The Annals of Statistics
1988, Vol. 16, No. 3, 1133–1140

A NOTE ON THE ASYMPTOTIC NORMALITY IN THE COX REGRESSION MODEL

BY E. ARJAS AND P. HAARA

University of Oulu

A general version of the Cox regression model is studied as a discrete time process consisting of the embedded experiments at the jump times. Simple conditions are given for the consistency and asymptotic normality of the estimator for the regression parameters.

1. Introduction. The standard reference for asymptotic normality results for the Cox (1972) regression model is Andersen and Gill (1982), henceforth abbreviated as AG. It improves in several ways on earlier results (see the discussion and references in AG), e.g., by allowing for nonidentically distributed individuals and time dependent random covariates. Prentice and Self (1983) extend the results of AG to more general relative risk functions. AG provides (Theorem 3.4) the limiting joint (normal) law of the regression coefficients in the relative risk function and an expression depending on the estimated cumulative baseline hazard.

Often, however, the interest is focused directly on the relative risks and the baseline hazard is viewed as a nuisance parameter of the model. Such an orientation is reflected in the strong emphasis on the Cox partial likelihood expression which does not involve the baseline hazard. But then the question arises whether, for the more restrictive goal of deriving asymptotic normality results for the regression coefficients (see Theorem 3.2 in AG), their conditions could perhaps be simplified and relaxed. [Note that the asymptotic normality of the regression coefficients alone is not sufficient to determine, for example, an approximate confidence interval for the t-year survival probability; see, e.g., Altman and Andersen (1986).]

A simplification can be achieved by directly considering the embedded experiments at failure times (in which the failing individual is picked from the risk set). Such an approach was used, in fact, by Naes (1982); it corresponds to the product form of the partial likelihood expression and leads to a simple martingale argument of a discrete time parameter. Naes, however, made the strong assumption that all individuals (each being represented by failure time, censoring time and covariates) are sampled from a fixed distribution. While we want to avoid such assumptions here, we need to postulate directly some of the properties Naes was able to derive from his special structure.

Compared to AG, we can drop the finite interval condition as well as assumptions concerning the uniform convergence of certain key quantities over the considered (unit) time interval. As a result, we need a slightly stronger

Received May 1987; revised December 1987.
AMS 1980 *subject classifications*. Primary 62F12; secondary 62M99.
Key words and phrases. Hazard-rate, regression, censoring, asymptotic normality, martingale.

Lindeberg condition and we also assume that the covariates are square-integrable. We are left with very simple conditions [Conditions C and L in Section 3] which strongly resemble those in Arjas and Haara (1987).

Considering the background, none of our results is in the least surprising; however, we feel that, because Cox's model is so widely applied in practice, it is useful to have these conditions and proofs on record. As in AG, we allow for repeated events: The counting processes can count more than one event.

2. Preliminaries. Consider a sequence of observations

$$\left((N_{nj}, Y_{nj}, \mathbf{Z}_{nj})_{1\le j\le n}\right)_{n\ge 1}$$

and a family of probability measures $(\mathbb{P}^{\beta})_{\beta\in\mathbb{R}^p}$ defined on a measurable space $(\Omega, \mathscr{F})$. We assume that for each $1 \le j \le n$ the counting process N_{nj} is adapted to a history $\mathbb{F}_n = (\mathscr{F}_{ns})_{s\ge 0}$ and has the $(\mathbb{P}^{\beta}, \mathscr{F}_n)$-intensity,

$$\lambda_{nj}(s) = Y_{nj}(s)\lambda_0(s)\exp\{\beta'\mathbf{Z}_{nj}(s)\}, \qquad s \ge 0.$$

All unexplained notation is as in AG. Let $\beta_0 \in \mathbb{R}^p$ be the unknown true value. We denote $\mathbb{P} = \mathbb{P}^{\beta_0}$, $\mathbb{E} = \mathbb{E}^{\beta_0}$, etc. Since all random quantities and the history $\mathbb{F}_n$ depend on the index n, we drop it until Section 3. We assume $N_j(\infty) < \infty$, $\mathbb{P}$-a.s., for each $1 \le j \le n$, $\beta \in \mathbb{R}^p$.

From here on we consider stochastic processes on $\mathbb{N} = \{1, 2, \dots\}$ rather than on $[0, \infty)$. Let $(T_v, X_v)_{v\ge 1}$ be the marked point process representation of $\mathbf{N} := (N_1, N_2, \dots, N_n)$, i.e., $T_1 < T_2 < \cdots$ are the failure times and X_v indexes the individual failing at T_v. If $T_v(\omega) = \infty$, then we define $X_v(\omega) := \Delta \notin \{1, 2, \dots, n\}$. Let $K := \max\{v \in \mathbb{N} | T_v < \infty\}$ be the total number of observed points. By assumption $K < \infty$, $\mathbb{P}$-a.s. For $v \in \mathbb{N}$ we define $\Delta N_j(v) := \Delta N_j(T_v)$, $Y_j(v-1) := Y_j(T_v)$, $\mathbf{Z}_j(v-1) := \mathbf{Z}_j(T_v)$, using the conventions $Y_j(\infty) = 0$ and $\mathbf{Z}_j(\infty) = \mathbf{0}$. Finally denote $\mathscr{G}_{v-1} := \mathscr{F}_{T_v-}$, $v \in \mathbb{N}$. Now, $\mathbb{G} := (\mathscr{G}_v)_{v\ge 0}$ is the basic history to which the sequence $\Delta N_j(v)$, $v \in \mathbb{N}$, is adapted for each $j \in \mathbb{N}$ and with respect to which our martingale structure is formed. Note that K is a stopping time in this history, for $\{K \le v\} = \{T_{v+1} = \infty\} \in \mathscr{F}_{T_{v+1}-} = \mathscr{G}_v$, $v \in \mathbb{N}$.

According to the previous definitions $\mathbf{Z}_{X_v}(v-1)$ is the covariate vector of the individual who has failed at T_v if $T_v < \infty$ and 0 otherwise. In order to simplify the presentation, we assume that all random variables $\|\mathbf{Z}_{X_v}(v-1)\|$, $v \ge 1$, are square-integrable w.r.t. each $\mathbb{P}^{\beta}$, $\beta \in \mathbb{R}^p$. For each $\beta \in \mathbb{R}^p$ we define

$$\begin{aligned} p_j(\beta, v-1) &:= \frac{Y_j(v-1)\exp\{\beta'\mathbf{Z}_j(v-1)\}}{\sum_{k=1}^n Y_k(v-1)\exp\{\beta'\mathbf{Z}_k(v-1)\}} \\ &= \mathbb{P}^{\beta}\left(\Delta N_j(v) = 1 | \mathscr{G}_{v-1}\right), \qquad 1 \le j \le n \end{aligned} \tag{1}$$

$$\left(\text{following the convention } \frac{0}{0} = 0\right),$$

$$p_{\Delta}(\beta, v-1) := 1 - \sum_{j=1}^{n} p_j(\beta, v-1),$$

$$\mathbf{E}(\boldsymbol{\beta}, v-1) := \sum_{j=1}^{n} \mathbf{Z}_j(v-1) p_j(\boldsymbol{\beta}, v-1) = \mathbf{E}^{\boldsymbol{\beta}}\big(\mathbf{Z}_{X_v}(v-1) | \mathscr{G}_{v-1}\big), \tag{2}$$

$$\begin{aligned} \mathbf{V}(\boldsymbol{\beta}, v-1) &:= \sum_{j=1}^{n} \mathbf{Z}_j^{\otimes 2}(v-1) p_j(v-1) - \mathbf{E}^{\otimes 2}(\boldsymbol{\beta}, v-1) \\ &= \operatorname{Cov}^{\boldsymbol{\beta}}\big(\mathbf{Z}_{X_v}(v-1) | \mathscr{G}_{v-1}\big). \end{aligned} \tag{3}$$

Using this notation, the Cox partial likelihood is

$$L(\boldsymbol{\beta}, \infty) := \prod_{v=1}^{\infty} p_{X_v}(\boldsymbol{\beta}, v-1) = \prod_{v \le K} p_{X_v}(\boldsymbol{\beta}, v-1). \tag{4}$$

From (1)–(4) we get, by a simple calculation, expressions for the log-likelihood, score and observed information (considered as processes on $t \in \mathbb{N} \cup \{\infty\}$):

$$\begin{aligned} l(\boldsymbol{\beta}, t) &:= \sum_{v=1}^{t} \log\big(p_{X_v}(\boldsymbol{\beta}, v-1)\big) \\ &= \sum_{v \le t} \sum_{j \le n} \log\big(p_j(\boldsymbol{\beta}, v-1)\big)\, \Delta N_j(v) \\ &= \sum_{v \le t} \Bigg[\sum_{j \le n} Y_j(v-1) \boldsymbol{\beta}' \mathbf{Z}_j(v-1)\, \Delta N_j(v) \\ &\qquad - \log\Bigg\{ \sum_{k=1}^{n} Y_k(v-1) \exp\{\boldsymbol{\beta}' \mathbf{Z}_k(v-1)\} \Bigg\} I_{\{v \le K\}} \Bigg] \end{aligned} \tag{5}$$

(defining $0 \cdot \infty = 0$);

$$\begin{aligned} \mathbf{U}(\boldsymbol{\beta}, t) &:= \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, t) \\ &= \sum_{v \le t} \sum_{j \le n} Y_j(v-1) \mathbf{Z}_j(v-1)\, \Delta M_j(\boldsymbol{\beta}, v), \end{aligned} \tag{6}$$

where $\Delta M_j(\boldsymbol{\beta}, v) := \Delta N_j(v) - p_j(\boldsymbol{\beta}, v-1)$ and

$$\begin{aligned} \mathbf{I}(\boldsymbol{\beta}, t) &:= -\frac{\partial^2}{\partial \boldsymbol{\beta}^2} l(\boldsymbol{\beta}, t) \\ &= \sum_{v \le t} V(\boldsymbol{\beta}, v-1). \end{aligned} \tag{7}$$

The process $\mathbf{U}(\boldsymbol{\beta}_0, \cdot)$ is a square-integrable $(\mathbb{P}^{\boldsymbol{\beta}_0}, \mathbb{G})$-martingale and by (2) and (6), we have

$$\mathbf{U}(\boldsymbol{\beta}_0, t) = \sum_{v \le t} \big\{ \mathbf{Z}_{X_v}(v-1) - E\big(\mathbf{Z}_{X_v}(v-1) | \mathscr{G}_{v-1}\big) \big\} \tag{8}$$

and

$$\langle \mathbf{U}(\beta_0, \cdot)\rangle(t) = \sum_{v \le t} \mathrm{Cov}(\Delta \mathbf{U}(\beta_0, v)|\mathscr{G}_{v-1})$$

$$(9) \qquad = \sum_{v \le t} \mathrm{Cov}\big(\mathbf{Z}_{X_v}(v-1)|\mathscr{G}_{v-1}\big) \quad [\text{by } (8)]$$

$$= \sum_{v \le t} \mathbf{V}(\beta_0, v-1) = \mathbf{I}(\beta_0, t) \quad [\text{by } (3) \text{ and } (7)].$$

REMARK 1. If $\mathbf{Z}_{X_v}(v-1)$, $v \ge 1$, are not assumed square-integrable, $\mathbf{E}(\beta, v-1)$ and $\mathbf{V}(\beta, v-1)$ in (2) and (3) cannot be interpreted as conditional moments. In that case $\mathbf{U}$ is a locally square-integrable martingale and the relationship (9) between $\langle \mathbf{U}\rangle$ and $\mathbf{I}$ can be established using localizing stopping times.

REMARK 2. It is also interesting to compare (8) with the corresponding expression in AG; the martingale differences in the former take the role of a predictable integrand in the martingale representation of the latter.

3. Asymptotic results. First we shall prove that for the *consistency* of the solution $\hat{\mathbf{B}}_n$ of $\mathbf{U}_n(\beta, \infty) = 0$ (which is *assumed* to exist, maximize $l(\beta, \infty)$ and be unique with probability tending to 1), Condition C will suffice.

CONDITION C. There are a sequence $c_n \to +\infty$, a neighbourhood $\mathscr{B}$ of β_0 and a bounded continuous matrix valued function Σ: $\mathscr{B} \to \mathbb{R}^{p \times p}$ such that

$$(10) \qquad c_n^{-1}\mathbf{I}_n(\beta, K_n) \to_p \Sigma(\beta)$$

uniformly over $\beta \in \mathscr{B}$. The matrix $\Sigma_0 := \Sigma(\beta_0)$ is positive definite.

It is no restriction to assume that $\mathscr{B}$ is an open ball in $\mathbb{R}^p$. Let

$$(11) \qquad \mathbf{Z}_{nj}^*(v-1) := \mathbf{Z}_{nj}(v-1) - \mathbb{E}\big(\mathbf{Z}_{n, X_{nv}}(v-1)|G_{n, v-1}\big)$$

and define

$$(12a) \quad S_n^*(\beta, t) := \sum_{v \le t} \log\left\{\frac{\sum_{j=1}^n Y_{nj}(v-1)\exp\{\beta' \mathbf{Z}_{nj}^*(v-1)\}}{\sum_{j=1}^n Y_{nj}(v-1)\exp\{\beta_0' \mathbf{Z}_{nj}^*(v-1)\}}\right\} I_{\{v \le K_n\}},$$

$$(12b) \quad \mathbf{A}_n^*(\beta, t) := \frac{\partial}{\partial \beta} S_n^*(\beta, t), \qquad t \in \mathbb{N}.$$

It is easy to see that the quantities defined in (1) and (3)–(7) remain unchanged if we use the centered covariate processes instead of the original ones. Then we have

$$(13a) \qquad \mathbf{A}_n^*(\beta, t) = \sum_{v \le t} (\mathbf{E}_n(\beta, v-1) - \mathbf{E}_n(\beta_0, v-1)),$$

$$(13b) \qquad \mathbf{I}_n(\beta, t) = \frac{\partial^2}{\partial \beta^2} S_n^*(\beta, t).$$

LEMMA 1. *Suppose Condition C holds. Then there is a twice differentiable function* $s\colon \mathscr{B} \to \mathbb{R}$ *such that*

$$c_n^{-1} S_n^*(\beta, K_n) \to_p s(\beta), \tag{14}$$

uniformly over $\beta \in \mathscr{B}$, *and*

$$\frac{\partial}{\partial \beta} s(\beta)|_{\beta=\beta_0} = 0, \qquad \frac{\partial^2}{\partial \beta^2} s(\beta) = \Sigma(\beta). \tag{15}$$

PROOF. Note that $A_n^*(\beta_0, K_n) \equiv 0$ by (13a). Thus by Condition C and (13) the sequence $c_n^{-1}\mathbf{A}_n^*(\beta, K_n)$, $n \geq 1$, of mappings from $\mathscr{B} \times \Omega$ to $\mathbb{R}^p$ satisfies the conditions of Lemma A in Arjas and Haara [(1987), page 17]. This implies the existence of a differentiable function $\mathbf{a}\colon \mathscr{B} \to \mathbb{R}^p$ such that

$$c_n^{-1}\mathbf{A}_n^*(\beta, K_n) \to_p \mathbf{a}(\beta) \quad \text{uniformly over } \beta \in \mathscr{B}$$

and

$$\frac{\partial}{\partial \beta}\mathbf{a}(\beta) = \Sigma(\beta).$$

Similarly, since $S_n^*(\beta_0, K_n) \equiv 0$, we can apply Lemma A to the sequence $c_n^{-1}S_n^*(\beta, K_n)$, $n \geq 1$. The claim follows from this. □

LEMMA 2 (Consistency). *Under Condition C,*

$$\hat{\beta}_n \to_p \beta_0. \tag{16}$$

PROOF. The proof is analogous to that of Lemma 3.1 in AG. Consider the average log-likelihood ratio process,

$$\begin{aligned} l_n(\beta, \beta_0; t) &:= c_n^{-1}(l_n(\beta, t) - l_n(\beta_0, t)) \\ &= c_n^{-1} \sum_{v \leq t} \left[\sum_{j \leq n} Y_{nj}(v-1)(\beta - \beta_0)'\mathbf{Z}_{nj}^*(v-1)\,\Delta N_{nj}(v) \right] \\ &\quad - c_n^{-1} S_n^*(\beta, t) \end{aligned}$$

[by (5), (12a) and the remark following (12)]. It has the $(\mathbb{P}, \mathscr{G}_n)$-compensator

$$\begin{aligned} k_n(\beta, \beta_0; t) &:= c_n^{-1} \sum_{v \leq t} \left[\sum_{j \leq n} Y_{nj}(v-1)(\beta - \beta_0)'\mathbf{Z}_{nj}^*(v-1) p_{nj}(\beta_0, v-1) \right] \\ &\quad - c_n^{-1} S_n^*(\beta, t) \\ &= -c_n^{-1} S_n^*(\beta, t) \quad [\text{by (11)}]. \end{aligned} \tag{17}$$

Then for each β the process

$$l_n(\beta, \beta_0; t) - k_n(\beta, \beta_0; t) = c_n^{-1}(\beta - \beta_0)'\mathbf{U}_n(\beta_0, t) \quad [\text{by (8), (11) and (17)}]$$

is a square-integrable martingale with variance process

$$\begin{aligned}(18)\qquad &\langle l_n(\beta, \beta_0; \cdot) - k_n(\beta, \beta_0; \cdot)\rangle(t) \\ &\quad = c_n^{-2}(\beta - \beta_0)'\langle \mathbf{U}_n(\beta_0, \cdot)\rangle(t)(\beta - \beta_0) \\ &\quad = c_n^{-1}(\beta - \beta_0)'c_n^{-1}\mathbf{I}_n(\beta_0, t)(\beta - \beta_0).\end{aligned}$$

By (17) and Lemma 1 it follows that

$$(19)\qquad k_n(\beta, \beta_0; \infty) \to_p -s(\beta), \qquad \beta \in \mathscr{B},$$

while by (18) and (10)

$$(20)\qquad c_n\langle l_n(\beta, \beta_0; \cdot) - k_n(\beta, \beta; \cdot)\rangle(\infty) \to_p (\beta - \beta_0)'\Sigma_0(\beta - \beta_0).$$

The inequality of Lenglart, together with (19) and (20), implies that

$$(21)\qquad l_n(\beta, \beta_0; \infty) \to_p -s(\beta).$$

The claim follows from (15) and (21) by the same convex analysis arguments as were used in the proofs of Lemma 3 in Arjas and Haara (1987) and Lemma 3.1 in AG. □

For the asymptotic normality of $\hat{\beta}_n$ we need

CONDITION L. The family $\{\|\mathbf{Z}_{nj}(s)\||Y_{nj}(s)|1 \le j \le n,\ s \in [0, \infty)\}$ is bounded by a random variable $\hat{Z}_n$ such that

$$(22)\qquad c_n^{-1/2}\hat{Z}_n \to_p 0.$$

LEMMA 3. *Under Conditions C and L,* $c_n^{-1/2}\mathbf{U}_n(\beta_0, K_n) \to_{\mathscr{D}} \mathbf{N}(\mathbf{0}, \Sigma_0)$.

PROOF. The proof is analogous to that of Theorem 4.1 in Naes (1982). By the Cramér–Wold device [Billingsley (1968), page 49], it suffices to fix a $\mathbf{w} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$ and prove

$$(23)\qquad c_n^{-1/2}\mathbf{w}'\mathbf{U}_n(\beta_0, K_n) \to_{\mathscr{D}} \mathbf{N}(\mathbf{0}, \mathbf{w}'\Sigma_0\mathbf{w}).$$

We apply the CLT of Dvoretzky (1972) for martingale difference arrays, given as Theorem 2.5(a) in Helland (1982), with $\xi_{nv} := (c_n\mathbf{w}'\Sigma_0\mathbf{w})^{-1/2}\mathbf{w}'\Delta\mathbf{U}_n(\beta_0, v)$, $n, v \in \mathbb{N}$, being the martingale differences w.r.t. to the sequence $(\mathbb{G}_n)_{n\in\mathbb{N}}$ of histories (recall that K_n is a $\mathbb{G}_n$-stopping time). Now

$$\begin{aligned}(24)\qquad \sum_{v=1}^{K_n}\mathbf{E}\left(\xi_{nv}^2|\mathscr{G}_{n,v-1}\right) &= (\mathbf{w}'\Sigma_0\mathbf{w})^{-1}c_n^{-1}\sum_{v=1}^{K_n}\mathbf{E}\left((\mathbf{w}'\Delta\mathbf{U}_n(\beta_0, v))^2|\mathscr{G}_{n,v-1}\right) \\ &= (\mathbf{w}'\Sigma_0\mathbf{w})^{-1}c_n^{-1}\sum_{v=1}^{K_n}\mathbf{w}'\mathrm{Cov}\left(\Delta\mathbf{U}_n(\beta_0, v)|\mathscr{G}_{n,v-1}\right)\mathbf{w} \\ &= (\mathbf{w}'\Sigma_0\mathbf{w})^{-1}\mathbf{w}'c_n^{-1}\mathbf{I}_n(\beta_0, K_n)\mathbf{w} \to_p 1\end{aligned}$$

by (9) and (10), which leaves us the Lindeberg condition

$$\sum_{v=1}^{K_n} \mathbf{E}\left(\xi_{nv}^2 I_{\{|\xi_{nv}|>\varepsilon\}} | \mathscr{G}_{n,v-1}\right) \to_p 0, \qquad \varepsilon > 0, \tag{25}$$

to verify.

Fix $\varepsilon > 0$. Choose $\delta > 0$ such that $(\mathbf{w}'\Sigma_0\mathbf{w})^{-1/2} 2\|\mathbf{w}\|\delta < \varepsilon$. Then the left side of (25) is 0 on the event $\{c_n^{-1/2}\hat{Z}_n < \delta\}$ since, by (2),

$$\begin{aligned}(c_n\mathbf{w}'\Sigma_0\mathbf{w})^{-1/2}\left|\mathbf{w}'\mathbf{Z}_{n,j}(v-1) - \mathbf{E}\left(\mathbf{w}'\mathbf{Z}_{n,X_{nv}}(v-1)|\mathscr{G}_{n,v-1}\right)\right| \\ \le (\mathbf{w}'\Sigma_0\mathbf{w})^{-1/2} 2\|\mathbf{w}\| c_n^{-1/2}\hat{Z}_n \quad (\text{cf. } H_{il}(t) \text{ in AG})\end{aligned}$$

for each $n, v \in \mathbb{N}$ and $1 \le j \le n$ such that $Y_{nj}(v-1) = 1$. But Condition L says $\mathbb{P}(c_n^{-1/2}\hat{Z}_n < \delta) \to 1$, so the probability that the left side of (25) is equal to 0 converges to 1. □

The following theorem is now true.

THEOREM (Asymptotic normality of $\hat{\beta}_n$). *Suppose that Conditions C and L are satisfied. Then, as* $n \to \infty$,

$$c_n^{1/2}(\hat{\beta}_n - \beta_0) \to_{\mathscr{D}} \mathbf{N}(\mathbf{0}, \Sigma_0^{-1}). \tag{26}$$

PROOF. Condition C implies that

$$c_n^{-1}\mathbf{I}_n(\beta_n^*, K_n) \to_p \Sigma_0$$

for any random sequence $(\beta_n^*)_{n\in\mathbb{N}}$ satisfying $\beta_n^* \to_p \beta_0$. The claim (26) now follows by the well-known Taylor-series argument as given, e.g., in Arjas and Haara [(1987), pages 12–13]. □

Acknowledgment. We are grateful to a referee for a number of comments that led to a further simplification of our proofs.

REFERENCES

ALTMAN, D. G. and ANDERSEN, P. K. (1986). A note on the uncertainty of a survival probability estimated from Cox's regression model. *Biometrika* **73** 722–724.

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

ARJAS, E. and HAARA, P. (1987). A logistic regression model for hazard: Asymptotic results. *Scand. J. Statist.* **14** 1–18.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

DVORETZKY, A. (1972). Central limit theorems for dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 513–535. Univ. California Press.

HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–94.

NAES, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scand. J. Statist.* **9** 107–115.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory of Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 803–813.

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
UNIVERSITY OF OULU
LINNANMAA
90570 OULU
FINLAND