sigma^{-1} + \frac{1}{p}\Sigma^{-1}\left(\frac{1-\alpha}{\alpha}K_0^2\right)\Sigma^{-1}.$$

Remark.

Although the value obtained here for $\sigma_{3,Ber}^2$ under the assumption of independent sampling is the same as that obtained in [5] under simple random sampling (SRS), the variance $\sigma_{2,Ber}^2$ under

independent sampling is slightly larger than the variance $\sigma_{2,\mathrm{SRS}}^2$ given in [5] under SRS, the difference being $\Sigma^{-1}\left(\frac{1-\alpha}{\alpha} K_0^2\right)\Sigma^{-1}$. However, when the true value $\theta_0=0$,

equation(44)

$$\sigma_{2,\mathrm{Ber}}^2 = \sigma_{3,\mathrm{Ber}}^2 = \Sigma^{-1} + \frac{1-\mathrm{p}}{\mathrm{p}}\Sigma^{-1}(1-\alpha)V_0\Sigma^{-1} = \Sigma^{-1} + \frac{1-\mathrm{p}}{\mathrm{p}}\Sigma^{-1}(\Gamma-\alpha V_1)\Sigma^{-1}$$

since $K_0=0$.

5. Estimator efficiency for the case-cohort sampling design

In this section, we consider a model where the asymptotic information, the asymptotic variance of the SP88 estimator $\widetilde{\theta}$, and the asymptotic variances of the CL99 estimators $\widetilde{\theta}_2$ and $\widetilde{\theta}_3$ are given by , , and , respectively. With these in hand, we are able to compare how close the variances of the maximum pseudolikelihood estimator of Prentice [16] and the CL99 estimators come to the theoretical lower bound; we find that the lower bound is not achieved in both cases.

Throughout this section we will assume time-independent covariates, and focus on the case $\theta_0=0$. In particular, we will compute the asymptotic efficiencies of the maximum pseudolikelihood estimator of Prentice [16] and the CL99 estimators under a model, labeled Model A, which satisfies the following assumptions.

1. All subjects are followed from time $t=0$ to either an exponential failure time with parameter λ, or to censoring according to a uniform distribution over (0,1).

2. The failure time and the censoring time are independent of covariates.

5.1. Efficiency of SP88 estimator

Condition 2.3 in Section 2 is clearly stronger than Conditions 3.2 and 3.6. Therefore, the following Corollary follows from Theorem 1 and Theorem 3 and Proposition 4.

Corollary 1.

Under Conditions 2.1‑2.5, 3.1, *and* 3.3‑3.5, *the asymptotic information lower bound when* $\theta_0=0$ *is given by* (11), *and the SP88 estimator* $\widetilde{\theta}$ *is a consistent estimator of* θ_0 *and has asymptotic distribution given by* (29). *Furthermore,* $\widetilde{\theta}$ *and the maximum pseudolikelihood estimator of Prentice* [16] *are asymptotically equivalent.*

We now compare the asymptotic variance lower bound derived in Section 2 to the asymptotic variance of the SP88 estimator $\widetilde{\theta}$ under Model A, where the assumptions of Corollary 1 are assumed to hold. For this special case, we have

Corollary 2.

1. *The asymptotic variance of* $\tilde{\theta}$ *in Model* A *when* $\theta_0=0$ *equals*

equation(45)

$$\tilde{\sigma}_{Ber}^2 = \Sigma^{-1}\left(1 + \frac{2(1-p)}{p} J_1(d)\right),$$

Where $J_1(d)$ *is given by* (3), $d=1-\exp\{-\lambda\}$ *is the probability of failure prior to time* $t=1$, *and* p *is, again, the probability that a risk subject is added to the sampled cohort at time* 0. *Here* $\Sigma = \text{var}(Z)[1 + d/\log(1-d)]$ *is the variance of the maximum partial likelihood estimator* $\hat{\theta}$ *for the full cohort case.*

2. *The asymptotic information* (11) *for* θ *in Model* A *when* $\theta_0=0$ *equals*

equation(46)

$$I_* = \Sigma(1 + (1-p) J_2(d)),$$

here $J_2(d)$ *is given by* (4). *Hence, the asymptotic variance lower bound is*

equation(47)

$$V_B^2 = \frac{1}{I_*} = \Sigma^{-1}\left(\frac{1}{1 + (1-p) J_2(d)}\right).$$

Proof.

From the assumptions, the first claim follows directly from Theorem 3 in Section 3 if we let $d=1 -\exp\{-\lambda\}$. The second part is a special case of Theorem 1 in Section 2 where $\overline{C}(t) = (1-t)$ for $0 \leqslant t < 1$ and equals 0 if $t \geqslant 1$. Therefore,

equation(48)

$$I_* = \text{var}(Z)\int_0^1 (1 + 2(1-p)\log \overline{G}(t) + (1-p)\log^2 \overline{G}(t))(1-t)\,dG(t).$$

Note that $G(t)=1-\exp\{-\lambda t\}$ from assumption 1 for Model A; using this relation (48) can be simplified to yield (46). □

Now define $e^{SP}(d,p)$, *the* (*asymptotic*) *efficiency* of the SP88 estimator, as a function of $d \in (0,1)$ and $p \in (0,1]$, to be the ratio of $\tilde{\sigma}_{Ber}^2$ to V_B^2, i.e.

equation(49)

$$e^{SP}(d, p) = \frac{\tilde{\sigma}_{Ber}^2}{V_B^2} = \left[1 + \frac{2(1-p)}{p} J_1(d)\right][1 + (1-p) J_2(d)],$$

for $J_1(d)$ and $J_2(d)$ given by and , respectively. We say the SP88 estimator $\tilde{\theta}$ in the case-cohort design is *asymptotically efficient* if $e^{SP}(d,p)=1$. Before we investigate the properties of the efficiency $e^{SP}(d,p)$, we present the following lemma for $J_1(d)$ and $J_2(d)$. The proof of the lemma is simple and therefore is omitted.

Lemma 2.

1. $J_1(d) \geqslant 0$ *for* $d \geqslant 0$ *and* $J_1(d)=0$ *if and only if* $d=0$. *Moreover*, $J_1(1)=1$ *and* $J_1(d)$ *is a strictly increasing function of* d *for* $d>0$;

2. $J_2(d)<0$ *for all* $d \in (0,1)$. *In addition*, $J_2(0)=J_2(1)=0$;

3. *for every* $d>0$, $J_2(d)+2J_1(d)>0$ *and* $1+J_2(d)>0$.

For each fixed $d \in (0,1)$, let $ed^{SP}(p)$ denote the efficiency (49) as a function of p only. Then we have the following proposition for $ed^{SP}(p)$.

Proposition 6.

1. $ed^{SP}(0)=\infty$;

2. $ed^{SP}(1)=1$;

3. $ed^{SP}(p)$ *is a strictly decreasing function for* $p \in (0,1)$;

4. $ed^{SP}(p)>1$ *for all* $d \in (0,1)$ *and* $p \in (0,1)$.

Proof.

Properties 1 and 2 are clear, and therefore 3 suffices to show 4. Thus we only need to prove that with each fixed $d \in (0,1)$, $ded^{SP}(p)/dp<0$ for $p \in (0,1)$. Fixing $d \in (0,1)$ and taking the derivative of $ed^{SP}(p)$ on p, we have

equation(50)

$$p^2 \frac{de_d^{SP}(p)}{dp} = 2p^2 J_2(d)\left(J_1(d) - \frac{1}{2}\right) - 2J_1(d)(1 + J_2(d)).$$

Note that $ded^{SP}(p)/dp<0$ if $J_1(d) - \frac{1}{2} \geqslant 0$ for any $p \in (0,1)$ from Lemma 2. Now if $J_1(d) - \frac{1}{2} < 0$,(50) can be written as

$$p^2 \frac{de_d^{SP}(p)}{dp} = -2(1-p^2) J_2(d) \left(J_1(d) - \frac{1}{2} \right) - (J_2(d) + 2 J_1(d)),$$

which is negative, again from Lemma 2. $\square$

Note that Property 2 above, for $p=1$, recovers the result in BHHW, that the MPLE is efficient when data is collected on all cohort individuals.

From Proposition 6 and the asymptotic equivalence of the maximum pseudolikelihood estimator with the SP88 estimator $\tilde{\theta}$ we conclude that the maximum pseudolikelihood estimator of Prentice [16] for the case-cohort design does not achieve the asymptotic lower bound.

5.2. Efficiency of CL99 estimators

We apply the same technique as in the previous subsection to compute the asymptotic efficiency of the CL99 estimator under our simplified model when $\theta_0=0$.

First note that when $\theta_0=0$, both CL estimators have the common asymptotic variance $\sigma_{3,Ber}^2$ given by (44).

Proposition 7.

Under the same assumptions given in Theorem 4, Model A and $\theta_0=0$, the asymptotic variance $\sigma_{3,Ber}^2$ of the CL99 estimators simplifies to

equation(51)

$$\sigma_{3,Ber}^2 = \Sigma^{-1} \left(1 - \frac{1-p}{p} J_2(d) \right),$$

where $J_2(d)$ is given by (4) and $d = P(X^0 \leqslant 1) = 1 - e^{-\lambda}$.

Proof.

Note that under Model A with $\theta_0=0$, W defined in (36) can be simplified as follows:

$$W = \int_0^1 (Z - E(Z)) 1(T \geqslant t) \, dA_0(t) = \lambda T(Z - E(Z)).$$

Therefore, since $K_0=0$ where K_0 is defined by (36),

$(1-\alpha) V_0 = E[W^2 1(\Delta=0)] = \lambda^2 \, \text{var}(Z) E[T^2 1(X^0 \geqslant Y)] = \lambda^2 \, \text{var}(Z) E[E(Y^2 1(X^0 \geqslant Y)|Y)] = \lambda^2 \, \text{var}(Z) E[Y^2 e^{-\lambda Y}]$, since X^0,

Y are independent.

Substituting $\lambda = -\log(1-d)$ a simple calculation shows that $(1-\alpha)V_0/\Sigma = -J_2(d)$. The result follows. $\square$

For each fixed $d \in (0,1)$, let $ed^{CL}(p)$ denote the relative efficiency of the CL99 estimators when compared to the asymptotic lower bound given by Corollary 2, as a function of p only, i.e.,

$$e_d^{CL}(p) = \frac{\sigma_{3,Ber}^2}{V_B^2} = \left[1 - \frac{(1-p)}{p} J_2(d)\right][1 + (1-p)J_2(d)], \quad \text{for fixed } d \in (0,1).$$

We have the following proposition for $ed^{CL}(p)$.

Proposition 8.

1. $ed^{CL}(0) = \infty$ $anded^{CL}(1) = 1$;

2. $ed^{CL}(p)$ is a strictly decreasing function $forp \in (0,1)$;

3. $ed^{CL}(p) > 1$ for all$d \in (0,1)$ and$p \in (0,1)$.

Proof.

It is sufficient to prove 2. Simple calculation gives

$$p^2 \frac{de_d^{CL}(p)}{dp} = (1-p^2)J_2(d)(1 + J_2(d)) < 0$$

since $J_2(d) < 0$ and $(1+J_2(d)) > 0$ from Lemma 2. $\square$

From Proposition 8 above, we find that the estimators proposed by Chen and Lo [5] in the case-cohort sampling design do not achieve the asymptotic lower bound.

5.3. Discussion

Although the maximum pseudolikelihood estimator of Prentice [16] and the estimators of Chen and Lo [5]are technically inefficient for all $p < 1$, for practical purposes one wants to find the distance between their asymptotic variances and the theoretical variance lower bound. Table 1 gives the asymptotic variance $\tilde{\sigma}_{Ber}^2$ of the SP88 estimator $\tilde{\theta}$, the common asymptotic variance $\sigma_{3,Ber}^2$ of the CL99 estimators $\tilde{\theta}_2$ and $\tilde{\theta}_3$, and the asymptotic lower bound V_B^2, each relative to the same asymptotic variance of MPLE $\hat{\theta}$, for various values of p ranging from 0.001 to 0.5 and $d = 0.001, 0.01, 0.1, 0.5$ and 0.8.

Table 1. Comparison of asymptotic variances of SP88 and CL99 estimators with asymptotic lower bound

Disease probability d	Sampling fraction p	Variance (SP88) $\widetilde{\sigma}^2_{\text{Ber}}$	Variance (CL99) $\sigma_{3,\text{Ber}}^{2}$	Lower bound V_{B}^{2}
0.001	0.001	1.666	1.666	1.001
	0.005	1.133	1.133	1.001
	0.010	1.066	1.066	1.001
	0.050	1.013	1.013	1.001
	0.100	1.006	1.006	1.001
	0.300	1.002	1.002	1.000
	0.500	1.001	1.001	1.000
0.010	0.001	7.682	7.666	1.007
	0.005	2.331	2.328	1.007
	0.010	1.662	1.661	1.007
	0.050	1.127	1.127	1.006
	0.100	1.060	1.060	1.006
	0.300	1.016	1.016	1.005
	0.500	1.007	1.007	1.003
0.100	0.001	69.947	68.150	1.072
	0.005	14.734	14.376	1.072
	0.010	7.833	7.654	1.071
	0.050	2.311	2.277	1.068
	0.100	1.621	1.605	1.064
	0.300	1.161	1.157	1.049
	0.500	1.069	1.067	1.035
0.500	0.001	411.889	345.606	1.526
	0.005	82.849	69.645	1.523
	0.010	41.719	35.150	1.519
	0.050	8.815	7.554	1.487
	0.100	4.702	4.105	1.450
	0.300	1.960	1.805	1.318
	0.500	1.411	1.345	1.208
0.800	0.001	818.837	541.781	2.178
	0.005	163.912	108.723	2.167
	0.010	82.047	54.591	2.155
	0.050	16.554	11.285	2.059
	0.100	8.368	5.872	1.950
	0.300	2.910	2.263	1.610
	0.500	1.819	1.541	1.371

One important feature we can see from Table 1 is that when d is small, the SP88 estimator and the CL99 estimators are nearly efficient for the case-cohort design with a sample of only 10‑30% or so of the whole cohort. That is, given the information gathered in a case-cohort design in this range of p, other estimators can only improve on the SP88 and the CL99 estimators by a small amount. We can also see from Table 1 how the SP88 estimator and the CL99 estimator are

far from efficient for small values of p; if these designs are necessary then one should perhaps look for alternative methods of estimation.

A main point to take away from Table 1 is that p should be at least as high as d to obtain reasonable efficiency. For example, if $d=0.5$, and we sample only $p=0.1\%$ of the whole cohort, the asymptotic variances of both the SP88 estimator and the CL99 estimators are far from the asymptotic lower bound. However, even though both estimators are far from efficient, one can see how the CL99 estimators greatly improve the SP88 estimator in this situation, with the CL99 estimator having 83.9% of the variance of the SP88 estimator in this particular situation. Lastly, in many studies, the disease probability d is known in advance approximately, and we see from Table 1 how this information would be valuable in helping us decide what fraction of the whole cohort to sample.

6. Concluding remarks

Neither the asymptotic variance of the maximum pseudolikelihood estimator of Prentice [16], nor that of the estimators proposed in [5] achieve the computed lower bound for the case-cohort design under the null $\theta_0=0$. In principle, the tools used here can also be applied in the case when $\theta_0 \neq 0$. However, when doing so a difficulty arises when we attempt to compute the solution $\beta^*(t)$ of the normal equation and obtain the orthogonal projection $A\beta^*(x)$ of $\rho(x)$ onto the closed space $\{ A\beta : \beta \in \mathscr{B} \}$ of $L^2(\nu)$, as was done in and when $\theta_0=0$. Following the same setup as given in Section 2, one can show that the normal equation to be solved in the general case when $\theta_0 \neq 0$ is given by

$$R\beta(t)\,\frac{M_0(t)}{\overline{G}(t)} - \int_0^t R\beta(s)\,\frac{M_0(s)}{\overline{G}(s)}\,\frac{dG(s)}{\overline{G}(s)} - (1-p)\int_0^t \frac{\int_s^\infty \beta g^{1/2}\,d\nu}{\overline{G}(s)}\,\frac{K_0(s)}{\overline{G}(s)}\,ds$$

$$= \frac{1}{2}\left[\frac{M_1(t)}{\overline{G}(t)} - \int_0^t \frac{M_1(s)}{\overline{G}(s)}\,\frac{dG(s)}{\overline{G}(s)} - (1-p)\int_0^t \frac{\log \overline{G}(s)}{\overline{G}(s)}\,K_1(s)\,ds\right],$$

where $M_i(t) = E[Z^i e^{\theta Z} 1(T > t)]$, $K_i(s) = (1/p)\int_{-\infty}^\infty z^i e^{2\theta z} f(s,0,1,z)dz$. Unfortunately, the solution to this equation, and therefore a variance lower bound for the case of nonzero θ_0 is not as forthcoming as in the null case.

In conclusion, although the maximum pseudolikelihood estimator of Prentice [16] and the estimators of Chen and Lo [5] do not achieve the asymptotic lower bound, they generally perform well in the case-cohort design with a large enough sampling fraction and small disease probabilities. It is our hope that our analysis of the asymptotic variance lower bound will provide better insight into the case-cohort sampling design in general.

References

[1] J.M. Begun, W.J. Hall, W. Huang, J.A. Wellner, Information and asymptotic efficiency in parametric-nonparametric models, Ann. Statist. 11 (1983) 432–452.

[2] J.M. Begun, J.A. Wellner, Asymptotic efficiency of relative risk estimates, in: P.K. Sen (Ed.), Contributions to Statistics: Essays in Honor of Norman L. Johnson, 1982.

[3] J. Boice, M. Blettner, R. Kleinermam, Radiation dose and leukemia risk in patients treated for cancer of the cervix, J. Natl. Cancer Inst. 79 (1987) 1295–1311.

[4] N.E. Breslow, J.H. Lubin, P. Marek, B. Langholz, Multiplicative models and cohort analysis, J. Amer. Statist. Assoc. 78 (1983) 1–12.

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

[6] Y.S. Chow, H. Teicher, Probability Theory: Independence, Interchangeability, Martingales, 3rd Edition, Springer, New York, 1997.

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

[8] R. Durrett, Probability: Theory and Examples, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1991.

[9] D. Garabrant, J. Held, B. Langholz, L. Bernstein, Mortality of aircraft manufacturing workers in Southern California, Amer. J. Indust. Med. 13 (1988) 683–693.

[10] P.E. Greenwood, W. Wefelmeyer, Efficiency of estimators for partially specified filtered models, Stochastic Process. Appl. 36 (1990) 353–370.

[11] J. Ha´jek, A characterization of limiting distributions of regular estimates, Z. Wahrsch. Verw. Gebiete. 14 (1970) 323–330.

[12] B. Langholz, Ø: Borgan, Counter-matching: a stratified nested case-control sampling method, Biometrika 82 (1995) 69–79.

[13] L. LeCam, On a theorem of J. Ha´jek, in: J. Jureckova (Ed.), Contributions to Statistics; Jaroslan Hájek Memorial Volume, Reidel, Dordrecht, 1979, pp. 119–135.

[14] F. Liddell, J. McDonald, D. Thomas, Methods of cohort analysis: appraisal by application to asbestos miners, J. Roy. Statist. Soc. A 140 (1977) 469–491.

[15] D.G. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.

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

[17] S.G. Self, R.L. Prentice, Asymptotic distribution theory and efficiency results for case-cohort studies, Ann. Statist. 16 (1988) 64–81.

[18] D.C. Thomas, Addendum to: methods of cohort analysis: appraisal by application to asbestos mining, in: F.D.K. Liddell, J.C. McDonald, D.C. Thomas (Eds.), J. Roy. Statist. Soc. A 140 (1977) 469–491.

[19] A. Whittemore, A. McMillan, Lung cancer mortality among US uranium miners: a reappraisal, J. Natl. Cancer Inst. 71 (1983) 489–499.

Further reading

P.K. Andersen, R.D. Gill, Cox's regression model for counting processes: a large sample study, Ann. Statist. 10 (1982) 1100–1120.