*The Annals of Statistics*
1992, Vol. 20, No. 4, 1903–1928

# ASYMPTOTIC THEORY FOR NESTED CASE-CONTROL SAMPLING IN THE COX REGRESSION MODEL

BY LARRY GOLDSTEIN[1] AND BRYAN LANGHOLZ[2]

*University of Southern California*

By providing a probabilistic model for nested case-control sampling in epidemiologic cohort studies, consistency and asymptotic normality of the maximum partial likelihood estimator of regression parameters in a Cox proportional hazards model can be derived using process and martingale theory as in Andersen and Gill. A general expression for the asymptotic variance is given and used to calculate asymptotic relative efficiencies relative to the full cohort variance in some important special cases.

**1. Introduction.** The Cox model for failure time data [Cox (1972)] specifies that the hazard for failure at time $t$ for an individual with covariate vector $Z(t) \in \mathbf{R}^d$ is

$$\lambda_0(t)\exp\{\beta_0' Z(t)\},$$

where $\lambda_0$ is the baseline hazard rate and $\beta_0$ is an unknown parameter in $\mathbf{R}^d$.

With $T_i$ the time of failure or censoring of the $i$th individual with covariate $Z_i$, the partial likelihood for a set of data is given by the product

$$L_f(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta' Z_i(T_i)\}}{\sum_{j \in R_i} \exp\{\beta' Z_j(T_i)\}} \right\}^{\delta_i},$$

where $\delta_i$ is the indicator that the failure of individual $i$ was observed and $R_i$ is the set of those at risk at the time of the $i$th failure [Cox (1972)]. Using a counting process model and martingale central limit theorems, Andersen and Gill (1982) show that $L_f$ may be treated analogously to a standard likelihood for inferences on $\beta_0$. This model is well suited for the analysis of cohort studies in medical research in which morbidity or mortality is the endpoint of interest and which involve the follow-up of a large number of subjects. It has become a standard method of evaluating clinical trials, disease prevention trials and epidemiologic cohort studies. Often assembling the covariant histories $Z_i$ for each member of the cohort is prohibitively expensive and methods in which covariate data need only be collected for a small sample of the cohort are highly desirable.

Nested case-control sampling [Thomas (1977)] is a popular method of sampling from a cohort and has been employed in many studies [e.g., Liddel, McDonald and Thomas (1977), Breslow, Lubin, Marek and Langholz (1983),

Received April 1990; revised January 1992.
[1]Research supported in part by NSF Grant DMS-90-05833.
[2]Research supported in part by National Cancer Institute Grants CA14089 and CA42949.
*AMS* 1980 *subject classifications*. Primary 62F12; secondary 62M99, 62D05.
*Key words and phrases*. Survival analysis, cohort sampling, martingale, censoring, efficiency.

Whittemore and MacMillan (1983), Boice, Blettner, Kleinerman, et al. (1987), Yeh, Yu, Mo, Luo, Tong and Henderson (1989)]. The expression $L_f$ is formally equivalent to the conditional logistic likelihood used for the analysis of matched case-control studies in which the failure at a given time is considered the *case* and the others at risk at that time are considered *controls*. Unlike matched case-control studies in which the matched sets are disjoint, in $L_f$ controls will generally appear in more than one risk set. Typically, there are a large number of controls per failure and it is well known [e.g., Breslow, Lubin, Marek and Langholz (1983)] that, for matched case-control studies, little added efficiency is realized with more than six or seven controls per case. Thus under this paradigm, it is natural to use a sample risk set $\tilde{R}_i$ of size $m$ consisting of the failure and a random sample of $m - 1$ controls instead of the $R_i$ in $L_f$, leading to the expression

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta' Z_i(T_i)\}}{\Sigma_{j \in \tilde{R}_i} \exp\{\beta' Z_j(T_i)\}} \right\}^{\delta_i}.$$

For example, in a study of mortality from esophagus cancer among the employees of a certain San Diego county aircraft manufacturing firm, Garabrant, Held, Langholz and Bernstein (1988) defined a cohort of 14,067 workers as all those who worked at the firm four years or more prior to December 31, 1982, with at least some of this time after January 1, 1958. This cohort was followed through December 31, 1982. For each subject the researchers needed to collect "basic information" to define the cohort: date and age of entry into and exit from the cohort (obtained from company records), mortality status at the end of the study (obtained through company, California Department of Motor Vehicles, Social Security or credit company records) and, if dead, the date and cause of death (obtained from the death certificate).

To explore the relationship between particular exposures and the occurrence of esophagus cancer, 4 age and sex matched controls were randomly selected for each of the 14 esophagus cases, yielding a valid sample from the cohort with close to full efficiency. For the $(1 + 4) \times 14 = 70$ subjects in this sample, detailed exposure histories were compiled by collecting job histories and ascertaining, through company records and interviews with employees at the plant, the types of substances used in the various manufacturing processes by an individual with such a job at a given time; an impossible task for the entire cohort of over 14,000. We return to this example and calculate the asymptotic relative efficiency of using such a sample relative to the entire cohort in Section 6.2.

Theoretical justification for the use of $L$ in the estimation of $\beta$ has thus far been limited to the partial likelihood arguments of Oakes (1981) and Prentice (1986a). However, conditions for the consistency and asymptotic normality of the log relative risk parameter based on $L$ have not previously been explored.

As in Andersen and Gill (1982), we associate with the $i$th individual a counting process $N_i$ with rate

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp\{\beta_0' Z_i(t)\}$$

such that $M_i$, with $dM_i = dN_i - \lambda_i$, is a martingale. The term $Y_i(t)$ is the indicator that the $i$th individual is at risk at time $t$; in particular, $N_i$ can only jump when $Y_i(t) = 1$.

Although our overall framework is somewhat less general than is considered by Andersen and Gill (1982), our conditions are easy to understand, apply to situations of great practical importance and vastly simplify the exposition; their relaxation is discussed in Section 7.1. We considers cohorts with a single stratum, but the results generalize to multiple strata in an obvious way. Whereas Prentice and Self (1983) allow for general relative risks of the form $r(\beta'Z)$, we restrict attention to relative risk of the form $\exp(\beta'Z)$.

In this paper we provide a probabilistic model for nested case-control sampling which allows for a theoretically rigorous study of this sampling design. We provide conditions to prove the consistency (Theorem 1) and asymptotic normality (Theorem 3) of $\hat{\beta}$, the maximum partial likelihood estimator based on $L$, and show that the inverse information, the second derivative of $\log\{L(\beta)\}$ at $\beta = \hat{\beta}$, is a consistent estimator of the inverse variance matrix of $\hat{\beta}$ (Theorem 2). We show that under our assumptions our formulas reduce to those of Andersen and Gill (1982) as the number of controls tends to infinity. Furthermore we obtain asymptotic relative efficiencies relative to the estimator based on $L_f$ for the special cases $\beta_0 = 0$ and of univariate binary exposure. These last examples generalize the results of Breslow and Patton (1979) and Breslow, Lubin, Marek and Langholz (1983).

**2. Notation and assumptions.** Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\{\mathscr{F}_t\}_{t\in[0,a)}$ a right continuous, nondecreasing family of sub-$\sigma$-algebras of $\mathscr{F}$ with $\mathscr{F}_0$ containing all $P$ null subsets of $\mathscr{F}$. Although we normalize our finite time interval to be $[0, 1]$, we take $a > 1$ so that right-hand limits at 1 exist. We assume the probability space supports copies of a counting process $N$, a covariate process $Z$, an observation process $Y$ and the sampling functions $\eta$ described in succeeding text. With $d$ a nonnegative integer, $Z$ and $Y$ are assumed to be predictable mappings from $\Omega \times [0, a)$ into $\mathbf{R}^d$ and $\{0, 1\}$, respectively. The process $Y$ is the indicator that an individual is at risk. The vector $\beta_0 \in \mathbf{R}^d$ will be fixed. Let $\lambda(t) = Y(t)\lambda_0(t)\exp(\beta_0'Z(t))$, the hazard function multiplied by the at risk indicator at time $t$ for an individual with covariate vector $Z(t)$, and

$$\Lambda(t) = \int_0^t \lambda(s)\, ds.$$

The following conditions are assumed to be in force throughout. The relaxation of these conditions is discussed in Section 7.1.

CONDITION 1. The process

$$M = N - \Lambda \text{ is a } \{\mathscr{F}_t\}_{t\in[0,a)} \text{ local martingale.}$$

CONDITION 2. The baseline hazard $\lambda_0$ is a deterministic measurable function bounded away from 0 and infinity.

CONDITION 3. The covariate process $Z(\cdot)$ is bounded.

Using that $N$ has only unit jumps, the preceding conditions imply that $M$ is a $\{\mathcal{F}_t\}_{t\in[0,a)}$ local square integrable martingale with quadratic variation $\Lambda$. Since the conditions imply also that $E\{\Lambda(t)\} < \infty$ for all $t \geq 0$, $M$ is an $\{\mathcal{F}_t\}_{t\in[0,a)}$ martingale and has a only a finite number of jumps in any bounded interval with probability 1 (see also Lemma 2).

It is worth mentioning that in the framework of some authors the intensity process is introduced under assumptions by the relation

$$\lim_{h\downarrow 0}\frac{1}{h}P\big(N(t+h)-N(t)=1|\mathcal{F}_t^-\big)=\lambda(t),$$

from which Condition 1 now follows [see, e.g., Aalen (1978) or Andersen and Borgan (1985)].

CONDITION 4. The process $Y$ is left continuous and has only finitely many jumps on any bounded interval. Furthermore, we assume that $[0,1]\subset\bigcup_{k=1}^{K} I_k$, where $I_1, I_2, \ldots, I_K$ are measurable subsets of $[0,1]$, and

$$\min_k P\big(\forall\, t\in I_k,\, Y(t)=1\big)>0. \tag{1}$$

In particular (1) implies that for every $t\in[0,1]$,

$$p(t)=P\big(Y(t)=1\big)>0$$

and is implied by

$$P\big(\forall\, t\in[0,1],\, Y(t)=1\big)>0. \tag{2}$$

Let $Z_Y(t)$ be a random vector with

$$P\big(Z_Y(t)\in B\big)=P\big(Z(t)\in B|Y(t)=1\big) \tag{3}$$

for $B$ a Borel subset of $\mathbf{R}^d$.

CONDITION 5. With $V(t)=\mathrm{Cov}(Z_Y(t))$, the matrix

$$V=\int_0^1 V(s)\lambda_0(s)\,ds$$

is positive definite.

We will denote this condition by $V>0$. This condition is necessary; see Lemma 3 in the Appendix.

The positive integer $n$ will denote the number of individuals. Let $I=\{1,2,\ldots,n\}$.

CONDITION 6. With respect to the measure $P$, the processes $(Z_i, Y_i)$, $i \in I$, are independent copies of the process $(Z, Y)$, with $\lambda_i$, $N_i$ and $M_i$ defined in the obvious analogous way.

Lemma 2 in the Appendix shows that the preceding assumptions imply that event times of $N_i$ and $N_j$ have no points in common for $i \neq j$.

In order to apply martingale central limit theorems, we need to associate nested case-control sampling with a predictable process. Informally, immediately after any failure or change in a subject's at risk status, a set of controls is drawn for every subject at risk. Thus, at any failure time a set of controls for the failed individual will already have been selected. Since the selection may be done by an adapted random mechanism, the resulting sampling processes (described in succeeding text) are predictable.

Define $R(t)$, the risk set at time $t+$, by

$$R(t) = \{j: Y_j(t+) = 1\}.$$

Let $n(t) = |R(t)|$, the number of individuals at risk at time $t+$.

Let $T_0 = 0$ and for $k \geq 1$,

$$T_k = \inf\{t > T_{k-1}: Y_i(t+) \neq Y_i(t) \text{ or } N_i(t) \neq N_i(t-) \text{ for some } i\};$$

that is, $T_1, T_2, \ldots$ is the ordered collection of event times of the $Y_i$ and $N_i$ processes. Event times of $\{N_i\}$ will be called *failure* times. Let $n_k = n(T_k)$ and $R_k = R(T_k)$.

If $i \in R_{k-1}$, let $P_{m,i}(R_{k-1})$ be the set of all subsets of size $m$ of $R_{k-1}$ that include $i$; there are $\binom{n_{k-1}-1}{m-1}$ such subsets. Let $\tilde{R}_{k,i}$ be independently and uniformly chosen from $P_{m,i}(R_{k-1})$; if $i \notin R_{k-1}$, we let $\tilde{R}_{k,i}$ be the empty set. The set $\tilde{R}_{k,i}$ lists the individuals that would serve in the risk set should individual $i$ be the failure at time $T_k$. Setting $\eta_{ij}(0) = 0$ arbitrarily, we note that the preceding construction makes

$$\eta_{ij}(t) = \sum_{k \geq 1} 1\{j \in \tilde{R}_{k,i}\} 1\{T_{k-1} < t \leq T_k\}$$

predictable.

The $\tilde{R}_{k,i}$ and the $\eta_{ij}$ are equivalent ways to represent the sampling process. Our choice in writing one over the other in what follows is a matter of convenience only.

At the *failure* time $T_k$, let $i_k$ denote the failed individual, that is, the unique index $i_k$ such that

$$N_{i_k}(T_k) \neq N_{i_k}(T_k-).$$

Setting $\tilde{R}_k = \tilde{R}_{k,i_k}$ for notational convenience, the nested case-control partial likelihood may be written

$$L(\beta) = \prod_k \left\{ \frac{\exp\{\beta' Z_{i_k}(T_k)\}}{\sum_{j \in \tilde{R}_k} \exp\{\beta' Z_j(T_k)\}} \right\},$$

where the product is taken over all $k$ such that $T_k$ is a failure time less than or equal to 1. Note that this likelihood accommodates situations where there may be multiple failure times associated with single individuals and reduces to the $L$ given previously when censoring occurs after the first failure.

We estimate $\beta_0$ by $\hat{\beta}$, a solution of the likelihood equation

$$\frac{d \log L(\beta)}{d\beta} = 0.$$

Here and in what follows, expressions of the form $0/0$ and $\binom{m}{j}$ for $m < j$ will be set to 1. For $a \in \mathbf{R}^d$, $|a|^2 = a'a$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa'$. For a matrix $D \in \mathbf{R}^{d \times d}$, let

$$\|D\| = \sup_{|a| \le 1} |Da|.$$

The symbols $\Rightarrow$ and $\to_p$ will denote convergence in distribution and in probability, respectively. Let $U = \{1, 2, \ldots, m\}$ and, for $T \subset I$, $Y_T = \prod_{j \in T} Y_j$.

**3. Preliminaries.** In this section, fix throughout and suppress an arbitrary $s \in [0, 1]$; for this $s$ write $Y = Y(s)$, $Z = Z(s)$ and $p = p(s)$. The following lemma is key.

LEMMA 1. *Let $\rho \in \{1, 2\}$ and $(Y_i, Z_i)$, $i \in I$, be independent copies of $(Y, Z)$ with $Z \in \mathbf{R}^d$, $Y \in \{0, 1\}$ and*

$$P(Y = 1) = p > 0.$$

*Let $R = \{j\colon Y_j = 1\}$, $P = \{T \subset R,\ |T| = m\}$ and $P_i = \{T \in P\colon i \in T\}$. With $T \in P$, let $w(T)$ be of the form*

$$w(T) = \left\{ \frac{\sum_{j \in T} Z_j \exp(\beta' Z_j)}{\sum_{j \in T} \exp(\beta' Z_j)} \right\}^{\otimes \rho} \tag{4}$$

*or*

$$w(T) = \left\{ \frac{\sum_{j \in T} Z_j^{\otimes \rho} \exp(\beta' Z_j)}{\sum_{j \in T} \exp(\beta' Z_j)} \right\} \tag{5}$$

*with $w(\varnothing) = 0$. Let $\mathscr{F} = \sigma\{Z_i\colon i \in I\}$ and $\mathscr{Y}$ be a sigma algebra containing $\sigma\{Y_i;\ i \in I\}$. Suppose that, given $\mathscr{Y}$, $\tilde{R}_i$ are mutually independent and uniform on $P_i$ and that $\mathscr{F}$ and $\sigma\{\tilde{R}_i;\ i \in I\}$ are conditionally independent given $\mathscr{Y}$. Let*

$$S_n = \frac{1}{n} \sum_{i=1}^{n} w(\tilde{R}_i) Y_i A_i,$$

*where $A_i$ is $\exp(\beta_0' Z_i)$ when $\rho = 2$ and may be either $\exp(\beta_0' Z_i)$ or $Z_i' \exp(\beta_0' Z_i)$ when $\rho = 1$.*

*Then* $S_n \to_p q$ *where*

$$q = pE\left\{ w(U)\frac{1}{m}\sum_{j\in U} A_j \middle| Y_U = 1 \right\}.$$

PROOF. We consider the case $\rho = 1$ and $A_i = \exp(\beta_0' Z_i)$. The other cases are analogous. Let $\mathscr{G} = \mathscr{F} \vee \mathscr{Y}$. With $S_{n,k}$ the $k$th component of $S_n$, it suffices to show that there exists a sequence $Q_n$ such that:

(a) $\mathrm{Var}\{S_{n,k}|\mathscr{G}\} \to_p 0$;
(b) $E\{S_{n,k}|\mathscr{G}\}/Q_{n,k} \to_p 1$;
(c) $Q_n \to_p q$.

Conditional on $\mathscr{G}$, every component of $S_n$ is the average of $n$ independent random variables. Let $Z(T) = \max_{j\in T}|Z_j|$. Using that

$$0 \le |w(T)| \le Z(T)$$

we have

$$\mathrm{Var}\{S_{n,k}|\mathscr{G}\} \le \frac{1}{n^2}\sum_{i=1}^{n} E\left\{ Y_i \exp\left(2|\beta_0|Z\left(\tilde{R}_i\right)\right) Z\left(\tilde{R}_i\right)^2 \middle| \mathscr{G} \right\}.$$

Taking expectation and using Condition 3 shows (a).

We next compute $E\{S_n|\mathscr{G}\}$.

First,

$$w\left(\tilde{R}_i\right)\exp(\beta_0' Z_i) Y_i = \sum_{T\in P_i} w(T)\exp(\beta_0' Z_i) Y_i 1\left(T = \tilde{R}_i\right),$$

so

$$E\left\{ w\left(\tilde{R}_i\right)\exp(\beta_0' Z_i) Y_i \middle| \mathscr{G} \right\} = \sum_{T\in P_i} w(T)\exp(\beta_0' Z_i) Y_i E\left\{ 1\left(T = \tilde{R}_i\right) \middle| \mathscr{G} \right\}.$$

By the conditional independence condition,

$$E\left\{ 1\left(T = \tilde{R}_i\right) \middle| \mathscr{G} \right\} = E\left\{ 1\left(T = \tilde{R}_i\right) \middle| \mathscr{Y} \right\} = \binom{|R|-1}{m-1}^{-1} Y_T 1(i \in T).$$

Hence, $E\{w(\tilde{R}_i)\exp(\beta_0' Z_i) Y_i|\mathscr{G}\}$ equals

$$\binom{|R|-1}{m-1}^{-1} \sum_{T\in P_i} Y_T w(T)\exp(\beta_0' Z_i),$$

and so

$$E\{S_n|\mathscr{G}\} = \frac{1}{n}\binom{|R|-1}{m-1}^{-1} \sum_{i=1}^{n}\sum_{T\in P_i} Y_T w(T)\exp(\beta_0' Z_i)$$

$$= \frac{1}{n}\binom{|R|-1}{m-1}^{-1} \sum_{|T|=m} \left\{ Y_T w(T) \sum_{i\in T} \exp(\beta_0' Z_i) \right\}. \tag{6}$$

Let

$$Q_n = \frac{1}{m} p^{-m+1} \binom{n}{m}^{-1} \sum_{|T|=m} \left\{ Y_T w(T) \sum_{i \in T} \exp(\beta_0' Z_i) \right\}. \tag{7}$$

Since

$$\frac{|R|}{n} \to_p p \quad \text{we have} \quad \frac{E\{S_{n,k} | \mathscr{G}\}}{Q_{n,k}} \to_p 1, \tag{8}$$

and therefore (b).

It remains only to show $Q_n \to_p q$.

First, note

$$\operatorname{Var}(Q_n) = c \binom{n}{m}^{-2} \sum_{|T|=m,\ |S|=m} \operatorname{Cov}\Big( Y_T w(T) \sum_{i \in T} e^{\beta_0' Z_i},\ Y_S w(S) \sum_{i \in S} \exp(\beta_0' Z_i) \Big).$$

For $T \cap S = \varnothing$, the preceding covariance is zero. There are $\binom{n}{2m}\binom{2m}{m}$ pairs $S,T$ that do not intersect; hence the preceding sum has

$$\binom{n}{m}^2 - \binom{n}{2m}\binom{2m}{m}$$

nonzero terms. Since by Condition 3 each of these covariance terms are bounded,

$$\operatorname{Var}\{Q_n\} \to 0 \tag{9}$$

is now implied by

$$\lim_{n \to \infty} \binom{n}{m}^{-2} \binom{n}{2m}\binom{2m}{m} = 1.$$

Last,

$$EQ_n = \frac{1}{m} p^{-m+1} E\left\{ Y_U w(U) \sum_{i \in U} \exp(\beta_0' Z_i) \right\}$$

$$= \frac{pE\{Y_U w(U)(1/m)\Sigma_{i \in U} \exp(\beta_0' Z_i)\}}{EY_U} = q. \qquad \square$$

COROLLARY 1. *In the special case* $\rho = 1$, $\beta = \beta_0$ *and* $A_i = \exp(\beta_0' Z_i)$,

$$E\{S_n | \mathscr{G}\} = \frac{1}{n} \sum_{i=1}^{n} Y_i Z_i \exp(\beta_0' Z_i).$$

PROOF. From (6) we have

$$E\{S_n|\mathscr{S}\} = \frac{1}{n}\binom{|R|-1}{m-1}^{-1} \sum_{T\in P}\sum_{i\in T} Z_i \exp(\beta_0' Z_i)$$

$$= \frac{1}{n}\sum_{i\in R} Z_i \exp(\beta_0' Z_i) = \frac{1}{n}\sum_{i=1}^{n} Y_i Z_i \exp(\beta_0' Z_i). \qquad \square$$

REMARK. The conclusion of the lemma and its application, which follows, involves convergence only for fixed times $s \in [0, 1]$. To emphasize the dependence of $q$ on $\beta$ and $s$, we will write $q(\beta, s)$ in succeeding text.

**4. Consistency of $\hat{\beta}$.** The proof that $\hat{\beta}$ is consistent depends on $C(\beta, t)$, the logarithm of the nested case-control Cox likelihood at time $t$:

$$C(\beta, t) = \sum_{i=1}^{n} \int_0^t \left[\beta' Z_i(s) - \log\left\{\sum_{j=1}^{n} \eta_{i,j}(s)\exp\big(\beta' Z_j(s)\big)\right\}\right] dN_i(s).$$

THEOREM 1. *The estimator $\hat{\beta}$ is consistent for $\beta$.*

PROOF. Let

$$S_i(\beta, t) = \sum_{j=1}^{n} \eta_{i,j}(t)\exp\big(\beta' Z_j(t)\big),$$

$$X(\beta, t) = n^{-1}(C(\beta, t) - C(\beta_0, t))$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_0^t \left[(\beta - \beta_0)' Z_i(s) - \log\left\{\frac{S_i(\beta, s)}{S_i(\beta_0, s)}\right\}\right] dN_i(s)$$

and

$$A(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} \int_0^t \left[(\beta - \beta_0)' Z_i(s) - \log\left\{\frac{S_i(\beta, s)}{S_i(\beta_0, s)}\right\}\right] \lambda_i(s)\, ds.$$

Then for each $\beta$, $X(\beta, \cdot) - A(\beta, \cdot)$ is a square integrable martingale with quadratic variation at time $t$ given by

$$\frac{1}{n^2}\sum_{i=1}^{n} \int_0^t \left[(\beta - \beta_0)' Z_i(s) - \log\left\{\frac{S_i(\beta, s)}{S_i(\beta_0, s)}\right\}\right]^2 \lambda_i(s)\, ds.$$

Taking expectation and using Condition 3 we see that the preceding quadratic variation at time 1 tends to zero as $n \to \infty$; therefore, by an inequality of Lenglart (1977), $X(\beta, 1)$ converges in probability to the same limit as $A(\beta, 1)$, as in the proof of Lemma 3.1 of Andersen and Gill (1982).

In what follows, interchanges of limiting operations are justified by a dominated convergence theorem. Hence,

$$\frac{\partial}{\partial\beta}A(\beta,1) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1\left[Z_i(s) - \frac{\sum_{j=1}^n \eta_{i,j}(s)Z_j(s)\exp\left(\beta'Z_j(s)\right)}{\sum_{j=1}^n \eta_{i,j}(s)\exp\left(\beta'Z_j(s)\right)}\right] \times Y_i(s)\exp\left(\beta_0'Z_i(s)\right)\lambda_0(s)\,ds,$$

and, using Lemma 1,

$$\lim_{n\to\infty}\frac{\partial}{\partial\beta}A(\beta,1) = \int_0^1\left[q(\beta_0,s) - q(\beta,s)\right]\lambda_0(s)\,ds \quad \text{in probability,}$$

where

$$q(\beta,s) = p(s)E\left\{\frac{\sum_{j\in U}Z_j(s)\exp\left(\beta'Z_j(s)\right)}{\sum_{j\in U}\exp\left(\beta'Z_j(s)\right)}\frac{1}{m}\sum_{j\in U}\exp\left(\beta_0'Z_j(s)\right)\middle| Y_U(s) = 1\right\}.$$

This demonstrates that $A(\beta,1)$ converges to a function with first derivative 0 at $\beta = \beta_0$.

With $\Gamma$ as in (11) in Section 5, the second derivative of the limit of $A(\beta,1)$ is equal to minus a nonnegative definite matrix for every $\beta$ and at $\beta_0$ to $-\Gamma$, shown in Lemma 4 of the Appendix to be negative definite. One can now follow the argument of Andersen and Gill (1982) to demonstrate that since $\hat\beta$ maximizes $X(\beta,1)$, $\hat\beta \to_p \beta_0$. □

**5. Asymptotic normality of $\hat\beta$.** In this section, we derive the asymptotic distribution of the estimator $\hat\beta$ and give a consistent estimator for $\Gamma$, the inverse of the estimators asymptotic covariance matrix.

First we describe the construction of a random variable $\hat Z$. Suppressing $s$, a fixed time and $\beta$, let $\mathbf{Z}_{Y,U} = (Z_{Y,1}, Z_{Y,2}, \ldots, Z_{Y,m})$, a vector of $m$ independent copies of the random variable $Z_Y$ with distribution given in (3). Let

$$p_j = \frac{\exp\left(\beta'Z_{Y,j}\right)}{\sum_{i\in U}\exp\left(\beta'Z_{Y,i}\right)} \tag{10}$$

and set $P(\hat Z = Z_{Y,j}|\mathbf{Z}_{Y,U}) = p_j$.

Let

$$\bar Z = \sum_{j\in U} p_j Z_{Y,j}.$$

Then

$$\operatorname{Cov}\left(\hat Z\,|\mathbf{Z}_{Y,U}\right) = \sum_{j\in U} Z_{Y,j}^{\otimes 2}p_j - \left(\sum_{j\in U} Z_{Y,j}p_j\right)^{\otimes 2} = \sum_{j\in U}\left(Z_{Y,j} - \bar Z\right)^{\otimes 2}p_j.$$

Define

$$\Gamma(\beta,t) = E\left\{\int_0^t p(s)\frac{1}{m}\sum_{j\in U}\exp\bigl(\beta_0' Z_{Y,j}(s)\bigr)\mathrm{Cov}\bigl(\hat{Z}(s)\big|\mathbf{Z}_{Y,U}\bigr)\lambda_0(s)\,ds\right\},$$

and

$$\Gamma = \Gamma(\beta_0,1). \tag{11}$$

The proof that $\Gamma$ is positive definite and therefore invertible, is given in Lemma 4 of the Appendix.

We will consider the score process

$$U(\beta,t) = \sum_{i=1}^n \int_0^t \left[Z_i(s) - \frac{\sum_{j=1}^n \eta_{i,j}(s)Z_j(s)\exp\bigl(\beta' Z_j(s)\bigr)}{\sum_{j=1}^n \eta_{i,j}(s)\exp\bigl(\beta' Z_j(s)\bigr)}\right] dN_i(s) \tag{12}$$

and the information process

$$\mathcal{I}(\beta,t) = \int_0^t \sum_{i=1}^n D_i(\beta,s)\,dN_i(s),$$

where

$$D_i(\beta,s) = \frac{\sum_{j=1}^n \eta_{i,j}(s)Z_j(s)^{\otimes 2}\exp\bigl(\beta' Z_j(s)\bigr)}{\sum_{j=1}^n \eta_{i,j}(s)\exp\bigl(\beta' Z_j(s)\bigr)} - \left(\frac{\sum_{j=1}^n \eta_{i,j}(s)Z_j(s)\exp\bigl(\beta' Z_j(s)\bigr)}{\sum_{j=1}^n \eta_{i,j}(s)\exp\bigl(\beta' Z_j(s)\bigr)}\right)^{\otimes 2}.$$

We now give a consistent estimator of $\Gamma$.

THEOREM 2. *For any* $\beta^* \to_p \beta_0$,

$$n^{-1}\mathcal{I}(\beta^*,1) \to_p \Gamma \quad \text{as } n \to \infty.$$

PROOF. Taking the derivative with respect to $\beta$ and using Condition 3, one may verify that

$$\sup_{i,|v|\le 1} \|D_i'(\beta,s)v\| < \infty$$

and hence there exists a $K$ such that

$$\bigl\|n^{-1}\mathcal{I}(\beta_1,1) - n^{-1}\mathcal{I}(\beta_2,1)\bigr\| \le K|\beta_1 - \beta_2|.$$

Therefore, since $|\beta^* - \beta_0| \to_p 0$ it suffices to show $n^{-1}\mathcal{I}(\beta_0,1) \to_p \Gamma$. Now as

$$\frac{1}{n}\int_0^t \sum_{i=1}^n D_i(\beta_0,s)\,dN_i(s) - \frac{1}{n}\int_0^t \sum_{i=1}^n D_i(\beta_0,s)\,d\Lambda_i(s)$$

is a square integrable martingale, an argument as in Theorem 1 using Lenglart's (1977) inequality shows the above converges to zero in probability.

Lemma 1 now shows that

$$\frac{1}{n}\int_0^t \sum_{i=1}^n D_i(\beta_0, s)\, d\Lambda_i(s) \to_p \Gamma. \qquad \square$$

We now derive the asymptotic distribution of the estimator $\hat{\beta}$.

THEOREM 3.

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \Rightarrow N(0, \Gamma^{-1}).$$

PROOF. With $U_k$ the $k$th component of the score $U$ as in (12) and $\mathscr{I}_k$ the $k$th row of $\mathscr{I}$, the usual Taylor series argument yields

$$U_k(\beta, 1) - U_k(\beta_0, 1) = -\mathscr{I}_k(\beta_k^*, 1)(\beta - \beta_0)$$

for some $\beta_k^*$ on the line segment between $\beta$ and $\beta_0$.

Substituting $\hat{\beta}$ for $\beta$, we have

$$n^{-1/2}U_k(\beta_0, 1) = \left\{n^{-1}\mathscr{I}_k(\beta_k^*, 1)\right\}n^{1/2}\left(\hat{\beta} - \beta_0\right).$$

Hence, using Theorem 2, it suffices to prove

$$n^{-1/2}U(\beta_0, 1) \Rightarrow \mathscr{N}(0, \Gamma). \tag{13}$$

Let

$$E_i(s) = \frac{\sum_{j=1}^n \eta_{i,j}(s) Z_j(s) \exp\left(\beta_0' Z_j(s)\right)}{\sum_{j=1}^n \eta_{i,j}(s) \exp\left(\beta_0' Z_j(s)\right)}$$

and

$$E(s) = \frac{\sum_{j=1}^n Y_j(s) Z_j(s) \exp\left(\beta_0' Z_j(s)\right)}{\sum_{j=1}^n Y_j(s) \exp\left(\beta_0' Z_j(s)\right)}.$$

Then

$$U(\beta_0, t) = \sum_{i=1}^n \int_0^t \left[Z_i(s) - E_i(s)\right] dN_i(s)$$

$$= \sum_{i=1}^n \left\{\int_0^t \left[Z_i(s) - E(s)\right] dN_i(s) + \int_0^t \left[E(s) - E_i(s)\right] dN_i(s)\right\} \tag{14}$$

$$= \sum_{i=1}^n \left\{\int_0^t \left[Z_i(s) - E(s)\right] dM_i(s) + \int_0^t \left[E(s) - E_i(s)\right] dM_i(s) \right.$$

$$\left. + \int_0^t \left[E(s) - E_i(s)\right] d\Lambda_i(s)\right\} \tag{15}$$

$$= \sum_{i=1}^n \left\{\int_0^t \left[Z_i(s) - E_i(s)\right] dM_i(s) + \int_0^t \left[E(s) - E_i(s)\right] d\Lambda_i(s)\right\}.$$

The term

$$A_t = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t [Z_i(s) - E_i(s)]\, dM_i(s)$$

is a stochastic integral of a predictable process against a martingale and is a martingale. We apply the martingale central limit theorem of Rebolledo (1980) as presented in Andersen and Gill (1982). By the independence condition

$$\langle M_i, M_j \rangle_t = 0;$$

hence

$$\langle A \rangle_t = \int_0^t \frac{1}{n} \sum_{i=1}^{n} [Z_i(s) - E_i(s)]^{\otimes 2} \lambda_i(s)\, ds.$$

Expanding the product we obtain an integrand with the four terms $Z_i^{\otimes 2} - Z_i E_i' - E_i Z_i' + E_i^{\otimes 2}$. Factoring out $\lambda_0$ and using the law of large numbers on the first term we have

$$\frac{1}{n} \sum_{i=1}^{n} Z_i^{\otimes 2} Y_i \exp(\beta_0' Z_i) \to pE\{Z^{\otimes 2} \exp(\beta_0' Z) | Y = 1\} = pE\{Z_Y^{\otimes 2} \exp(\beta_0' Z_Y)\}$$

$$= pE\left\{ m Z_{Y,i}^{\otimes 2} p_i \frac{1}{m} \sum_{j \in U} \exp(\beta_0' Z_{Y,j}) \right\}$$

$$= pE\left\{ \sum_{j \in U} Z_{Y,j}^{\otimes 2} p_j \frac{1}{m} \sum_{j \in U} \exp(\beta_0' Z_{Y,j}) \right\}.$$

Using Lemma 1, we see that the remaining three terms all converge to the same limit. For example, $(1/n)\Sigma_{i=1}^n E_i Y_i Z_i' \exp(\beta_0' Z_i)$ converges to

$$pE\left\{ \left( \frac{\Sigma_{j \in U} Z_{Y,j} \exp(\beta_0' Z_{Y,j})}{\Sigma_{j \in U} \exp(\beta_0' Z_{Y,j})} \right) \frac{1}{m} \sum_{j \in U} Z_{Y,j}' \exp(\beta_0' Z_{Y,j}) \right\}$$

$$= pE\left\{ \left( \frac{\Sigma_{j \in U} Z_{Y,j} \exp(\beta_0' Z_{Y,j})}{\Sigma_{i \in U} \exp(\beta_0' Z_{Y,i})} \right)^{\otimes 2} \frac{1}{m} \sum_{j \in U} \exp(\beta_0' Z_{Y,j}) \right\}$$

$$= pE\left\{ \left( \sum_{j \in U} Z_{Y,j} p_j \right)^{\otimes 2} \frac{1}{m} \sum_{j \in U} \exp(\beta_0' Z_{Y,j}) \right\}.$$

Adding the last three terms with appropriate signs, we see that the integrand is $\lambda_0$ multiplied by

$$pE\left\{ \frac{1}{m} \sum_{j \in U} \exp(\beta_0' Z_{Y,j}) \left( \sum_{j \in U} Z_{Y,j}^{\otimes 2} p_j - \left( \sum_{j \in U} Z_{Y,j} p_j \right)^{\otimes 2} \right) \right\};$$

hence

$$\langle A \rangle_t \to_p \Gamma(\beta_0, t) \quad \text{as } n \to \infty.$$

Next, we verify the Lindeberg condition [I.4 in Andersen and Gill (1982)]. Letting

$$Z_{ij}(s) - E_{ij}(s)$$

be the $j$th component of the vector $[Z_i(s) - E_i(s)]$, we have that

$$(16)\quad \int_0^1 \frac{1}{n} \sum_{i=1}^{n} \left[ Z_{ij}(s) - E_{ij}(s) \right]^2 \lambda_i(s) 1\left\{ |Z_{ij}(s) - E_{ij}(s)| > \sqrt{n}\,\varepsilon \right\} ds \to_p 0$$

since the expectation of the preceding integral converges to zero by Conditions 3 and 6. Hence, by the theorem of Rebolledo (1980) as in Andersen and Gill (1982), we have

$$A_1 \Rightarrow N(0, \Gamma).$$

To complete the proof of (13) it remains to demonstrate that

$$B_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^1 \left[ E(s) - E_i(s) \right] d\Lambda_i(s) \to_p 0.$$

Let

$$\mathscr{G} = \sigma\{ Z_i, Y_i, N_i; i = 1, \ldots, n \};$$

that is, $\mathscr{G}$ contains all information except that of sampling. With

$$a_k(s) = \{E(s) - E_k(s)\} Y_k(s) \exp(\beta_0' Z_k(s)) \lambda_0(s),$$

write

$$E\left\{ \| B_1 \|^2 \middle| \mathscr{G} \right\} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \int_0^1 \int_0^1 E\left\{ a_i(s_1)' a_j(s_2) \middle| \mathscr{G} \right\} ds_1\, ds_2.$$

By Corollary 1 we have, suppressing $s$,

$$\sum_{k=1}^{n} E\{a_k | \mathscr{G}\} = 0.$$

Conditional on $\mathscr{G}$, the only random parts of $E_i$ and $E_j$ are those of sampling; thus $E_i$ and $E_j$ are conditionally independent given $\mathscr{G}$ when $i \neq j$. In particular,

$$E\{a_i' a_j | \mathscr{G}\} = E\{a_i' | \mathscr{G}\} E\{a_j | \mathscr{G}\} \quad \text{for } i \neq j.$$

Hence,

$$E\left\{ \| B_1 \|^2 \middle| \mathscr{G} \right\} = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \int_0^1 \left\{ E\{a_i' a_i | \mathscr{G}\} - E\{a_i' | \mathscr{G}\} E\{a_i | \mathscr{G}\} \right\} ds_1\, ds_2.$$

Let $\mathbf{T}$ and $\mathbf{F}$, respectively, stand for the collection of event and failure times less than or equal to 1, with time 1 included in both sets. Using the independence of sampling between event times, we have that $E\{\| B_1 \|^2 | \mathscr{G}\}$ equals

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{T}} \int_{T_j}^{T_{j+1}} \int_{T_j}^{T_{j+1}} \left\{ E\{a_i' a_i | \mathscr{G}\} - E\{a_i' | \mathscr{G}\} E\{a_i | \mathscr{G}\} \right\} ds_1\, ds_2.$$

From this expression we see that the contribution to the double integral over time representing $E\{\|B_1\|^2|\mathscr{S}\}$ comes from the diminishing quadrants that intersect the diagonal $s_1 = s_2$.

Using Conditions 2 and 3, $E\{\|B_1\|^2|\mathscr{S}\}$ may be majorized by a random variable not depending on $i$ or $s$ times:

$$\sum_{T} \Delta T_j^2,$$

where $\Delta T_j = T_{j+1} - T_j$ is taken over consecutive event times in $\mathbf{T}$.

Hence, to show $B_1 \to_p 0$ as $n \to \infty$ it suffices to demonstrate that

$$E\left\{ \sum_{\mathbf{T}} \Delta T_j^2 \right\} \to 0 \quad \text{as } n \to \infty.$$

Note that since for $0 \le a \le b$ one has $a^2 + (b-a)^2 \le b^2$, the preceding equation will follow from

$$E\left\{ \sum_{\mathbf{F}} \Delta' T_j^2 \right\} \to 0 \quad \text{as } n \to \infty, \tag{17}$$

where now

$$\Delta' T_j = T_{k(j+1)} - T_{k(j)} \tag{18}$$

is taken over successive *failure* times in $\mathbf{F}$, $T_{k(j)}$ being the $j$th failure time.

Let

$$\bar{\lambda}(s) = \sum_{i=1}^{n} \lambda_i(s)$$

and, for $\varepsilon > 0$,

$$A_\varepsilon = \left\{ \exists\, s \in [0,1],\, \bar{\lambda}(s) < \varepsilon n \right\}.$$

For every $k = 1, 2, \ldots, K$, Condition 4 and an elementary large deviation argument [see, for example, Billingsley (1986), Theorem 9.1] on the $n$ independent Bernoulli variables $1\{\forall\ s \in I_k,\ Y_i(s) = 1\}$ show that for $\varepsilon$ sufficiently small

$$P\left(\exists\, s \in I_k,\, \bar{\lambda}(s) < \varepsilon n\right) < Ce^{-\gamma_k n}$$

with $\gamma_k > 0$, and now summing over $k$, that therefore, with $\gamma = \min_k \gamma_k > 0$,

$$P(A_\varepsilon) < Ce^{-\gamma n} \quad \text{with } \gamma > 0.$$

With $C$ not necessarily the same at each occurrence, taking $\Delta' T_j$ over failure times as in (18) we have the bound on the tail probability of $\Delta' T_j$:

$$P\left(\{\Delta' T_j > t\} \cap A_\varepsilon^c\right) \le Ce^{-\varepsilon n t}.$$

Using

$$E\left[\Delta' T_j^2\right] = E\left[\Delta' T_j^2\, 1(A_\varepsilon)\right] + E\left[\Delta' T_j^2\, 1(A_\varepsilon^c)\right]$$

and that times are confined to the unit interval, implying $\Delta' T_j^2 \le 1$, $E[\Delta' T_j^2]$ is

majorized by

$$Ce^{-\gamma n} + \frac{C}{n^2} = O\left(\frac{1}{n^2}\right).$$

As the sum in (17) is over $O(n)$ terms, this proves (17), and so the theorem. □

**6. Three special cases.** We consider three important special cases below. The first case we consider is the limit of the variance $\Gamma$ as $m \to \infty$. The second is the computation of asymptotic relative efficiency (ARE) against using the full cohort when $\beta_0 = 0$. The third case is that of a univariate binary exposure.

6.1. $m \to \infty$. Under our hypotheses the inverse of the full cohort covariance of Andersen and Gill [(1982), condition D, page 1105] may be written

$$\Sigma = \int_0^1 \left\{ \frac{E\left[Y(s)Z(s)^{\otimes 2}\exp(\beta_0' Z(s))\right]}{E\left[Y(s)\exp(\beta_0' Z(s))\right]} - \left\{ \frac{E\left[Y(s)Z(s)\exp(\beta_0' Z(s))\right]}{E\left[Y(s)\exp(\beta_0' Z(s))\right]} \right\}^{\otimes 2} \right\} \times E\left[Y(s)\exp(\beta_0' Z(s))\right]\lambda_0(s)\, ds. \tag{19}$$

Using $E\{Y(s)Z(s)^{\otimes 2}\exp(\beta_0' Z(s))\} = p(s)E\{Z_Y^{\otimes 2}(s)\exp(\beta_0' Z_Y(s))\}$ and similar formulas for all other terms, dropping $s$, the integrand is seen to be equal to

$$\left\{ \frac{E\left[Z_Y^{\otimes 2}\exp(\beta_0' Z_Y)\right]}{E\left[\exp(\beta_0' Z_Y)\right]} - \left\{ \frac{E\left[Z_Y \exp(\beta_0' Z_Y)\right]}{E\left[\exp(\beta_0' Z_Y)\right]} \right\}^{\otimes 2} \right\} pE\left[\exp(\beta_0' Z_Y)\right].$$

To see that the limit as $m \to \infty$ of $\Gamma$ yields the same result, note first that

$$\frac{1}{m}\sum_{i=1}^{m} \exp(\beta_0' Z_{Y,i}) \to_p E\left[\exp(\beta_0' Z_Y)\right] \quad \text{as } m \to \infty,$$

and therefore, using (10) that

$$\sum_{j \in U} Z_{Y,j}^{\otimes 2} p_j = \frac{(1/m)\Sigma_{j=1}^m Z_{Y,j}^{\otimes 2}\exp(\beta_0' Z_{Y,j})}{(1/m)\Sigma_{i=1}^m \exp(\beta_0' Z_{Y,i})} \to_p \frac{E\left[Z_Y^{\otimes 2}\exp(\beta_0' Z_Y)\right]}{E\left[\exp(\beta_0' Z_Y)\right]} \quad \text{as } m \to \infty.$$

A similar computation for the other terms of $\Gamma$ in (11) yields the equivalence of $\Sigma$ and the limiting $\Gamma$.

6.2. $\beta_0 = 0$. When $\beta_0 = 0$, $\exp(\beta_0' Z_{Y,j}(s)) \equiv 1$ so that $\mathrm{Cov}(\hat{Z}(s)|\mathbf{Z}_{Y,U})$ is the uniform finite sampling covariance matrix

$$\frac{1}{m}\sum_{j=1}^{m}\left(Z_{Y,j} - \bar{Z}_Y\right)^{\otimes 2}$$

and

$$\frac{1}{m}\sum_{j=1}^{m}\exp(\beta_0' Z_{Y,j}) = 1.$$

Thus

$$\Gamma = E\int_0^1 p(s)\frac{1}{m}\sum_{j=1}^{m}\left(Z_{Y,j} - \bar{Z}_Y\right)^{\otimes 2}\lambda_0(s)\,ds$$

$$= \frac{m-1}{m}\int_0^1 p(s)V(s)\lambda_0(s)\,ds = \frac{m-1}{m}\Sigma,$$

where $V(s) = \mathrm{Cov}(Z_Y(s))$ and $\Sigma$ is the score variance for the full cohort. Thus, for $1:m-1$ matching, the asymptotic relative efficiency of nested case-control sampling relative to the full cohort for $d = 1$ is $\Sigma/\Gamma = (m-1)/m$, independent of censoring and covariate distributions. Thus, for the Aircraft manufacturing employees study of Garabrant, Held, Langholz and Bernstein (1988), the $1:4$ matching ratio could be expected to have efficiency of about $4/5 = 80\%$ relative to the full cohort when $\beta_0 = 0$. Breslow and Patton (1979) derive this result for binary exposures based on the matched case-control study paradigm. Our approach provides a formal proof and generalizes their result.

6.3. *Univariate binary exposure.*

(*A*) *General expression.* Let $Z_1, Z_2, \ldots, Z_n$ be processes indicating "exposure" or "nonexposure" by taking on the value 1 or 0, respectively. We suppose that the conditions in Section 2 are satisfied, implying that $\hat{\beta}$ is consistent and has an asymptotic normal distribution. With $T(s) = \sum_{k=1}^{m} Z_{Y,k}(s)$,

$$\mathrm{Var}\left(\hat{Z}(s)|\mathbf{Z}_{Y,U}\right) = \frac{T(s)e^{\beta_0}(m - T(s))}{\left(m - T(s) + T(s)e^{\beta_0}\right)^2}.$$

Since $\sum_{k=1}^{m}\exp(Z_{Y,k}\beta_0) = (m - T(s) + T(s)e^{\beta_0})$, we have

$$\Gamma = E\left\{\int_0^1 p(s)\frac{T(s)e^{\beta_0}(m - T(s))}{m\left(m - T(s) + T(s)e^{\beta_0}\right)}\lambda_0(s)\,ds\right\}.$$

Let $P(Z_Y(s) = 1) = \pi(s)$. Noting that $T(s)$ has a binomial$(m, \pi(s))$ distribution and suppressing $s$,

$$\frac{1}{m}E\left\{\frac{Te^{\beta_0}(m-T)}{m-T+Te^{\beta_0}}\right\}$$
$$= \frac{1}{m}\sum_{t=0}^{m}\binom{m}{t}\pi^t(1-\pi)^{m-t}\left\{\frac{te^{\beta_0}(m-t)}{m-t+te^{\beta_0}}\right\}$$
$$= (1-\pi)\sum_{t=1}^{m-1}\binom{m-1}{t}\pi^t(1-\pi)^{m-t-1}\left\{\frac{te^{\beta_0}}{m-t+te^{\beta_0}}\right\}.$$

Thus

$$\Gamma = \int_0^1 p(s)(1-\pi(s))$$
$$\times \sum_{t=1}^{m-1}\binom{m-1}{t}\pi(s)^t(1-\pi(s))^{m-t-1}\left\{\frac{te^{\beta_0}}{m-t+te^{\beta_0}}\right\}\lambda_0(s)\,ds. \tag{20}$$

The full cohort asymptotic information may be calculated using (19), yielding

$$\Sigma = \int_0^1 p(s)\frac{\pi(s)e^{\beta_0}(1-\pi(s))}{1-\pi(s)+\pi(s)e^{\beta_0}}\lambda_0(s)\,ds, \tag{21}$$

from which the asymptotic relative efficiency may be computed.

(*B*) *Constant probability of exposure.* Breslow, Lubin, Marek and Langholz (1983) consider the situation where $\pi(s) \equiv \pi_0$ and, again using the matched case-control study paradigm, derive expressions for the nested case-control sample and full cohort expected information. The nested case-control expression is

$$\frac{1-\pi_0}{1-\pi_0+\pi_0e^{\beta_0}}\sum_{t=0}^{m-1}\binom{m-1}{t}\pi_0^t(1-\pi_0)^{m-t-1}\frac{te^{\beta_0}}{m-t+te^{\beta_0}}. \tag{22}$$

When $\pi(s) \equiv \pi_0$ in (20) and (21), each of these expressions differ from Breslow, Lubin, Marek and Langholz by a factor

$$= \left[1-\pi_0+\pi_0e^{\beta_0}\right]\int_0^1 p(s)\lambda_0(s)\,ds. \tag{23}$$

This is because their expressions give expected information per *failure*, whereas (20) and (21) are per *subject*. Expression (23) is the conversion factor, which gives the expected failures per subject. Of course this factor makes no difference in the AREs so that the curves of Breslow, Lubin, Marek and Langholz (1983) apply.

The condition that $\pi(s)$ is constant over time may be approximately true for rare diseases when censoring does not depend on exposure status and covariates are fixed over time as in the intervention trial design described in the next

section. For many cohorts, even when studying a rare disease, $\pi(s)$ may change drastically over time because of differential censoring patterns for exposed and unexposed individuals or because exposure status may change over time. A simple multiple event situation with constant $\pi(s)$ is when the $N_i$ are homogeneous Poisson processes with intensities $\lambda_0 \exp(\beta_0' Z_i)$, $Z_i$ constant in time (corresponding to exposure either occurring at time zero or not at all) and no censoring $[Y_i(s) \equiv 1]$.

(*C*) *Idealized intervention trial with fixed binary exposure.* In intervention trials with a disease outcome, individuals are randomly assigned to "treated" or "untreated" groups and are followed for a fixed period of time to assess the differential disease rate between the two groups. Prentice (1986) and Self and Prentice (1988) discuss the great potential cost savings of efficient sampling schemes for such cohorts. Although in practice there may be some "loss to follow up," a convenient approximation to this type of cohort is the "idealized intervention trial" in which individuals enter the study of time 0 and (1) are censored after the getting the disease of interest or (2) are on study over the entire study period.

Letting $P(Z(0) = 1) = \pi_0$ and $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$,

$$p(s) = (1 - \pi_0)\exp\{-\Lambda_0(s)\} + \pi_0 \exp\{-\Lambda_0(s)e^{\beta_0}\}$$

and

$$\pi(s) = \pi_0 \exp\{-\Lambda_0(s)e^{\beta_0}\}/p(s),$$

AREs may be calculated by substituting these values into (19) and (20).

The AREs for the idealized intervention trial were calculated for various $\pi_0$, $m$ and $e^{\beta_0}$; representative results are given in Table 1. The AREs with constant $\pi(s)$, as described in Section 6.3(b), are given for comparison. IMSL-PC FORTRAN routine DBINPR was used to calculate the binomial probabilities in (20) and (22) and routine DBDAG was used to evaluate the integrals in (20) and (21). The constant $\pi(s)$ asymptotic relative efficiencies serve as a reasonable approximation to those of the intervention trial for relative risks close to 1 or $\pi_0$ small. However, for $\pi_0 \geq 0.5$ for which all constant $\pi(s)$ AREs are nondecreasing in relative risk, the intervention trial AREs peak and decrease so that, for large relative risks, there is a significant disparity between the two.

## 7. Discussion.

7.1. *Conditions.* The conditions of Section 2 allow for easy interpretation. Indeed, because of our simplified framework, we have been able to provide a necessary condition for the consistent estimation of the parameter vector in Lemma 3.

The restriction to bounded covariates may be weakened to the moment condition (4.2) of Andersen and Gill (1982) at the expense of some technicality such as that of Andersen and Gill (1982) Appendix III.

TABLE 1

*Asymptotic relative efficiencies of nested case-control sampling relative to the full cohort for the idealized intervention trial cohort with exponential failure rates [Section 6.3(c)] and for cohorts with constant probability of exposure over time, and $\lambda_0 = 1$ [Breslow, Lubin, Marek and Langholz (1983) or Section 6.3(b)]*

| Relative risk | Intervention trial | | | $\pi_0(s)$ Constant | | |
|---|---|---|---|---|---|---|
| | $\pi_0 = 0.01$ | 0.5 | 0.9 | $\pi_0 = 0.1$ | 0.5 | 0.9 |
| | | | **1:1 Matching** | | | |
| 1 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| 2 | 0.36 | 0.47 | 0.62 | 0.37 | 0.50 | 0.63 |
| 4 | 0.24 | 0.38 | 0.63 | 0.26 | 0.50 | 0.74 |
| 8 | 0.15 | 0.30 | 0.53 | 0.19 | 0.50 | 0.81 |
| | | | **1:8 Matching** | | | |
| 1 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 |
| 2 | 0.82 | 0.88 | 0.93 | 0.83 | 0.90 | 0.93 |
| 4 | 0.72 | 0.85 | 0.94 | 0.76 | 0.92 | 0.96 |
| 8 | 0.61 | 0.80 | 0.90 | 0.69 | 0.94 | 0.98 |

The restriction imposed by (1) in Condition 4 on the censoring process $Y$ is a weakened version of (2) adopted by Andersen and Gill [(1982), equation 4.3, page 1111] for the independent identically distributed case. We have weakened this condition in order to accommodate examples such as a 5 year follow up of a 50–60 year age cohort; in such a study, no single individual can span the entire interval [50, 60]. However, weakening the condition still further to $\inf_t P(Y(t) = 1) > 0$ may not suffice. For consider the intervals $I_0 = \{1\}$, $I_1 = [0, 1/2)$, $I_2 = [1/2, 3/4), \ldots$. Let $X_0, X_1, \ldots$ be independent Bernoulli($p$), $0 < p < 1$, $p + q = 1$, and suppose

$$Y(s) = \sum_{k=0}^{\infty} \mathbf{1}\{s \in I_k\} X_k \quad \text{and} \quad \bar{Y}(s) = \sum_{i=1}^{n} Y_i(s).$$

Then $P(\forall\, s \in [0, 1],\ \bar{Y}(s) \neq 0) = \prod_{k=0}^{\infty} P(\forall\, s \in I_k\ \bar{Y}(s) \neq 0) = 0$, so that

$$P\left(\exists\, k\colon \bar{Y}(s) = 0\ \forall\, s \in I_k\right) = 1$$

(this will be true for infinitely many $k$, in fact) and the argument in Theorem 3 breaks down despite that $\inf_t P(Y(t) = 1) = p > 0$.

Our results apply in more generality than the independent case alone. For example, if the covariate processes are dependent due to a "common effect" $X_0$, conditioning will obtain independence. For example, if the covariates are of the form $Z_i = f(X_i, X_0)$ where $X_0, X_1, \ldots, X_n$ are independent and $X_1, \ldots, X_n$ are equal in distribution under a measure $Q$, our results apply with respect to the conditioned measure $P(A) = Q(A|X_0)$. For a randomization at time $t = 0$ of the members of a cohort into two groups of equal size according to a vector

of indicators $X_0$, with the different groups receiving different exposure levels or treatments, conditioning on $X_0$ will again leave $Z_i$ independent but, in this instance, no longer indentically distributed. We now outline a set of conditions that may be applied in the independent but nonidentically distributed case.

With $(Z_1, Y_1), (Z_2, Y_2), \ldots, (Z_n, Y_n)$ independent, we adopt Condition 2 on the baseline hazard, and the natural extensions of Conditions 1, and 3; that the processes $M_i = N_i - \Lambda_i$ are $\{\mathcal{F}\}_{t\in[0,a)}$ local martingales, and that the covariate processes are bounded. Condition 4 on the censoring may be relaxed to the existence of a constant $C > 0$ such that $\forall\ i\ \min_k P(\forall\ t \in I_k Y_i(t) = 1) \geq C$. Let $p_i(s) = P(Y_i(s) = 1)$.

To extend Lemma 1 to the nonidentically distributed case, "asymptotic stability" conditions need be imposed. All that is required is control on sums of the form

$$S_n = \frac{1}{n}\sum_{i=1}^{n} w(\tilde{R}_i)Y_i A_i$$

for $w(T)$ as in (4) or (5) with $\rho \in \{0, 1, 2\}$ and $A_i$ either $\exp(\beta_0' Z_i)$ or $Z_i' \exp(\beta_0' Z_i)$ as in Lemma 1.

Since (6) [and other versions of (6) for other choices of $\rho$ and $A_i$] is obtained by conditioning on the covariates, (6) holds as in Lemma 1 with no change. Insisting that the limit of $S_n$ exists for all $s$ in the case $\rho = 0$, $\beta_0 = 0$ (and as usual suppressing $s$ in the notation) is equivalent to the condition

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} p_i = p. \tag{24}$$

It follows that $|R|/n \to_p p$, hence (8) holds with $Q_n$ as in (7) unchanged. With no additional assumptions one has (9), that is, that $\mathrm{Var}\{Q_n\} \to 0$. Hence, all that is further required for asymptotic stability is the appropriate convergence of the means $EQ_n$ for the various choices of $\rho$ and $A_i$.

Some of these conditions on the convergence of the $EQ_n$ are related (in mean) to the Andersen and Gill (1982) conditions on the convergence of asymptotic sums at $\beta = \beta_0$. To see the parallel, for $m \leq n$ let $A = \{\mathbf{k}: \mathbf{k} = (k_1, k_2, \ldots, k_n), |\mathbf{k}| = m, k_i \in \{0, 1\}\}$, where $|\mathbf{k}| = \sum_{i=1}^{n} k_i$. Using the identity

$$\left(\frac{1}{n}\sum_{j=1}^{n} p_j\right)^m - \frac{m!}{n^m}\sum_{|T|=m}\prod_{j\in T} p_j = \frac{1}{n^m}\sum_{\substack{|\mathbf{k}|=m\\ \mathbf{k}\notin A}}\binom{m}{\mathbf{k}}\prod_{i=1}^{n} p_i^{k_i},$$

one can show that (24) implies

$$\frac{m!}{n^m}\sum_{|T|=m}\prod_{j\in T} p_j = \frac{m!}{n^m}\sum_{|T|=m} P(Y_T = 1) \to p^m \quad \text{as } n \to \infty. \tag{25}$$

Now computing $EQ_n$ at $\beta = \beta_0$ with $w(T)$ as in (5) and $A_i = \exp(\beta_0' Z_i)$, say, yields

$$EQ_n = \frac{1}{m} p^{-m+1} \binom{n}{m}^{-1} \sum_{|T|=m} \sum_{i \in T} E\{Y_i Z_i^{\otimes \rho} \exp(\beta_0' Z_i)\} P\{Y_{T-\{i\}} = 1\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\{Y_i Z_i^{\otimes \rho} \exp(\beta_0' Z_i)\} \left\{ \frac{n}{m} p^{-m+1} \binom{n}{m}^{-1} \sum_{T \in P_i} P(Y_{T-\{i\}} = 1) \right\}.$$

Using (25) with $m$ replaced by $m - 1$ shows the second factor converges to 1; hence assuming the convergence of $EQ_n$ at $\beta = \beta_0$ in this case is equivalent to assuming that the averages of means of the form

$$\frac{1}{n} \sum_{i=1}^{n} E\{Y_i Z_i^{\otimes \rho} \exp(\beta_0' Z_i)\}$$

converges as $n \to \infty$, as in Section 3 of Andersen and Gill (1982).

Using the extension of boundedness Condition 3 and the convergence of the sums $S_n$, the Lindeberg condition (16) follows easily. All that remains is to guarantee that the inverse asymptotic covariance matrix is positive definite. Using the hypothesis that the $EQ_n$ converge, letting

$$v(s) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E\left\{ \left( Z_i(s)^{\otimes 2} - \left\{ \frac{\sum_{j \in \tilde{R}_i} Z_j(s) \exp(\beta_0' Z_j(s))}{\sum_{j \in \tilde{R}_i} \exp(\beta_0' Z_j(s))} \right\}^{\otimes 2} \right) \times Y_i(s) \exp(\beta_0' Z_i(s)) \right\},$$

we require

$$\Gamma = \int_0^1 v(s) \lambda_0(s)\, ds$$

be positive definite.

7.2. *Matched case-control studies.* Our results have implications for the analysis of matched case-control studies. After the introduction of Cox's partial likelihood approach to estimation in the proportional hazards model, Prentice and Breslow (1978) and Prentice (1986a) (essentially) showed that the "inversion" of the prospective conditional logistic model used for age matched case-control studies yields the nested case-control partial likelihood $L$, indicating that $\beta$ in the proportional hazards model is the parameter being estimated in time matched case-control studies. Our work gives conditions under which this is true and shows that nested case-control sampling is an appropriate way to view matched case-control studies. It is worthwhile to compare this approach with that often used when developing theory related to the analysis of matched case-control data. Data are viewed as "prospective," treating the covariates as fixed and the disease status as random. The consistency and

asymptotic normality of this estimator is based on standard likelihood theory under the assumption that the matched sets are independent. This is not completely satisfactory when cases and controls are age (or, more generally, time) matched since this situation admits the possibility that individuals may be eligible to serve as controls in multiple matched sets and/or controls in a given matched set and may go on to become cases later on. In that framework, the rare disease assumption must be invoked to circumvent these difficulties.

7.3. *Comparison with case-cohort asymptotics.* It is interesting to compare our approach to assessing the asymptotic behavior of nested case-control sampling to that of Self and Prentice (1988) for case-cohort asymptotics. Let $\tilde{C}$ be the set of indices of those in the sampled subcohort, a random sample from $\{1, 2, \ldots, n\}$. The case-cohort score function as given in Prentice (1986) may be written as our (14) with

$$E_i(s) = \frac{\sum_{j \in \tilde{C} \cup \{i\}} Y_j(s) Z_j(s) \exp\left(\beta_0' Z_j(s)\right)}{\sum_{j \in \tilde{C} \cup \{i\}} Y_j(s) \exp\left(\beta_0' Z_j(s)\right)}.$$

Since $|\tilde{C}|$ is assumed to be stochastically proportional to $n$, the number sampled in each risk set will increase proportionally to $n$. To show consistency, the law of large numbers is used to show the convergence of $E_i(s)$ and $E(s)$ to the same limit so that the case-cohort analog to the second term in (15) converges to zero in probability. This is in contrast to nested case-control sampling for which the size of the sampled risk set is fixed at $m$ and a finite sampling result, Lemma 1, is used to account for the dependence on $i$. Further, the case-cohort analog to the third term in (15) yields the sampling induced covariance $\Delta$ in the Self and Prentice (1988) expression for the variance of the score (their Theorem 3) whereas in nested case-control sampling this term vanishes.

Considering the simplicity of the case-cohort sampling technique, the ARE formulas given by Self and Prentice [(1988), Section 5] are surprisingly complex compared to the nested case-control formulas derived here. Especially noteworthy is the $\beta_0 = 0$ situation. Unlike the very simple nested case-control expression derived in Section 5.2, which depends only on matching ratio, the case-cohort expression depends on covariate, failure and censoring distributions as well as the proportion of the cohort sampled.

7.4. *Cost-efficiency considerations.* Often, efficiency calculations are used to determine the sampling design which will give the best efficiency for a fixed cost. If the main component of cost is per *individual* in the sample, care must be taken that the various designs yield the same expected sample size [Langholz and Thomas (1990)].

In nested case-control sampling individuals may be sampled repeatedly and failures can be picked as controls so that the proportion of the cohort sampled is less than $m$ times the proportion of failures. The proportion used in the

special case of the idealized intervention trial of Section 6.2(c) can be shown to be $1 - (1 - p_f)^m$ where $p_f$ is the overall probability of failure.

7.5. *Further directions.* In this work we have focused on the estimation of $\beta_0$, mostly for the sake of making efficiency calculations for nested case-control sampling. It would also be of interest to consider estimation of the underlying cumulative hazard function as well.

## APPENDIX

Lemmas 2 and 3 may be used to verify the claims made in Section 2; Lemma 4 demonstrates that the matrix $\Gamma$ as defined in (11) is positive definite.

LEMMA 2. *Let* $T = \Lambda(1)$. *For* $0 \le t \le T$, *define*

$$\sigma_t = \inf\{s: \Lambda(s) \ge t\}.$$

*Then the process* $\tilde{N}$, *defined by*

$$\tilde{N}(t) = N(\sigma_t),$$

*is a Poisson process with rate* 1 *for* $t \in [0, T]$. *Consequently, with* $T_1, T_2, \ldots$ *the event times of* $N$, $\Lambda(T_1), \Lambda(T_2), \ldots$ *have the same joint distribution as* $\tilde{T}_1, \tilde{T}_2, \ldots$, *the event times of a Poisson process of rate* 1.

We omit the proof, as it follows easily by a stopping time argument using Theorem 6.2 in Chapter 2 of Ikeda and Watanabe (1981).

We now demonstrate that the condition $V > 0$ is necessary to prove consistency of $\hat{\beta}$.

LEMMA 3. *If* $V$ *is not positive definite, then* $\beta_0$ *is unidentifiable.*

PROOF. If $V$ is not positive definite, then there exists $a \in \mathbf{R}^d$, $a \ne 0$, such that $a'Va = 0$, that is

$$\int_0^1 a' \operatorname{Cov}(Z_Y(s))a\lambda_0(s)\,ds = 0.$$

Hence, with $c(s) = EZ_Y(s)$ we have

$$P(a'Z_Y(s) = a'c(s)) = 1$$

for almost all $s$. Hence, on the event $Y(s) = 1$, for almost all $s$ and all scalars $\alpha \in \mathbf{R}$,

$$\begin{aligned}\lambda(s) &= Y(s)\lambda_0(s)\exp(\beta_0'Z(s)) \\ &= Y(s)\left[\lambda_0(s)\exp(-\alpha a'c(s))\right]\exp((\alpha a + \beta_0)'Z(s))\end{aligned}$$

with probability 1. Of course, the preceding holds trivially on the event $Y(s) = 0$. Since modifying $\lambda$ on a set of measure zero does not change the

distribution of failure times, the models $(\lambda_0 \exp(-\alpha a'c), \beta_0 + \alpha a)$, $\alpha \in \mathbf{R}$, are indistinguishable; hence $\beta_0$ is not identifiable. □

We remark that by using the absolute continuity of the distribution of failure times with respect to Lebesgue measure, Lemma 2 may be used to show, for the Cox model specifically, the nonidentifiability of $\beta_0$ is exhibited in the Cox full and nested case-control likelihoods by $L(\beta) = L(\beta + \alpha a)$; hence if $\hat{\beta}$ maximizes $L(\beta)$, then so does $\hat{\beta} + \alpha a$ for any $\alpha \in \mathbf{R}$.

Recall the definition of $\Gamma$ in (11). The next lemma combined with the previous shows that the condition $V > 0$ is both necessary and, in terms of $\Gamma$, sufficient for the Cox nested case-control likelihood procedure to succeed.

LEMMA 4. *The matrix $\Gamma$ is positive definite whenever $V$ is positive definite.*

PROOF. Suppose that $\Gamma$ is not positive definite. Then there exists an $a \in \mathbf{R}^d$, $a \neq 0$, such that $a'\Gamma a = 0$. Hence, for almost every $s$,

$$E\left[p\frac{1}{m}\sum_{j\in U}\exp(\beta_0' Z_{Y,j})\sum_{i\in U}\left[a'\left(Z_{Y,i} - \bar{Z}\right)\right]^2 p_i\right] = 0$$

and so

$$P\left[\sum_{i\in U}\left\{a'\left(Z_{Y,i} - \bar{Z}\right)\right\}^2 p_i = 0\right] = 1.$$

Since the preceding is a sum of positive terms,

$$P\left[a'\left(Z_{Y,j} - \bar{Z}\right) = 0 = a'\left(Z_{Y,i} - \bar{Z}\right)\right] = 1.$$

Hence,

$$P\left[a'Z_{Y,j} = a'Z_{Y,i}\right] = 1 \quad \text{for almost all } s,$$

contradicting the positive definiteness of $V$. □

**Acknowledgments.** The authors thank Professor Duncan Thomas and the referees for helpful suggestions, many of which have been incorporated into the paper to advantage.

The first author would also like to thank the Statistics Department at the University of British Columbia for hospitality while this paper was undergoing revision.

## REFERENCES

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ANDERSEN, P. K. and BORGAN, O. (1985). Counting processes models for life history data: A review. *Scand. J. Statist.* **12** 97–140.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

BILLINGSLEY, P. (1986). *Probability and Measure*. Wiley, New York.

BOICE, J. D., BLETTNER, M., KLEINERMAN, R. A., ET AL. (1987). Radiation dose and leukemia risk in patients treated for cancer of the cervix. *Journal of the National Cancer Institute* **79** 1295–1311.

BRESLOW, N. E. and DAY, N.E. (1987). *Statistical Methods in Cancer Research*, Volume 2: *The Design and Analysis of Cohort Studies*. Lyon/International Agency for Research on Cancer, Lyon, France.

BRESLOW, N. E. and PATTON, J. (1979). Case-control analysis of cohort studies. In *Energy and Health* (N. Breslow and A. Whittemore, eds.) 226–242. SIAM, Philadelphia.

BRESLOW, N. E., LUBIN, J. H., MAREK, P. and LANGHOLZ, B (1983). Multiplicative models and cohort analysis. *J. Amer. Statist. Assoc.* **78** 1–12.

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

GARABRANT, D. H., HELD, J., LANGHOLZ, B. and BERNSTEIN, L. (1988). Mortality of aircraft manufacturing workers in Southern California. *American Journal of Industrial Medicine* **13** 683–693.

IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam.

LANGHOLZ, B. and THOMAS, D. C. (1990). Nested case control and case-cohort methods of sampling from a cohort: A critical comparison. *American Journal of Epidemiology* **131** 169–176.

LENGLART, E. (1977). Relation de Domination entre deux Processus. *Ann. Inst. H. Poincaré* **13** 171–179.

LIDDEL, F. D. K., MCDONALD, J. C. and THOMAS, D. C. (1977). Methods for cohort analysis: Appraisal by application to asbestos mining, *J. Roy. Statist. Soc. Ser. A* **140** 469–483.

OAKES, D. (1981). Survival times: Aspects of partial likelihood (with discussion). *Internat. Statist. Rev.* **49** 235–264.

PRENTICE, R. L. (1986a). On the design of synthetic case-control studies. *Biometrics* **42** 301–310.

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

PRENTICE, R. L. and BRESLOW, N. E. (1978). Retrospective studies and failure time models. *Biometrika* **65** 153–158.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebeite* **51** 269–286.

SELF, S. G. and PRENTICE, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16** 64–81.

THOMAS, D. C. (1977). Addendum to a paper by F. D. K. Liddel, J. C. McDonald and D. C. Thomas. *J. Roy. Statist. Soc. Ser. A* **140** 483–485.

WHITTEMORE, A. S. and MACMILLAN, A. (1983). Lung cancer mortality among U.S. uranium miners: A reappraisal. *Journal of the National Cancer Institute* **71** 489–499.

YEH, F-S., YU, M. C., MO, C-C., LUO, S., TONG, M. J. and HENDERSON, B. E. (1989). Hepatitis B virus, aflatoxins, and hepatocellular carcinoma in southern Guangxi, China. *Cancer Research* **76** 117–123.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
1042 WEST 36TH PLACE
LOS ANGELES, CALIFORNIA 90089-1113

DEPARTMENT OF PREVENTIVE MEDICINE
UNIVERSITY OF SOUTHERN CALIFORNIA
2025 ZONAL AVENUE
LOS ANGELES, CALIFORNIA 90033-9987