The Annals of Statistics
1988, Vol. 16, No. 1, 64–81

ASYMPTOTIC DISTRIBUTION THEORY AND EFFICIENCY RESULTS FOR CASE–COHORT STUDIES[1]

BY STEVEN G. SELF AND ROSS L. PRENTICE

Fred Hutchinson Cancer Research Center and University of Washington

A case–cohort design was recently proposed [Prentice (1986)] as a means of reducing cost in large epidemiologic cohort studies. A "pseudolikelihood" procedure was described for relative risk regression parameter estimation. This procedure involves covariate data only on subjects who develop disease and on a random subset of the entire cohort. In contrast, the usual partial likelihood estimation procedure requires covariate histories on the entire cohort. Accordingly, a case–cohort design may affect cost saving, particularly with large cohorts and infrequent disease occurrence. Asymptotic distribution theory for such pseudolikelihood estimators, along with that for corresponding cumulative failure rate estimators, are presented here. Certain asymptotic efficiency expressions relative to full-cohort estimators are developed and tabulated in situations of relevance to the design of large-scale disease prevention trials. The theoretical developments make use of martingale convergence results and finite population convergence results.

1. Introduction. Epidemiologic studies and disease prevention trials often involve the follow-up of a large cohort of subjects, a small fraction of whom will develop the disease endpoint, or endpoints, of interest during a prescribed follow-up period. Study objectives typically involve estimation of the relationship between disease rates and individual exposures, characteristics or randomization assignments. In fact, the assembly of the histories of such "covariates" will often constitute a major study cost. Consequently, there has been recent interest in designs that involve collection of the raw covariate data on all study subjects, but the processing of such only on some sampling basis. For example, in a study of nutrient intakes in relation to the incidence of selected chronic diseases, one may obtain diet records and blood samples on all cohort members, perhaps several times during cohort follow-up. The diet records may simply be stored with detailed hand coding and nutrient intake analysis taking place on a sample basis. Similarly, blood serum samples may be frozen at an appropriate temperature with expensive biochemical analysis to determine nutrient levels to take place on a sample basis.

The relative risk, or relative risk process, is defined as the ratio of the (instantaneous) disease rate given a general covariate history to that given some standard history. It provides a natural approach to the modeling and understanding of the dependence of disease rates on aspects of the preceding covariate history. In the presence of a large cohort with infrequent disease events, the efficiency with which relative risk parameters may be estimated depends strongly

Received July 1986; revised May 1987.
[1]This work was supported by NIH Grants GM-24472 and CA-34847.
AMS 1980 *subject classifications*. Primary 62E20; secondary 62G05, 60G15.
Key words and phrases. Case–cohort design, counting process, Cox regression, pseudolikelihood, relative risk regression, time-dependent covariates.

on the number of subjects experiencing failure, but the marginal contribution from subjects not developing disease is small. In considering covariate sampling procedures, it is then natural to consider designs on which covariate histories are assembled for all subjects developing disease, the cases in epidemiologic terminology, along with a subset of the subjects not developing disease. One approach to doing so involves the selection of an independent random sample (with replacement) of subjects at risk but without disease (i.e., control subjects) at each distinct failure time [e.g., Liddell, McDonald and Thomas (1977) and Breslow, Lubin, Marek and Langholtz (1983)], which gives rise to a partial likelihood approach to relative risk regression parameter estimation [Oakes (1981)]. This approach leads to poorer efficiency results [Whittemore and McMillan (1982)] than does the odds ratio estimator in corresponding situations but under simple case–control sampling with unmatched controls (and binary response rather than time to response). Suboptimal efficiency may result from the rather artificial usage of a selected control subject only in relation to the subjects time-matched case(s), even though the subject may quite suitably serve as a member of the "control" group at other time points as well.

The case–cohort design, recently described by Prentice (1986), avoids this problem by selecting a subcohort randomly from the entire cohort, which then provides a comparison group at all disease occurrence times. This design also allows the comparison group to be selected in advance of cohort follow-up, a distinct advantage in the prevention trial context since the subcohort can be used for monitoring the achievement of intervention goals and for other purposes. Also, in contrast to the control sample in a time-matched case–control design, the subcohort provides a natural comparison group for a range of disease endpoints.

Prentice (1986) proposed a "pseudolikelihood" procedure for the relative risk parameter along with heuristic procedures for parameter estimation. A corresponding estimator was also given for the cumulative baseline failure rate, for which no estimation procedure currently exists under time-matched case–control sampling. It was also noted that the martingale convergence results, which proved so advantageous in developing the asymptotic properties of the maximum partial likelihood estimator and related quantities under full-cohort sampling [Andersen and Gill (1982) and Prentice and Self (1983)], were not sufficient here since certain generating σ-algebras were not nested.

This paper develops asymptotic distribution theory for the case–cohort maximum pseudolikelihood estimator and related quantities using a combination of martingale and finite population convergence results. Corresponding asymptotic efficiency expressions are developed for relative risk parameter estimation. Efficiency calculations are provided for certain cohort configurations and subcohort sampling fractions.

2. Case–cohort estimators. Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a complete probability space, and let $\{\mathscr{F}_t, t \in [0,1]\}$ be an increasing right-continuous family of sub σ-algebras of $\mathscr{F}$ defined so that $\mathscr{F}_t$ consists of failure (disease) time and covariate histories up to time t and censoring histories to time t^+, for all subjects in a cohort C of

size n. All processes discussed in this work are adapted to $\{\mathcal{F}_t,\ t \in [0,1]\}$. More explicitly, let $N = (N_1, \ldots, N_n)$ be the multivariate counting process defined so that N_i counts failures on the ith subject at times $t \in [0,1]$. Assume each N_i to have totally inaccessible jump times and $N_i(1)$ to be almost surely finite. The censoring process $Y = (Y_1, \ldots, Y_n)$, with left-continuous sample paths, is defined so that $Y_i(t) = 1$ if the ith subject is "at risk" for observable failure at time t and $Y_i(t) = 0$ otherwise.

Each counting process N_i can be uniquely decomposed into the sum of its cumulative intensity process Λ_i and a local square integrable martingale M_i, so that

$$\begin{aligned} N_i(t) &= \Lambda_i(t) + M_i(t) \\ &= \int_0^t \lambda_i(u)\, du + M_i(t), \quad \text{all } t, \end{aligned}$$

assuming Λ_i to be absolutely continuous. Under the usual independent failure time and independent censorship assumptions, the intensity process λ_i possesses a standard hazard process interpretation [e.g., Self and Prentice (1982)]. A relative risk model [Cox (1972), Andersen and Gill (1982) and Prentice and Self (1983)] for the hazard process then yields

$$\lambda_i(t) = Y_i(t)\lambda_0(t) r\{\beta_0' Z_i(t)\}, \quad \text{all } (i,t), \tag{2.1}$$

where λ_0 and r are fixed functions, $\beta_0' = (\beta_{01}, \ldots, \beta_{0q})$ is a relative risk parameter to be estimated and $Z_i'(t) = \{Z_{i1}(t), \ldots, Z_{iq}(t)\}$, with sample paths that are left-continuous with right-hand limits, is a regression vector consisting of data-analyst-defined functions of $\mathcal{F}_t$.

The maximum partial likelihood estimator $\hat{\beta}$ of β is defined as a solution of $\partial \log L(\beta, 1)/\partial\beta = 0$, where

$$\begin{aligned} \log L(\beta, t) = &\sum_{i \in C} \int_0^t \log r\{\beta' Z_i(u)\}\, dN_i(u) \\ &- \int_0^t \log\left[\sum_{l \in C} Y_l(u) r\{\beta' Z_l(u)\}\right] d\bar{N}(u), \end{aligned} \tag{2.2}$$

with $\bar{N} = N_1 + \cdots + N_n$.

Here we consider the properties of an "estimator" $\tilde{\beta}$, defined as a solution to $\partial \log \tilde{L}(\beta, 1)/\partial\beta = 0$, where

$$\begin{aligned} \log \tilde{L}(\beta, t) = &\sum_{i \in C} \int_0^t \log r\{\beta' Z_i(u)\}\, dN_i(u) \\ &- \int_0^t \log\left[\sum_{l \in \tilde{C}} Y_l(u) r\{\beta' Z_l(u)\}\right] d\bar{N}(u), \end{aligned} \tag{2.3}$$

where $\tilde{C}$ is a subset of size $\tilde{n} \leq n$ that is randomly selected from the entire cohort C. Note that the definition of $\tilde{\beta}$ involves covariate data only on subjects that fail ($N_i(1) > 0$) and on members of the subcohort $\tilde{C}$. Note also that calculation of $\tilde{\beta}$ requires a system for identifying the counting process jump times in the entire cohort C, but it does not require a cohort roster or even an

enumeration of C. As will be shown, $\tilde{\beta}$ defined previously differs slightly from, but is quite generally asymptotically equivalent to, the maximum pseudolikelihood estimator defined in Prentice (1986).

Corresponding to the maximum partial likelihood estimator $\hat{\beta}$ is a natural estimator

$$\hat{\Lambda}(t) = \int_0^t \left[\sum_{i \in C} Y_i(u) r\{\hat{\beta}' Z_i(u)\} \right]^{-1} d\bar{N}(u) \tag{2.4}$$

of the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$. Here we consider the asymptotic properties instead of

$$\tilde{\Lambda}(t) = \tilde{n} n^{-1} \int_0^t \left[\sum_{i \in \tilde{C}} Y_i(u) r\{\tilde{\beta}' Z_i(u)\} \right]^{-1} d\bar{N}(u), \tag{2.5}$$

which again involves covariate data only on subcohort members and cases. Note, however, that calculation of $\tilde{\Lambda}$ does require the cohort size n.

3. Asymptotic distribution theory for $\tilde{\beta}$ and $\tilde{\Lambda}$. The estimators defined previously can be shown to have asymptotic normal distributions under quite unrestrictive conditions. Naturally, the necessary conditions for the desired asymptotic convergence of $(\hat{\beta}, \hat{\Lambda})$ are also necessary for $(\tilde{\beta}, \tilde{\Lambda})$. Some additional conditions are also required to ensure a sufficiently rapid convergence of certain subcohort averages to their full-cohort counterparts. In order to display these conditions, the notation of Prentice and Self (1983), hereafter referred to as PS, will be extended as follows: Denote

$$S^{(j)}(\beta, t) = n^{-1} \sum_{l \in C} Y_l(t) x_l^{\{j\}}(\beta, t), \qquad \tilde{S}^{(j)}(\beta, t) = \tilde{n}^{-1} \sum_{l \in \tilde{C}} Y_l(t) x_l^{\{j\}}(\beta, t),$$

$$j = 0, \ldots, 6,$$

where

$$x_l^{(0)}(\beta, t) = r\{\beta' Z_l(t)\},$$

$$x_l^{(1)}(\beta, t) = Z_l(t) r^{(1)}\{\beta' Z_l(t)\},$$

$$x_l^{(2)}(\beta, t) = Z_l(t)^{\otimes 2} u^{(1)}\{\beta' Z_l(t)\}^2 r\{\beta' Z_l(t)\},$$

$$x_l^{(3)}(\beta, t) = Z_l(t)^{\otimes 2} r^{(2)}\{\beta' Z_l(t)\},$$

$$x_l^{(4)}(\beta, t) = [u\{\beta' Z_l(t)\} - u\{\beta_0' Z_l(t)\}] r\{\beta_0' Z_l(t)\},$$

$$x_l^{(5)}(\beta, t) = Z_l(t) u^{(1)}\{\beta' Z_l(t)\} r\{\beta_0' Z_l(t)\},$$

and

$$x_l^{(6)}(\beta, t) = Z_l(t)^{\otimes 2} u^{(2)}\{\beta' Z_l(t)\} r\{\beta_0' Z_l(t)\}.$$

Also define

$$E(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t), \qquad \tilde{E}(\beta, t) = \tilde{S}^{(1)}(\beta, t)/\tilde{S}^{(0)}(\beta, t)$$

and

$$V(\beta, t) = S^{(2)}(\beta, t)/S^{(0)}(\beta, t) - E(\beta, t)^{\otimes 2},$$

$$\tilde{V}(\beta, t) = \tilde{S}^{(2)}(\beta, t)/\tilde{S}^{(0)}(\beta, t) - \tilde{E}(\beta, t)^{\otimes 2}.$$

In these expressions $a^{\otimes 2}$ denotes the $q \times q$ matrix with (i, j) element $a_i a_j$ for any $a' = (a_1, \ldots, a_q)$, $r^{(1)}(x) = dr(x)/dx$, $r^{(2)}(x) = dr^{(1)}(x)/dx$, $u(x) = \log r(x)$, $u^{(1)}(x) = du(x)/dx$ and $u^{(2)}(x) = du^{(1)}(x)/dx$.

Conditions A–F that follow were shown in Prentice and Self (1983) to ensure the asymptotic convergence of $\hat{\beta}$ and $\hat{\Lambda}$:

A (Finite interval). $\int_0^1 \lambda_0(t)\, dt < \infty$.

B (Asymptotic stability). There exists a neighborhood $\mathscr{B}$ of β_0 and functions $s^{(0)}, \ldots, s^{(6)}$ defined on $\mathscr{B} \times [0, 1]$ such that

$$\sup_{\substack{t \in [0,1] \\ \beta \in \mathscr{B}}} \|S^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \to_P 0, \qquad j = 0, \ldots, 6.$$

C (Alternate Lindeberg condition). $n^{-1/2} \sup_{i,t} \|Z_i(t) u^{(1)}\{\beta_0' Z_i(t)\}\| \to_P 0$.

D (Asymptotic regularity conditions). $s^{(j)}(\cdot, t)$ are continuous functions of $\beta \in \mathscr{B}$, uniformly in $t \in [0, 1]$ and $s^{(j)}$ are bounded on $\mathscr{B} \times [0, 1]$, for $j = 0, \ldots, 6$. Also $s^{(0)}$ is bounded away from zero and the matrix

$$\Sigma = \int_0^1 v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t)\, dt$$

is positive definite, where $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$ and $e = s^{(1)}/s^{(0)}$. Finally, $s^{(0)}(\beta, t)$ and $s^{(4)}(\beta, t)$ are assumed to be twice differentiable with respect to β on $\mathscr{B} \times [0, 1]$.

E (Asymptotic stability of observed information matrix).

$$\sup_{\beta \in \mathscr{B}} \int_0^1 n^{-2} \sum_{l \in C} \|Z_l(t)\|^4 u^{(2)}\{\beta' Z_l(t)\}^2 Y_l(t) r\{\beta_0' Z_l(t)\} \lambda_0(t)\, dt \to_P 0.$$

F (Regression function positivity). $r\{\beta' Z_i\}$ is locally bounded away from zero on $\mathscr{B}$ for all $i = 1, \ldots, n$.

Additional conditions are also required to ensure the desired asymptotic behavior of certain subcohort averages:

G (Stability of subcohort averages).

(i) (Nontrivial subcohort). $\tilde{n} n^{-1} \to_P \alpha$ for some $\alpha \in (0, 1)$.

(ii) (Asymptotic normality of subcohort averages at β_0). For $\varepsilon > 0$

$$\sup_t n^{-1} \sum_{l \in C} Y_l(t) r\{\beta_0' Z_l(t)\}^2 I_{\{n^{-1/2} Y_l(t) r\{\beta_0' Z_l(t)\} > \varepsilon\}} \to_P 0,$$

$$\sup_t n^{-1} \sum_{l \in C} Y_l(t) \|Z_l(t)\|^2 r\{\beta_0' Z_l(t)\}^2 I_{\{n^{-1/2} Y_l(t) \|Z_l(t)\| r^{(1)}\{\beta_0' Z_l(t)\} > \varepsilon\}} \to_P 0,$$

and the sequence of distributions of $n^{1/2}\{\tilde{E}(\beta_0, t) - E(\beta_0, t)\}$ is tight on the product space of left-continuous functions with right-hand limits equipped with the product Skorohod topology.

(iii) (Asymptotic stability and regularity of covariance function). There exists a neighborhood $\mathscr{B}$ of β_0 and functions $q^{(j)}(\beta, t, w)$, $j = 0, 1, 2$, defined on $\mathscr{B} \times [0,1]^2$ such that $q^{(j)}(\cdot, t, w)$ are continuous functions of $\beta \in \mathscr{B}$ uniformly in $(t, w) \in [0,1]^2$, the $q^{(j)}$ are bounded on $\mathscr{B} \times [0,1]^2$ and

$$\sup_{\substack{(t,w)\in[0,1]^2 \\ \beta\in\mathscr{B}}} \|Q^{(j)}(\beta, t, w) - q^{(j)}(\beta, t, w)\| \to_P 0, \qquad j = 0, 1, 2,$$

where

$$Q^{(0)}(\beta, t, w) = n^{-1} \sum_{l\in C} Y_l(t) Y_l(w) x_l^{(0)}(\beta, t) x_l^{(0)}(\beta, w),$$

$$Q^{(1)}(\beta, t, w) = n^{-1} \sum_{l\in C} Y_l(t) Y_l(w) x_l^{(1)}(\beta, t) x_l^{(1)}(\beta, w)^T$$

and

$$Q^{(2)}(\beta, t, w) = n^{-1} \sum_{l\in C} Y_l(t) Y_l(w) x_l^{(0)}(\beta, t) x_l^{(1)}(\beta, w).$$

Moreover, $\sup_{n\geq 1} E[Q^{(j)}(\beta, t, w)]$, $j = 0, 1, 2$, are bounded sequences.

(iv) (Asymptotic stability of subcohort averages).

$$\sup_{\substack{t\in[0,1] \\ \beta\in\mathscr{B}}} \|\tilde{S}^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \to_P 0, \qquad j = 0, \ldots, 3,$$

and

$$\sup_{\substack{(t,w)\in[0,1]^2 \\ \beta\in\mathscr{B}}} \|\tilde{Q}^{(j)}(\beta, t, w) - q^{(j)}(\beta, t, w)\| \to_P 0, \qquad j = 0, 1, 2.$$

Note that $\mathscr{B}$ appearing in conditions B, F and G should be regarded as the intersection of β_0-neighborhoods for which these three conditions are individually satisfied. In these conditions $\|\cdot\|$ refers to the Euclidean norm and convergence properties involve $n \to \infty$.

Conditions G(i)–G(iii) are required for the application of a central limit theorem for random sampling without replacement from a finite population to the processes $\tilde{S}^{(0)}$ and $\tilde{S}^{(1)}$. In particular, G(iii) ensures the convergence of the finite sample covariance functions to that of the limiting distribution. Condition G(iv) is required for the consistency of $\tilde{\beta}$ and the consistency of estimators of the variance of $\tilde{\beta}$.

Now let us prove

LEMMA 3.1 (Consistency of $\tilde{\beta}$). $\tilde{\beta} \to_P \beta_0$.

PROOF. As in PS set

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

$$A(\beta, t) = \int_0^t \left[S^{(4)}(\beta, w) - \log\{S^{(0)}(\beta, w)/S^{(0)}(\beta_0, w)\} S^{(0)}(\beta_0, w)\right] \lambda_0(w)\, dw.$$

Also define

$$\tilde{X}(\beta, t) = n^{-1}\{\log \tilde{L}(\beta, t) - \log \tilde{L}(\beta_0, t)\}.$$

Then

$$\begin{aligned}\tilde{X}(\beta, t) - A(\beta, t) = \{X(\beta, t) - A(\beta, t)\} \\ -n^{-1}\int_0^t \{\log \tilde{S}^{(0)}(\beta, w) - \log S^{(0)}(\beta, w) \\ -\log \tilde{S}^{(0)}(\beta_0, w) + \log S^{(0)}(\beta_0, w)\}\, d\bar{N}(w).\end{aligned}$$

As shown in PS, $X(\beta, t) - A(\beta, t)$ converges in probability to zero under conditions A–F. The integral in the preceding expression can be written as

$$\begin{aligned}n^{-1}\int_0^t \{\log \tilde{S}^{(0)}(\beta, w) - \log S^{(0)}(\beta, w) - \log \tilde{S}^{(0)}(\beta_0, w) \\ +\log S^{(0)}(\beta_0, w)\}\, d\bar{M}(w) \\ +\int_0^t \{\log \tilde{S}^{(0)}(\beta, w) - \log S^{(0)}(\beta, w) \\ -\log \tilde{S}^{(0)}(\beta_0, w) + \log S^{(0)}(\beta_0, w)\} S^{(0)}(\beta_0, w)\lambda_0(w)\, dw.\end{aligned}$$

The second of these two integrals converges in probability to zero in view of conditions A and B, regularity conditions D and G(iv). By virtue of the continuity of $r(\cdot)$ and the left continuity of $Z(\cdot)$, the integrand of the first integral is predictable and locally bounded for each $\beta \in \mathscr{B}$, so that this integral is itself a locally square integrable martingale with variance process

$$\begin{aligned}\mathscr{B}(\beta, t) = n^{-1}\int_0^t \{\log \tilde{S}^{(0)}(\beta, w) - \log S^{(0)}(\beta, w) - \log \tilde{S}^{(0)}(\beta_0, w) \\ +\log S^{(0)}(\beta_0, w)\}^2 S^{(0)}(\beta, w)\lambda_0(w)\, dw,\end{aligned}$$

which again converges in probability to zero on the basis of conditions A, B, D and G(iv). It follows that $\tilde{X}(\beta, t)$ converges in probability to the same limit as does $A(\beta, t)$ for each $\beta \in \mathscr{B}$. $A(\beta, t)$ was shown by PS to converge in probability to a concave function with unique maximum at $\beta = \beta_0$, hence the argument of Andersen and Gill (1982), Appendix 2, shows $\tilde{\beta} \to_P \beta_0$. □

The "score" process corresponding to (2.3) can be written as

$$\begin{aligned}n^{-1/2}\tilde{U}(\beta_0, t) &= n^{-1/2}\,\partial \log \tilde{L}(\beta_0, t)/\partial\beta_0 \\ &= \int_0^t n^{-1/2} \sum_{l\in C} \left[Z_l(w)u_l^{(1)}(w) - \tilde{E}(w)\right] dN_l(w) \\ &= \int_0^t n^{-1/2} \sum_{l\in C} \left[Z_l(w)u_l^{(1)}(w) - E(w)\right] dM_l(w) \\ &\quad - \int_0^t n^{-1/2}\{\tilde{E}(w) - E(w)\}\, d\bar{\Lambda}(w) \\ &\quad - \int_0^t n^{-1/2} \sum_{l\in C} \left[\tilde{E}(w) - E(w)\right] dM_l(w),\end{aligned} \tag{3.1}$$

where $\bar{\Lambda} = \Lambda_1 + \cdots + \Lambda_n$ and the notation $u_i^{(1)}(w) = u^{(1)}\{\beta_0' Z_i(w)\}$, $\tilde{E}(w) =$

$\tilde{E}(\beta_0, w)$ along with other similar notation, is used here and subsequently. The third term on the right-hand side of (3.1) is a local square integrable martingale with respect to the filtration $\{\tilde{\mathcal{F}}(t),\ t \in [0,1]\}$, where $\tilde{\mathcal{F}}(t) = \mathcal{F}(t) \vee \sigma(\tilde{C})$, with $\sigma(\tilde{C})$ denoting the σ-algebra of possible subcohort selections. This martingale has covariation process

$$\int_0^t \{\tilde{E}(w) - E(w)\}^{\otimes 2} S^{(0)}(w)\lambda_0(w)\, dw,$$

which converges to zero by the convergence properties implied by conditions B and G(iv). Hence the third term in (3.1) converges in probability to zero. The first term is the score process for the full-cohort analysis and is a local square integrable martingale with respect to the filtration $\{\mathcal{F}(t),\ t \in [0,1]\}$. As shown in PS, this term converges in distribution to a Gaussian process. The second term involves the difference between subcohort and full-cohort means. When viewed conditionally on $\mathcal{F}(1)$, the randomness in this term arises solely from the choice of the subcohort $\tilde{C}$. The following simple proposition will be used to show that the first and second terms of (3.1) converge jointly to independent Gaussian random variables. The proof of this proposition is given in the Appendix.

PROPOSITION 1. *Let* $\mathbf{X}_n = (X_{1n}, \ldots, X_{nn})$ *and* $\boldsymbol{\delta}_n = (\delta_{1n}, \ldots, \delta_{nn})$ *be independent random variables such that:*

(1) $\boldsymbol{\delta}_n$ *is a vector of* $\tilde{n}$ *ones and* $n - \tilde{n}$ *zeros, each possible configuration of zeros and ones is equally likely and* $\tilde{n}n^{-1} \to \alpha \in (0,1)$.

(2) *For some scalar functions of* $\mathbf{X}_n$, $f_{in}(\mathbf{X}_n)$, *and for any* $\varepsilon > 0$,

$$n^{-1} \sum_{i=1}^{n} [\, f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I_{\{|f_{in}(\mathbf{X}_n) - f_{\cdot n}(X_n)| > n^{1/2}\varepsilon\}} \to_P 0$$

and $S_{fn}^2 \to_P \sigma_f^2 > 0$, *where* $f_{\cdot n}(\mathbf{X}_n) = n^{-1}\sum_{i=1}^n f_{in}(\mathbf{X}_n)$ *and*

$$S_{fn}^2 = n^{-1} \sum_{i=1}^{n} [\, f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2.$$

(3) *The scalar functions of* $\mathbf{X}_n$, $g_n(\mathbf{X}_n)$, *converge in distribution to a Gaussian random variable with mean zero and variance* σ_g^2.

Then for $h_n(\mathbf{X}_n, \boldsymbol{\delta}_n) = n^{1/2}[\tilde{n}^{-1}\sum_{i=1}^n \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]$, $\{g_n(X_n), h_n(X_n, \boldsymbol{\delta}_n)\}$ *converge in distribution to a bivariate Gaussian random variable with mean zero and covariance matrix given by*

$$\begin{bmatrix} \sigma_g^2 & 0 \\ 0 & (1-\alpha)\alpha^{-1}\sigma_f^2 \end{bmatrix}.$$

We are now in a position to prove

THEOREM 3.1 (Asymptotic normality of the score statistic).

$$n^{-1/2}\tilde{U}(\beta_0, 1) \to_D N(0, \Sigma + \Delta)$$

with

$$\Delta = \int_0^1 \int_0^1 G(\beta_0, x, w) S^{(0)}(x) S^{(0)}(w) \lambda_0(x) \lambda_0(w)\, dx\, dw,$$

$$\begin{aligned} G(\beta_0, x, w) = (1-\alpha)\alpha^{-1}\Big[& \{s^{(0)}(x)s^{(0)}(w)\}^{-1} h^{(1)}(\beta_0, x, w) \\ & + \{s^{(0)}(x)s^{(0)}(w)\}^{-2} s^{(1)}(x) s^{(1)}(w)^T h^{(0)}(\beta_0, x, w) \\ & - s^{(0)}(x)^{-1} s^{(0)}(w)^{-2} s^{(1)}(w) h^{(2)}(\beta_0, w, x) \\ & - s^{(0)}(w)^{-1} s^{(0)}(x)^{-2} s^{(1)}(x) h^{(2)}(\beta_0, x, w)\Big], \end{aligned}$$

where $h^{(j)}(\beta_0, x, w)$ *are given by*

$$h^{(0)}(\beta, x, w) = q^{(0)}(\beta, x, w) - s^{(0)}(\beta, x) s^{(0)}(\beta, w),$$

$$h^{(1)}(\beta, x, w) = q^{(1)}(\beta, x, w) - s^{(1)}(\beta, x) s^{(1)}(\beta, w)^T,$$

and

$$h^{(2)}(\beta, x, w) = q^{(2)}(\beta, x, w) - s^{(0)}(\beta, x) s^{(1)}(\beta, w).$$

PROOF. Straightforward algebra and conditions B and G(iv) give

$$\begin{aligned} n^{1/2}\{\tilde{E}(t) - E(t)\} &= -n^{1/2}\{\tilde{S}^{(0)}(t) - S^{(0)}(t)\} E(t) \\ &\quad + n^{1/2}\{\tilde{S}^{(1)}(t) - S^{(1)}(t)\} \tilde{S}^{(0)}(t)^{-1} \\ &= -n^{1/2}\{\tilde{S}^{(0)}(t) - S^{(0)}(t)\}\, e(t) \\ &\quad + n^{1/2}\{\tilde{S}^{(1)}(t) - S^{(1)}(t)\} s^{(0)}(t)^{-1} + o_p(1). \end{aligned}$$

Now consider application of Proposition 1, where X_{in} represents $\{Y_i(u), N_i(u), Z_i(u); 0 \le u \le 1\}$, $f_{in}(\mathbf{X}_n)$ represents a linear combination of elements of $Y_i(t) r\{\beta_0' Z_i(t)\}$ and $Y_i(t) Z_i(t) r^{(1)}\{\beta_0' Z_i(t)\}$ and $g_n(\mathbf{X}_n)$ represents linear combinations of elements of the full-cohort score process all evaluated at a finite number of fixed time points in $[0, 1]$. Condition (1) of Proposition 1 follows from G(i) and the fact that the subcohort is selected using simple random sampling without replacement. Condition (2) is easily shown to follow from G(ii) by repeated application of the inequality used by Andersen and Gill

$$|a-b|^2 I_{\{|a-b|>\varepsilon\}} \le 4|a|^2 I_{\{|a|>\varepsilon/2\}} + 4|b|^2 I_{\{|b|>\varepsilon/2\}}$$

and the convergence properties implied by G(iii) and G(iv). Condition (3) follows from the convergence of the full-cohort score process to a continuous Gaussian process with limiting covariance process given by

$$\int_0^t \{s^{(2)}(w)/s^{(0)}(w) - e(w)^{\otimes 2}\} s^{(0)}(w) \lambda_0(w)\, dw,$$

which equals $\hat{\Sigma}$ at $t = 1$. Now using G(iv) and the Cramér–Wold device as in Aalen (1977), we have convergence of the finite-dimensional distributions of the full score process and $n^{1/2}\{\tilde{E}(t) - E(t)\}$ to independent Gaussian distributions. The tightness condition G(ii) implies weak convergence of these processes. The

covariance function of the limiting process for $n^{1/2}\{\tilde{E}(t) - E(t)\}$ is shown to be $G(\beta_0, x, w)$ by a straightforward but tedious calculation of the covariance process conditional on $\mathscr{F}(1)$, the use of the convergence properties G(iii) and G(iv) and the equivalence of the covariance of the limiting process and the limiting finite sample covariance implied by G(iii). The fact that $\bar{\Lambda}$ is absolutely continuous with respect to Lebesgue measure and that linear functionals of the Gaussian processes are Gaussian implies that the first two terms in expression (3.1) converge to a mean zero Gaussian random variable with covariance $\not{\Sigma} + \Delta$. □

The asymptotic normality of $\tilde{\beta}$ now follows from Theorem 3.1 and the usual Taylor series expansions.

THEOREM 3.2 (Asymptotic normality of $\tilde{\beta}$).

$$n^{1/2}(\tilde{\beta} - \beta_0) \to_D N(0, \not{\Sigma}^{-1} + \not{\Sigma}^{-1}\Delta\not{\Sigma}^{-1}).$$

PROOF. A Taylor expansion of $\partial \log \tilde{L}(\beta, 1)/\partial\beta$ about β_0, evaluated at $\tilde{\beta}$, gives

$$n^{-1/2}\,\partial \log \tilde{L}(\beta_0, 1)/\partial\beta_0 = \{-n^{-1}\,\partial^2 \log \tilde{L}(\beta_*)/\partial\beta_*^2\} n^{1/2}(\tilde{\beta} - \beta_0),$$

where β_* is between $\tilde{\beta}$ and β_0, whence it is sufficient to show that

$$n^{-1}\,\partial^2 \log \tilde{L}(\beta_*)/\partial\beta_*^2 \to_P \Sigma,$$

for any random β_* such that $\beta_* \to_P \beta_0$.

As in PS define, for any $\beta \in \mathscr{B}$,

$$K(\beta, t) = \int_0^t n^{-1} \sum_{i=1}^{n} \left[X(\beta, w) - Z_i(w)^{\otimes 2} u^{(2)}\{\beta' Z_i(w)\} \right] dN_i(w),$$

where

$$X(\beta, w) = S^{(3)}(\beta, w)/S^{(0)}(\beta, w) - S^{(1)}(\beta, w)^{\otimes 2}/S^{(0)}(\beta, w)^2.$$

Also let $\tilde{X}(\beta, w)$ be similarly defined in terms of replacing each $S^{(j)}$ by $\tilde{S}^{(j)}$. One can then write

$$\begin{aligned} -n^{-1}\,\partial^2 \log \tilde{L}(\beta, t)/\partial\beta^2 - K(\beta, t) \\ = \{n^{-1}\,\partial^2 \log L(\beta, t)/\partial\beta^2 - K(\beta, t)\} \\ + n^{-1} \int_0^t \{\tilde{X}(\beta, w) - X(\beta, w)\}\, d\bar{N}(w). \end{aligned}$$

The expression on the left-hand side was shown to converge in probability to zero under conditions A–F in PS, whereas that on the right-hand side can be shown to converge to zero by G(iv) and the fact that $n^{-1}\bar{N}(1)$ is bounded in probability. It follows that $-n^{-1}\,\partial^2 \log \tilde{L}(\beta, 1)/\partial\beta^2$ and $K(\beta, 1)$ converge in probability to the same matrix for any $\beta \in \mathscr{B}$, and hence, since $\tilde{\beta} \to_P \beta_0$, that

$$-n^{-1}\,\partial^2 \log \tilde{L}(\beta_*, 1)/\partial\beta_*^2 \to_P \not{\Sigma},$$

for any random β_* between $\tilde{\beta}$ and β_0. □

Application of Theorem 3.2 requires consistent estimators of $\mathcal{I}$ and Δ. The preceding shows $-n^{-1}\,\partial^2 \log \tilde{L}(\tilde{\beta}, 1)/\partial\tilde{\beta}^2$ to be a consistent estimator of $\mathcal{I}$. Alternatively, and presumably preferably, one may consider the positive semidefinite matrix $\tilde{\mathcal{I}}(\tilde{\beta})$ as estimator of $\mathcal{I}$, where

$$\tilde{\mathcal{I}}(\beta) = n^{-1}\int_0^1 \tilde{V}(\beta, t)\, d\bar{N}(t).$$

A simple argument of the type just given for $-n^{-1}\,\partial^2 \log \tilde{L}(\tilde{\beta}, 1)/\partial\tilde{\beta}^2$ in conjunction with G(iv) shows $\tilde{\mathcal{I}}(\tilde{\beta})$ to be consistent for Σ.

As estimator of Δ consider $\tilde{\Delta}(\tilde{\beta})$, where

$$\tilde{\Delta}(\beta) = n^{-2}\int_0^1\int_0^1 \tilde{G}(\beta, x, w)\, d\bar{N}(x)\, d\bar{N}(w),$$

where $\tilde{G}$ is defined as G in Theorem 3.1 with each $s^{(j)}$ and $h^{(j)}$ replaced by $\tilde{S}^{(j)}$ and $\tilde{H}^{(j)}$ and $\tilde{\alpha} = \tilde{n}n^{-1}$. To show that $\tilde{\Delta}(\tilde{\beta})$ is consistent for Δ, it is sufficient to show $\tilde{G}(\tilde{\beta}, x, w) \to_p G(\beta_0, x, w)$, uniformly in (x, w). The uniform convergence of $\tilde{G}(\tilde{\beta}, x, w)$ follows from the convergence of $\tilde{S}^{(0)}$, $\tilde{S}^{(1)}$ and $\tilde{Q}^{(j)}$, $j = 0, 1, 2$, as described in G(iv) the convergence of $\tilde{H}^{(i)}$ to $h^{(i)}$, $i = 0, 1, 2$, uniformly on $\mathcal{B} \times [0, 1]^2$. This result plus the consistency of $\tilde{\beta}$ gives the required convergence of $\tilde{G}(\tilde{\beta}, t, w)$. The consistency of $\tilde{\Delta}(\tilde{\beta})$ then follows by noting that $n^{-1}\bar{N}(t)$ converges uniformly to $\int_0^t s^{(0)}(x)\lambda_0(x)\, dx$, that $n^{-1}\bar{N}(1)$ is bounded in probability and that G resides in the product space of left-continuous functions.

THEOREM 3.3 [Weak convergence of $n^{1/2}(\tilde{\Lambda} - \Lambda_0)$]. *$n^{1/2}(\tilde{\beta} - \beta_0)$ and $n^{1/2}(\tilde{\Lambda} - \Lambda_0)$ converge weakly and jointly to Gaussian random variables with mean zero. The limiting covariance matrix of $n^{1/2}(\tilde{\beta} - \beta_0)$ is $\mathcal{I}^{-1} + \mathcal{I}^{-1}\Delta\mathcal{I}^{-1}$, where $\mathcal{I}$ and Δ are given in Theorem 3.2. The limiting covariance between $n^{1/2}(\tilde{\Lambda}(u) - \Lambda_0(u))$ and $n^{1/2}(\tilde{\Lambda}(t) - \Lambda_0(t))$ is given by*

$$\int_0^{u\wedge t}\{S^{(0)}(w)\}^{-1}\lambda_0(w)\, dw + \int_0^u e'(w)\lambda_0(w)\, dw(\mathcal{I}^{-1} + \mathcal{I}^{-1}\Delta\mathcal{I}^{-1})\int_0^t e(w)\lambda_0(w)\, dw$$

and the limiting covariance between $n^{1/2}(\tilde{\beta} - \beta_0)$ and $n^{1/2}(\tilde{\Lambda}(t) - \Lambda_0(t))$ is given by

$$(\mathcal{I}^{-1} + \mathcal{I}^{-1}\Delta\mathcal{I}^{-1})\int_0^t e(w)\lambda_0(w)\, dw.$$

PROOF. From Theorems 3.1 and 3.2, $n^{1/2}(\tilde{\beta} - \beta_0)$ may be written as

$$\mathcal{I}^{-1}n^{-1/2}U(\beta_0, 1) - \mathcal{I}^{-1}\int_0^1 n^{1/2}\{\tilde{E}(t) - E(t)\}\lambda_0(t)\, dt + o_p(1).$$

Following Andersen and Gill (1982) and using the convergence properties of the $\tilde{S}^{(j)}$ and $S^{(j)}$ processes described previously, it is straightforward to show that

$n^{1/2}(\tilde{\Lambda}(t) - \Lambda_0(t))$ equals

$$
\begin{aligned}
&\int_0^t \{S^{(0)}(w)\}^{-1} n^{1/2}\, d\overline{M}(w) + n^{1/2}\big(\tilde{\beta} - \beta_0\big)^T \\
&\qquad \times \int_0^t e(w)\lambda_0(w)\, dw + o_p(1) \\
&= \int_0^t \{S^{(0)}(w)\}^{-1} n^{1/2}\, d\overline{M}(w) \\
&\quad + n^{-1/2} U(\beta_0, 1)^T \Sigma^{-1} \int_0^t e(w)\lambda_0(w)\, dw \\
&\quad - \int_0^1 n^{1/2}\{\tilde{E}(t) - E(t)\}^T \lambda_0(w)\, dw\, \Sigma^{-1} \\
&\qquad \times \int_0^t e(w)\lambda_0(w)\, dw + o_p(1).
\end{aligned} \tag{3.2}
$$

PS have shown the weak convergence of $\int_0^t \{S^{(0)}(w)\}^{-1} n^{-1/2}\, d\overline{M}(w)$ and $n^{-1/2}U(\beta_0, t)$ to independent continuous Gaussian martingales with limiting covariance functions

$$\int_0^t \{S^{(2)}(w)/S^{(0)}(w) - e(w)^{\otimes 2}\} S^{(0)}(w)\lambda_0(w)\, dw$$

and

$$\int_0^t \{S^{(0)}(w)\}^{-1} \lambda_0(w)\, dw,$$

respectively. Now, using the same arguments as in Theorem 3.1, it can be shown that the first three terms on the right-hand side of (3.2) converge weakly to mutually independent continuous Gaussian processes. Using the orthogonality relationships and the covariance calculations given above and in Theorem 3.1, the stated limiting covariance structure for $n^{1/2}(\tilde{\beta} - \beta_0)$ and $n^{1/2}(\tilde{\Lambda} - \Lambda_0)$ is easily calculated. By virtue of G(iii) and G(iv), these covariances are readily estimated by replacing population quantities by their natural sample estimators. □

4. Relationship of $\tilde{\beta}$ to previous work. Prentice (1986) proposed a pseudolikelihood estimator of β for the case–cohort setting that is slightly different than $\tilde{\beta}$. He proposes estimating β by the solution to the equation

$$
\begin{aligned}
0 = \int_0^1 \sum_{l \in C} \Bigg[& Z_l(w) r^{(1)}\{\beta' Z_l(w)\} \\
& - \frac{\tilde{S}^{(1)}(\beta, w) + \tilde{n}^{-1} \Sigma_{i \in D(w)} Z_i(w) r^{(1)}(\beta' Z_i(w))}{\tilde{S}^{(0)}(\beta, w) + \tilde{n}^{-1} \Sigma_{i \in D(w)} r(\beta' Z_i(w))} \Bigg] dN_l(w),
\end{aligned} \tag{4.1}
$$

where $D(w) = \{i\colon\ i \notin \tilde{C},\ N_i(w) \neq N_i(w-)\}$. Thus the only difference between Prentice's estimator and $\tilde{\beta}$ is that the "comparison risk set" at time w includes all subcohort members at risk at w plus any individuals who are not in the

subcohort but who are observed to fail at time w. Absolutely continuous failure rates imply that at most one individual will be in $D(w)$ at any time. It is straightforward to show that Prentice's estimator and $\tilde{\beta}$ are asymptotically equivalent provided an individual's contributions to $S^{(1)}$ and $S^{(0)}$ are asymptotically negligible in the sense of G(ii).

Prentice also proposed an estimator of the variance that is somewhat different than the estimator proposed here. He decomposes the variance of the "score" statistic $n^{-1/2}\tilde{U}(\beta_0, 1)$ into two terms: The first term is n^{-1} times the sum of variances of each risk set's contribution to the score statistic; and the second term is n^{-1} times the sum of covariances between score contributions from pairs of different risk sets. Prentice suggests estimating the first term in his variance expression by $\tilde{\Sigma}(\tilde{\beta})$ with the minor (and asymptotically negligible) difference that the sums he uses in computing the $\tilde{S}^{(j)}$'s are over subcohort members plus members of $D(w)$. Thus, to demonstrate consistency of Prentice's variance estimator, it must be shown that his second term converges to $\Delta(\beta_0)$.

Prentice writes n^{-1} times the second term in his variance estimator as

$$2n^{-1} \sum_{j \in C \setminus \tilde{C}} N_j(1) \sum_{\{k:\, t_k < t_j\}} v_{kj}, \tag{4.2}$$

where $C \setminus \tilde{C}$ denotes the set of individuals who are not in the subcohort, v_{kj} denotes the covariance between score contributions of the kth and jth risk set and the t_k's represent the failure times. Using Prentice's notation, the v_{kj}'s are given by the expression

$$-\left(B_k + b_{jk} - b_{ik}\right)'\left(c_{ij} - B_j R_j^{-1}\right) r_{ij} R_j^{-1} \left(R_k + r_{jk} - r_{ik}\right)^{-1}, \tag{4.3}$$

where $r_{li} = Y_l(t_i) r(\beta_0' Z_l(t_i))$, $b_{li} = Y_l(t_i) Z_l(t_i) r^{(1)}(\beta_0' Z_l(t_i))$, $c_{li} = r^{-1}(\beta_0' Z_l(t_i))$, B_j and R_j denote sums of b_{lj} and r_{lj}, where all sums are taken over subcohort members plus $D(t_j)$. Expression (4.3) may be decomposed into five terms:

$$\begin{aligned} & R_k^{-1} R_j^{-1} \textstyle\sum b_{ik}' b_{ij} + B_k' B_j R_k^{-2} R_j^{-2} \textstyle\sum r_{ij} r_{ik} \\ & \quad - B_j' R_k^{-1} R_j^{-2} \textstyle\sum b_{ik} r_{ij} - B_k' R_k^{-2} R_j^{-1} \textstyle\sum b_{ij} r_{ik} \\ & \quad + R_j^{-1} R_k^{-1} \textstyle\sum \left[\left(b_{ik} - b_{ij}\right)' c_{ij} + \left(b_{jk} - b_{ik}\right)' B_j R_j^{-1}\right] \\ & \qquad \times r_{ij}\left(r_{ik} - r_{jk}\right)\left(R_k + r_{jk} - r_{ik}\right)^{-1}. \end{aligned} \tag{4.4}$$

Again noting the asymptotic negligibility of including $D(t_j)$ in the summations, $n - \tilde{n}$ times the first four terms in (4.4) converges to the four terms comprising $G(\beta_0, t_j, t_k)$ in Theorem 3.1, whereas $n - \tilde{n}$ times the fifth term in (4.4) converges to zero. Replacing v_{kj} by $(n - \tilde{n})^{-1} G(\beta_0, t_k, t_j)$ in expression (4.2) and rewriting (4.2) as a double integral with respect to counting process gives

$$2\int_0^1 \int_0^{t-} G(\beta_0, x, w) n^{-1}\, d\bar{N}(x) \left(n - \tilde{n}\right)^{-1} \sum_{i \in C \setminus \tilde{C}} dN_i(w). \tag{4.5}$$

Since $n^{-1}\bar{N}(t)$ and $(n - \tilde{n})^{-1} \Sigma_{i \in C \setminus \tilde{C}} N_i(t)$ converge uniformly to the same monotone, continuous limiting function $\int_0^t s^{(0)}(x) \lambda_0(x)\, dx$ and since G can be

proximated uniformly by a step function on a subset of $[0,1]^2$, which has Lebesgue measure arbitrarily close to 1, it follows that expression (4.5) converges in probability to $\Delta(\beta_0)$.

5. Asymptotic relative efficiency. In making the decision to use a case–cohort instead of a full-cohort design, the cost savings must be balanced against a loss in statistical efficiency. In this section we compute asymptotic efficiencies of the case–cohort pseudolikelihood estimator $\tilde{\beta}$ relative to the partial likelihood estimator $\hat{\beta}$ in the special case of a single binary covariate and exponential relative risk function. We assume that the covariate takes value one in a fraction Π of the entire cohort and takes value zero in fraction $1 - \Pi$. We also assume that all subjects in the cohort are followed from time zero to either failure or to the end of the study at $t = 1$.

For fixed scalar covariates, the asymptotic relative efficiency is given by the expression

$$(1 + \Delta\Sigma^{-1})^{-1}, \tag{5.1}$$

where

$$\Delta = 2(1-\alpha)\alpha^{-1}\int_0^1\int_0^t E\left[Y(u)Y(t)\{Z - e(u)\}\right. \\ \left.\times\{Z - e(t)\}\exp(2\beta Z)\right] d\Lambda_0(u)\, d\Lambda_0(t) \tag{5.2}$$

and

$$\Sigma = \int_0^1 E\left[Y(t)\{Z - e(t)\}^2\exp(\beta Z)\right] d\Lambda_0(t). \tag{5.3}$$

Under the above assumptions, the integrand in (5.2) may be written as

$$\Pi\exp\{2\beta - \Lambda_0(t)e^{\beta}\}\{1 - e(u)\}\{1 - e(t)\} \\ +(1-\Pi)\exp\{-\Lambda_0(t)\}e(u)e(t) \tag{5.4}$$

and the integrand in (5.3) may be written as

$$\Pi\exp\{\beta - \Lambda_0(t)e^{\beta}\}\{1 - e(t)\}^2 + (1-\Pi)\exp\{-\Lambda_0(t)\}e^2(t), \tag{5.5}$$

where

$$e(t) = \frac{\Pi\exp\{\beta - \Lambda_0(t)e^{\beta}\}}{\Pi\exp\{\beta - \Lambda_0(t)e^{\beta}\} + (1-\Pi)\exp\{-\Lambda_0(t)\}}. \tag{5.6}$$

Substituting (5.4), (5.5) and (5.6) into (5.2) and (5.3), performing the inner integration in (5.2) explicitly and making a change in variables gives

$$\Delta = 2(1-\alpha)\alpha^{-1}\Big\{\Pi e^{2\beta}\int[w - G(w)][1 - p(w)]e^{-we^{\beta}}\, dw \\ +(1-\Pi)\int G(w)p(w)e^{-w}\, dw\Big\} \tag{5.7}$$

and

$$\Sigma = \Pi e^{\beta}\int [1 - p(w)]^2 e^{-we^{\beta}}\, dw + (1 - \Pi)\int p^2(w) e^{-w}\, dw, \tag{5.8}$$

where the limits of integration are from 0 to $\Lambda_0(1)$ and

$$G(w) = \begin{cases} w - \log\left[\Pi e^{\beta} + (1 - \Pi)e^{w(e^{\beta}-1)}\right] \\ \qquad + \log\left[1 + \Pi(e^{\beta} - 1)\right], & \text{for } \beta \neq 0, \\ w, & \text{for } \beta = 0, \end{cases} \tag{5.9}$$

and

$$p(w) = \Pi e^{\beta}\left[\Pi e^{\beta} + (1 - \Pi)e^{w(e^{\beta}-1)}\right]^{-1}. \tag{5.10}$$

When $\beta = 0$, the expressions for Δ and Σ simplify considerably. In this case, the ARE may be written

$$\left\{1 + 2(1 - \alpha)\alpha^{-1}\left[1 + (1 - d)\, d^{-1}\log(1 - d)\right]\right\}^{-1}, \tag{5.11}$$

where d denotes the probability of failure prior to the end of the study in the group with $Z = 0$ [i.e., $d = 1 - \exp\{-\Lambda_0(1)\}$]. A Taylor expansion of $\log(1 - d)$ shows that (5.11) is approximately $\{1 + (1 - \alpha)\alpha^{-1} d\}^{-1}$ if d is small. It is interesting to note that this is precisely the ARE of the case–cohort estimator of the log–odds ratio γ at $\gamma = 0$ ($\beta = 0$), where

$$e^{\gamma} = \frac{\left\{1 - (1 - d)^{e^{\beta}}\right\}(1 - d)^{e^{\beta}}}{d/(1 - d)},$$

based only on a binary response that indicates for each subject whether or not the failure time exceeds unity. When $\beta \neq 0$, the integration in (5.7) and (5.8) may be easily done numerically. Table 1 gives values for the ARE when $\Pi = 0.5$ for two values of β and two failure probabilities d. The values are tabulated as a function of the relative numbers of cases and noncases (controls) that are

TABLE 1
Asymptotic relative efficiency[a] *(% increase in standard error)*

Case–control	Overall disease probability	Subcohort fraction	Relative risk	
			1 ($\beta = 0$)	2 ($\beta = 0.693$)
1 : 1	0.05	0.053	0.52 (0.39)	0.55 (0.35)
	0.10	0.111	0.55 (0.35)	0.57 (0.32)
1 : 3	0.05	0.158	0.79 (0.13)	0.80 (0.12)
	0.10	0.333	0.83 (0.10)	0.84 (0.09)
1 : 5	0.05	0.263	0.88 (0.07)	0.89 (0.06)
	0.10	0.555	0.92 (0.04)	0.93 (0.04)

[a] Two-sample problem with equal group sizes, t = time on trial, no loss to follow-up, no staggered entry.

expected by the end of the study. These "case–control" ratios are translated into subcohort fractions α, as a function of failure probability in Table 1.

As seen in Table 1, the ARE increases slightly as β increases. As expected, the ARE decreases as the number of "controls" per case decreases.

The theoretical values in Table 1 may be compared to the results of Prentice's simulation study. In that work, a full-cohort size of 500 was used, with "case–control" ratios of 1 : 1 (subcohort to 55) and 1 : 5 (subcohort of 275) and overall disease probability of 0.10.

For a relative risk of 1, Prentice observed sample relative efficiencies of 0.66 and 0.93 at 1 : 1 and 1 : 5 case-to-control ratios, respectively. This can be compared to AREs of 0.55 and 0.92 from the first column of Table 1. It is interesting to note that the ratios of the average of the variance estimator in the simulation study were 0.54 and 0.93, respectively, virtually identical to the corresponding AREs.

At a relative risk of 2, Prentice observed sample relative efficiencies of 0.59 and 0.92 at 1 : 1 and 1 : 5 case-to-control ratios, respectively. These can be compared to AREs of 0.57 and 0.93 from the second column of Table 1. The ratios of the corresponding variance estimator averages from the simulation study are 0.58 and 0.91, respectively. It seems reasonable to conclude that there is little evidence for any important difference between asymptotic relative efficiencies and the corresponding actual relative efficiencies under the simulation conditions, particularly since the sample relative efficiencies were based on a rather small simulation study.

6. Discussion. The conditions for asymptotic normality of the case–control estimators that are described above are fairly complicated and do not easily lend themselves to simple interpretations. It would be useful to describe a few, more restrictive but more easily understood conditions under which the desired results obtain. The following five conditions are sufficient for the asymptotic properties of the full-cohort estimator to obtain (conditions A–F): (1) the relative risk function $r(\cdot)$ is twice continuously differentiable and in a neighborhood of β_0, $r\{\beta' Z_i(t)\}$ is locally bounded away from zero for all $i = 1, \ldots, n$; (2) $\{(N_i, Y_i, Z_i); 1 \leq i \leq n\}$ are independent, identically distributed processes with the Z_i left-continuous with right limits and bounded almost surely; (3) random, noninformative right censoring with a positive probability of not being censored prior to time 1; (4) finite interval (condition A); and (5) $\hat{\Sigma}$ is positive definite. Condition B follows from (2) by Andersen and Gill's Theorem III.1 about the strong law of large numbers in Banach space. Conditions C and E follow from the boundedness of the Z_i. Condition D follows from (1) and (3) by arguments that parallel Andersen and Gill's Theorem 4.1. Note, however, that noninformative censoring is an additional assumption here that is required to assure regularity of the limiting functions $s^{(j)}$ in β. Conditions A and F correspond to (4) and (1), respectively. Two additional conditions may be added to these five to obtain the asymptotic results for the case–cohort estimator: (6) nontrivial subcohort [condition G(i)] and (7) tightness of the sequence of distributions on $n^{1/2}[\tilde{E}(t) - E(t)]$, $j = 0, 1$. Conditions G(iii), G(iv) and the Lindeberg conditions in G(ii) follow

from (2)–(4), using similar arguments as before. The tightness condition is the least intuitive and most difficult to verify. In general, the tightness condition would be expected to fail if the $Z_i(t)$'s were fluctuating wildly in t.

A useful generalization of the previous theory would be to allow for stratification. Suppose K strata are defined in the cohort. From the kth stratum of size n_k, a subcohort of size $\tilde{n}_k$ is sampled, where the subcohort fraction $\tilde{n}_k/n_k$ may vary among strata. The stratified analysis would proceed by forming stratum specific functions $\log \tilde{L}_k(\beta, t)$ as given by (2.3) except where the summations are restricted to be over individuals in the kth stratum. The "estimator" $\tilde{\beta}$ is then defined as the solution to the equation $\partial \Sigma \log \tilde{L}_k(\beta, 1)/\partial\beta = 0$. Provided that the regularity conditions given previously hold for each stratum, the limiting distribution of $n^{1/2}\, \partial \Sigma \log \tilde{L}_k(\beta_0, 1)/\partial\beta$ is Gaussian with mean zero and variance $\dot{\Sigma} + \bar{\Delta}$ where $\dot{\Sigma} = \Sigma \dot{\Sigma}_k$ and $\bar{\Delta} = \Sigma \Delta_k$ with $\dot{\Sigma}_k$ and Δ_k being stratum-specific versions of $\dot{\Sigma}, \Delta$ given in Theorem 3.1. Thus the results of Theorem 3.2 are obtained with the limiting variance given by $\bar{\dot{\Sigma}}^{-1} + \bar{\dot{\Sigma}}^{-1} \Delta \bar{\dot{\Sigma}}^{-1}$.

It is interesting to compare the AREs in Table 1 to AREs of a synthetic case–control design. In this design, K time-matched controls are randomly selected for comparison with each case. The ARE at $\beta = 0$ for this design is given by $K/K + 1$. Thus for K equal to 1, 3 and 5, the AREs are 0.50, 0.75 and 0.83, respectively. As seen from Table 1, these AREs are dominated by the case–cohort design with the greatest improvement in efficiency coming at the larger disease probabilities. As speculated by Prentice, this is due to the fact that with increasing disease probability, the size of the comparison set for each case increases with the case–cohort design but is constant ($= K$) for the synthetic case–control design even though the total number of cases and controls used in the two analyses are equal.

APPENDIX

PROOF OF PROPOSITION 1. Pick $\varepsilon > 0$. For $\tau, \delta > 0$, let $\mathscr{S}_{n\delta\tau}$ denote the set on which

$$n^{-1} \sum_{i=1}^{n} [\, f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I_{\{|f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)| > n^{1/2}\tau\}} < \delta$$

and

$$|S_{fn}^2 - \sigma_f^2| < \delta.$$

Let (g_n, h_n) denote $(g_n(\mathbf{X}_n), h_n(\mathbf{X}_n, \delta_n))$ and write their joint distribution function as

$$\begin{aligned} P\{g_n \le w, h_n \le v\} &= P\{g_n \le w, h_n^* \le v_n^*\} \\ &= E_{\mathbf{X}_n}\Big\{ I_{\{g_n \le w, \mathbf{X}_n \in \mathscr{S}_{n\delta\tau}\}} P[\, h_n^* \le v_n^* | \mathbf{X}_n] \Big\} \\ &\quad + E_{\mathbf{X}_n}\Big\{ I_{\{g_n \le w, \mathbf{X}_n \notin \mathscr{S}_{n\delta\tau}\}} P[\, h_n^* \le v_n^* | \mathbf{X}_n] \Big\}, \end{aligned} \tag{A1.1}$$

where $h_n^* = h_n \tilde{n}^{1/2}(n - \tilde{n})^{-1/2} S_{fn}^{-1}$ and $v_n^* = v\tilde{n}^{1/2}(n - \tilde{n})^{-1/2} S_{fn}^{-1}$. By Hájek's

theorem on sampling without replacement from a finite population as stated in Cochran (1977), pages 39–40, for any $\varepsilon > 0$, a $\tau > 0$ and $\delta > 0$ may be chosen such that for large enough n and for $\mathbf{X}_n \in \mathscr{S}_{n\delta\tau}$, $P\{h_n^* \leq v_n^* | \mathbf{X}_n\}$ is within ε of $\Phi(v_n^*)$ where $\Phi(\cdot)$ denotes the standard unit Gaussian cumulative distribution function. By continuity of $\Phi(\cdot)$ and the convergence properties implied by conditions (1) and (2), v_n^* may be replaced by $v^* = v\alpha^{1/2}(1-\alpha)^{-1/2}\sigma_f^{-1}$ in this approximation. By the convergence properties implied by condition (2), the probability of the set $\mathscr{S}_{n\delta\tau}$ may be bounded below by $1 - \varepsilon$ for large enough n. Thus the second term on the right-hand side of (A1.1) is bounded by ε. In addition, this implies that $E_{\mathbf{X}_n}\{I_{\{g_n \leq w,\ \mathbf{X}_n \in \mathscr{S}_{n\delta\tau}\}}\}$ is within ε of $P\{g_n \leq w\}$, which, in turn, is within ε of $\Phi(w/\sigma_g)$ for large enough n. Thus, by taking n large enough, $P\{g_n \leq w,\ h_n \leq v\}$ can be made arbitrarily close to $\Phi(w/\sigma_g)\Phi(v^*)$. □

REFERENCES

AALEN, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. *Z. Wahrsch. verw. Gebiete* **38** 261–277.

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

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

COCHRAN, W. G. (1977). *Sampling Techniques*. Wiley, New York.

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

LIDDELL, F. D. K., MCDONALD, J. C. and THOMAS, D. C. (1977). Methods for cohort analysis: Appraisal by application to asbestos mining (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 469–490.

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

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

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.

SELF, S. G. and PRENTICE, R. L. (1982). Commentary on Andersen and Gill's distribution theory for the Cox regression model. *Ann. Statist.* **10** 1121–1124.

WHITTEMORE, A. S. and MCMILLAN, A. (1982). Analyzing occupational cohort data: Application to U.S. uranium miners. In *Environmental Epidemiology: Risk Assessment* (R. L. Prentice and A. S. Whittemore, eds.) 65–81. SIAM, Philadelphia.

FRED HUTCHINSON CANCER RESEARCH CENTER
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104