The Annals of Statistics
1990, Vol. 18, No. 3, 1203–1220

ASYMPTOTIC PROPERTIES OF A GENERAL CLASS OF NONPARAMETRIC TESTS FOR SURVIVAL ANALYSIS[1]

BY MICHAEL P. JONES AND JOHN CROWLEY

University of Iowa and Fred Hutchinson Cancer Research Center and University of Washington

Many of the popular nonparametric test statistics for censored survival data used in two-sample, s-sample trend and single continuous covariate situations are special cases of a general statistic, differing only in the choice of covariate-based label and weight function. Formulated in terms of counting processes and martingales this general statistic, standardized by the square root of its consistent variance estimator, is shown to be asymptotically normal under the null hypothesis and under a sequence of contiguous hazard alternatives that includes both relative and excess risk models. As an application to some specific cases of the general statistic, the asymptotic relative efficiencies of the Brown, Hollander and Korwar modification of the Kendall rank statistic, the Cox score statistic and the generalized logrank statistic of Jones and Crowley are investigated under relative and excess risk models. Finally, an example is given in which the Cox score test is not as efficient as the generalized logrank test in the presence of outliers in the covariate space.

1. Introduction. Tarone and Ware (1977) introduced a class of s-sample statistics for right-censored survival data that includes the logrank test of Mantel (1966) and the generalized Wilcoxon procedures of Gehan (1965) and Breslow (1970). Subsequently, this class has been generalized to include the two-sample statistics of Efron (1967), Peto and Peto (1972) and Harrington and Fleming (1982), a group of procedures optimal against time-transformed location alternatives of the survival function developed by Gill (1980, Section 5.3) and the s-sample linear rank statistics of Prentice (1978). More recently, Jones and Crowley (1989) introduced a more general class of single-covariate nonparametric tests for right-censored survival data that includes the Tarone–Ware two-sample class, the Cox (1972) score test, the Tarone (1975) and Jonckheere (Gehan, 1965) s-sample trend statistics, the Brown, Hollander and Korwar (1974) modification of the Kendall rank statistic, the linear rank statistics of Prentice (1978), the logit rank statistic of O'Brien (1978) and several new procedures. This class can be generalized to include the Tarone–Ware s-sample class.

Within the last 10 years the theory of martingales applied to counting processes has found widespread use in the study of asymptotic properties of survival analysis procedures. Resulting theorems have quite general results

Received July 1987; revised July 1989.
[1]Work supported by PHS Fellowship T32 CA09168-11 and CA39065.
AMS 1980 *subject classifications*. Primary 63E20; secondary 60G44, 62G20.
Key words and phrases. Failure-time data, censoring, nonparametrics, martingale, counting process, relative risk, excess risk, Cox score test, Kendall rank test, logrank, asymptotic relative efficiency, covariate distributions, robustness.

and their proofs are both intuitive and straightforward. Aalen (1975, 1978) introduced the use of martingale theory to survival analysis. Gill (1980) followed up Aalen's work with an in-depth study of the general class of two-sample statistics. Since then the s-sample problem [Andersen, Borgan, Gill and Keiding (1982)] and Cox regression [Andersen and Gill (1982), Prentice and Self (1983)] have been studied using martingale theory. Andersen and Borgan (1985) give an excellent review of further applications of counting processes to life history data. Here, in a similar fashion, we shall study the large sample properties of the class of single-covariate statistics. In Section 2 we shall introduce the W statistic and its variance estimator. Section 3 contains a theorem concerning the weak convergence of W and the consistency of its variance estimator. In Section 4 we derive the asymptotic distribution of W under contiguous relative and excess risk alternatives and compare the asymptotic relative efficiencies (ARE) of the Cox score test, Kendall rank statistic and generalized logrank test (to be defined in the next section). In Section 5, the AREs of these tests are compared when there are outliers in the covariate space.

2. The W statistic; martingale framework. In this section we shall state the hazard model, give the hypotheses of interest and formulate the general class of nonparametric tests proposed by Jones and Crowley (1989) in the framework of martingales and counting processes. The necessary mathematical background can be found in Chapter 2 of Gill (1980). First we must fix some notation and definitions. Let us assume there is given a sequence of fixed complete probability spaces $(\Omega^{(n)}, \mathscr{F}^{(n)}, P^{(n)})$ along with a family of right-continuous, nondecreasing complete sub-σ-algebras $\{\mathscr{F}_t^{(n)}\colon 0 \le t < \infty\}$ which correspond to the history of the survival study up to and including time t, $n = 1, 2, \dots,$ where n is the sample size of the study. In this paper we assume an individual can fail at most once. Let T_j and C_j represent the times to failure and censoring, respectively, for the jth individual. The at-risk indicator $Y_j(t)$ is a predictable process which assumes the value 1 when $t \le T_j \wedge C_j$ and 0 otherwise. Define the observed failure counting process $N_j(t) = I[T_j \le t, C_j \ge T_j]$, where I is the indicator function. A dot in place of a subscript will be used to denote summation over that subscript, e.g., $N_.(t) = \sum N_j(t)$. Throughout this paper we allow only a single covariate $X_j(t)$, which is assumed to be an adapted, locally bounded, left-continuous process with right-hand limits. We shall always be working under the general random censorship model, i.e., for any $h > 0$, $P[T \in [t, t+h), C \in [t, t+h) | \underline{X}(t)] = P[T \in [t, t+h) | \underline{X}(t)] \cdot P[C \in [t, t+h) | \underline{X}(t)]$, where $\underline{X}(t) = \{X(u)\colon 0 \le u \le t\}$. The hazard of failure

$$\lambda\left[t|\underline{X}(t)\right] = \lim_{h \downarrow 0} h^{-1} P\left[T \in [t, t+h) | T \ge t, \underline{X}(t)\right] \tag{1}$$

under the random censorship model can equivalently be written as

$$Y(t+)\lambda\left[t+|\underline{X}(t)\right] = \lim_{h \downarrow 0} h^{-1} P\left[N(t+h) - N(t) = 1 | \mathscr{F}_t\right].$$

In Section 3.1 the hazard function will be extended to having both continuous and discrete components. To allow for contiguous hazard alternatives, we shall allow the hazard λ and cumulative hazard Λ to depend on n. Throughout we shall assume the covariate has been properly constructed so that the jth individual's hazard at time t is a function of $X_j(t)$, i.e, $d\Lambda_j^n(t) = d\Lambda^n[t|\underline{X}_j(t)] = d\Lambda^n[t|X_j(t)]$. Under the above assumptions and certain regularity conditions it follows from Theorem 2.5.1 of Dolivo (1974) that

$$M_j(t) = N_j(t) - \int_0^t Y_j(s)\, d\Lambda^n\big[s|X_j(s)\big] \tag{2}$$

are martingales over $[0,\infty)$. Furthermore, they are square-integrable martingales and

$$\langle M_i, M_j\rangle = I[i=j]\int Y_j(s)\, d\Lambda^n\big[s|X_j(s)\big]. \tag{3}$$

Let $\mathcal{X}$ represent the space of potential covariate paths or some subspace of it and let Λ_x be the cumulative hazard for the path $x \in \mathcal{X}$. The null hypothesis would be written formally as H_0: $\Lambda_x^n = \Lambda$ over $[0,\infty)$ for all $x \in \mathcal{X}$ and all n and the contiguous hazard alternative as H_{ca}: $\sup|d\Lambda_x^n - d\Lambda| \to 0$ as $n \to \infty$, where the sup is taken over $[0,\infty)$ and all $x \in \mathcal{X}$. The type of alternative hypothesis considered in this paper is one in which the hazard rate $d\Lambda_x$ is monotone in x. One note of caution is necessary here. Internal covariates $X_j(s)$ [as defined by Kalbfleisch and Prentice (1980)] which predict the outcome $N_j(t)$ for some $t > s$ should be excluded from consideration here, since $M_j(t)$ need not be a martingale.

The test statistic is based on assigning covariate-derived quantitative labels to each individual study. Define the label process $Z_j(t) = g_t(X_j(t)|\mathscr{F}_{t-})$, where g_t is a monotone $\mathscr{F}_{t-}$-measurable function such that $Z_j(t)$ is also locally bounded and left-continuous with right-hand limits. The covariates $\{X_j(t), j = 1,\dots,n\}$ may actually be unobserved at time t, but the labels $\{Z_j(t), j = 1,\dots,n\}$ must be known. The most obvious choice is $Z_j(t) = X_j(t)$. As another example, letting $r_j(t)$ be the rank of $X_j(t)$ among the covariates at risk $\{X_l(t): Y_l(t) = 1\}$, we might select $Z_j(t) = r_j(t)/Y_\cdot(t)$. Other candidates for $Z_j(t)$ appear in Jones and Crowley (1989). As a general class of statistics let us propose

$$W_n(t) = n^{-1/2}\int_0^t w(s)\sum_{j=1}^n Y_j(s)\big[Z_j(s) - \bar{Z}(s)\big]\, dN_j(s), \tag{4}$$

where $\bar{Z}(s) = Y_\cdot^{-1}(s)\Sigma Y_j(s)Z_j(s)$ is the average label in the risk set at time s and $w(s)$ is a locally bounded, predictable weight function. Note that under H_0, $d\Lambda_j(s) = Y_j(s)\, d\Lambda(s)$ so that by (2), $W_n(t)$ is itself a local martingale, which is sufficient motivation for basing a test on (4). If there is a single failure at time t, then $dW_n(t)$ is equal to the weight $w(t)n^{-1/2}$ times the deviation of the failing individual's label from the average label of those at risk at t. In the two-sample problem where $Z_j(t) = 0$ or 1, by using more traditional notation, $W_n(t) = n^{-1/2}\,\Sigma w_i(O_i - E_i)$, where O_i is the observed number of failures in group 2 at the ith failure time, E_i is its expected number and the summation is over all failure times up to and including t. In this setting, $W_n(t)$ is the logrank test or Cox score test when $w = 1$ but is the Gehan test when $w = Y_./n$. In the more general covariate setting $W_n(t)$ is the Cox score test when $Z_j = X_j$ and $w = 1$; $W_n(t)$ is the Brown, Hollander and Korwar (1974) modification of the Kendall rank statistic when $Z_j = r_j/Y_.$ and $w = Y_./n$. As proposed by Jones and Crowley (1989), let us define the generalized logrank test to be $W_n(t)$ when $Z_j = r_j/Y_.$ and $w = 1$. In the two-sample problem this is the usual logrank test or Cox score test. However, in the general covariate setting, it differs from the Cox score test through its use of the ranks of the covariates at risk. The proposed variance estimator for $W_n(t)$ is

$$(5)\quad V_n(t) = n^{-1}\int_0^t w^2(s)\sum_{j=1}^n Y_j(s)\left[Z_j(s) - \bar{Z}(s)\right]^2 \frac{Y_.(s) - \Delta N_.(s)}{Y_.(s) - 1}\frac{dN_.(s)}{Y_.(s)},$$

where $\Delta N(s) = N(s) - N(s-)$. Allowing a discrete component to the hazard function (cf. Section 3.1) and thereby tied failure times so that $\Delta N_.(s) \geq 1$, $V_n(t)$ is an unbiased estimator for $\mathrm{Var}(W_n(t))$ under H_0 for bounded processes $\{wY_j(Z_j - \bar{Z})\}$. This follows from arguments similar to Propositions 3.2.2 and 3.3.1 of Gill (1980).

3. Weak convergence; consistency of the variance estimator. In this section we shall state sufficient conditions for the weak convergence of a more general version of $W_n(t)$ to a normal process and for the consistency of its variance estimator, both under a contiguous sequence of alternative hypotheses. This will allow us to calculate asymptotic relative efficiencies in Section 4. For now the hazard is assumed to be continuous; in Section 3.1 some notes are given for hazards having both continuous and discrete components.

For convenience let us generalize the W_n statistic (4) to

$$(6)\qquad W_n(t) = n^{-1/2}\int_0^t \sum_{j=1}^n H_j(s)\, dN_j(s),$$

where H_j is a locally bounded predictable process subject to the two constraints for $t \in [0, \infty)$: $H_j(t) = Y_j(t)H_j(t)$ and $H_.(t) = 0$. H_j will often depend on n. As a variance estimator for $W_n(t)$, let us use

$$(7)\qquad V_n(t) = n^{-1}\int_0^t \sum_{j=1}^n H_j^2(s)\frac{Y_.(s) - \Delta N_.(s)}{Y_.(s) - 1}\frac{dN_.(s)}{Y_.(s)}.$$

Substitution of $H_j = wY_j(Z_j - \bar{Z})$ into (6) and (7) yields (4) and (5). Using (2), the statistic can be decomposed as $W_n(t) = E_n(t) + U_n(t)$, where

$$(8)\qquad E_n(t) = n^{-1/2}\int_0^t \sum_{j=1}^n H_j(s)\, d\Lambda^n\left[s|X_j(s)\right],$$

$$(9)\qquad U_n(t) = n^{-1/2}\int_0^t \sum_{j=1}^n H_j(s)\, dM_j(s).$$

Under H_0, $E_n(t) \equiv 0$. Since $U_n(t)$ is a sum of integrals of locally bounded predictable processes with respect to square-integrable martingales, $U_n(t)$ is itself a local square-integrable martingale. When the $\{H_j\}$ are bounded, $U_n(t)$ is a martingale and by the martingale property $E[U_n(t)] = 0$ for all t and hence $E(W_n) = E(E_n)$. E_n can be thought of as the trend component and U_n as the error or martingale component of W_n.

The following theorem deals with the asymptotic normality of U_n and the consistency of V_n under H_{ca}.

THEOREM 3.1. *Assume there exists an interval I of the form $[0, u)$ or $[0, u]$ such that the following conditions hold.*

CONDITION 3.1.1. (a) $\Lambda(t) < \infty$ *for all* $t \in I$.

(b) *The H_{ca} model holds, i.e.,*

$$\sup_{j, [0,\infty)} \left| d\Lambda_j^n - d\Lambda \right| \to_p 0 \quad \text{as } n \to \infty$$

(c) *There exists a function y, such that* $\inf_{t \in I} y(t) > 0$ *and for all* $t \in I$,

$$\sup_{[0,t]} \left| \frac{Y_\cdot}{n} - y \right| \to_p 0 \quad \text{as } n \to \infty$$

(d) *There exists a left-continuous, nonnegative function v with right-hand limits such that for all $t \in I$,*

$$\sup_{[0,t]} \left| Y_\cdot^{-1} \sum_{j=1}^{n} H_j^2 - v \right| \to_p 0 \quad \text{as } n \to \infty$$

and v is bounded over all closed subintervals of I and zero off I.

(e)
$$\sup_{j,[0,t]} n^{-1/2} |H_j| \to_p 0 \quad \text{as } n \to \infty,\ t \in I.$$

CONDITION 3.1.2. *If $u \notin I$, then assume*

(a)
$$\int_I yv \, d\Lambda < \infty.$$

(b)
$$\lim_{t \uparrow u} \limsup_{n \to \infty} P\left\{ n^{-1} \int_{(t,u]} \sum_{j=1}^{n} H_j^2 \, d\Lambda_j^n > \varepsilon \right\} = 0 \quad \text{for all } \varepsilon > 0.$$

There exists a function v_1, bounded on $[0, u]$, such that for all $\varepsilon > 0$,

(c)
$$\sup_{[0,u]} \left| Y_\cdot^{-1} \sum_{j=1}^{n} H_j^2 - v_1 \right| \to_p 0.$$

(d)
$$\lim_{t \uparrow u} \limsup_{n \to \infty} P\left\{ \int_{(t,u]} v_1 \left[n^{-1} \sum_{j=1}^{n} Y_j \, d\Lambda_j^n \right] > \varepsilon \right\} = 0.$$

CONDITION 3.1.3. *If $u < \infty$, then assume*

(a) $$n^{-1} \int_{(u,\infty)} \sum_{j=1}^{n} H_j^2 \, d\Lambda_j^n \to_p 0 \quad \text{as } n \to \infty.$$

There exists a function v_2, bounded on $[0, \infty)$, such that for all $\varepsilon > 0$,

(b) $$\sup_{[0,\infty)} \left| Y_\cdot^{-1} \sum_{j=1}^{n} H_j^2 - v_2 \right| \to_p 0 \quad \text{as } n \to \infty.$$

(c) $$\limsup_{n \to \infty} P\left\{ \int_{(u,\infty)} v_2 n^{-1} \sum_{j=1}^{n} Y_j \, d\Lambda_j^n > \varepsilon \right\} = 0.$$

Then under Condition 3.1.1, $U_n \to_d$ *(a zero mean Gaussian process with independent increments and variance function $\int_0^t yv \, d\Lambda$ in $D(I)$). Furthermore, for each $t \in I$,*

$$\sup_{s \in [0,t]} \left| V_n(s) - \int_0^s yv \, d\Lambda \right| \to_p 0 \quad \text{as } n \to \infty.$$

Adding Condition 3.1.2 *extends these results to* $[0, u]$ *and adding Condition* 3.1.3 *extends them to* $[0, \infty)$.

The proof of this theorem follows in a straightforward manner from well-known martingale methods, in particular Rebelledo's martingale limit theorem [cf. Theorem 2.4.1 of Gill (1980)]. We will briefly comment on the conditions of the theorem; we consider the special case $H_j = wY_j(Z_j - \bar{Z})$. The interval I is critical to the theorem. Since $\{t | v(t) > 0\} \subset \{t | y(t) > 0\}$, it seems reasonable to consider, as long as $\lim w > 0$, $I = \{t | v(t) > 0, \Lambda(t) < \infty\}$, where Λ is the baseline cumulative hazard, i.e., the one associated with the covariable path $X \equiv 0$. In the two-sample problem, I reduces to Gill's [(1980) page 92] interval $\{t | y_1(t) \wedge y_2(t) > 0\}$, where $y_i(t)$ is the limiting proportion of group i individuals still at risk at t. The function $y(t)$ of Condition 3.1.1(c) is just the proportion of individuals at risk at time t. Since $Y_\cdot^{-1} \sum H_j^2 = w^2 Y_\cdot^{-1} \sum Y_j (Z_j - \bar{Z})^2$, the function $v(t)$ of Condition 3.1.1(d) is the limit of the square of the weight process times the sample variance of the Z labels at risk at time t. The weight function w is typically a uniformly bounded process so that Condition 3.1.1(e) reduces to

(10) $$n^{-1/2} \sup_{j, [0,t]} Y_j |Z_j| \to_p 0,$$

which is equivalent to the Lindeberg-type condition of Andersen and Gill [(1982), Section 4(c)]. As they point out (10) is easily verified for (a) bounded $\{Z_j\}$, (b) $\{Z_j\}$ which are bounded by random variables having a bounded rth moment, some $r > 2$, and (c) i.i.d. $\{Z_j\}$ with the bounded second moment condition $E \sup YZ^2 < \infty$, where the supremum is taken over $[0, t]$.

Several choices for the label Z_j were introduced by Jones and Crowley (1989), many of which lead to tests which are in common use. When $Z_j = X_j$, the raw covariate, (10) follows if (a), (b) or (c) is true. Labels based on ranked covariates are dealt with easily. Let $r_j(t)$ be the rank of $X_j(t)$ among those at risk at t, i.e, $\{X_l(t)|Y_l(t) = 1\}$. The rank label $Z_j(t) = r_j(t)/Y.(t) \leq 1$ and so satisfies (10). If $\{Z_j\}$ represent the expected or approximate order statistics from a distribution G_t, then (10) is satisfied by choosing G_t such that (a), (b) or (c) is satisfied; e.g., $Z_j(t) = \Phi^{-1}\{(r_j(t) - \frac{1}{2})/Y.(t)\}$ and $Z_j(t) = \text{logit}\{(r_j(t) - \frac{1}{2})/Y.(t)\}$, where Φ is the normal distribution function [cf. O'Brien (1978)]. Finally, if $Z_j(t) = \Psi(X_j(t) - \bar{X}(t))$, where Ψ downweights outlier as is done in M-estimation procedures, then Ψ must be chosen so that (10) is satisfied.

Conditions 3.1.2 and 3.1.3 state that contributions to W_n and V_n from the complement of I are asymptotically negligible. The condition $H.= 0$ is not used in Theorem 3.1. Note that under H_0, $E_n \equiv 0$ so that $W_n(t)/\sqrt{V_n(t)}$ is asymptotically a $N(0, 1)$ random variable for $t \in I$, $[0, u]$ or $[0, \infty)$ depending on which set of conditions holds.

3.1. *Mixed hazard functions.* So far the hazard of failure has been assumed to be continuous. It is of some interest, albeit theoretical, that the martingale framework of Section 2 and the theorem of Section 3 can be extended to a more general hazard. In particular, consider a mixed hazard characterized by $d\Lambda^n[t|X(t)] = \lambda_c^n[t|X(t)]\,dt + \lambda_d^n[t|X(t)]$, where the continuous component λ_c^n is defined by (1) and the discrete component $\lambda_d^n[t|X(t)]$ by $P[T = t|T \geq t, X(t)]$. Under the general random censorship model this is equivalent to

$$Y(t)\lambda_d^n[t|X(t)] = P[\Delta N(t) = 1|\mathscr{F}_{t-}].$$

Although λ_c^n as given by (1) is totally general, complete generality of λ_d^n is not easy to deal with. Let us consider the special model

$$\lambda_{dj}^n(t) = \Delta\Lambda_d(t) + \Delta\Lambda_d^*(t)C_j^n, \tag{11}$$

where Λ_d and Λ_d^* are both increasing, positive step functions independent of n, $\Delta\Lambda_d = C_j^n = 0$ whenever $\Delta\Lambda_d^* = 0$, $\Lambda_d^* < \infty$ whenever $\Lambda_d < \infty$ and $\sup C_j^n \to 0$ a.s., where the sup is over j and $t \in [0, \infty]$. Under the mixed hazard using (11), $M_j(t)$ as defined by (2) is still a martingale; however, the covariation process $\langle M_i, M_j\rangle$ needs correction by substituting $(1 - \Delta\Lambda_j^n)\,d\Lambda_j^n$ for $d\Lambda_j^n$ in (3). Theorem 3.1 now holds for this more general hazard model if one makes the same $(1 - \Delta\Lambda)d\Lambda$ substitution into Conditions 3.1.2 and 3.1.3 and the resulting variance function of the theorem. The proof is now more complex than before and involves defining all our processes on a transformed time axis, establishing asymptotic normality on the new axis and then using a Skorohod-type construction to get normality on the original time axis. The details of the proof are given in a technical report [Jones and Crowley (1988)] which is patterned after Theorem 4.2.1 of Gill (1980).

4. Asymptotic relative efficiencies under relative and excess risk models. In this section we shall study the Pitman asymptotic relative efficiencies of several members of the W_n class of procedures under a contiguous sequence of hazard alternatives. For simplicity, failure hazards are assumed to be continuous. Let us consider the following model for contiguous hazard alternatives:

$$H_{ca}: \lambda_j^n(t) = \lambda(t)\left[\alpha(t) + \beta n^{-1/2} X_j(t) + O(n^{-1})\right]. \tag{12}$$

The baseline hazard is $\lambda(t)\alpha(t)$. When $\alpha(t) \equiv 1$, $\lambda_j^n(t)$ defines a relative risk (RR) model as linear when the $O(n^{-1})$ term is zero and as exponential when $O(n^{-1})$ is the remainder of the Taylor series expansion of $\exp[\beta n^{-1/2} X_j(t)]$. When $\lambda(t) \equiv 1$, $\lambda_j^n(t)$ defines an excess risk (ER) model.

We next need to find the limiting distribution of $W_n(t)/\sqrt{V_n(t)}$ under (12). Recall that $W_n(t) = E_n(t) + U_n(t)$, where from (8), (12) and the constraint $H. = 0$,

$$\begin{aligned} E_n(t) &= n^{-1/2} \int_0^t \sum_1^n H_j(u)\left[\lambda_j^n(u) - \lambda(u)\alpha(u)\right] du \\ &= \beta n^{-1} \int_0^t \sum_1^n H_j(u) X_j(u) \lambda(u)\, du + o_p(1). \end{aligned}$$

Using the usual form, $H_j = wY_j(Z_j - \bar{Z})$, note that $n^{-1}\sum H_j X_j = w[Y./n]C_{XZ}$ and $n^{-1}\sum H_j^2 = w^2[Y./n]C_{ZZ}$, where $C_{XZ}(t)$ is the sample covariance of the covariate X with the label Z at time t among those at risk. $C_{XZ}(t)$ uses the weight $1/Y.(t)$ rather than $1/(Y.(t) - 1)$. We will assume enough regularity conditions about the distributions of X and Z among those at risk at time t that the sample covariances converge in probability to bounded functions, viz. $C_{XZ} \to \sigma_{XZ}$ and $C_{ZZ} \to \sigma_{ZZ}$. We will also assume $w \to w_0$ in probability. Then assuming Condition 3.1.1 of Theorem 3.1 and noting $v = w_0^2 \sigma_{ZZ}$ in Condition 3.1.1(d), it follows from (13) and (14) that $W_n(t)/\sqrt{V_n(t)} \to_d N[\mu(t)/\sigma(t), 1]$, where

$$\mu(t) = \beta \int_0^t w_0(u) y(u) \sigma_{XZ}(u) \lambda(u)\, du, \tag{13}$$

$$\sigma^2(t) = \int_0^t w_0^2(u) y(u) \sigma_{ZZ}(u) \lambda(u) \alpha(u)\, du, \tag{14}$$

for all $t \in I$. Conditions 3.1.2 and 3.1.3 extend this range to $[0, u]$ and $[0, \infty)$, respectively.

Let $R_j^*(t) = Y.^{-1}(t) r_j(t)$, where $r_j(t) =$ rank of $X_j(t)$ among the risk set $\{l\colon Y_l(t) = 1\}$. The three versions of W_n to be considered here are the Cox (1972) score test (COX) ($Z_j = X_j, w = 1$), the generalized logrank test (GL) recently proposed by Jones and Crowley (1989) ($Z_j = R_j^*, w = 1$) and the Brown, Hollander and Korwar (1974) modification of the Kendall rank correlation (K) ($Z_j = R_j^*, w = Y./n$). By way of definition the efficacy of the GL test is

$\mu_{\mathrm{GL}}^2(t)/\sigma_{\mathrm{GL}}^2(t)$. The ARE of the GL test to the COX test is the ratio of their efficacies.

From now on let us assume that the covariates are fixed and at time zero represent an i.i.d. sample from some distribution. The submodels of (12) to be considered here for $\lambda_j^n(t)$ are

Model A: $\lambda\left(1 + \beta X_j/\sqrt{n}\right)$,

Model B: $2\lambda t\left(1 + \beta X_j/\sqrt{n}\right)$,

Model C: $2\lambda t + \lambda\beta X_j/\sqrt{n}$.

Model A is both a RR and an ER model. Its background component $\lambda(t)\alpha(t) = \lambda$ and its covariate component $\lambda\beta X_j/\sqrt{n}$ are constant over time. Model B is a RR model in which both the background and covariate components of the hazard are increasing in t. Model C is an ER model in which the background increases in t while the covariate component is constant in t. We shall model the censoring distribution through the continuous hazard $\psi_j(t) = \psi(\phi + \gamma X_j)$, which can be viewed either as a RR or ER model. If $\gamma > 0$, individuals with large X values are at greater risk of being censored.

In order to calculate $\mu(t)$ and $\sigma^2(t)$ from (13) and (14), we need to derive y, σ_{XX}, σ_{XR^*} and $\sigma_{R^*R^*}$. From (A.18) of Lehmann (1975), $\sigma_{R^*R^*} = 1/12$ for continuous covariates. Since $\lambda^n(t|X) = \lambda(t)(\alpha(t) + \beta n^{-1/2}X) \to \lambda(t)\alpha(t)$, we may for the computations of σ_{XX} and σ_{XR^*} consider the distribution $P_t(x)$ of the covariates at risk at time t in a model with hazard $\lambda(t)\alpha(t)$. Loosely speaking,

$$
\begin{aligned}
dP_t(x) &= P\left(X = x \text{ at } t | T \geq t, C \geq t\right) \\
&= \frac{P\left(T \geq t, C \geq t | X = x \text{ at } 0\right) dP_0(x)}{\int P\left(T \geq t, C \geq t | X = u \text{ at } 0\right) dP_0(u)} \\
&= \frac{S_T(t|x)\, S_C(t|x)\, dP_0(x)}{\int S_T(t|u)\, S_C(t|u)\, dP_0(u)},
\end{aligned} \tag{15}
$$

assuming a general random censorship model. S_T and S_C are the conditional survival functions for T and C. Define $\zeta(t) = \exp\{-\int_0^t \lambda(u)\alpha(u)\, du - \psi\phi t\}$ and $\eta(t) = \gamma\psi t$.

Then $S_T(t|x)S_C(t|x) = \zeta(t)\exp(-\eta(t)x)$ and hence from (15),

$$
dP_t(x) = \frac{e^{-\eta(t)x}\, dP_0(x)}{\int e^{-\eta(t)u}\, dP_0(u)}. \tag{16}
$$

Using (16) we can derive the variance function $\sigma_{XX}(t)$. The asymptotic proportion $y(t)$ of individuals still at risk at t is equal to the denominator of (15), i.e.,

$$
y(t) = \zeta(t) \int e^{-\eta(t)x}\, dP_0(x).
$$

Note that $P_t(x)$ is independent of $\alpha(t)$ and ϕ. If $\gamma = 0$, then $\eta(t) = 0$ and $y(t) = \zeta(t)$ and $P_t(x) = P_0(x)$.

Derivation of $\sigma_{XR^*}(t)$ is a little more involved. Let $X_1^0(t), \dots, X_{Y.}^0(t)$ represent the covariates at risk at t, indexed arbitrarily, and let $r_1^0(t), \dots, r_{Y.}^0(t)$ be the corresponding ranks. Let $X_{(1)}^0(t), \dots, X_{(Y.)}^0(t)$ be the order statistics. Then for continuous covariates

$$\begin{aligned} C_{xr}(t) &= Y_\cdot^{-1}(t) \sum_{j=1}^{Y.(t)} \left[X_j^0(t) - \bar{X}^0(t)\right]\left[r_j^0(t) - \bar{r}^0(t)\right] \\ &= Y_\cdot^{-1}(t) \sum_{j=1}^{Y.(t)} jX_{(j)}^0(t) - \tfrac{1}{2}[Y.(t) + 1]\bar{X}^0(t), \end{aligned}$$

which contains the form of an L-estimate. From page 368 of Lehmann (1983) concerning the limit of an L-estimate, it follows that

$$\frac{C_{xr}(t)}{Y.(t)} \to_p \sigma_{XR^*}(t) = \int_{P_t^{-1}(0)}^{P_t^{-1}(1)} yP_t(y)\, dP_t(y) - \tfrac{1}{2}E_tX,$$

where E_tX is the expectation of $P_t(x)$.

We are finally able to investigate the AREs for the COX, GL and K tests for models A, B and C and we do so for binary, uniform and exponential covariates. The appropriate values of y, σ_{XX}, σ_{XR^*} and $\sigma_{R^*R^*}$ for use in (13) and (14) can be found in the Appendix for these three covariate distributions.

First let us consider binary covariates. Substituting the values from the Appendix into (13) and (14), we find that the COX and GL tests have the same efficacy; since these tests are equivalent in the two-sample problem, this serves as a partial check on our calculations. In Model A if we let $\lambda = \beta = 1$ and let $\phi = 1$, $\gamma = 0$ in $\psi_j(t)$ (i.e., two-sample problem with equal censoring) and then compute the efficacies of COX, GL and K for $\psi = 0, 1, 2, 2.5, 3$ and 4, we arrive at Figure 5.2.1 of Gill (1980), but off by a scale factor of $p(1 - p)$. This scale factor is explained by Gill's use of a different standardizing constant from ours.

Next let us consider covariates which are $U(a, b)$ random variables at $t = 0$. The integrand terms of (13) and (14) are given in the Appendix. In the special case that $\eta(t) = 0$, which happens when $\psi = 0$ (no censoring) or $\gamma = 0$ (censoring independent of X), the efficacies of the COX and GL tests are equal. Hence in this setup for uniform covariates, the generalized logrank is asymptotically as efficient as the COX score test under H_{ca} given in (12). The reason for this is that when $\eta(t) = 0$, $P_t(x)$ is constant over time, i.e., $U(a, b)$; hence the expected order statistic is $(\bar{Y}(t) + 1)^{-1}$ times its rank so that the efficacies are equal.

For initially $U(0, 3)$ covariates and parameters ψ, ϕ, γ and α set to 1, the efficacies of the three tests are plotted in Figure 1 for Models A and C. The efficacy of each test is a function of time, with the interpretation that at time t all of the individuals still at risk are censored. The efficacies for Model B were

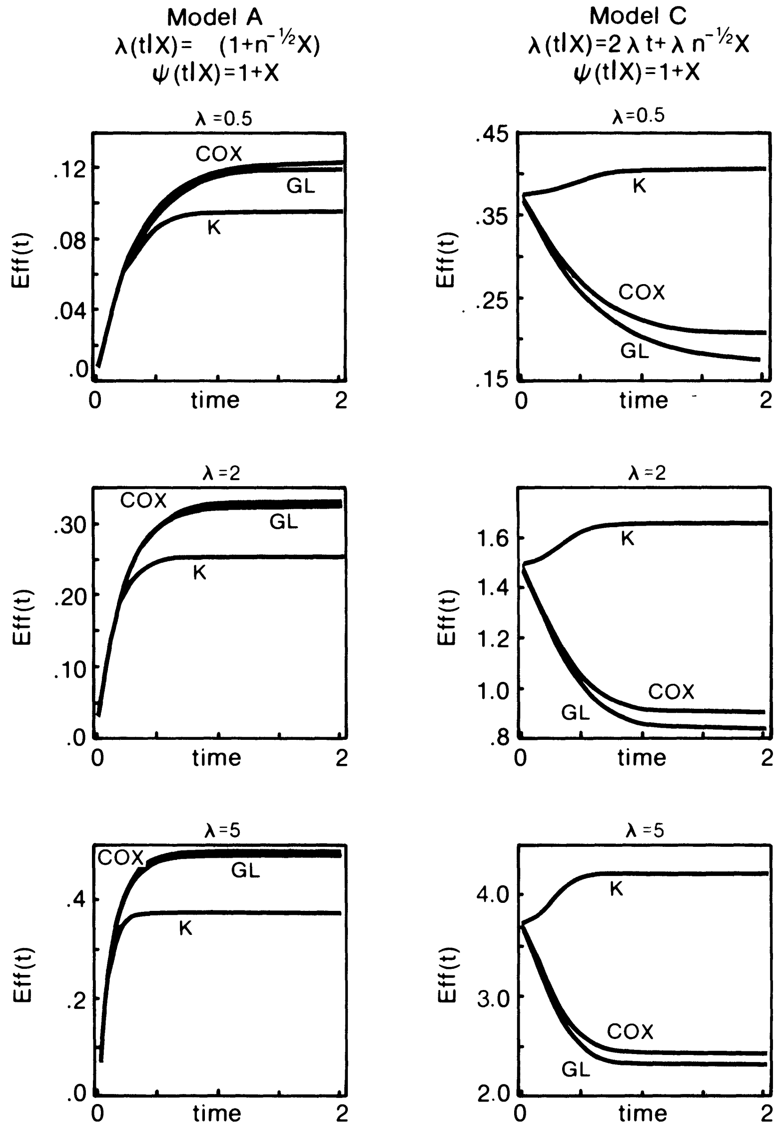

FIG. 1. *Efficacies for the Cox score* (COX), *generalized logrank* (GL) *and Kendall* (K) *statistics.* $\lambda(t|X)$ *is the failure hazard,* $\psi(t|X)$ *is the censoring hazard and* X *is a uniform*(0, 3) *covariate.*

nearly identical in character to those of Model A and were therefore omitted. Although β was set to 1 for the efficacy plots, its value is irrelevant since it cancels out in the ARE of two tests. In either model when the parameter λ is increased, the failure hazard increases relative to the censoring hazard so that more individuals fail. This increases the effective sample size and hence the power of each test. This can be seen in Figure 1, where increasing λ results in increasing efficacy. The left column of Figure 1 reflects the fact that the COX test was designed for proportional hazards models, such as Model A. The GL test is nearly as efficient. In the right column of Figure 1, K is the most powerful test for Model C whereas the efficacies of COX and GL are monotone decreasing. The reason for this is as follows. When t is near zero, the background component $2\lambda t$ of Model C is small relative to the covariate

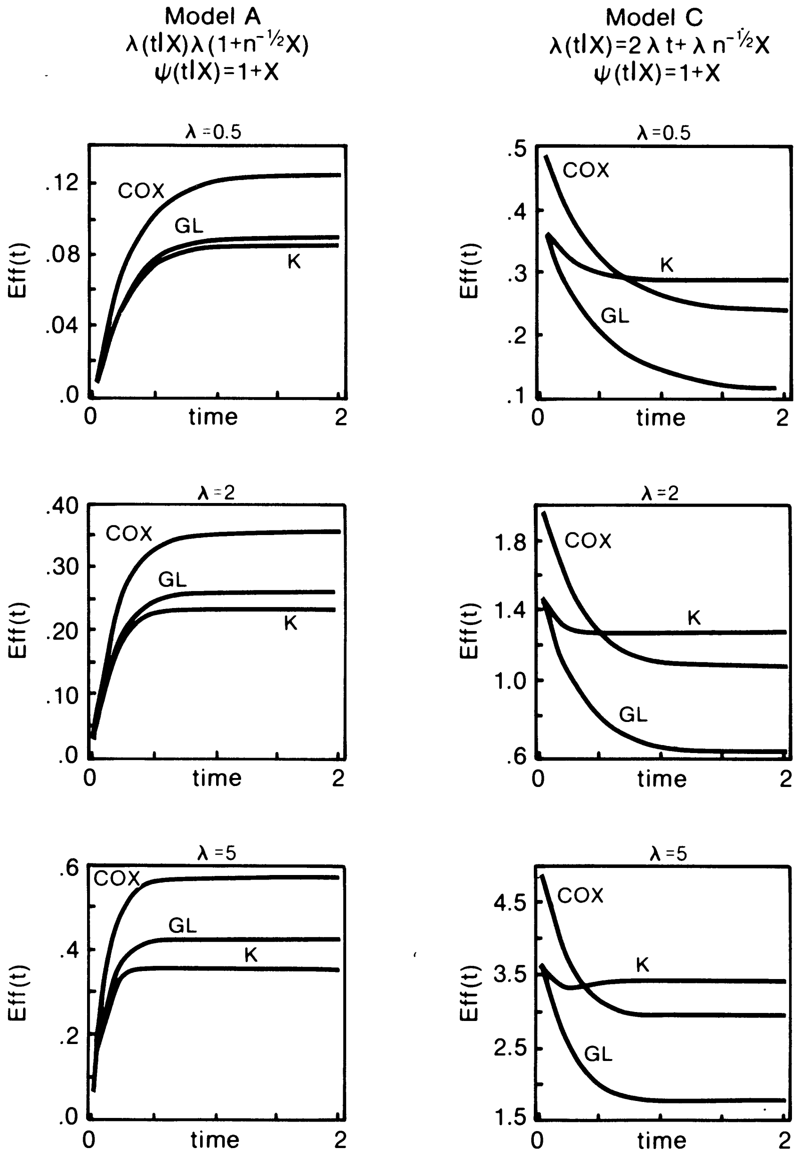

FIG. 2. *Efficacies for the Cox score* (COX), *generalized logrank* (GL) *and Kendall* (K) *statistics.* $\lambda(t|X)$ *is the failure hazard,* $\psi(t|X)$ *is the censoring hazard and* X *is an exponential*(1) *covariate.*

component $\lambda\beta X/\sqrt{n}$ of the hazard so that the hazard–covariate association is strong for small t. But as t increases, the background component begins to dominate the covariate component, so that the hazard–covariate association gets weaker as t increases. The COX and GL tests, which weight failures equally over time, are affected by this decreasing association over time whereas K, which weights the early failure times, has a better chance of detecting the association.

Finally, let us consider covariates which are initially exponential with parameter θ. The integrands of (13) and (14) are again calculated from values in the Appendix. If $\eta(t) = 0$, as in cases of no censoring or censoring independent of X, then the ARE of GL is 75% that of COX. For initially exponential ($\theta = 1$) covariates and parameters α, ψ, ϕ and γ set to 1, the efficacies of the three tests are plotted in Figure 2 for Models A and C. The comments above for uniform covariates are applicable here as well. There is however one real difference for Model C; namely for exponential covariates the COX efficacy dominates that of K for awhile, but then the relationship is reversed. We explain this reversal by analogy with the initially uniform covariate situation. As discussed above, the covariate information is completely contained in the ranks when the covariates are uniformly distributed through time, i.e., when $\eta(t) \equiv 0$. As $\eta(t)$ increases, the covariate distribution at time t, (A.1) gets further and further from uniform so that the ranks contain less and less of the covariate information. In fact what one sees in efficacy plots for initially uniform covariates similar to Figure 1 if one increases either ψ, γ or t of $\eta(t) = \psi\gamma t$ is a decrease in the ARE of GL to COX.

5. Study of robustness of W_n. In this section we shall study the effect that outliers in the measured covariate have on the AREs of various versions of the W_n statistic. Suppose that the hazard function is again given by (12), but that X is observed with some contamination, i.e., $X_c = X + \delta J$ is observed in place of X, where at time $t = 0$, $P(J = 0) = 1 - P(J = 1) = 1 - \varepsilon$ and J is independent of X. The bulk of the measured covariates is generated from a $P_0(x)$ distribution; however, $100\varepsilon\%$ of the X_c's arise from a contaminating distribution $P_0(x - \delta)$ for which the true hazard given by (12) is less than expected for X_c, $\delta > 0$. These mismeasured covariates are therefore outliers for $\delta \neq 0$.

For W_n based on the true covariate X, $W_n(t)/\sqrt{V_n(t)} \to_d N(\mu(t)/\sigma(t), 1)$, where $\mu(t)$ and $\sigma^2(t)$ are defined in (13) and (14). Replacing $\mu(t)$ and $\sigma^2(t)$ by $\mu_c(t)$ and $\sigma_c^2(t)$, the same results holds for W_n based on the measured covariate X_c. When the label function $Z = X_c$, $\mu_c(t) = \mu(t)$ and

$$\sigma_c^2(t) = \sigma^2(t) + \delta^2 \int_0^t w_0^2 y \sigma_{JJ} \lambda \alpha .$$

When the label function $Z(t) = R_c^*(t) = Y_\cdot^{-1}(t) r_c(t)$, where $r_c(t)$ is the rank of X_c among those at risk at t,

$$\mu_c(t) = \beta \int_0^t y w_0 \{\sigma_{X_c, R_c^*} - \delta \sigma_{J, R_c^*}\} \lambda$$

TABLE 1
AREs of various versions of W_n under H_{ca} for contaminated uniform covariates ($\varepsilon = 0.05$).

	ARE (% of COX)				
δ	**COX**	**GL**	**K**	**SCOX**	**SGL**
		Constant Hazard Model			
0	100	100	75	89	89
1	100	129	96	89	114
2	100	269	201	89	239
3	100	502	377	89	446
		Excess Risk Model			
0	100	100	182	138	138
1	100	129	234	138	178
2	100	269	488	138	371
3	100	502	912	138	693

and $\sigma_c^2(t) = \sigma^2(t)$, where σ_{X_c, R_c^*} and σ_{J, R_c^*} are the asymptotic covariances of X_c and J with R_c^*.

We consider the special case in which $P_0(X) = U(0,1)$. Two failure hazard models are used: a constant failure hazard model in which $\lambda(t|X) = 1 + \beta n^{-1/2}X$ and an ER model $\lambda(t|X) = t + \beta n^{-1/2}X$. A constant censoring hazard $\psi(t|X) = 1$ is assumed which is independent of both X and J. The components of $\mu_c(t)$ and $\sigma_c^2(t)$ are not given here but can be easily computed. Table 1 contains the asymptotic relative efficiencies (at $t = \infty$) for the constant and ER models for the COX, GL, K and the survival-weighted procedures SCOX [$w_0(t) = S(t), Z = X_c$] and SGL [$w_0(t) = S(t), Z = R_c^*$] tests when $\varepsilon = 0.05$ and δ varies.

For uniform covariates the GL test is 100% as efficient as the COX test when $\delta = 0$ and $\psi(t|X)$ is independent of X. For $\delta \geq 1$ and $\varepsilon > 0$, the Pitman efficacies of the tests based on $Z = R_c^*$ remain unchanged as δ varies, whereas the efficacies of COX and SCOX based on $Z = X_c$ decrease as δ increases. Hence, the greater the deviation δ from the true covariate, the greater the increase in efficiency of the rank-based procedures relative to the covariate-based procedures. Not surprisingly, the survival-weighted procedures SCOX and SGL perform better under the ER model. Neither do quite as well as K in this particular setting.

6. Discussion. By various choices of the weight process w and label process Z, the statistic W_n represents a broad class of nonparametric testing procedures, many of which are reported in the literature [cf. Jones and Crowley (1989)]. The advantages of having such a broad class are threefold: First, the already well studied two-sample procedures may suggest statistics to be used in more general settings, e.g., the generalized logrank. Second, there is a single central limit theorem for the entire class; asymptotic normality for a particular version of W_n is a corollary to the main theorem. Third, the ARE of any pair of tests (applicable to the same setting) in this class can be calculated.

Since W_n is the Cox score test when the covariate X is chosen as the label Z, it seems reasonable to compare the present work with that of Andersen and Gill (1982) and to ask the question of whether the statistic W_n and its large sample properties fall out of the Cox model. Under the null hypothesis W_n is a local martingale. Furthermore, it is consistent under the general alternative of ordered hazards under certain regularity conditions [Jones (1986)]. (By ordered hazards we mean that if individual j has a larger covariate than individual i, then he also has a greater hazard.) The hazard function is quite general. The Cox model, on the other hand, assumes a RR form to the hazard. In order to derive W_n from the Cox model, one would first assume the hazard to be $\lambda_0(t)\exp(\beta w(t)Z(t))$. Equation (4) arises from differentiating the log partial likelihood with respect to β and then setting $\beta = 0$. If $Z_j(t)$ is the covariate $X_j(t)$ or even a grouping of the covariate, all works well. However, many choices of the label Z are not covered. For example, if $Z_j(t)$ is based on the rank of $X_j(t)$ among the covariates at risk, then the Cox model would be forced to model the jth individual's hazard $\lambda_j(t)$ as a function of his ranking within the risk set in order to motivate W_n; i.e., a person's hazard of failure is determined by factors totally independent of him. This $\lambda_0(t)\exp(\beta w(t)Z(t))$ is obviously not a hazard and therefore cannot be used to make up a likelihood, not even an approximate one. $M(t)$ based on such a function need not be a martingale either, e.g., $M(t)$ given by equation (2.2) of Andersen and Gill (1982). The same is also true for labels such as $Z_j(t) = \Psi(X_j(t) - \bar{X}(t))$, for some function Ψ.

The standardized statistic $W_n(t)/\sqrt{V_n(t)}$ is asymptotically normal under conditions stated in Theorem 3.1. This theorem holds quite generally, including the case of mixed hazard functions as pointed out in Section 3.1. Theorem 3.1 serves as a necessary foundation for the ARE studies of Sections 4 and 5. In Section 4 the AREs of three versions of W_n (COX, GL and K) under both RR and ER failure hazard models are calculated. As expected, the COX test performs the best when the true hazard is of RR form. When the true hazard is of the ER form with $\alpha(t)$ increasing in t, K is frequently more powerful than the COX test. The superiority of the K test under the ER model is due to its early weighting of failure times when the hazard–covariate association is the strongest as $\alpha(t)$ is increasing in t. If the background component were decreasing in time, e.g., $\alpha(t) = \alpha/(1 + t)$, then a set of weights which emphasizes later failure times should be used, e.g., $w(t) = 1 - Y.(t)/n$ or $1 - \hat{F}(t)$. The performance of the COX test could be enhanced by using such weights as well. Using the method of Lemma 5.2.1 of Gill (1980), it can be shown that the optimal weight for $W_n(t)$ based on $Z(t) = X(t)$ is $w(t) \propto 1/\alpha(t)$. Unfortunately, $\alpha(t)$ is unknown. Weights that depend on $Y.(t)$, such as K, are quite sensitive to the censoring distribution and should be avoided when there is much censoring. A good choice of weight function can be made from the Harrington and Fleming (1982) $G^{\rho\gamma}$ class in which $w(t) = [1 - \hat{F}(t)]^{\rho}[\hat{F}(t)]^{\gamma}$. Choice of ρ and γ depends upon the general trend in $\alpha(t)$.

Compared with the COX test, the GL test has reasonably similar behavior for uniform covariates (Figure 1), is less efficient for exponential covariates (Figure 2) and is more efficient and of particular practical value in those cases where covariate outliers are present (Section 5).

APPENDIX

In this Appendix the values of $dP_t(x)$ of (16) and of the functions $y(t)$, $\sigma_{XX}(t)$ and $\sigma_{XR^*}(t)$ and $\sigma_{R^*R^*}(t)$, defined in the text just after (16), are given for binary, uniform and gamma covariates.

Initial binary covariates. Suppose $p_0(x) = \text{Bernoulli}(p)$. Then

$$\begin{aligned} y(t) &= \zeta(t)q(t), \\ p_t(i) &= q^{-1}(t)\left[pe^{-\eta(t)}\right]^i(1-p)^{1-i}, \qquad i = 0,1, \\ \sigma_{XX}(t) &= p_t(0)p_t(1), \end{aligned}$$

where $q(t) = 1 - p + p\exp(-\eta(t))$. It is straightforward to show that

$$\begin{aligned} \sigma_{XR^*}(t) &= \tfrac{1}{2}p_t(0)p_t(1), \\ \sigma_{R^*R^*}(t) &= \tfrac{1}{4}p_t(0)p_t(1). \end{aligned}$$

Initial gamma covariates. Suppose $p_0(x) = \Gamma(\alpha, \theta)$, the gamma density. Note that exponential(θ) = $\Gamma(1, \theta)$. Then

$$\begin{aligned} y(t) &= \zeta(t)\left[\frac{\theta}{\theta + \eta(t)}\right]^\alpha, \\ p_t(x) &= \Gamma(\alpha, \theta + \eta(t)), \\ \sigma_{XX}(t) &= \alpha[\theta + \eta(t)]^{-2}, \\ \sigma_{XR^*}(t) &= [4(\theta + \eta(t))]^{-1}, \\ \sigma_{R^*R^*}(t) &= 1/12. \end{aligned}$$

Initial uniform covariates. Suppose $p_0(x) = (b-a)^{-1}I[a < x < b]$. Let $\rho(t) = \exp(-\eta(t)a) - \exp(-\eta(t)b)$. Then

$$\text{(A.1)} \qquad \begin{aligned} y(t) &= \begin{cases} \zeta(t)\rho(t)/\eta(t)(b-a), & \text{if } \eta(t) \neq 0, \\ \zeta(t), & \text{if } \eta(t) = 0, \end{cases} \\ p_t(x) &= \begin{cases} \eta(t)e^{-\eta(t)x}\rho^{-1}(t)I[a < x < b], & \text{if } \eta(t) \neq 0, \\ p_0(x), & \text{if } \eta(t) = 0. \end{cases} \end{aligned}$$

$p_t(x)$ is the truncated exponential density over (a, b), where $-\infty < a < b < \infty$.

Also,

$$\sigma_{XX}(t) = \begin{cases} \eta^{-2}(t) + (b-a)^2 \rho^{-2}(t) e^{-\eta(t)(a+b)}, & \text{if } \eta(t) \neq 0, \\ (b-a)^2/12, & \text{if } \eta(t) = 0, \end{cases}$$

$$\sigma_{R^*R^*}(t) = 1/12.$$

For $\eta(t) = 0$, $\sigma_{XR^*}(t) = (b-a)/12$. For $\eta(t) \neq 0$,

$$\sigma_{XR^*}(t) = \frac{e^{-2a\eta(t)} - e^{-2b\eta(t)} - 2\eta(t)(b-a)e^{-n(t)(a+b)}}{4\eta(t)[e^{-a\eta(t)} - e^{-b\eta(t)}]^2}. \tag{A.2}$$

If we let $a = 0$ and $b = \infty$ and replace $\eta(t)$ by $\theta + \eta(t)$, then the truncated exponential $p_t(x)$ becomes $\Gamma(1, \theta + \eta(t))$ and (A.2) becomes $[4(\theta + \eta(t))]^{-1}$ corresponding to $\sigma_{XR^*}(t)$ given above for initial gamma($\alpha = 1$) covariates.

Acknowledgment. We would like to thank the reviewers for several comments that helped improve the presentation of this paper.

REFERENCES

AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. thesis, Dept. Statist., Univ. California, Berkeley.

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

ANDERSEN, P. K. and BORGAN, O. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist* **12** 97–158.

ANDERSEN, P. K., BORGAN, O., GILL, R. D. and KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes with applications to censored survival data. *Internat. Statist. Rev.* **50** 219–258.

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. (1970). A generalized Kruskal–Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57** 579–594.

BROWN, B. W., JR., HOLLANDER, M. and KORWAR, R. M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In *Reliability and Biometry: Statistical Analysis of Lifelength* (F. Proschan and R. J. Serfling, eds.) 327–354. SIAM, Philadelphia.

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

DOLIVO, F. (1974). Counting processes and integrated conditional rates: a martingale approach with application to detection. Technical Report, College of Engineering, Univ. Michigan.

EFRON, B. (1967). A two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–854.

GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52** 203–223.

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Mathematical Centre Tracts* **124** Mathematisch Centrum, Amsterdam.

JONES, M. P. (1986). A general class of nonparametric trend tests for survival analysis. Ph.D. thesis, Univ. Washington.

JONES, M. P. and CROWLEY, J. (1988). Asymptotic properties of a general class of nonparametric trend tests for survival analysis. Technical Report No. 88–1, Dept. Preventive Medicine, Univ. Iowa.

JONES, M. P. and CROWLEY, J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics* **45** 157–170.

HARRINGTON, D. P. and FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69** 533–546.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden–Day, San Francisco.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemoth. Rep.* **50** 163–170.

O'BRIEN, P. C. (1978). A nonparametric test for association with censored data. *Biometrics* **34** 243–250.

PETO, R. and PETO, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *J. Roy. Statist. Soc. Ser. A* **135** 185–206.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179.

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.

TARONE, R. E. (1975). Tests for trend in life table analysis. *Biometrika* **62** 679–682.

TARONE, R. E. and WARE, J. (1977). On distribution-free test for equality of survival distributions. *Biometrika* **64** 156–160.

DEPARTMENT OF PREVENTIVE MEDICINE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

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