The Annals of Statistics
1991, Vol. 19, No. 3, 1434–1456

CONDITIONAL RANK TESTS FOR RANDOMLY CENSORED DATA

BY ARNOLD JANSSEN

University of Siegen

The present paper derives various survival tests and their optimality results for randomly censored lifetime data by an extension of familiar rank test arguments. The approach is based on local asymptotic normal models which have a natural interpretation in terms of hazard rates. In particular, the description of classical rank tests by hazard rates may be of separate interest. As an application of the new methods, a justification of conditional survival tests is given also under unequal censoring distributions. It turns out that censoring is a nuisance phenomenon which asymptotically drops out. Conditional tests are exact permutation tests which are shown to be equivalent to their unconditional counterparts. They have all kinds of optimality properties and can be recommended for applications at least in those cases when a model with equal censorship cannot be excluded under the null hypothesis.

1. Introduction. A standard situation in applied survival analysis can be described as follows. Assume that X_{1i}, $i = 1, \dots, n$, denote independent random variables standing for survival times of individuals which are not totally observable. Let X_{2i} be independent censoring variables for $i = 1, \dots, n$, which are also independent from X_{1i}. Under random right censoring it is assumed that only

$$X_i = \min(X_{1i}, X_{2i}) \quad \text{and} \quad \Delta_i = 1_{\{X_{1i} \le X_{2i}\}}, \tag{1.1}$$

the indicator of the event $\{X_{1i} \le X_{2i}\}$, are observable for $i = 1, \dots, n$. In practice one is interested in statistical inference for the lifetimes X_{1i} only. Here often generalized rank tests are used whose introduction requires the following notation. Throughout, let the variables X_i be continuously distributed for $i = 1, \dots, n$. Consider the ordered observations

$$X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$$

of the sample (1.1) and let D_{ni} denote the antirank of the ith observation, that is, $X_{i:n} = X_{D_{ni}}$ and define antirank vectors as

$$D_n = (D_{ni})_{i=1,\dots,n}.$$

Introduce also

$$\Delta^{(n,i)} = \Delta_{D_{ni}}, \qquad \Delta^{(n)} = (\Delta^{(n,i)})_{i=1,\dots,n}.$$

Notice that $\Delta^{(n,i)} = 1$ if $X_{i:n}$ is uncensored, that is, a lifetime is observed.

Received March 1989; revised June 1990.

AMS 1980 *subject classifications*. Primary 62G20; secondary 62G10

Key words and phrases. Conditional rank tests, censored data, survival analysis, Pitman efficiency, hazard rates.

The theoretical justification of generalized rank tests based on $(D_n, \Delta^{(n)})$ was deeply influenced by the counting process approach of Aalen (1978). Following these lines the asymptotic theory reached a high level. The asymptotic behaviour of survival tests was clarified by Gill (1980), Leurgans (1984) and further authors whose papers can be found in the survey articles of Andersen, Borgan, Gill and Keiding (1982) and Andersen and Borgan (1985). Two sample goodness-of-fit tests were proposed by Schumacher (1984).

By a different approach based on a local asymptotic normal approximation (LAN), Neuhaus (1988) and Janssen (1989) showed that the antiranks and censoring indicators $(D_n, \Delta^{(n)})$ are asymptotically sufficient within a model with given local alternatives. For these reasons, Neuhaus (1988) proposed rank tests for censored data which are similar to the classical score rank tests of Hájek and Šidák (1967) under uncensored data. Below we will pursue these fruitful ideas. Earlier Albers and Akritas (1987) also adapted the classical rank test procedures. They studied ordinary rank tests for the censored and uncensored portion of the data separately. A comparison of their results and the present approach is contained in Neuhaus (1988). In the sequel two main results are obtained.

1. The standard survival tests and their optimality properties can be obtained by an extension of the classical rank test theory of Hájek and Šidák (1967). In addition it is explained that the commonly used local alternatives have a natural explanation in terms of hazard rates which is extremely helpful for a practical motivation of the results. The interpretation of classical (uncensored) rank tests given in terms of hazard rates may be of separate interest.
2. At finite sample size conditional tests are proposed which can improve the familiar survival tests. It is shown that these tests approximate the desired level also under local (contiguous) unequal censoring distributions although they are in principle constructed for equal censoring conditions under the null hypothesis.

Roughly speaking, the accuracy of the conditional tests has the following reason. Within a local model including different censorship the influence of the lifetime distributions and the censoring distributions becomes asymptotically independent, Janssen (1989). Conditional tests are permutation tests with exact critical values. Permutation tests with estimated variances were earlier treated under equal censorship by Andersen, Borgan, Gill and Keiding [(1982), Section 3.5], with their martingale methods.

What is the difference between the powerful counting process methods and the present approach? Here the influence of the censoring distributions is given by nonparametric nuisance parameters which asymptotically drop out, see also Remark 3.1(b). Similarly the k-sample problem is treatable; see Janssen (1991). In a forthcoming paper it is shown that the methods apply to discrete observations when ties are present. Recent Monte Carlo results show that conditional tests can be recommended in practice whenever the censoring procedures do not differ drastically; see Janssen and Brenner (1991).

2. Conditional survival tests. In the literature the following well-motivated test statistics T_n are frequently applied. Define

$$(2.1)\quad T_n = T_n(D_n, \Delta^{(n)}) = \sum_{i=1}^{n} w_n(i)\Delta^{(n,i)}\left(c_{nD_{ni}} + \sum_{j=1}^{i-1} \frac{c_{nD_{nj}}}{n+1-i}\right),$$

where c_{ni} denote given regression coefficients and $w_n(i)$ are certain random weights specified later. Notice that the weights can be used to make the later tests sensitive for certain directions of alternatives. In order to motivate the choice of T_n in (2.1), consider a two sample problem with n_1 individuals in the first group (treatment group) and $n_2 = n - n_1$ individuals in the second group (control group) indicated by the index set $S_2 = \{n_1 + 1, \dots, n\}$. If we define

$$(2.2)\qquad c_{ni} = n^{-1/2}\left(1_{S_2}(i) - n_2/n\right),$$

it is easy to see that the conditional expectation of $c_{nD_{ni}}$ equals

$$(2.3)\qquad E\left(c_{nD_{ni}} \middle| D_{n1}, \dots, D_{ni-1}\right) = -\sum_{j=1}^{i-1} \frac{c_{nD_{nj}}}{n+1-i}$$

under i.i.d. $X_1, \dots, X_n$ and thus (2.1) measures the deviation from the i.i.d. hypothesis.

In the sequel we will consider the nonparametric hypothesis

$$(2.4)\qquad H_0\colon X_{11}, \dots, X_{1n} \text{ are i.i.d. with d.f. } F_1(\cdot, 0),$$

where $F_1(\cdot, 0)$ is a completely unknown continuous distribution function (d.f.). Throughout, the d.f. of the censoring variable X_{21} is specified by a substochastic distribution $F_2(\cdot, 0)$ on $(-\infty, \infty]$ with $\lim_{x\to\infty} F_2(x, 0) \le 1$, which is an unknown nuisance parameter. Let $H(\cdot)$ be the d.f. of X_1 under H_0, namely

$$(2.5)\qquad 1 - H(x) = (1 - F_1(x, 0))(1 - F_2(x, 0)),$$

which is assumed to be continuous and let

$$(2.6)\qquad p(u) = E(\Delta_1 | H(X_1) = u), \qquad u \in (0, 1),$$

denote the regression function of Δ_1 w.r.t. $H(X_1) = u$. By means of the regression function $p(\cdot)$ the distribution of the pivoted sample $(H(X_1), \Delta_1)$ can be written as

$$(2.7)\quad \mathscr{L}(H(X_1), \Delta_1)(A \times B) = \int_A (1_B(1)p(u) + 1_B(0)(1 - p(u)))\, du$$

for Borel sets A and B. Notice that the consideration of pivoted samples is motivated by invariance properties of rank tests.

Throughout, the following notation is used. Let λ denote the uniform distribution on $(0,1)$ and 1 the indicator function. We write $\to_{L_2}$ for the convergence in $L_2((0,1), \lambda)$ and $\to_\lambda$ for convergence in probability. Let $[x] := \sup\{k \in \mathbb{Z}\colon k \le x\}$ denote the entire function. The functions $u \to w_n(1 + [nu])$ and so on are always denoted on $(0, 1)$. The normal distribution with mean a and covariance Γ is denoted by $N(a, \Gamma)$ and Φ is the standard normal d.f.

In connection with k-sample problems, more general regression coefficients as in (2.2) are needed. Consider the following regularity conditions.

REGRESSION COEFFICIENTS. Let c_{ni} be regression coefficients such that the usual Noether conditions are satisfied:

$$\max_{1 \le i \le n} |c_{ni}| \to 0, \qquad \sum_{i=1}^{n} c_{ni} = 0 \quad \text{for each } n \tag{2.8}$$

and

$$\sum_{i=1}^{n} c_{ni}^2 \to c > 0 \tag{2.9}$$

as $n \to \infty$. Sometimes we need the condition

$$n^{1/2} c_{n1+[nu]} \to_{L_2} c(u) \tag{2.10}$$

as $n \to \infty$ for some $L_2(0,1)$ function $c(\cdot) \neq 0$, which is more restrictive than (2.9). Notice that in case of (2.2) the condition (2.10) holds whenever

$$n_1/n \to \gamma \in (0,1). \tag{2.11}$$

WEIGHT FUNCTIONS [Assumption (W)]. Let $w_n(i)$ denote random weights such that $w_n(i)$ is measurable w.r.t. the σ-field $\sigma(\Delta^{(n,j)}\colon\ j < i)$. For i.i.d. lifetimes $X_{11}, \dots, X_{1n}$ and i.i.d. censoring variables $X_{21}, \dots, X_{2n}$ with d.f.'s $F_1(\cdot, 0)$ and $F_2(\cdot, 0)$ assume that there exists an $L_2(0,1)$ function $w = w_{F_1, F_2}$, which may depend on $F_1(\cdot, 0)$ and $F_2(\cdot, 0)$, such that

$$\frac{1}{n} \sum_{i=1}^{n} E\Big(\big(w_n(i) - w(H(X_{i:n}))\big)^2\Big) \to 0 \quad \text{as } n \to \infty. \tag{2.12}$$

DISCUSSION. (a) The condition (2.12) is a natural extension of the classical L_2-convergence assumption for weights of Hájek and Šidák (1967) since for nonrandom weights the condition (2.12) is equivalent to the convergence of

$$u \mapsto w_n(1 + [nu]) \tag{2.13}$$

to $w(u)$ in $L_2(0,1)$ as $n \to \infty$. This result can be deduced from Lemma D of the Appendix if one first considers $w_n(i) = E(w(U_{i:n}))$.

(b) Next we will discuss random weights $w_n(i) := \tilde{w}_n(X_{i:n})$ which arise from a given predictable process $\tilde{w}_n(t)$. The condition (2.12) is satisfied if for instance $\tilde{w}_n(t)$ and $w(H(t))$ are uniformly bounded and

$$\sup_{t \in K} |\tilde{w}_n(t) - w(H(t))| \to 0 \tag{2.14}$$

in probability for each compact subset K of $\{s\colon H(s) < 1\}$. In practice, often weight functions

$$\tilde{w}_n(X_{i:n}) = \rho_n\big(\hat{F}_n(X_{i:n}-)\big) \tag{2.15}$$

are used for deterministic functions ρ_n which are based on the Kaplan–Meier estimator

$$\hat{F}_n(X_{i:n}-) = 1 - \prod_{j<i}\left(\frac{n-j}{n+1-j}\right)^{\Delta^{(n,j)}} \tag{2.16}$$

of the pooled sample. Here $\hat{F}_n(t-)$ denotes the left-sided limit at t. Notice that the Kaplan–Meier estimator satisfies

$$\sup_{t\le t_0}\left|\hat{F}_n(t) - F_1(t,0)\right| \to 0 \tag{2.17}$$

in probability whenever $H(t_0) < 1$; see Gill [(1980), Theorem 4.1.1]. Thus condition (2.12) holds for predictable processes $\tilde{w}_n$ of the form (2.15) when ρ_n is uniformly bounded, continuous and uniformly convergent on compact subsets of $(0,1)$. Under these assumptions, the statistic $T_n(D_n, \Delta^{(n)})$ (2.1) is asymptotically normal distributed. Similar results were obtained by Gill (1980) but the present proof is an extension of the classical proof for ordinary rank tests. Special attention is also devoted to the asymptotic normality of the conditional statistic $T_n(D_n, \delta)$ given $\Delta^{(n)} = \delta$ which is needed for conditional tests.

THEOREM 2.1. *Assume that the conditions* (2.8), (2.9) *and Assumption* (W) *hold. Under i.i.d. survival distributions and i.i.d. censoring distributions with common d.f.'s* $F_1(\cdot,0)$ *and* $F_2(\cdot,0)$, *respectively, the statistic* T_n *has the following properties.*

(a) $T_n \to N(0,\sigma^2)$ *in distribution as* $n\to\infty$, *where the variance*

$$\sigma^2 = c\int_0^1 w^2(u)p(u)\,du \tag{2.18}$$

is determined by the limit score function $w = w_{F_1,F_2}$ *and the regression function* $p(\cdot)$, *see* (2.6).

(b) *In addition assume* (2.10) *and let* $\overline{T}_n(D_n, \Delta^{(n)})$ *be a further statistic with* $\overline{T}_n \to T_n \to 0$ *in probability as* $n\to\infty$. *Let* $F_{n,\delta}(t)$ *denote the distribution function of* $\mathscr{L}(\overline{T}_n(D_n,\delta))$ *under uniformly distributed antiranks for fixed* δ. *Then*

$$\sup_{t\in\mathbb{R}}\left|F_{n,\Delta(n)(\omega)}(t) - \Phi(t/\sigma)\right| \to 0 \tag{2.19}$$

in probability as $n\to\infty$, *whenever* $\sigma^2 > 0$.

The proofs are presented in Section 5.

In practical situations, Theorem 2.1 often holds for each reasonable pair $(F_1(\cdot,0), F_2(\cdot,0))$ of d.f.'s. However, the application of T_n as a test statistic requires a consistent estimator V_n for σ^2 in order to get a distribution-free statistic. Various consistent estimators V_n are discussed in Gill (1980) for the two sample case. This consideration leads to the unconditional (asymptotic level α) tests

$$\varphi_n = \begin{cases} 1, & T_n V_n^{-1/2} > u_{1-\alpha}, \\ 0, & T_n V_n^{-1/2} \leq u_{1-\alpha}, \end{cases} \tag{2.20}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of Φ. Monte Carlo results show that the true level of φ_n may differ from the nominal level α; see Section 4, Latta (1981) or Janssen and Brenner (1991). In order to improve the unconditional tests, often permutation tests based on T_n are applied at finite sample size. Notice that those permutation tests where T_n is standardized by its permutation variance can be handled with the methods of Andersen, Borgan, Gill and Keiding [(1982), Section 3.5], under equal censoring conditions. We will study conditional tests $\tilde{\varphi}_n$ in detail which are exact permutation tests. They are motivated by the fact that D_n and $\Delta^{(n)}$ are independent under i.i.d. survival times and equal censoring distributions. For a given statistic $T_n(D_n, \Delta^{(n)})$, conditional tests are defined as follows.

1. In a first step we observe $\Delta^{(n)} = (\delta_1, \dots, \delta_n) = \delta$.
2. For fixed δ, we apply the level α test

$$\tilde{\varphi}_{n,\delta}(D_n) = \begin{cases} 1, & T_n(D_n, \delta) > c_n(\alpha, \delta), \\ \gamma(\alpha, \delta), & T_n(D_n, \delta) = c_n(\alpha, \delta), \qquad \gamma(\alpha,\delta) \in [0,1], \\ 0, & T_n(D_n, \delta) < c_n(\alpha, \delta), \end{cases} \tag{2.21}$$

where the critical value $c_n(\alpha, \delta)$ is the $(1-\alpha)$ quantile of the usual rank statistic $T_n(\cdot, \delta)$ under uniformly distributed antiranks.

Under equal censoring distributions, $\tilde{\varphi}_n = \tilde{\varphi}_{n,\Delta}(n)(D_n)$ is an exact level α test and it reduces to an ordinary rank test whenever no censoring is present. In addition, $\tilde{\varphi}_n$ is always distribution free and no estimator V_n of σ^2 is needed. As it is discussed in the Introduction, point 2 and in Sections 3 and 4, conditional tests also work under locally different censoring distributions. Together with the finite sample results this fact now justifies its application in practice when strongly different censoring conditions can be excluded. Conditional tests of the form (2.21) were proposed by Neuhaus (1988).

Usually statistics T_n given by (2.1) will be considered in (2.21) but also conditional tests based on $\overline{T}_n$ [see Theorem 2.1(b)] or $T_n/V_n^{1/2}$ work well. As a direct consequence of Theorem 2.1(b), we see that conditional tests are asymptotically equivalent to their unconditional counterparts.

COROLLARY 2.1. *Under the assumptions of Theorem* 2.1(b), *we obtain*

$$\tilde{\varphi}_{n,\Delta}(n)(D_n) - 1_{[u_{1-\alpha},\infty)}(T_n/\sigma) \to 0 \tag{2.22}$$

in probability as $n \to \infty$.

3. Conditional survival tests under local alternatives. In modern nonparametric statistics the quality of tests is usually compared under local alternatives arising from a fixed but arbitrary curve of d.f.'s; see Pfanzagl and Wefelmeyer (1982) or Strasser (1985). In this spirit we will specify the survival model, compare with Janssen (1989). For related considerations concerning the estimation of the survival functions, see van der Vaart (1988).

At the beginning assume that $\Theta \subset \mathbb{R}$ is an open neighbourhood of the origin and let $(P_{j\vartheta})_{\vartheta \in \Theta}$ be a family of distributions on $\mathbb{R}$ for $j = 1$, on $(-\infty, \infty]$ for $j = 2$, respectively, with d.f.'s $F_j(\cdot, \vartheta)$. Below the curve $P_{1\vartheta}$ is used to introduce a model for the survival times where as an extension of the two sample case different conditions for the individuals i are expressed by regression coefficients c_{ni} with (2.8) and (2.9). The second curve investigates unknown censoring distributions. Here a further triangular array d_{ni} of regression coefficients is introduced which again satisfies the regularity conditions (2.8) and (2.9). The nowadays commonly known local consideration of the asymptotic test theory motivates the consideration of local parameters $(s, t) \in \mathbb{R}^2$ which are used to compensate the influence of increasing sample sizes. In view of these principles, define the joint distribution of the survival times and censoring distributions under (s, t) (sufficiently small) by

$$\mathscr{L}(X_{11}, \dots, X_{1n}, X_{21}, \dots, X_{2n}) = \bigotimes_{i=1}^{n} P_{1sc_{ni}} \otimes \bigotimes_{i=1}^{n} P_{2td_{ni}}, \tag{3.1}$$

where sc_{ni} describes the survival condition of the ith individual and td_{ni} its censoring procedure. Denote by

$$Q_{nst} := \mathscr{L}(X_1, \Delta_1, X_2, \Delta_2, \dots, X_n, \Delta_n), \tag{3.2}$$

the distribution of our observations under s and t. Assume that X_i has a continuous distribution for $i = 1, \dots, n$. This model is rich enough to describe locally unequal censorship.

Assume the following standard regularity assumptions for the underlying curves. Let $\vartheta \to P_{j\vartheta}$ be 2-differentiable at zero with a derivative $\dot{L}_j \in L_2(P_{j0})$ for $j = 1, 2$, that is,

$$\left\| 2\left(\frac{dP_{j\vartheta}}{dP_{j0}}\right)^{1/2} - 2 - \vartheta \dot{L}_j \right\|_{L_2(P_{j0})} = o(|\vartheta|) \tag{3.3}$$

and

$$P_{j\vartheta}\left(\left\{\frac{dP_{j0}}{dP_{j\vartheta}} = 0\right\}\right) = o(|\vartheta|^2) \quad \text{as } \vartheta \to 0. \tag{3.4}$$

Notice that $\dot{L}_j$ is a tangent vector in the sense of Pfanzagl and Wefelmeyer (1982). As usual $d\nu/d\mu$ denotes the density of the μ-absolutely continuous part of ν and $\|\cdot\|_{L_2}$ indicates the usual L_2-norm. From Janssen (1989), we recall that

$$(s, t) \mapsto \mathscr{L}(X_1, \Delta_1 | P_{1s} \otimes P_{2t}) \tag{3.5}$$

has the L_2-derivative

$$(\psi_1(x,\delta),\psi_2(x,\delta))^T \tag{3.6}$$
$$:= \left(\gamma_1(x)\delta + \frac{\int_{(x,\infty)}\dot{L}_1\,dP_{10}}{1-F_1(x,0)}, -\gamma_2(x)\delta + \dot{L}_2(x)\right)^T$$

on $\mathbb{R}\times\{0,1\}$ with

$$\begin{aligned}\gamma_1(x) &:= \dot{L}_1(x) - \frac{\int_{(x,\infty)}\dot{L}_1\,dP_{10}}{1-F_1(x,0)},\\ \gamma_2(x) &:= \dot{L}_2(x) - \frac{\int_{[x,\infty]}\dot{L}_2\,dP_{20}}{1-F_2(x-,0)}.\end{aligned} \tag{3.7}$$

For practical purposes survival models are often specified by hazard rates and not by d.f.'s. In this setting $\gamma_1(\cdot)$ has an interesting interpretation. If we define the hazard rate of $P_{1\vartheta}$ w.r.t. P_{10} by

$$\lambda_\vartheta := \frac{dP_{1\vartheta}}{dP_{10}}\bigg/(1-F_1(\cdot,\vartheta)), \tag{3.8}$$

the function $\gamma_1(\cdot)$ is the derivative of the hazard rate ratio, that is,

$$\left(\frac{\lambda_\vartheta}{\lambda_0}-1\right)\bigg/\vartheta \to \gamma_1(\cdot) \tag{3.9}$$

is convergent in P_{10} probability as $\vartheta \to 0$; see Janssen (1989), proof of Lemma 2. Notice that for μ-dominated families, $\gamma_1(\cdot)$ is the derivative of the cumulative hazard functions $d(-\log(1-F_1(\cdot,\vartheta)))/d\mu$ at zero, which was earlier considered by Gill [(1980), page 117]. Obviously (3.8) is not restricted to survival d.f. ($F_1(0,\vartheta)=0$) and (3.3)–(3.9) works for arbitrary distributions on $\mathbb{R}$.

In a first step a solution of the underlying test problem along a fixed curve was given in Janssen (1989). For the test problem with nuisance (censoring) parameter $t \in \{t \in \mathbb{R}: td_{ni} \in \Theta \text{ for } i=1,\dots,n\} =: M_n$

$$\{Q_{nst}: s \le 0, t \in M_n\} \quad \text{against} \quad \{Q_{nst}: s > 0, t \in M_n\} \tag{3.10}$$

the (upper) level α test based on the central sequence

$$Z_n^{(1)} := \sum_{i=1}^n c_{ni}\psi_1(X_i,\Delta_i) \tag{3.11}$$

of the local asymptotic normal (LAN) expansion is asymptotically optimal within the class of asymptotic level α tests, see also (5.28). Let now T_n denote the statistic (2.1) with a limit weight function $w(u) = w_{F_1,F_2}(u)$. The crucial point for the optimality of survival tests is to show by classical rank test methods that

$$Z_n^{(1)} - T_n \to_{Q_{nst}} 0, \tag{3.12}$$

whenever $w(\cdot)$ coincides with the derivative of the hazard ratio of the pivoted sample, that is,

$$w_{F_1, F_2}(u) = \gamma_1(H^{-1}(u)). \tag{3.13}$$

The general result is presented in Theorem 3.1 also when (3.13) is violated. Further interesting facts should be mentioned before the results are stated. In order to give a practical interpretation, the asymptotic moments of T_n are later expressed in terms of hazard rates. The results are given for the conditional tests $\tilde{\varphi}_n$ since we may not have a consistent estimator V_n under general regression coefficients. If such an estimator exists, we may replace $\tilde{\varphi}_n$ by φ_n and in particular we obtain the optimality results of Gill (1980) for the two sample case. For practical purposes we remark that $\tilde{\varphi}_n$ is an asymptotical level α test under local unequal contiguous censorship as long as (2.18) is positive.

THEOREM 3.1. *Let the conditions of Theorem* (2.1) *be fulfilled. For a family Q_{nst} given in* (3.2) *we obtain the following assertions.*

(a) $$\mathscr{L}(T_n | Q_{nst}) \to N\left(cs \int w(u)\gamma_1(H^{-1}(u))p(u)\,du, c\int w^2(u)p(u)\,du\right)$$

in distribution as $n \to \infty$, where γ_1 denotes the underlying derivative of the hazard rates ratio (3.9) *and H^{-1} is specified in* (2.5).

(b) *Under the additional assumption* (2.10) *we obtain*

$$E_{Q_{nst}}\tilde{\varphi}_n \to 1 - \Phi\left(u_{1-\alpha} - \frac{c^{1/2}s\int w(u)\gamma_1(H^{-1}(u))p(u)\,du}{\left(\int w^2(u)p(u)\,du\right)^{1/2}}\right) \tag{3.14}$$

as $n \to \infty$ [$u_{1-\alpha}$ is the $(1-\alpha)$ quantile of Φ] whenever (2.18) *is positive. In particular notice that for each nuisance parameter t,*

$$E_{Q_{n0t}}\tilde{\varphi}_n \to \alpha \quad \text{as } n \to \infty. \tag{3.15}$$

(c) *The conditional tests $\tilde{\varphi}_n$ are asymptotically optimal among the class of level α test for the test problem* (3.10) *if and only if there exists some $d \geq 0$ such that*

$$\left(dw(u) - \gamma_1(H^{-1}(u))\right)p(u) = 0 \quad \lambda \text{ almost everywhere}. \tag{3.16}$$

Let ARE denote the asymptotic relative efficiency in the sense of Hájek and Šidák (1967), page 267, for the composite test problem (3.10) and the conditional test $\tilde{\varphi}_n$ given by the weight function $w(\cdot)$, that is, compared with the optimal test 100(1 − ARE) percent of the observations are wasted if the test under consideration is used. Routine calculations show that whenever the mean of the normal distribution in Theorem 3.1(a) is nonnegative for $s > 0$,

$$\text{ARE} = \left(\int w(u)\gamma_1(H^{-1}(u))p(u)\,du\right)^2 \times\left(\int \gamma_1^2(H^{-1}(u))p(u)\,du \int w^2(u)p(u)\,du\right)^{-1} \tag{3.17}$$

is just $\cos^2\theta$ of the angle θ between the limit weight function $w(\cdot)$ and the derivative of the hazard ratio $\gamma_1(H^{-1}(\cdot))$ in the L^2-space w.r.t. $p(\cdot)\lambda$. However, for a simple test problem $\{Q_{n00}\}$ against $\{Q_{nst}\}$ in general even better tests as those given by (3.13) can be obtained; see Janssen (1989).

REMARK 3.1. (a) In the case of the two sample problem (2.2), Theorem 3.1(a) was earlier obtained by Gill [(1980), 5.2.14, page 106], under his assumptions.

(b) Within our model dQ_{n0t}/dQ_{n00} may be different from 1 caused by the influence of the nuisance parameter t. Here our model differs from the model of Gill (1980), page 32 and page 118, who always assumed that for an i.i.d. sample $(X_{1i})_{i=1,\dots,n}$ the corresponding distributions P and P' are equal in his notation; see Gill [(1980), (3.18)].

(c) Our point of view yields another look at the classical theory of rank tests without censoring ($p(u) \equiv 1$) which is completely included. Notice that for deterministic weights $w_n(i)$ with (2.13) and under $\Delta^{(n,i)} \equiv 1$, the statistic (2.1) equals

$$\tilde{T}_n = \sum_{i=1}^{n} c_{nD_{ni}}\left(w_n(i) - \sum_{j=1}^{i} \frac{w_n(j)}{n+1-j}\right), \tag{3.18}$$

where

$$\begin{aligned} a_n(1+[nu]) &:= w_n(1+[nu]) - \sum_{j=1}^{1+[nu]} \frac{w_n(j)}{n+1-j} \\ &\to_{L_2} w(u) - \int_0^u \frac{w(v)}{1-v}\,dv =: a(u). \end{aligned} \tag{3.19}$$

By Lemma 5.2, the common approach via classical scores $a_n(i)$ and the methods relying on the present weights $w_n(i)$ are equivalent. Notice also that $w(\cdot) \to a(\cdot)$ is an isometry from $L_2(\lambda)$ into $\{g \in L_2(\lambda)\colon \int g\,d\lambda = 0\}$; see Efron and Johnstone (1990) and Ritov and Wellner (1988). For both representations, optimality properties can be discussed. The related tests are optimal along a given curve $F_1(\cdot, \vartheta)$ with tangent $\dot{L}_1$ and hazard ratio derivative γ_1 if and only if

$$da(\cdot) = \dot{L}_1(F_1^{-1}(\cdot, 0)) \tag{3.20}$$

for some $d \geq 0$ or equivalently

$$dw(\cdot) = \gamma_1(F_1^{-1}(\cdot, 0)), \tag{3.21}$$

and also the ARE can be expressed by hazard rates (3.17). It is our feeling that the consideration in terms of hazard rates (3.21) gives a more natural nonparametric explanation for the optimality of linear rank tests than (3.20). We argue as follows. Common rank tests with score function $a(\cdot)$ (as the Wilcoxon test) are often optimal for a location family $F_1(\cdot, \vartheta)$. But in nonparametrics $F_1(\cdot, 0)$ is unknown and a further d.f. $\tilde{F}_1(\cdot, 0)$ may be the true basis point. Obviously, there exists a curve $\tilde{F}_1(\cdot, \vartheta)$ where the given test is optimal. However, $\tilde{F}_1(\cdot, \vartheta)$ is in general no longer a location family and its meaning is here unclear. But a proper interpretation can be given in terms of hazard rates (3.21): The normalized derivative of the hazard ratio is (positive) proportional to the weight function. The dependence of (3.21) on the unknown basis point $\tilde{F}_1(\cdot, 0)$ has a self-normalizing effect which yields a quite natural model. A distribution-free description can be given in the context of the pivoted sample $F_1(X_1, 0)$ whose derivative of the hazard ratio equals $\gamma_1(F_1^{-1}(\cdot, 0))$ whenever X_1 is distributed according to $F_1(\cdot, \vartheta)$.

(d) Following the ideas of Neuhaus (1987) we note that our approach suggests a further procedure for generalized rank tests with estimated scores. We propose to use an estimator for the optimal weights $w_n(i) = a_n^{(1)}(i) - a_n^{(2)}(i)$ of (5.4) and to insert this estimator in the test statistic T_n. What we are doing here is to estimate the underlying hazard rates ratio derivative $\gamma_1(H^{-1}(u))$. Notice that estimators for $a_n^{(1)}(i)$ and $a_n^{(2)}(i)$ are available from the paper of Neuhaus (1987).

Finally, we will consider various practical examples where the model is specified in terms of hazard rates. As mentioned in Remark 3.2(c) this approach has the advantage that hazard rate alternatives can be attached at each basis point $F_1(\cdot, 0)$ which can be regarded as a nuisance parameter.

EXAMPLE 3.1. Consider survival distributions as in (3.3)–(3.7). Then we obtain the following optimality results under semiparametric alternatives given by nuisance parameters $F_1(\cdot, 0)$ and $F_2(\cdot, 0)$.

(a) For a constant derivative $\gamma_1(x) = d > 0$ of the ratio of the hazard rates the conditional log-rank test with $w_n(i) = 1$ is optimal.

(b) For $w_n(i) = 1 - E(U_{i:n}) = (n + 1 - i)/(n + 1)$ the conditional version of the test of Gehan and Wilcoxon is optimal whenever

$$\gamma_1(H^{-1}(u)) = d(1 - u), \qquad d > 0. \tag{3.22}$$

This test is a test for differences in the hazard rates for small lifetimes.

(c) Consider the derivatives

$$\frac{d}{d\vartheta}\left(\frac{\lambda_\vartheta(x)}{\lambda_0(x)} - 1\right)\bigg|_{\vartheta=0} = \gamma_1(x) = (1 - F_1(x, 0))^\rho F_1(x, 0)^\kappa, \tag{3.23}$$

$$\rho \geq 0, \kappa \geq 0,$$

where $d\lambda_\vartheta(x)/d\lambda_0(x)$ denotes the hazard rate ratio of $F_1(\cdot, \vartheta)$ w.r.t. $F_1(\cdot, 0)$.

Then the choice of

$$(3.24)\quad w_n(i) = \left(1 - \hat{F}_n(X_{i:n}-)\right)^{\rho} \hat{F}_n(X_{i:n}-)^{\kappa} \qquad \text{[see (2.14)–(2.17)]}$$

leads to optimal conditional tests. This example includes a conditional version of the test of Harrington and Fleming (1982) for $\kappa = 0$ and of Prentice's (1978) version of the Wilcoxon test ($\rho = 1, \kappa = 0$). Notice that for $\rho > 0$ and $\kappa = 0$, the function $\gamma_1(x)$ may be understood as a model for the deviation of the hazard rates for small lifetimes. The case $\rho > 0$ and $\kappa > 0$ corresponds to a deviation of central lifetimes and $\gamma = 0$ and $\kappa > 0$ stands for a deviation of the hazard rates for long lifetimes. We see that in the parts (a) and (c), $F_1(\cdot, 0)$ and $F_2(\cdot, 0)$ are arbitrary nuisance parameters within the semiparametric model. In the case of part (b), however, $F_1(\cdot, 0)$ and $F_2(\cdot, 0)$ are connected via (3.22) and $H(\cdot)$, see (2.5).

4. Monte Carlo results. The recommendation of conditional tests also under (not too extreme) unequal censorship will be supported by the subsequent Monte Carlo results. Earlier Monte Carlo simulations were done by Latta (1981) and for more details see Janssen and Brenner (1991).

In the sequel let us always consider the two sample problem with regression coefficients (2.2) at sample size $n = 10$ and $n_1 = n_2 = 5$. Later we will study the unconditional (abbreviated by unc) and conditional (con) versions of the tests given in Example (3.1). Let LR denote the log-rank test, GW the Gehan–Wilcoxon test and PW the Prentice version of the Wilcoxon test. In connection with the unconditional test φ_n, we always took the estimator (3.3.11) of Gill (1980) as estimator V_n of the variance, which is usually applied in practice. The following results were obtained during a simulation with 50,000 Monte Carlo steps on a VAX 88 computer.

The calculations in Table 1 show that the unconditional tests are not accurate here under i.i.d. survival times and i.i.d. censoring distributions for the nominal level $\alpha = 0.02$, 0.05, 0.1 and 0.2. Notice that under the null hypothesis the model is completely specified by the regression function $p(\cdot)$, see (2.7). The results in Table 1 motivate the investigation of conditional tests. Next we show that conditional tests yield reasonable results (not only asymptotically) also under different censoring distributions for the null hypothesis.

TABLE 1

The true level of unconditional tests under the null hypothesis and equal censorship

	$p(u) = u$			$p(u) = \min(1, 1.5u^2)$		
α	LR	GW	PW	LR	GW	PW
0.02	0.002	0.001	0.002	0.001	0.001	0.001
0.05	0.043	0.036	0.043	0.039	0.033	0.040
0.10	0.115	0.115	0.115	0.113	0.112	0.114
0.20	0.262	0.263	0.251	0.310	0.304	0.296

TABLE 2
True level of the tests φ_n and $\tilde{\varphi}_n$ at the nominal level $\alpha = 0.1$ (under different censoring distributions)

Censoring distribution		**LR**		**GW**		**PW**	
Group 1	**Group 2**	**unc**	**con**	**unc**	**con**	**unc**	**con**
F_b	F_d	0.168	0.092	0.007	0.086	0.017	0.089
F_d	F_b	0.193	0.076	0.008	0.094	0.030	0.087
F_a	F_f	0.239	0.102	0.006	0.105	0.025	0.101
F_f	F_a	0.219	0.091	0.002	0.092	0.013	0.094
F_c	F_e	0.104	0.074	0.004	0.068	0.008	0.070
F_e	F_c	0.138	0.052	0.004	0.082	0.023	0.072

Assume throughout that the lifetime distributions $(X_{1i})_{i=1,\dots,n}$ are i.i.d. uniformly distributed on $(0, 1)$. The censoring distribution will be taken i.i.d. within the first group $(X_{2i})_{i \le 5}$ and i.i.d. in the second group $(X_{2i})_{i>5}$, respectively, but their distributions differ in both groups. We used the following kind of censoring distributions. Choose the Koziol–Green model

$$1 - F_2(x, 0) = (1 - x)^{1/p_0 - 1} \tag{4.1}$$

and denote by F_a, F_b and F_c the distributions (4.1) with $p_0 = \frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$, respectively. Define

$$F_d(x) = x^2, \qquad F_e(x) = x^3, \qquad 0 < x < 1,$$

and let F_f be the distribution function on $(0, 1)$ such that the density is proportional to $(x - \frac{1}{2})^2$. Later we consider tests at level $\alpha = 0.1$. Table 2 gives by a Monte Carlo simulation the true level of the tests under different censoring distributions. The Monte Carlo study shows that in most of all cases the conditional tests are up to a few exceptional cases (where the conditional tests are too conservative) closer to the desired level $\alpha = 0.1$ as their unconditional counterparts.

5. The proofs. In the sequel we will reduce the test statistic T_n to a form which can be handled by classical rank test arguments. Rewrite (2.1) as

$$T_n = \sum_{i=1}^{n} c_{nD_{ni}} \left(w_n(i)\Delta^{(n,i)} + \sum_{j=i+1}^{n} \frac{w_n(j)\Delta^{(n,j)}}{n+1-j} \right). \tag{5.1}$$

The survival statistic T_n will be compared with

$$\begin{aligned} \tilde{T}_n &= \sum_{i=1}^{n} c_{nD_{ni}} \left(\tilde{w}_n(i)\Delta^{(n,i)} + \sum_{j=i+1}^{n} \frac{\tilde{w}_n(j) p_{nj}}{n+1-j} \right) \\ &= \sum_{i=1}^{n} c_{nD_{ni}} \left(\tilde{w}_n(i)\Delta^{(n,i)} - \sum_{j=1}^{i} \frac{\tilde{w}_n(j) p_{nj}}{n+1-j} \right), \end{aligned} \tag{5.2}$$

where by definition

$$(5.3)\qquad p_{nj} := E\big(p(U_{j:n})\big) \quad\text{and}\quad \tilde{w}_n(j) := E\big(w(U_{j:n})\big)$$

arise from uniformly distributed order statistics $U_{j:n}$ and the limit weight function $w(\cdot)$. If we now define scores $a_n^{(j)}(i)$ for $j = 1, 2$ by

$$(5.4)\qquad a_n^{(1)}(i) - a_n^{(2)}(i) := \tilde{w}_n(i) \quad\text{and}\quad a_n^{(2)}(i) := -\sum_{j=1}^{i} \frac{\tilde{w}_n(j) p_{nj}}{n+1-j},$$

we see that

$$(5.5)\qquad \tilde{T}_n = \sum_{i=1}^{n} c_{nD_{ni}}\big(a_n^{(1)}(i)\Delta^{(n,i)} + a_n^{(2)}(i)(1 - \Delta^{(n,i)})\big).$$

Rank statistics given by (5.5) were carefully studied by Neuhaus (1988) and Janssen (1989) whenever limit score functions $a^{(j)}(\cdot) \in L_2(0,1)$ exist, that is,

$$(5.6)\qquad a_n^{(j)}(1 + [nu]) \to_{L_2} a^{(j)}(u) \quad \text{as } n \to \infty \quad \text{for } j = 1, 2.$$

The reduction of T_n requires the following:

LEMMA 5.1. (a) *Consider nonrandom weights $w_n(i)$ such that* (2.13) *converges in $L_2(0,1)$ to $w(\cdot)$. Then*

$$(5.7)\qquad \sum_{i=1}^{1+[nu]} \frac{w_n(i)}{n+1-i} \to_{L_2} \int_0^u \frac{w(v)}{1-v}\, dv.$$

In particular

$$(5.8)\qquad \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} \frac{w_n(j)}{n+1-j}\right)^2 \to \int_0^1 \left(\int_0^u \frac{w(v)}{1-v}\, dv\right)^2 du \quad \textit{as } n \to \infty$$

and the scores (5.4) *satisfy* (5.6) *with*

$$(5.9)\qquad a^{(1)}(u) - a^{(2)}(u) = w(u) \quad \textit{and} \quad a^{(2)}(u) = -\int_0^u \frac{w(v)p(v)}{1-v}\, dv.$$

(b) *Define*

$$b_1(u) = \gamma_1(H^{-1}(u)) \quad \textit{and} \quad b_2(u) = \frac{\int_{(H^{-1}(u),\infty)} \dot{L}_1\, dP_{10}}{1 - F_1(H^{-1}(u), 0)}$$

with $\psi_1(H^{-1}(u), \delta) = \delta b_1(u) + b_2(u)$, where ψ_1 is as in (3.6). *Whenever $\gamma_1(H^{-1}(\cdot))p(\cdot) \in L_2(0,1)$, the function $b_2(\cdot)$ is square integrable too and*

$$(5.10)\qquad b_2(u) = -\int_0^u \frac{\gamma_1(H^{-1}(v))p(v)}{1-v}\, dv.$$

PROOF. (a) In the sequel we use the abbreviation $\overline{w}_n(u) := w_n(1 + [nu])$. Then Lemma C of the Appendix implies

$$(5.11)\qquad \int_0^u \frac{\overline{w}_n(v)}{1-v}\, dv \to_{L_2} \int_0^u \frac{w(v)}{1-v}\, dv.$$

Uniform integrability shows that

$$\int_0^u \frac{\overline{w}_n(v)}{1-v}\,dv - \int_0^{[nu]/n} \frac{\overline{w}_n(v)}{1-v}\,dv \to_{L_2} 0.$$

Thus it remains to show that

$$\int_0^{[nu]/n} \frac{\overline{w}_n(v)}{1-v}\,dv - \sum_{i=1}^{[nu]} \frac{w_n(i)}{n+1-i} \to_{L_2} 0 \tag{5.12}$$

since $\overline{w}_n(v)/(n-[nv]) \to_{L_2} 0$. Notice that the absolute value of (5.12) equals

$$\begin{aligned} &\left|\sum_{i=1}^{[nu]} w_n(i)\left\{\int_{(i-1)/n}^{i/n} \frac{1}{1-v}\,dv - \frac{1}{n+1-i}\right\}\right| \\ &\qquad \le \sum_{i=1}^{[nu]} |w_n(i)|\left\{\frac{1}{n-i} - \frac{1}{n+1-i}\right\} \\ &\qquad = \sum_{i=1}^{[nu]} \frac{|w_n(i)|}{(n-i)(n+1-i)} \le \int_0^u \frac{|\overline{w}_n(v)|}{1-v}\,dv. \end{aligned} \tag{5.13}$$

For fixed $u<1$, we obtain

$$\sum_{i=1}^{[nu]} \frac{|w_n(i)|}{(n-i)(n+1-i)} \to 0 \tag{5.14}$$

if we apply (5.11) to the scores $|w_n(i)|/(n-i)1_{[1,[nu]]}(i)$. Since the convergence is dominated by an L_2 convergent sequence, we see that (5.13) converges to zero in $L_2(0,1)$.

(b) Note that it is enough to prove the result for the pivoted samples $(H(X_i),\Delta_i)$. Thus we may suppose that $P_{1\vartheta}$ and P_{20} are distributions on $(0,1)$ and $(0,1]$, respectively, such that $H(x)=x$ for $0\le x\le 1$ and $\gamma_1(u)$ is the derivative of the hazard rates ratio of $P_{1\vartheta}$ for $u\in(0,1)$. Differentiation of (2.5) then leads to

$$1 = f_1(u)(1-F_2(u,0)) + f_2(u)(1-F_1(u,0)), \qquad 0<u<1, \tag{5.15}$$

where f_i denotes the density of the absolutely continuous part of $F_i(\cdot,0)$. Next we prove that $F_1(\cdot,0)$ is absolutely continuous with density f_1. If not we see by taking integrals on both sides of (5.15) and using integration by parts that

$$\begin{aligned} 1 &< \int_0^1 (1-F_2(u,0))\,dP_{10}(u) + \int_0^1 (1-F_1(u,0))\,dP_{20}(u) \\ &= 2 - F_1(1,0)F_2(1,0) = 1, \end{aligned} \tag{5.16}$$

since $F_2(u,0)<1$ on $(0,1)$. This is the desired contradiction. Consequently, the definition of b_2 and (3.6) yield

$$(d/du)b_2(u) = -\gamma_1(u)f_1(u)/(1-F_1(u,0)). \tag{5.17}$$

Using the joint distribution of (X_1, Δ_1), it is easy to see that

$$(d/du)b_2(u) = -\gamma_1(u)p(u)/(1-u) \tag{5.18}$$

λ-almost everywhere, which implies

$$b_2(u) = -\int_0^u \frac{\gamma_1(u)p(u)}{1-u}\,du + c. \tag{5.19}$$

On the other hand, $b_2(u) \to 0$ holds as $u \downarrow 0$ since

$$\int \dot{L}_1\,dP_{10} = 0. \tag{5.20}$$

□

In the special situation $H(x) = F_1(x, 0)$, that is, $p(\cdot) \equiv 1$, Lemma 5.1(b) is a consequence of Efron and Johnstone [(1990), Lemma 1] and Ritov and Wellner [(1988), Proposition 2.1].

LEMMA 5.2.

$$T_n - \tilde{T}_n \to_{Q_{n00}} 0.$$

PROOF. In a first step the lemma is proved for $w_n(i) = \tilde{w}_n(i)$; see (5.3). Since D_n and $\Delta^{(n)}$ are independent, the classical variance formula of Hájek and Šidák [(1967), page 61], yields for $n > 1$ and some $K > 0$,

$$\begin{aligned}
&\operatorname{Var}\big(T_n - \tilde{T}_n\big)\\
&\quad= \frac{1}{n-1}\sum_{i=1}^n c_{ni}^2 \sum_{i=1}^n E\Bigg(\Bigg(\sum_{j=i+1}^n \tilde{w}_n(j)\frac{\Delta^{(n,j)} - p_{nj}}{n+1-j}\\
&\qquad\qquad - \frac{1}{n}\sum_{k=1}^n \sum_{j=k+1}^n \tilde{w}_n(j)\frac{\Delta^{(n,j)} - p_{nj}}{n+1-j}\Bigg)^2\Bigg)\\
&\quad\le \frac{K}{n}\sum_{i=1}^n E\Bigg(\Bigg(\sum_{j=1}^i \tilde{w}_n(j)\frac{\Delta^{(n,j)} - p_{nj}}{n+1-j} - \frac{1}{n}\sum_{k=1}^n\sum_{j=1}^k \tilde{w}_n(j)\frac{\Delta^{(n,j)} - p_{nj}}{n+1-j}\Bigg)^2\Bigg)\\
&\quad\le \frac{K}{n}\sum_{i=1}^n E\Bigg(\Bigg(\sum_{j=1}^i \tilde{w}_n(j)\frac{\Delta^{(n,j)} - p_{nj}}{n+1-j}\Bigg)^2\Bigg) \qquad (5.21)\\
&\quad= \frac{K}{n}\sum_{i=1}^n \operatorname{Var}\Bigg(\sum_{j=1}^i \tilde{w}_n(j)\frac{\Delta^{(n,j)}}{n+1-j}\Bigg)\\
&\quad= \frac{K}{n}\sum_{i=1}^n\sum_{j=1}^i \Bigg(\tilde{w}_n(j)^2\frac{\operatorname{Var}(\Delta^{(n,j)})}{(n+1-j)^2}\Bigg)\\
&\qquad+ \frac{K}{n}\sum_{i=1}^n \sum_{\substack{j=1\\k=1\\k\ne j}}^i \frac{\tilde{w}_n(j)\tilde{w}_n(k)\operatorname{Cov}(\Delta^{(n,k)}, \Delta^{(n,j)})}{(n+1-j)(n+1-k)}.
\end{aligned}$$

Next we will prove that the first term of the upper bound of (5.21) vanishes as $n \to \infty$. Notice that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{\tilde{w}_n(j)^2}{(n+1-j)^2} = \frac{1}{n}\sum_{j=1}^{n}\frac{\tilde{w}_n(j)^2}{n+1-j} \to 0 \quad \text{as } n \to \infty.$$

The last assertion can be proved as follows. For $\delta < 1$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=[n\delta]+1}^{n}\tilde{w}_n(i)^2 = \int_\delta^1 w^2(u)\,du,$$

which becomes arbitrarily small for $\delta \uparrow 1$. Moreover observe that

$$\frac{1}{n}\sum_{i=1}^{[n\delta]}\frac{\tilde{w}_n(i)^2}{n+1-i} \le \frac{1}{n}\frac{1}{n+1-[n\delta]}\sum_{i=1}^{n}\tilde{w}_n(i)^2 \to 0$$

as a consequence of the L_2 convergence of $u \to \tilde{w}_n(1+[nu])$. The covariance terms in the upper bound of (5.21) can be treated as follows. Lemma B of the Appendix implies

$$\left|\mathrm{Cov}(\Delta^{(n,k)}, \Delta^{(n,j)})\right| \le \left(\mathrm{Var}(p(U_{k:n}))\left(\mathrm{Var}(p(U_{j:n}))\right)\right)^{1/2}$$

for $k \neq j$. Consequently,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\k=1\\k\neq j}}^{i}\frac{\tilde{w}_n(j)\tilde{w}_n(k)\mathrm{Cov}(\Delta^{(n,k)}, \Delta^{(n,j)})}{(n+1-j)(n+1-k)}$$

$$\le \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}|\tilde{w}_n(j)|\frac{\mathrm{Var}(p(U_{j:n}))^{1/2}}{n+1-j}\right)^2 \to 0 \quad \text{as } n \to \infty$$

by (5.8) and Lemma D since $u \to |\tilde{w}(1+[nu])|\mathrm{Var}(p(U_{1+[nu]:n}))^{1/2}$ converges in $L_2(0,1)$ to zero which completes the proof of the first step. In the second step we assume that $w_n(i)$ satisfies the condition (2.12). Define

$$W_n := \sum_{i=1}^{n} c_{nD_{ni}}\left(\tilde{w}_n(i)\Delta^{(n,i)} + \sum_{j=i+1}^{n}\frac{\tilde{w}_n(j)\Delta^{(n,j)}}{n+1-j}\right),$$

$$S_n := \sum_{i=1}^{n} c_{nD_{ni}}\left(w(H(X_{i:n}))\Delta^{(n,i)} + \sum_{j=i+1}^{n}\frac{w(H(X_{j:n}))\Delta^{(n,j)}}{n+1-j}\right).$$

Since $\mathrm{Var}(W_n - \tilde{T}_n) \to 0$, it suffices to prove that

$$\mathrm{Var}(T_n - S_n) \to 0 \tag{5.22}$$

and

$$\mathrm{Var}(S_n - W_n) \to 0. \tag{5.23}$$

Later we will repeatedly make use of the independence of D_n and $(\Delta^n, (X_{i:n})_{i=1,\dots,n})$. The proof of (5.22) splits into two steps. If we set $\overline{w}_n(i) := w_n(i) - w(H(X_{i:n}))$, we get similarly to (5.21) for $n > 1$,

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i=1}^{n} \overline{w}_n(i)\Delta^{(n,i)} c_{nD_{ni}}\right)\\
&\quad = \frac{1}{n-1}\sum_{i=1}^{n} c_{ni}^2 E\left(\sum_{i=1}^{n}\left(\overline{w}_n(i)\Delta^{(n,i)} - \frac{1}{n}\sum_{j=1}^{n}\overline{w}_n(j)\Delta^{(n,j)}\right)^2\right) \qquad (5.24)\\
&\quad \le \frac{K}{n} E\left(\sum_{i=1}^{n}\overline{w}_n(i)^2\right) \to 0
\end{aligned}$$

if we take (2.12) into account. Similar as in (5.21), we see that

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{n} c_{nD_{ni}} \sum_{j=i+1}^{n} \frac{\overline{w}_n(j)\Delta^{(n,j)}}{n+1-j}\right) &\le \frac{K}{n}\sum_{i=1}^{n} E\left(\left(\sum_{j=1}^{i}\overline{w}_n(j)\frac{\Delta^{(n,j)}}{n+1-j}\right)^2\right)\\
&\le \frac{K}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{\left(E\left(\overline{w}_n(j)^2\Delta^{(n,j)2}\right)\right)^{1/2}}{n+1-j}\right)^2 \qquad (5.25)\\
&\le \frac{K}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{\left(E\left(\overline{w}_n(j)^2\right)\right)^{1/2}}{n+1-j}\right)^2.
\end{aligned}$$

Notice that the condition (2.12) is equivalent to the convergence of $u \mapsto (E(\overline{w}_n(1 + [nu])^2))^{1/2}$ to zero in $L_2(0,1)$. Thus (5.25) converges to zero by assertion (5.8) and (5.22) follows from (5.24) and (5.25). In order to give the proof of (5.23), we substitute $\overline{w}_n(i)$ by

$$w_n'(i) := w(H(X_{i:n})) - \tilde{w}_n(i).$$

As in (5.24), we see that

$$\operatorname{Var}\left(\sum_{i=1}^{n} w_n'(i)\Delta^{(n,i)} c_{nD_{ni}}\right) \le \frac{K}{n}\sum_{i=1}^{n}\operatorname{Var}(w(U_{i:n})) \to 0 \qquad (5.26)$$

which was proved in Lemma D. Similar to (5.25) we conclude as before that

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i=1}^{n} c_{nD_{ni}} \sum_{j=i+1}^{n} \frac{w_n'(j)\Delta^{(n,j)}}{n+1-j}\right) \qquad (5.27)\\
&\quad \le \frac{K}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{\left(\operatorname{Var}\left(w(U_{j:n})\right)\right)^{1/2}}{n+1-j}\right)^2 \to 0. \qquad \square
\end{aligned}$$

REMARK. The proof of Lemma 5.2 only uses (2.12) and the assumption that $w_n(i)$ is a function of $(X_{j:n}, \Delta^{(n,j)})_{j\le n}$.

Now we are in the position to put everything together.

PROOF OF THEOREM 3.1. The present proof uses the language of local asymptotic normality (LAN) of Strasser [(1985), Chapter 13]. Notice that (3.5) and (3.6) yield LAN for Q_{nst} [see Janssen (1989), page 124 and Theorem 1], that is,

$$\log(dQ_{nst}/dQ_{n00}) = (s,t)\left(Z_n^{(1)}, Z_n^{(2)}\right)^T - \left(\sigma_1^2 s^2 + \sigma_2^2 t^2\right)/2 + R_{nst}, \tag{5.28}$$

where in addition to (3.1),

$$Z_n^{(2)} := \sum_{i=1}^{n} d_{ni}\psi_2(X_i, \Delta_i) \tag{5.29}$$

and

$$\sigma_1^2 = c\int \psi_1^2\, dQ_{100}, \qquad \sigma_2^2 = \lim_{n\to\infty} \sum_{i=1}^{n} d_{ni}^2 \int \psi_2^2\, dQ_{100}. \tag{5.30}$$

Moreover,

$$\mathscr{L}\left(\left(Z_n^{(1)}, Z_n^{(2)}\right) \middle| Q_{n00}\right) \to N\left[0, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right] \tag{5.31}$$

in distribution and $\sup_{(s,t)\in K} Q_{n00}(|R_{nst}| > \varepsilon) \to 0$ for each $\varepsilon > 0$ and all compact subsets K of $\mathbb{R}^2$.

Next consider the limit score function $w(\cdot)$ and define $h(x, \delta) := a_1(H(x))\delta + a_2(H(x))(1-\delta)$ via (5.9). Then $\int h\, dQ_{100} = 0$ and

$$W_n := \sum_{i=1}^{n} c_{ni} h(X_i, \Delta_i) \tag{5.32}$$

is a central sequence at the basis point Q_{100} in direction of the tangent vector $(h, 0)$ w.r.t. the local parameter space $\mathbb{R}^2$. Theorem 3 of Janssen (1989) shows that $W_n - \tilde{T}_n \to_{Q_{n00}} 0$ and by Lemma 5.2 the underlying statistic T_n is a central sequence in direction $(h, 0)$ also, that is,

$$W_n - T_n \to_{Q_{n00}} 0. \tag{5.33}$$

Since $\int h\psi_2\, dQ_{100} = 0$ [see Janssen (1989), page 120], arguments based on the third lemma of Le Cam prove that

$$\mathscr{L}(T_n | Q_{nst}) \to N(\mu, \tilde{\sigma}^2) \tag{5.34}$$

in distribution given by the moments (5.35) and (5.36), see Janssen (1989). With the notation of Lemma 5.1(b), we obtain

$$\mu = sc\int h\psi_1\, dQ_{100} = sc\int \left(a_1 b_1 p + a_2 b_2 (1-p)\right) du, \tag{5.35}$$

$$\tilde{\sigma}^2 = c\int h^2\, dQ_{100} = c\int \left(a_1^2 p + a_2^2(1-p)\right) du; \tag{5.36}$$

see Janssen [(1989), pages 113 and 118]. An application of Lemma E yields the result of part (a). The test $1_{[u_{1-\alpha},\infty)}(T_n/\sigma)$ satisfies the assertions of Theorem 3.1(b) and (c). The same assertions also hold by Corollary 2.1 for the condi tional tests $\tilde{\varphi}_n$ if Theorem 2.1(b) is proved. However, Lemma 1 of Janssen (1989) shows that (2.19) is valid for $\overline{T}_n = \tilde{T}_n$.

The proof of Theorem 2.1(b) now follows from our final lemma.

LEMMA 5.3. *Assume that* (2.19) *holds for* T'_n *and let* $\overline{T}_n(D_n, \Delta^{(n)})$ *denote a further sequence of statistics with* $\overline{T}_n - T'_n \to 0$ *in* Q_{n00} *probability. Then* (2.19) *remains true for* $\overline{T}_n$.

PROOF. Since D_n and $\Delta^{(n)}$ are independent, we may choose in addition to $\Delta^{(n)}$: $(\Omega, \mathscr{A}, P) \to \{0,1\}^n$ a new probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ and uniformly distributed random variables $\tilde{D}_n$ from $\tilde{\Omega}$ into the set of permutations of $\{1, \dots, n\}$. Let us abbreviate $\overline{T}_n(\omega, \tilde{\omega}) := \overline{T}_n(\tilde{D}_n(\tilde{\omega}), \Delta^{(n)}(\omega))$ and introduce similarly $T'_n(\omega, \tilde{\omega})$. By our assumptions,

$$T'_n(\omega, \tilde{\omega}) - \overline{T}_n(\omega, \tilde{\omega}) \to_{P \otimes \tilde{P}} 0. \tag{5.37}$$

Passing to subsequences we may assume that (5.37) is almost everywhere convergent. By a proper choice of further subsequences we may again assume that

$$\tilde{\omega} \to T'_n(\omega, \tilde{\omega}) \tag{5.38}$$

is asymptotically $N(0, \sigma^2)$ distributed for $\omega \in N$ with $P(N) = 1$. Set $A = \{(\omega, \tilde{\omega})\colon T'_n(\omega, \tilde{\omega}) - \overline{T}_n(\omega, \tilde{\omega}) \to 0\}$ and $M = \{\omega\colon \tilde{P}(\{\tilde{\omega}\colon (\omega, \tilde{\omega}) \in A\}) = 1\}$. Thus $P(M) = 1$ and $\tilde{\omega} \mapsto T_n(\tilde{\omega}, \omega)$ is asymptotically $N(0, \sigma^2)$ distributed for $\omega \in M \cap N$. The subsequence criterion for convergence in probability finishes the proof. □

APPENDIX

In the sequel we will deduce certain results for the ordered values $(X_{i:n}, \Delta^{(n,i)})_{i=1,\dots,n}$ of the sample (1.1) which are needed as technical tools. The proofs are closely related to the approach of Hájek and Šidák (1967) and are mostly left to the reader. Assume throughout that $(X_{ji})_{i=1,\dots,n}$ are i.i.d. for each $j = 1, 2$ such that (X_1, Δ_1) has a density

$$(x, \delta) \to f(x, \delta) \tag{A1}$$

on $\mathbb{R} \times \{0,1\}$ with respect to the product of the Lebesgue and counting measure.

LEMMA A. (i) $(X_{i:n}, \Delta^{(n,i)})_{i=1,\dots,n}$ *has the joint density*

$$n! \prod_{i=1}^{n} f(x_i, \delta_i) 1_{\{y_1 < y_2 < \cdots < y_n\}}(x_1, \dots, x_n).$$

(ii) *Let* $t\colon (\mathbb{R} \times \{0,1\})^n \to \mathbb{R}$ *denote an integrable function. Then for each permutation* $r = (r_1, \dots, r_n)$ *of* $\{1, \dots, n\}$, *we obtain*

$$E\big(t((X_i, \Delta_i)_i) \big| R_n = r\big) = E\Big(t\Big(\big(X_{r_i:n}, \Delta^{(n, r_i)}\big)_i\Big)\Big).$$

Moreover it is easy to see that R_n and $(X_{1:n}, \dots, X_{n:n}, \Delta^{(n,1)}, \dots, \Delta^{(n,n)})$ are independent (also in the case when no densities exist). Consider now as in (2.7) the pivoted standard model $(H(X_i), \Delta_i)$ with the density

$$(u, \delta) \to \delta p(u) + (1-\delta)(1-p(u)) \tag{A2}$$

for $(u, \delta) \in (0,1) \times \{0,1\}$. Let $U_1, \dots, U_n$ denote i.i.d. random variables with a uniform distribution on $(0,1)$.

LEMMA B. (i) $E(\Delta^{(n,i)}) = E(p(U_{i:n}))$.
(ii) *For each pair* $1 \le i < j \le n$, *we obtain*

$$E(\Delta^{(n,i)}\Delta^{(n,j)}) = E\big(p(U_{i:n})p(U_{j:n})\big).$$

PROOF. (ii) Set $x = (x_1, \dots, x_n)$. Then $(X_{1:n}, \dots, X_{n:n}, \Delta^{(n,i)}, \Delta^{(n,j)})$ has the density

$$(x, \delta_i, \delta_j) \to n!\big(\delta_i p(x_i) + (1-\delta_i)(1-p(x_i))\big)$$
$$\times\big(\delta_j p(x_j) + (1-\delta_j)(1-p(x_j))\big)1_{\{y_1<y_2<\cdots<y_n\}}(x).$$

For the proof use Lemma A(i) and carry out integration over the free components. Thus

$$E(\Delta^{(n,i)}\Delta^{(n,j)}) = \int\int f_{ij}(x_i, x_j)p(x_i)p(x_j)\,dx_i\,dx_j,$$

where

$$f_{ij}(x_i, x_j) = n!\int 1_{\{y_1<y_2<\cdots<y_n\}}(x)\prod_{\substack{k\neq i\\ k\neq j}} dx_k$$

is the joint density of $(U_{i:n}, U_{j:n})$. The proof of formula (i) is similar. □

LEMMA C. *Assume that* $\varphi_n \to_{L_2} \varphi_0$ *in* $L_2(0,1)$. *Then*

$$w_n(u) := \int_0^u \varphi_n(v)/(1-v)\,dv$$

belongs to $L_2(0,1)$ *and* $w_n \to_{L_2} w_0$.

Lemma C was proved by Khmaladze (1981); see also Efron and Johnstone (1990) and Ritov and Wellner (1988).

LEMMA D. *Assume that* $w \in L_2(0,1)$. *Then we obtain*

$$\frac{1}{n}\sum_{i=1}^{n} \operatorname{Var}(w(U_{i:n})) \to 0. \tag{A3}$$

We see that $u \to (\operatorname{Var}(w(U_{1+[nu]:n})))^{1/2}$ converges to zero in $L_2(0,1)$.

PROOF. Set $a_n(i) = E(w(U_{i:n}))$. From Hájek and Šidák [(1967), page 157], we obtain

$$\frac{1}{n}E\left(\sum_{i=1}^{n}\left(w(U_{i:n}) - a_n(i)\right)^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left(w(U_i) - a_n(R_{ni})\right)^2\right)$$

$$= E\left(\left(w(U_1) - a_n(R_{n1})\right)^2\right) \to 0. \qquad \square$$

For the reformulation of the moments of T_n, we need the following result. Again there is for $p(\cdot) \equiv 1$ some overlap with the papers of Efron and Johnstone (1990) and Ritov and Wellner (1988) who calculated the Fisher information in terms of the derivative of the hazard function.

LEMMA E. *Assume that w_1 and w_2 are two functions on* $(0,1)$ *such that $w_j(\cdot)p(\cdot)^{1/2}$ is square integrable for a regression function $p(\cdot)$. Define*

$$a_1^{(j)}(u) - a_2^{(j)}(u) = w_j(u), \qquad a_2^{(j)}(u) = -\int_0^u \frac{w_j(v)p(v)}{1-v}\,dv \tag{A4}$$

for $j = 1, 2$.

Then

$$\int a_1^{(1)}(u)a_1^{(2)}(u)p(u)\,du + \int a_2^{(1)}(u)a_2^{(2)}(u)(1-p(u))\,du$$

$$= \int w_1(u)w_2(u)p(u)\,du.$$

PROOF. (a) First we prove that $(1-u)a_2^{(1)}(u)a_2^{(2)}(u) \to 0$ as $u \downarrow 0$ or $u \uparrow 1$. Since Lebesgue's theorem can be used for the first limit we may restrict ourselves to the case $u \uparrow 1$. Notice that

$$(1-u)a_2^{(1)}(u)^2 \le (1-u)\left(\int_0^u \frac{|w_1(v)|p(v)}{1-v}\,dv\right)^2$$

$$\le \int_u^1\left(\int_0^s \frac{|w_1(v)|p(v)}{1-v}\,dv\right)^2 ds \to 0$$

by Lemma C of the Appendix. A similar result holds for $a_2^{(2)}(u)$. Now integration by parts of $-a_2^{(1)}w_2p(\cdot)$ shows the result. $\square$

Acknowledgments. The author thanks M. Schumacher, S. Leurgans and H. Milbrodt for helpful comments and H. Witting for his interest and encouragement. I am grateful to my students S. Brenner and A. Philippus for their help during the Monte Carlo simulation.

REFERENCES

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

ALBERS, W. and AKRITAS M. G. (1987). Combined rank tests for the two-sample problem with randomly censored data. *J. Amer. Statist. Assoc.* **82** 648–655.

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. and KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Internat. Statist. Rev.* **50** 219–258.

EFRON, B. and JOHNSTONE, I. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18** 38–62.

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

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

HARRINGTON, D. P. and FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69** 553–566.

JANSSEN, A. (1989). Local asymptotic normality for randomly censored data with applications to rank tests. *Statist. Neerlandica* **43** 109–125.

JANSSEN, A. (1991). Optimal k-sample tests for randomly censored data. *Scand. J. Statist.* To appear.

JANSSEN, A. and BRENNER, S. (1991). Monte Carlo results for conditional survival tests under randomly censored data. *J. Statist. Comput. Simulation*. To appear.

KHMALADZE, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.

LATTA, R. B. (1981). A Monte Carlo study of some two-sample rank tests with censored data. *J. Amer. Statist. Assoc.* **76** 713–719.

LEURGANS, S. (1984). Asymptotic behavior of two-sample rank tests in the presence of random censoring. *Ann. Statist.* **12** 572–589.

NEUHAUS, G. (1987). Linear rank tests with estimated score functions for the two-sample problem with randomly censored data. In *Proceedings Fourth Prague Symposium on Asymptotic Statist.* To appear.

NEUHAUS, G. (1988). Asymptotically optimal rank tests for the two-sample problem with randomly censored data. *Comm. Statist. Theory Methods* **17** 2037–2058.

PFANZAGL, J. and WEFELMEYER, W. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statist.* **13**. Springer, New York.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 165–179.

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales, and the Cox model. In *Statistical Inference from Stochastic Processes* (N. U. Pabhu, ed.) **80** 191–219. *Contemporary Math.* Amer. Math. Soc, Providence, R. I.

SCHUMACHER, M. (1984). Two-sample tests of Cramér-von Mises and Kolmogorov–Smirnov type for randomly censored data. *Internat. Statist. Rev.* **52** 263–281.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

VAN DER VAART, A. W. (1988). Statistical estimation in large parameter spaces. *CWI Tract* **44** Centrum voor Wiskunde en Informatica, Amsterdam.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SIEGEN
HÖLDERLINSTR. 3
D-59 SIEGEN
GERMANY