ESTIMATES OF RELATIVE RISK

by

Janet M. Begun[1]

University of North Carolina

Summary

For the two-sample problem with proportional hazard functions, we consider estimation of the constant of proportionality, known as relative risk, using complete uncensored data. For this very special case of Cox's (1972) regression model for survival data, we find a two-step estimate which is asymptotically equivalent to Cox's partial likelihood estimate, and we show that both estimates are asymptotically optimal (in the sense of minimum asymptotic variance) among all regular rank estimates of relative risk.

[1]Supported by NSF grant 81-00748.

Key Words and Phrases: Relative risk, proportional hazards model, Lehmann alternative, Hâjek's representation theorem.

AMS 1970 subject classification: Primary 62G05; secondary 62F05, 62F15.

1. Introduction

Cox (1972) proposed the following nonparametric regression model for the analysis of censored data:

$$r(t|\underline{z}) = \exp(\underline{\beta}'\underline{z})\ r_o(t)$$

where $r(t|\underline{z})$ is the hazard function at time t for an individual with vector $\underline{z}$ of regressor variables, vector $\underline{\beta}$ of regression parameters, and common (unknown) hazard function $r_o(t)$. The maximum partial likelihood estimate of $\underline{\beta}$ was derived by Cox (1972, 1975). Asymptotic consistency and normality of this estimate was recently established by Tsiatis (1981).

Here, we focus attention on a special case of Cox's model. Specifically, we assume that there is only one regressor variable, that this variable takes values 0 and 1 only, and the observations are uncensored. The constant of proportionality $\theta \equiv \exp\{\beta\}$ is called relative risk. For this two-sample, uncensored case of Cox's model, we propose a family of nonparametric rank estimators of relative risk. One member of this family, a two-step estimate, is asymptotically equivalent to Cox's estimate of θ. We show that both the two-step and Cox's estimate are asymptotically optimal (in the sense of minimum asymptotic variance) among all regular rank estimates of relative risk.

Aside from containing the two-step estimate, this separate family of estimates serves several purposes. First, members of this family can be used as preliminary estimates in constructing the two-step estimate. Second, since estimates in this family are regular, the collection of regular rank estimates of relative risk is nonempty. Third, even when the underlying assumption of proportionality of hazard functions fails to hold, estimates in this family satisfy our definition of a "measure of relative risk" (see Section 4).

Let F and G denote continuous distribution functions (dfs) on $\mathbb{R}^+ \equiv [0,\infty)$ with survivor functions $\bar{F} \equiv 1 - F$ and $\bar{G} \equiv 1 - G$. By a proportional hazards model (PHM) we mean that the cumulative hazard functions corresponding to the continuous dfs F and G are proportional:

$$-\log \bar{G} = -\theta \log \bar{F};$$

the constant of proportionality $0 < \theta < \infty$ is called relative risk. Equivalently, F and G satisfy the Lehmann alternative $\bar{G} = \bar{F}^{\theta}$.

With the additional assumption that F and G are absolutely continuous with densities f and g respectively, the PHM is equivalent to the special case of Cox's model described above. Specifically, the hazard functions corresponding to F and G are proportional: $g/\bar{G} = \theta f/\bar{F}$. This additional assumption of absolute continuity is not used in this paper.

Let $X_1 < \cdots < X_m$ and $Y_1 < \cdots < Y_n$ denote the order statistics of two independent samples from F and G respectively. Denote the corresponding *left-continuous* empirical dfs by F_m and G_n respectively. Let $W_1 < \cdots < W_N$ denote the order statistics of the combined sample of $N \equiv m + n$ observations and denote the associated left-continuous empirical df by H_n. Thus, $H_n = \lambda_N F_m + (1 - \lambda_N) G_n$ where $\lambda_N \equiv m/N$. For $\lambda \in [0,1]$, set $H_\lambda \equiv \lambda F + (1-\lambda)G$.

Let $R_1 < \cdots < R_m$ denote the ordered ranks of the X's in the combined sample; set $\underline{R} \equiv (R_1, \ldots, R_m)$. Define the risk vectors $\underline{M} \equiv (M_1, \ldots, M_N)$ and $\underline{N} \equiv (N_1, \ldots, N_N)$ by $M_k \equiv m(1-F_m(W_k))$ and $N_k \equiv n(1-G_n(W_k))$ for $k = 1, \ldots, N$. In Cox's (1972) terminology, M_k of the X's and N_k of the Y's are "at risk" at time W_k.

Denote the identity function and Lebesgue measure on $[0,1]$ by I. Unless noted otherwise, all integrals are over $\mathbb{R}^+ \equiv [0,\infty)$. For real numbers a and b, set $a \wedge b \equiv \min\{a,b\}$.

2. Estimates of relative risk.

If F is continuous and $\bar{G} = \bar{F}^{\theta}$, then $dG/dF = \theta\bar{F}^{\theta-1}$ and hence

(2.1) $$(1-F)dG = \theta(1-G)dF.$$

A natural starting point for estimating θ in the PHM is the identity

(2.2) $$\theta = \int J(F,G)(1-F)dG/\int J(F,G)(1-G)dF,$$

where $J: [0,1]^2 \to \mathbb{R}$ is any score function for which

(2.3) $$0 < \int |J(F,G)|(1-G)dF = \int_0^1 |J(1-I,\ 1-I^{\theta})| I^{\theta} dI < \infty .$$

For a suitable collection of score functions J (defined below), a family $\{\hat{\theta}_J\}$ of nonparametric estimates of θ is obtained by replacing F and G in (2.2) with F_m and G_n respectively:

$$\hat{\theta}_J = \int J(F_m,G_n)(1-F_m)dG_n/\int J(F_m,G_n)(1-G_n)dF_m.$$

Remark 2.1 The estimate $\hat{\theta}_J$ may be expressed in terms of the ordered ranks of the observations. Under the PHM the distribution of the rank vector, and hence of the rank estimate $\hat{\theta}_J$, depends on θ only and not on the continuous df F. (See Lehmann (1953), p. 26.) The convergence in our asymptotic results is uniform for all continuous dfs F. We indicate this by saying that the convergence holds "uniformly in F."

Remark 2.2 With the identity score function $J \equiv 1$, (2.2) becomes

$$\theta = \int (1-F)dG/\int (1-G)dF = P\{Y<X\}/P\{X<Y\}$$

where $X \sim F$ is independent of $Y \sim G$. The corresponding estimate is $\hat{\theta}_{J\equiv 1} \underset{=}{\text{a.s.}}$ $(1-\hat{U})/\hat{U}$ where $\hat{U} = \int (1-G_n)dF_m$ is the Mann-Whitney-Wilcoxon statistic. Consequently, $\hat{\theta}_{J\equiv 1}$ is an attractive estimate since, regardless of whether the PHM holds, it *always* consistently estimates a natural quantity of interest. (Efron (1967) discusses an analogue of $\hat{U}$ for arbitrarily right censored data.)

The purpose of the score function J in the rank estimate $\hat{\theta}_J$ is to assign weights to the ranks, giving heaviest weight to those ranks which are most sensitive to the value of θ. When the relative risk $\theta > 1$, the df F is stochastically larger than the df G, implying that the X's tend to be larger than the Y's. This general tendency is reflected in the ordering of the largest observations, X_m and Y_n, from the two samples.

Proposition 2.3 Under a PHM,

$$P\{Y_n < X_m \mid \theta\} \to \begin{cases} \lambda & \text{if} \quad \theta = 1 \\ 1 & \text{if} \quad \theta > 1 \\ 0 & \text{if} \quad \theta < 1 \end{cases}$$

uniformly in F as m, n $\to \infty$ such that $\lambda_N \to \lambda \in (0,1)$.

Since the value of the parameter θ determines, in probability, the ordering of the largest observations from the two samples, the score function J is allowed to grow unbounded as both of its coordinates approach one. The shape of the score function is motivated by the locally most powerful rank test (LMPRT) of $\theta = \theta_\circ$ versus $\theta > \theta_\circ$. Specifically, the LMPRT is based on the statistic $T_N(\theta_\circ)$ where

$$T_N(\theta) \equiv \theta \int J_{*N}(F_m, G_n)(1-G_n)dH_N \tag{2.4}$$

and

$$J_{*N}(s,t) \equiv \{\lambda_N(1-s) + \theta(1-\lambda_N)(1-t)\}^{-1} \text{ for } (s,t) \in [0,1]^2. \tag{2.5}$$

To introduce a suitable collection of score functions, define the bounding function R by

$$R(s,t) \equiv \{(1-s) + (1-t)\}^{-1} \text{ for } (s,t) \in [0,1]^2.$$

Definition 2.4 Let $\mathcal{J}(\alpha_o)$ denote the collection of all real-valued functions J defined on $[0,1]^2$, excluding the function $J \equiv 0$, which satisfies the following conditions:

(J1) J is differentiable on $[0,1)^2$.

(J2) There are constants $M > 0$ and $\alpha < \alpha_o$ for which $|J| \leq MR^{\alpha}$ on $[0,1]^2$ and the first partial derivatives of J are bounded (in absolute value) by $MR^{\alpha+1}$ on $[0,1)^2$.

(J3) The first partial derivatives of J are continuous on $[0,1)^2$.

The power α of the bounding function R is called the *growth rate* of J. The condition $|J| \leq MR^{\alpha}$ ensures the integrability condition (2.3) only if $\alpha < 2$. Our proof of strong consistency of the estimator $\hat{\theta}_J$ requires J to satisfy the differentiability condition (J1) and the boundedness condition (J2) with $\alpha_o = 2$. To prove asymptotic normality, we impose stronger smoothness and boundedness conditions by requiring $J \in \mathcal{J}(3/2)$. The strengthened boundedness requirements correspond to the usual restriction of a finite variance for asymptotic normality, as opposed to just a first moment for almost sure convergence.

Theorem 2.5 Suppose J satisfies (J1) and (J2) with growth rate $\alpha < 2$. Then, under a PHM, $\hat{\theta}_J \to_{a.s.} \theta$ as $m,n \to \infty$ uniformly in F.

Theorem 2.6 Suppose $J \in \mathcal{J}(3/2)$ and $\lambda_N \to \lambda \in [0,1]$ as $m,n \to \infty$. Then, under a PHM,

$(mn/N)^{\frac{1}{2}} (\hat{\theta}_J - \theta) \to_d N(0, \theta^2\sigma^2(J))$ as $m,n \to \infty$ uniformly in F, where $\sigma^2(J) \equiv \sigma^2(J,\theta,\lambda)$ is given by

$$(2.6) \qquad \sigma^2(J) \equiv \frac{\int J^2(F,G)(1-F)(1-G)dH_\lambda}{\{\int J(F,G)(1-F)dG\} \cdot \{\int J(F,G)(1-G)dF\}} .$$

To stabilize asymptotic variance, define $\beta \equiv \log \theta$ and $\hat{\beta}_J \equiv \log \beta_J$.

Corollary 2.7 Under the conditions of Theorem 2.2,

$$(mn/N)^{1/2}(\hat{\beta}_J - \beta) \to_d N(0, \sigma^2(J))$$

as $m,n \to \infty$ uniformly in F.

Remark 2.8 In practice J usually depends on N through λ_N, and we replace J by a sequence $\{J_N\}$. Theorems 2.5 and 2.6 and Corollary 2.7 apply to the corresponding sequence of estimates provided $\{J_N\}$ satisfies our boundedness condition (J2) uniformly in N (for N sufficiently large). Generally, this condition requires that $m,n \to \infty$ such that $\lambda_N \to \lambda \in (0,1)$. For example, Crowley (1975) suggested the sequence of score functions $J_{CN}(F,G) \equiv \{\lambda_N(1-F) + (1-\lambda_N)(1-G)\}^{-1}$ which satisfy condition (J2) uniformly in N with $\alpha=1$ and $M = (\lambda_* \wedge 1 - \lambda^*)^{-1}$ provided $0 < \lambda_* \le \lambda_N \le \lambda^* < 1$.

The change of variables $1-F = I$ in (2.6) shows that $\sigma^2(J) \equiv \sigma^2(J,\theta,\lambda)$ does *not* involve the continuous df F. This observation, combined with Remark 2.1, shows that both the finite sample and asymptotic distributions of $\hat{\theta}_J$ and $\hat{\beta}_J$ are independent of F. In fact, $\sigma^2(J) \equiv \sigma^2(J,\theta,\lambda)$ depends only on the score function J, the parameter θ, and the limiting proportion λ.

Proposition 2.9 Under the PHM with $0 < \lambda < 1$, $\sigma^2(J)$ is minimized over score functions J in $\mathcal{J}(3/2)$ uniformly in F by J_* where

$$J_*(s,t) \equiv \{\lambda(1-s) + \theta(1-\lambda)(1-t)\}^{-1} \text{ for } (s,t) \in [0,1]^2.$$

Furthermore,

$$\begin{aligned} (2.7) \qquad \sigma^2(J_*) &= \{\textstyle\int J_*(F,G)(1-F)dG\}^{-1} \\ &= \{\textstyle\int_0^1 [\lambda+\theta(1-\lambda)I^{(\theta-1)/\theta}]^{-1}\, dI\}^{-1} \equiv \sigma_*^2 \ . \end{aligned}$$

The inverse of the second expression for $\sigma^2(J_*) \equiv \sigma_*^2$ in (2.7) appears in equation (4.9) of Efron (1977) as the asymptotic information number for $\beta \equiv \log \theta$ from Cox's partial likelihood function. By the Delta method, $\theta^2\sigma_*^2$ is the asymptotic variance of $\hat{\theta}_c \equiv \log \hat{\beta}_c$ where $\hat{\beta}_c$ is the root of Cox's partial likelihood function.

Although the asymptotic variance of $\hat{\theta}_{J_*}$ is also $\theta^2\sigma_*^2$, $\hat{\theta}_{J_*}$ is not an estimate of θ since the optimal score function J_* involves the unknown parameter θ. A two-step estimate is required: First obtain a preliminary (consistent) estimate $\hat{\theta}_o$; then estimate θ by $\hat{\theta}_{\hat{J}_*}$ where

$$\hat{J}_*(s,t) \equiv \{\lambda(1-s) + \hat{\theta}_o(1-\lambda)(1-t)\}^{-1} \text{ for } (s,t) \in [0,1]^2.$$

(An obvious choice for the preliminary estimate is $\hat{\theta}_{J\equiv 1}$ because it is consistent, easy to compute, and always interpretable. See Remark 2.2.) Provided the preliminary estimate is consistent, the two-step estimate is asymptotically equivalent to Cox's estimate $\hat{\theta}_c$, and asymptotically achieves the minimum asymptotic variance $\theta^2\sigma_*^2$ for our family of estimates $\{\hat{\theta}_J\}$ indexed by J in $J(3/2)$.

<u>Theorem 2.10</u> Suppose $\hat{\theta}_o$ is a consistent estimate of θ and $\lambda_N \to \lambda \in (0,1)$ as $m,n \to \infty$. Under the PHM,

$$(mn/N)^{1/2}(\hat{\theta}_{\hat{J}_*} - \theta) \to_d N(0,\theta^2\sigma_*^2) \text{ as } m,n \to \infty.$$

Furthermore, if $\hat{\theta}_o \to_p \theta$ uniformly in F, then the convergence in distribution is also uniform in F.

<u>Remark 2.11</u> Our requirement that $\hat{\theta}_o$ be consistent is weaker than the usual requirement that the preliminary estimate in a two-step approximation to the root of a likelihood equation be $n^{-1/4}$ consistent.

The obvious extension of our family of estimates of θ to censored data substitutes the Kaplan-Meier estimates of F and G in place of the empirical dfs. However, this extension to censored data is not satisfactory because the asymptotic variance of the censored data estimates is very sensitive to the "heaviness" of the censoring. For further discussion and an alternative extension, see Begun and Reid (1981).

3. Regular rank estimates of θ.

Both the two-step estimate and Cox's estimate of θ are asymptotically optimal (in the sense of minimum asymptotic variance) among the family of nonparametric rank estimates $\{\hat{\theta}_J\}$ indexed by J in $\mathcal{J}(3/2)$. Naturally, the next question is whether these estimates are also asymptotically optimal among *all* regular *rank* estimates of θ. This question leads us to consider the Fisher information for estimating θ based on the rank vector. Under the PHM, the distribution of the rank vector depends on the parameter θ but not on the continuous df F. Thus the reduction to rank estimates eliminates dependence on the (unspecified) df F and produces a parametric setting with only the parameter θ. An expression for the Fisher information can be written explicitly using the probability distribution of the rank vector $\underline{R}$:

$$P\{\underline{R} = \underline{r}|\theta\} = P\{\underline{M} = \underline{m}|\theta\} = P\{\underline{N} = \underline{n}|\theta\} = m!n!\theta^n / \prod_{k=1}^{N} (m_k + \theta n_k) \tag{3.1}$$

(cf. Savage (1956), p. 599). However, the dependence among members of the rank vector complicates its evaluation. Instead of calculating the Fisher information directly, we address the question of optimality by studying the asymptotic behavior of the likelihood ratio of the rank vector when the alternative hypothesis is a local alternative. Applying Hájek's (1970) representation theorem shows that both our two-step estimate and Cox's estimate achieve full asymptotic efficiency among all regular rank estimates of θ.

The condition of regularity of estimates of θ is implied by uniform (in θ) convergence in law of the estimates; it excludes superefficient estimates. For $h \in \mathbb{R}$, set $\theta_N \equiv \theta + h(mn/N)^{-\frac{1}{2}}$ for $m \geq m_h$ and $n \geq n_h$, where m_h and n_h denote the smallest pair of integers so that $m \geq m_h$ and $n \geq n_h$ implies $\theta_N > 0$.

Definition 3.1 (Hajek (1970), p. 324) A rank estimate E_N of θ, is called *regular* if, for all $x \in \mathbb{R}$,

$$P\{(mn/N)^{\frac{1}{2}}(E_N-\theta_N) \leq x\,|\,\theta_N\} \to L(x)$$

at all continuity points $x \in \mathbb{R}$ of the limiting df L.

Proposition 3.2 The rank estimates, $\hat{\theta}_c$, $\hat{\theta}_{\hat{J}_*}$, and $\hat{\theta}_{J_N}$ with $J_N \in J(3/2)$ uniformly in N, are regular estimates of θ, provided $\lambda_N \to \lambda \in (0,1)$.

Define the likelihood ratio of the rank vector $\underline{R} = \underline{r}$ by

$$L_N(h,\underline{r}) \equiv P\{\underline{R} = \underline{r} \mid \theta_N\}/P\{\underline{R} = \underline{r} \mid \theta\}$$

for $m \geq m_h$ and $n \geq n_h$. Thus $L_N(h,\underline{r})$ measures the likelihood of the observed rank vector under the PHM with parameter θ relative to a local alternative which converges to θ at the rate $(mn/N)^{-\frac{1}{2}}$.

Theorem 3.3 For all $h \in \mathbb{R}$, $m \geq m_h$, and $n \geq n_h$, the likelihood ratio of the rank vector $\underline{R}$ may be expressed as

$$L_N(h,\underline{R}) = \exp\{h\Delta_N(R) - \tfrac{1}{2}h^2\Gamma_N + T_N(h,\underline{R})\} \tag{3.2}$$

where $\Gamma_N \equiv (\theta^2\sigma_{*N}^2)^{-1}$, $\Gamma \equiv (\theta^2\sigma_*^2)^{-1}$, and

$$\sigma_{*N}^{-2} \equiv \int_0^1\{\lambda_N + \theta(1-\lambda_N)^{(\theta-1)/\theta}\}^{-1}dI$$

$$\to \int_0^1\{\lambda + \theta(1-\lambda)I^{(\theta-1)/\theta}\}^{-1}dI \equiv \sigma_*^{-2},$$

and where, under the PHM with parameter θ,

$$\Delta_N(\underline{R}) \to_d N(0,\Gamma) \text{ and}$$

$$T_N(h,\underline{R}) \to_p 0 \text{ for all } h \in \mathbb{R}$$

as $m,n \to \infty$ such that $\lambda_N \to \lambda \in (0,1)$.

In other words, for all $h \in \mathbb{R}$ under the PHM with parameter θ, the likelihood ratio of the rank vector converges in distribution to a random variable Z for which $\log Z \sim N(-\frac{1}{2}h^2\Gamma, h^2\Gamma)$. By LeCam's first lemma (Hájek and Sidak (1967), p. 203) the probability measures $P\{\cdot|\theta\}$ and $P\{\cdot|\theta_N\}$ are contiguous for all $h \in \mathbb{R}$.

Hájek (1970) proved a general representation theorem for the limiting law of regular estimates of a k-dimensional vector of parameters. For the special case of regular rank estimates of the parameter θ in a PHM, Hájek's hypotheses are verified in Theorem 3.3. The following result is a direct application of Hájek's theorem.

<u>Theorem 3.4</u> Suppose E_N is a regular rank estimate of θ with limiting df L; i.e., for all $x \in \mathbb{R}$,

$$P\{(mn/N)^{\frac{1}{2}}(E_N-\theta_N) \le x|\theta_N\} \to L(x)$$

at all continuity points of the df L. Then

$$L(x) = \int_{-\infty}^{\infty} \Phi(x-u|\Gamma^{-1})dK(u)$$

where $\Phi(\cdot|\Gamma^{-1})$ denotes the normal df with mean zero and variance Γ^{-1}, and $K(u)$ is a df on $\mathbb{R}$ that depends on the choice of the estimate E_N.

In other words, if E_N is a regular rank estimate of θ, then under the PHM with parameter θ_N,

$$(mn/N)^{\frac{1}{2}}(E_N-\theta_N) \to_d N(0,\Gamma^{-1}) + A$$

for all $h \in \mathbb{R}$, where $N(0,\Gamma^{-1}) = N(0,\theta^2\sigma_*^2)$ and A are independent and A has df K. Furthermore, the random variable A depends on the choice of the estimate. In particular, the asymptotic variance of a regular rank estimate is bounded below by $\Gamma^{-1} \equiv \theta^2\sigma_*^2$. Since both our two-step estimate and Cox's estimate are regular rank estimates with asymptotic variances which achieve this lower bound, then both are asymptotically optimal among all regular rank estimates of θ.

4. Measures of relative risk.

In the spirit of Bickel and Lehmann's (1975) work on descriptive statistics for nonparametric models, consider the problem of comparing two dfs when the corresponding hazard functions are *not* necessarily proportional. Let $\mathcal{F}$ denote the collection of all continuous dfs on $\mathbb{R}^+ \equiv [0, \infty)$.

Definition 4.1 For *any* two continuous dfs $F, G \in \mathcal{F}$, a measure of relative risk $\theta(F,G)$ is defined as a nonnegative-valued functional satisfying the following conditions.

(i) If $\bar{G} = \bar{F}^{\theta}$ with $0 < \theta < \infty$ then $\theta(F,G) = \theta$.

(ii) (Order) If $F_1 <_s F_2$ then $\theta(F_1,G) \le \theta(F_2,G)$ for all $G \in \mathcal{F}$; and if $G_1 <_s G_2$ then $\theta(F,G_1) \ge \theta(F,G_2)$ for all $F \in \mathcal{F}$. (Here $<_s$ denotes stochastic ordering.)

(iii) (Symmetry) For all $F,G \in \mathcal{F}$, $\theta(G,F) = [\theta(F,G)]^{-1}$.

The following three examples show that conditions (i)-(iii) are independent.

(a) The functional $\theta(F,G) = \int \bar{F}(x)dx / \int \bar{G}(x)dx = EX/EY$ where $X \sim F$ and $Y \sim G$ satisfies (ii) and (iii) but not (i);

(b) The functional $\theta(F,G) = [-2 \log \bar{G}(y) + \log \bar{G}(x)]/[-2 \log \bar{F}(y) + \log \bar{F}(x)]$ where $0 < x < y < \infty$ satisfies (i) and (iii) but not (ii);

(c) finally, $\theta(F,G) = \int[-\log \overline{G}/ -\log \overline{F}]d\mu$, where μ is a fixed probability measure on $\mathbb{R}^+$, satisfies (i) and (ii) but not (iii).

For a fixed score function J on $[0,1]^2$, define the functional $\theta(F,G;J)$ on $\mathcal{F} \times \mathcal{F}$ by

$$\theta(F,G;J) \equiv \int J(F,G)(1-F)dG/\int J(F,G)(1-G)dF.$$

In addition to conditions (J1) - (J3) given in Definition 2.1, define the following conditions.

(J4) For all $(s,t) \in [0,1]^2$, $J(s,t) = J(t,s)$.

(J5) For all $(s,t) \in [0,1]^2$, $J(s,t)(1-s)$ is nonincreasing in s and $J(s,t)(1-t)$ is nonincreasing in t.

<u>Proposition 4.2</u> If J satisfies (J4) and (J5), then $\theta(F,G;J)$ is a measure of relative risk.

Both the identity score function $J \equiv 1$ and the score function $J_c(F,G) = \{\lambda(1-F) + (1-\lambda)(1-G)\}^{-1}$ suggested by Crowley (1975) satisfy conditions (J4) and (J5). (For J_c, we interchange the roles of λ and $1-\lambda$ when we interchange those of F and G.)

Estimate $\theta(F,G;J)$ by $\theta(F_m,G_n;J)$. To describe the asymptotic behavior of these estimates, we introduce the following notation. Define the continuous measure K by

$$dK \equiv (1-F)dG - \theta(F,G;J)(1-G)dF;$$

and notice from (2.1) that, under a PHM, $dK = 0$. Set $J_1(s,t) \equiv \partial J(s,t)/\partial s$ and $J_2(s,t) \equiv \partial J(s,t)/\partial t$. Finally, denote the functions $J(F,G)$, $J_1(F,G)$, and $J_2(F,G)$ as simply as J, J_1, and J_2, respectively.

<u>Theorem 4.3</u> If J satisfies (J1)-(J5) with growth rate $\alpha \leq 1$ then

$$(F_m,G_n;J) \rightarrow_{a.s.} \theta(F,G;J) \text{ and}$$

$$(mn/N)^{1/2}[\theta(F_m,G_n;J) - \theta(F,G;J)] \rightarrow_d N(0,\mathcal{T}(J))$$

as $m,n \rightarrow \infty$ such that $\lambda_N \rightarrow \lambda \in (0,1)$, where

$$
\begin{aligned}
T(J) \equiv \{&\lambda\int J^2(1-F)^2 dG + (1-\lambda)\theta^2(F,G;J)\int J^2(1-G)^2 dF \\
&-2\lambda\theta(F,G;J)\int[\int_0^{\bullet} J dF]J dK + 2(1-\lambda)\int[\int_0^{\bullet} J dG]J dK \\
&+2\lambda\theta(F,G;J)\int[\int_0^{\bullet} J dF]J_2(1-G)dK - 2(1-\lambda)\int[\int_0^{\bullet} J dG]J_1(1-F)dK \\
&-2\lambda\int[\int_0^{\bullet} J_2 dK]J dK \quad -2(1-\lambda)\int[\int_0^{\bullet} J_1 dK]J dK \\
&+2\lambda\int[\int_0^{\bullet} J_2 dK]J_2(1-G)dK + 2(1-\lambda)\int[\int_0^{\bullet} J_1 dK]J_1(1-F)dK \\
&-\lambda[\int J_2(1-G)dK]^2 - (1-\lambda)[\int J_1(1-F)dK]^2\}/[\int J(1-G)dF]^2.
\end{aligned}
$$

For F and G absolutely continuous with densities f and g, Kalbfleisch and Prentice (1981) introduced the functionals

$$\eta \equiv \eta(F,G;\mu) \equiv \frac{\int (g/\bar{G})\{(f/\bar{F}) + (g/\bar{G})\}^{-1} d\mu}{\int (f/\bar{F})\{(f/\bar{F}) + (g/\bar{G})\}^{-1} d\mu}$$

where μ is a fixed df on $\mathbb{R}^+$. The numerator (denominator) of η can be interpreted as the ratio of the hazard function of G (F) to the total hazard function, weighted with respect to $d\mu$. For fixed μ, the functional $\eta(F,G;\mu)$ satisfies conditions (i) and (iii) for a measure of relative risk, but not condition (ii). However, Kalbfleisch and Prentice (1981) focus attention on the dfs

$$\mu_{KP} = 1-\bar{F}^{\gamma}\bar{G}^{\gamma} \text{ with } 0 < \gamma < \infty$$

which depend on F and G. Then $\eta(F,G;\mu_{KP})$ can be written as $\theta(F,G;J_{KP})$ where $J_{KP}(F,G) \equiv \gamma\, \bar{F}^{\gamma-1}\, \bar{G}^{\gamma-1}$. For $0 < \gamma < \infty$ the score function J_{KP} satisfies (J4) and (J5), and hence, by Proposition 4.2, $\eta(F,G;\mu_{KP}) = \theta(F,G;J_{KP})$ is a measure of relative risk. Although J_{KP} with $\frac{1}{2} \leq \gamma \leq 1$ does not satisfy the conditions of Theorem 4.2, the asymptotic variance of $\eta(F_m,G_n;\mu_{KP})$ (when suitably centered and normalized) derived from equation (8) of Kalbfleisch and Prentice (1981) is equal to $T(J_{KP})$. Kalbfleish and Prentice (1981) also extend their results to censored data.

5. Proofs

We outline the proofs of the main results. Further details are given in Begun (1981).

Proof of Proposition 2.3 Note that $P\{Y_n < X_m | \theta\} = \int G^n dF^m$. When $\theta = 1$, write $P\{Y_n < X_m | \theta\} = \int F^n dF^m = \lambda_N \to \lambda$. When $\theta > 1$, use the change of variables $F = 1 - (x/m)$ and apply the dominated convergence theorem. When $\theta < 1$, write the PHM as $\bar{F} = \bar{G}^{1/\theta}$ and note that $1/\theta > 1$. □

Proof of Theorem 2.5 By straightforward algebra,

(5.1) $$\hat{\theta}_J - \theta = \{\tilde{K}_N - \tilde{K} - \theta(K_N - K)\}/ K_N$$

where

$$\tilde{K}_N \equiv \tilde{K}_N(J) \equiv \int J(F_m, G_n)(1-F_m)dG_n,$$

$$K_N \equiv K_N(J) \equiv \int J(F_m, G_n)(1-G_n)dF_m,$$

(5.2) $$\tilde{K} \equiv \tilde{K}(J) \equiv \int J(F,G)(1-F)dG, \text{ and}$$

$$K \equiv K(J) \equiv \int J(F,G)(1-G)dF.$$

Using Wellner's (1977a, Theorem 1) strengthened version of the Glivenko-Cantelli theorem, and Wellner's (1977b, Theorem 1) almost sure "nearly linear" bounds for an empirical df, one can show that

(5.3) $$\tilde{K}_N - \tilde{K} \to_{a.s.} 0 \text{ and } K_N - K \to_{a.s.} 0,$$

and hence, by (5.1), that $\hat{\theta}_J - \theta \to_{a.s.} 0$. □

Proof of Theorem 2.6 We replace the two independent empirical processes $m^{1/2}(F_m - F)$ and $n^{1/2}(G_n - G)$ with two specially constructed empirical processes $U_m(F)$ and $V_n(G)$ which converge to two special independent processes in the strong sense that, for any $\delta > 0$,

$$\sup_{0 \le x < \infty} |U_m(F(x)) - U(F(x))|/(1-F(x))^{\frac{1}{2}-\delta} \to_p 0 \text{ as } m \to \infty \text{ and}$$

(5.4)

$$\sup_{0 \le x < \infty} |V_n(G(x)) - V(G(x))|/(1-G(x))^{\frac{1}{2}-\delta} \to_p 0 \text{ as } n \to \infty.$$

The special processes U(F) and V(G) are independent, mean-zero Gaussian processes on $\mathbb{R}^+$ with continuous sample paths and covariance structure

$$\text{Cov}[U(F(s), U(F(t))] = F(s) \wedge F(t) - F(s)F(t) \text{ and}$$

$$\text{(5.5)} \qquad \text{Cov}[V(G(s)), V(G(t))] = G(s) \wedge G(t) - G(s)G(t) \text{ for } s, t \in \mathbb{R}^+.$$

Since the original processes have the same distributions as the special ones, all conclusions concerning convergence in distribution also hold for the original processes. (See the Appendix of Shorack (1972).)

Replacing the original processes with the special ones, and using (5.4) and Shorack's (1972, Lemma A3) linear bounds for an empirical df, one can show that

$$\text{(5.6)} \qquad ((mn/N)^{\frac{1}{2}}(K_N - K),\ (mn/N)^{\frac{1}{2}}(\tilde{K}_N - \tilde{K}))$$

$$\to_p ((1-\lambda)^{\frac{1}{2}}A + \lambda^{\frac{1}{2}}B - \lambda^{\frac{1}{2}}C + (1-\lambda)^{\frac{1}{2}}D,$$

$$(1-\lambda)^{\frac{1}{2}}\tilde{A} + \lambda^{\frac{1}{2}}\tilde{B} - (1-\lambda)^{\frac{1}{2}}\tilde{C} + \lambda^{\frac{1}{2}}\tilde{D})$$

as $m,n \to \infty$ so that $\lambda_N \to \lambda \in [0,1]$, where

$$A \equiv \int J_1(F,G)U(F)(1-G)dF,\quad \tilde{A} \equiv \int J_1(F,G)U(F)(1-F)dG,$$

$$B \equiv \int J_2(F,G)V(G)(1-G)dF,\quad \tilde{B} \equiv \int J_2(F,G)V(G)(1-F)dG,$$

$$C \equiv \int J(F,G)V(G)dF,\qquad \tilde{C} \equiv \int J(F,G)U(F)dG,$$

$$D \equiv -\int U(F)d[J(F,G)(1-G)],\quad \tilde{D} \equiv -\int V(G)d[J(F,G)(1-F)],$$

$J_1(s,t) \equiv \partial J(s,t)/\partial s$ and $J_2(s,t) \equiv \partial J(s,t)/\partial t$. By (2.1), $\tilde{A} - \theta A = 0$ and $\tilde{B} - \theta B = 0$, so that, by (5.1), (5.3), and (5.6),

$$(mn/N)^{\frac{1}{2}}(\hat{\theta}_J - \theta) \to_p \{-(1-\lambda)^{\frac{1}{2}}\tilde{C} + \lambda^{\frac{1}{2}}\tilde{D} + \theta\lambda^{\frac{1}{2}}C - \theta(1-\lambda)^{\frac{1}{2}}D\}/K \equiv Z_J.$$

The limiting random variable Z_J is normally distributed with mean zero and

(5.7) $$\text{Var } Z_J = \theta\{(1-\lambda)\text{Var }(\tilde{C} + \theta D) + \lambda\theta^2\text{Var }(C + \theta^{-1}D)\}/K\tilde{K}$$

since $\theta = K/\tilde{K}$. Combining calculations (2), (4), and (5) of the Appendix with (2.1) yields

(5.8) $$\text{Var }(\tilde{C} + \theta D) = \theta\int J^2(F,G)(1-F)(1-G)dG.$$

Since the PHM can be written as $\overline{F} = \overline{G}^{1/\theta}$, then by interchanging the roles of F and G and replacing θ with $1/\theta$, it follows that

(5.9) $$\text{Var }(C + \theta^{-1}\tilde{D}) = \theta^{-1}\int J^2(F,G)(1-F)(1-G)dF.$$

From (5.7), (5.8), and (5.9) we have that $\text{Var } Z_J = \theta^2\sigma^2(J)$ where $\sigma^2(J)$ is given by (2.6). □

<u>Proof of Proposition 2.9</u> Since $H_\lambda \equiv \lambda F + (1-\lambda)G$ and $\overline{G} = \overline{F}^{\theta}$, then $dH_\lambda = [\lambda + \theta(1-\lambda)(1-F)^{\theta-1}]dF$ so that

(5.10) $$(1-F)dH_\lambda = J_*^{-1}(F,G)dF.$$

Applying (5.10) to (2.6) gives (2.7).

Using the Cauchy-Schwarz inequality, we have that, for all J in $\mathcal{J}(3/2)$,

$$\begin{aligned}
&[\int J(F,G)(1-F)dG][\int J(F,G)(1-G)dF] \\
&\quad = \theta[\int J(F,G)(1-G)dF]^2 \text{ by (2.1)} \\
&\quad \leq \theta[\int J^2(F,G)(1-G)J_*^{-1}(F,G)dF][\int J_*(F,G)(1-G)dF] \\
&\quad = [\int J^2(F,G)(1-F)(1-G)dH_\lambda][\int J_*(F,G)(1-F)dG] \text{ by (2.1) and (5.10)}
\end{aligned}$$

or, after rearranging terms, that $\sigma^2(J_*) \leq \sigma^2(J)$. □

Proof of Theorem 2.10 With the notation of (5.2), the spirit of (5.1), and the identity

$$(5.11)\qquad \hat{J}_*(s,t) - J_*(s,t) = (\hat{\theta}_\circ - \theta)J_\circ(s,t)$$

where $J_\circ(s,t) \equiv -(1-\lambda)(1-t)\, J_*(s,t)\, \hat{J}_*(s,t)$ for $(s,t)\epsilon\ [0,1]^2$, we find that

$$\begin{aligned}(mn/N)^{\frac{1}{2}}(\hat{\theta}_{\hat{J}_*}-\theta) &= (mn/N)^{\frac{1}{2}}(\hat{\theta}_{J_*}-\theta) + (mn/N)^{\frac{1}{2}}(\hat{\theta}_{\hat{J}_*}-\hat{\theta}_{J_*})\\ &= (mn/N)^{\frac{1}{2}}(\hat{\theta}_{J_*}-\theta) + (\theta_\circ-\theta)\{(mn/N)^{\frac{1}{2}}[\tilde{K}_N(J_\circ) - \tilde{K}(J_\circ)]\\ &\quad - \theta(mn/N)^{\frac{1}{2}}[K_N(J_\circ) - K(J_\circ)] - K_N(J_\circ)(mn/N)^{\frac{1}{2}}(\hat{\theta}_{J_*}-\theta)\}/K_N(\hat{J}_*).\end{aligned}$$

By Theorem 2.6, $(mn/N)^{\frac{1}{2}}(\hat{\theta}_{J_*} - \theta) \to_d N(0,\ \theta^2\sigma_*^2)$. Since $\hat{\theta}_\circ$ is consistent, then $\hat{\theta}_\circ - \theta = o_p(1)$.

Given $\varepsilon > 0$, there is a subset $\Omega_\circ \equiv \Omega_\circ(\varepsilon,\theta)$ having $P\{\Omega_\circ\} > 1-\varepsilon$ on which $|\hat{\theta}_\circ-\theta| < \theta/2$ for m and n sufficiently large. On the subset $\Omega_\circ$ with m and n sufficiently large, $J_\circ$ is in $\mathcal{J}(3/2)$ uniformly in N. Hence, by Remark 2.8 and (5.6),

$$(mn/N)^{\frac{1}{2}}\ (\tilde{K}_N(J_\circ) - \tilde{K}(J_\circ)) = O_p(1) \text{ and } (mn/N)^{\frac{1}{2}}\ (K_N(J_\circ) - K(J_\circ)) = O_p(1);$$

and by (5.3), $K_N(J_\circ) = O_p(1)$ and

$$\begin{aligned}K_N(\hat{J}_*) &= [K_N(\hat{J}_*) - K_N(J_*)] + K_N(J_*)\\ &= (\hat{\theta}_\circ - \theta)K_N(J_\circ) + K_N(J_*) \text{ by (5.10)}\\ &= o_p(1)O_p(1) + O_p(1) = O_p(1). \quad \square\end{aligned}$$

Proof of Theorem 3.3 Before expanding the likelihood ratio, we introduce several useful identities. In the spirit of (5.10), with J_{*N} defined in (2.5) and $H \equiv \lambda_N F + (1-\lambda_N)G$, we have that

$$(5.12)\qquad (1-F)dH = J_{*N}^{-1}(F,G)dF.$$

Hence, it follows that

(5.13) $$\int J_{*N}(F,G)(1-G)dH = \int(1-G)(1-F)^{-1}dF = \int(1-F)^{\theta-1}dF = \theta^{-1}$$

and that

(5.14) $$n(mn/N)^{-1}\theta^{-2} - \lambda_N^{-1}(1-\lambda_N)\int J_{*N}^2(F,G)(1-G)^2dH$$

$$= (\lambda_N\theta)^{-1}\int J_{*N}(F,G)(1-G)[1-\theta(1-\lambda_N)(1-G)J_{*N}(F,G)]dH \text{ by (5.13)}$$

$$= \theta^{-1}\int J_{*N}(F,G)(1-G)(1-F)J_{*N}(F,G)dH$$

$$= \Theta^{-2}\int J_{*N}(F,G)(1-F)dG \quad \text{by (5.12) and (2.1)}$$

$$= \theta^{-2}\int_0^1\{\lambda_N + \theta(1-\lambda_N)I^{(\theta-1)/\theta}\}^{-1}dI \equiv (\theta_2\sigma_{*N}^2)^{-1} \equiv \Gamma_N$$

by the change of variables $1-G = I$. Finally, for $\ell = 1$ and 2,

(5.15) $$(mn/N)^{-\ell/2}\sum_{k=1}^N (N_k/(M_k + \theta N_k))^\ell$$

$$= \lambda_N^{-1}(1-\lambda_N)^{\ell-1}(mn/N)^{(2-\ell)/2}\int J_{*N}^\ell(F_m,G_n)(1-G_n)^\ell dH_N.$$

Using (3.1), the expansion $\log(1+x) = x-x^2/2 + x^3/3(1+\nu)^3$ where $\nu \equiv \nu(x)$ is between zero and x, (5.14) and (5.15), we can express the likelihood ratio $L_N(h,\underline{R})$ in the form of (3.2) with

$$\Delta_N(\underline{R}) \equiv -\lambda_N^{-1}(mn/N)^{1/2}\{\int J_{*N}(F_m,G_n)(1-G_n)dH_N - \int J_{*N}(F,G)(1-G)dH\},$$

$$T_N(h,\underline{R}) \equiv \tfrac{1}{2}h^2T_{N1}(\underline{R}) + h^3 T_{N2}(h,\underline{R})/3 + h^3T_{N3}(h,\underline{R})/3,$$

$$T_{N1}(R) \equiv \lambda_N^{-1}(1-\lambda_N)\{\int J_{*N}^2(F_m,G_n)(1-G_n)^2dH_N - \int J_{*N}^2(F,G)(1-G)^2dH\},$$

$$T_{N2}(h,\underline{R}) \equiv (mn/N)^{-3/2}\sum_{k=1}^N [N_k/(M_k+\theta N_k)(1+\nu_{N2}(k))]^3, \text{ and}$$

$$T_{N3}(h,\underline{R}) \equiv n(mn/N)^{-3/2}\theta^{-3}/(1+\nu_{N1})^3$$

where ν_{N1} is between zero and $h(mn/N)^{-1/2}\theta^{-1}$ and $\nu_{N2}(k)$ is between zero and $h(mn/N)^{-1/2}\, N_k/(M_k+\theta N_k)$ for $k = 1, \ldots, N$.

Using the idea of Remark 2.8 and the techniques applied in the proof of Theorem 2.6, one can show that $\Delta_N(\underline{R}) \to_d \Gamma Z_{J_*}$ provided $\lambda_N \to \lambda \epsilon (0,1)$ where $Z_{J_*} \sim N(0,\Gamma^{-1})$ and $\Gamma \equiv (\theta^2 \sigma_*^2)^{-1}$. Using the idea of Remark 2.8 and the techniques mentioned in the proof of Theorem 2.5, one can show that $T_{N1}(\underline{R}) \to_p 0$ provided $\lambda_N \to \lambda \in (0,1)$. For all $k = 1, \ldots, N$, since $N_k/(M_k + \theta N_k) \le \theta^{-1}$, then $\nu_{N1} \wedge \nu_{N2}(k) \ge -\frac{1}{2}$ for all m and n sufficiently large. Hence, $T_{N2}(h,\underline{R}) \to 0$ and $T_{N3}(h,R) \to 0$ provided $\lambda_N \to \lambda \in (0,1)$. □

Proof of Proposition 3.2 Using Remark 2.8 and the special processes of Theorems 2.6 and 3.3, we conclude that, with $J_N \in \mathcal{J}(3/2)$ uniformly in N and $h \in \mathbb{R}$,

$$((mn/N)^{\frac{1}{2}}(\hat{\theta}_{J_N} - \theta),\ \log L_N(h,\underline{R}))' \to_p (Z_J,\ h\Gamma Z_{J_*} - \tfrac{1}{2}h^2\Gamma)' \equiv \underline{Z}$$

under a PHM with parameter θ provided $\lambda_N \to \lambda \in (0,1)$. Then $\underline{Z}$ has a bivariate normal distribution with

$$E\{\underline{Z}\} = \begin{pmatrix} 0 \\ -\frac{1}{2}h^2\Gamma \end{pmatrix} \quad \text{and} \quad \mathrm{Cov}\{\underline{Z}\} = \begin{pmatrix} \theta^2\sigma^2(J) & h \\ h & h^2\Gamma \end{pmatrix}$$

by (1), (3), and (5) of the Appendix. By LeCam's third lemma (cf. Hájek and Šidak (1967), p. 208), $\hat{\theta}_{J_N}$ is a regular rank estimate of θ.

A similar argument with Theorem 2.6 replaced by Theorem 2.10 establishes the regularity of the two-step estimate.

With uncensored data, Cox's estimate of θ is simply the Maximum Likelihood Estimate based on the ranks. Using a Taylor expansion of the score function of the ranks, one can show that $(mn/N)^{\frac{1}{2}}(\hat{\theta}_c - \theta) \to_d Z_{J_*}$ provided $\lambda_N \to \lambda \in (0,1)$. The regularity of $\hat{\theta}_c$ follows the LeCam's third lemma provided $\lambda_N \to \lambda \in (0,1)$. □

Proof of Theorem 4.3. The arguments for convergence are similar to those given in the proofs of Theorems 2.5 and 2.6. The evaluation of $T(J)$ involves calculations similar to those given in the Appendix. We omit the details. □

APPENDIX

Suppose $\overline{G} = \overline{F}^{\,\theta}$ with $0 < \theta < \infty$ and F continuous. Suppose J and $\tilde{J}$ are in $J(3/2)$, and denote the functions $J(F,G)$ and $\tilde{J}(F,G)$ as simple J and $\tilde{J}$, respectively. We repeatedly use (2.1), (5.5), and the identity $F = 1-(1-F)$.

(1) $\mathrm{Cov}[\int J\, U(F)dG, \int \tilde{J}\, U(F)dG]$

$$= \iint J(u)\tilde{J}(v)[F(u) \wedge F(v) - F(u)F(v)]dG(u)dG(y)$$

$$= \int[\int_0^{\bullet} J\, F\, dG]\, \tilde{J}(1-F)dG + \int[\int_{\bullet}^{\infty} J(1-F)dG]\, \tilde{J}\, F\, dG$$

$$= \int[\int_0^{\bullet} JdG]\, \tilde{J}(1-F)dG + \int[\int_0^{\bullet}\tilde{J}dG]\, J(1-F)dG$$

$$-[\int J(1-F)dG][\int\tilde{J}(1-F)dG].$$

(2) $$\mathrm{Var}[\int J\, U(F)dG] = 2\int[\int_0^{\bullet} JdG]\, J(1-F)dG - [\int J(1-F)dG]^2$$
by (1) with $\tilde{J} = J$.

(3) $\mathrm{Cov}[U(F)d[J(1-G)], \int \tilde{J}\, U(F)dG]$

$$= \iint \tilde{J}(v)[F(u) \wedge F(v) - F(u)F(v)]\, d[J(u)(1-G(u)]dG(v)$$

$$= \int\{\int_0^{\bullet} F\, d[J(1-G)]\}\tilde{J}(1-F)dG$$

$$+ \int\{\int_{\bullet}^{\infty} (1-F)d[J(1-G)]\}\tilde{J}\, F\, dG$$

$$= \int [J \tilde{J} F(1-F)(1-G)dG$$

$$- \int [\int_0^{\bullet} J(1-G)dF]\tilde{J}(1-F)dG$$

$$- \int J \tilde{J} F(1-F)(1-G)dG$$

$$+ \int [\int_{\bullet}^{\infty} J(1-G)dF]\tilde{J}dG$$

$$- \int [\int_{\bullet}^{\infty} J(1-G)dF]\tilde{J}(1-F)dG$$

$$= \int [\int_0^{\bullet} \tilde{J}dG]J(1-G)dF - \theta[\int J(1-G)dF][\int \tilde{J}(1-G)dF]$$

(4) $\mathrm{Cov}[\int U(F)d[J(1-G)], \int J\, U(F)dG]$

$$= \int [\int_0^{\bullet} JdG]J(1-G)dF - \theta[\int J(1-G)dF]^2 \text{ by (3) with } \tilde{J} = J.$$

(5) $\mathrm{Var}[\int U(F)d[J(1-G)]]$

$$= \mathrm{Var}[J(X)(1-G(X))] \text{ where } X \sim F$$

$$= \theta^{-1}\int J^2(1-F)(1-G)dG - [\int J(1-G)dF]^2.$$

Acknowledgement. This paper is based on part of the author's Ph.D. dissertation written under the direction of Jon A. Wellner at the University of Rochester. The author thanks the referees for their help with the revision.

REFERENCES

BEGUN, J.M. (1981) A class of rank estimators of relative risk. *Institute of Statistics Mimeo Series #1381*, Department of Statistics, University of North Carolina at Chapel Hill.

BEGUN, J.M. and REID, N. (1981). Estimating the relative risk with censored data. *Technical Report No. 81-16*, Institute of Applied Mathematics and Statistics, University of British Columbia.

BICKEL, P.J. and LEHMANN, E.L. (1975). Descriptive Statistics for nonparametric models I and II. *Ann. Statist.*, *3*, 1038-1069.

COX, D.R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B.*, *34*, 187-202.

COX, D.R. (1975). Partial Likelihood. *Biometrika*, *62*, 269-276.

CROWLEY, J. (1975). Estimation of relative risk in survival studies. *Technical Report No. 423*, Department of Statistics, University of Wisconsin.

EFRON, B. (1967). The two-sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. and Prob.*, 831-853. University of California Press.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.*, *72*, 557-565.

HÁJEK, J. (1970). A characterization of the limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, *14*, 323-330.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

KALBFLEISCH, J.D. and PRENTICE, R.L. (1981). Estimation of the average hazard ratio. *Biometrika*, *68*, 105-112.

LEHMANN, E.L. (1953). The power of rank tests. *Ann. Math. Statist.* *24*, 23-42.

SAVAGE, I.R. (1956). Contribution to the theory of rank order statistics: the two-sample case. *Ann. Math. Statist.*, *27*, 590-615.

SHORACK, G.R. (1972). Functions of order statistics. *Ann. Math. Statist.*, *43*, 412-427.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Ann. Statist.*, *9*, 93-108.

WELLNER, J.A. (1977a). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.*, *5*, 473-480.

WELLNER, J.A. (1977b). A law of the iterated logarithm for functions of order statistics. *Ann. Statist.*, *5*, 481-494.