Some Tests for Comparing Cumulative Incidence Functions and Cause-Specific Hazard Rates

Emad-Eldin A. A. ALY, Subhash C. KOCHAR, and Ian W. MCKEAGUE*

We consider the competing risks problem with the available data in the form of times and causes of failure. In many practical situations (e.g., in reliability testing) it is important to know whether two risks are equal or whether one is "more serious" than the other. We propose some distribution-free tests for comparing cumulative incidence functions and cause-specific hazard rates against ordered alternatives without making any assumptions on the nature of dependence between the risks. Both the censored and the uncensored cases are studied. The performance of the proposed tests is assessed in a simulation study. As an illustration, we compare the risks of two types of cancer mortality (thymic lymphoma and reticulum cell carcinoma) in a strain of laboratory mice.

KEY WORDS: Competing risks; Counting processes; Distribution-free tests; Ordered alternatives; Right-censored data.

1. INTRODUCTION

In the competing risks model, a unit is exposed to several risks at the same time, but it is assumed that the eventual failure of the unit is due to only one of these risks, which is called a "cause of failure." Let a unit be exposed to two risks and let the notional (or latent) lifetimes of the unit under these two risks be denoted by X and Y. In general, X and Y are dependent. Also, being lifetimes, they are nonnegative. We only observe (T, δ), where $T = \min(X, Y)$ is the time of failure and $\delta = 2 - I(X \leq Y)$ is the cause of failure. Here $I(A)$ is the indicator function of the event A. We assume that $P(X = Y) = 0$.

On the basis of the competing risks data, it is often useful to distinguish between the following alternatives: (a) the two risks are equal, and (b) one risk is greater than the other, within the environment in which the two risks are acting simultaneously. Such comparisons can be made in terms of the *cumulative incidence function*,

$$F_j(t) = P[T \leq t, \delta = j],$$

corresponding to each cause j.

Such comparisons are useful in many practical situations in industrial engineering and reliability life testing. Suppose that either of two components in a series system can be replaced to improve overall system reliability. A reasonable approach is to compare estimates of F_1 and F_2 and to replace the second component in preference to the first, say, if there is evidence to reject $F_1 = F_2$ in favor of $F_1 < F_2$. Similarly, to compare the quality of two types of components, they may be tested in pairs (cf. Froda 1987). The experiment is terminated as soon as either component fails. This experimental design identifies weak components early on, thus saving valuable time and accelerating the experiment. From the competing risks data, one would like to test whether the two components are of the same quality (i.e., $F_1 = F_2$) against the ordered alternative that the first component (say) is of better quality (i.e., $F_1 < F_2$). In the biomedical setting, the comparison of cumulative incidence among different types of failure may be useful when selecting the appropriate treatment for a patient (see Gray 1988). Benichou and Gail (1990) stressed the importance of cumulative incidence *estimation* in this context.

In this article we propose some tests for comparing cumulative incidence functions. Our tests are less subjective than inspection of estimates of the cumulative incidence functions alone. We first consider a test of the null hypothesis H_0: $F_1(t) = F_2(t)$, $t \geq 0$ against the ordered alternative

$$H_1: F_1(t) \leq F_2(t), \qquad t \geq 0,$$

with strict inequality for some t. Here H_1 says that risk Y is "more serious" than risk X. Note that there is often no reason to expect a priori that the cumulative incidence functions F_1 and F_2 are equal (except, say, when they represent two identical components in a series system), but this is the natural choice of the null hypothesis for the ordered alternative H_1.

In some applications it is of interest to base the comparison of risks on the *cause-specific hazard rate* (CSHR),

$$g_j(t) = f_j(t)/S_T(t),$$

where the F_j are assumed to have subdensities $f_j(t)$ and $S_T(t) = P[T > t] = 1 - F_1(t) - F_2(t)$ is the survival function of T. CSHR's provide detailed information on the extent of each type of risk at each time t. In the case where X and Y are independent, g_1 and g_2 reduce to the hazard rates corresponding to the marginal distributions of X and Y. Prentice et al. (1978) showed that in general only probabilities expressible as functions of g_1 and g_2 may be estimated from the observable data (T, δ). Because the cumulative incidence functions can be expressed as $F_j(t) = \int_0^t g_j(u) S_T(u)\, du$, the null hypothesis H_0 is equivalent to $g_1(t) = g_2(t)$, $t \geq 0$.

We introduce a second test of H_0 that is tailored to the ordered alternative

$$H_2: g_1(t) \leq g_2(t), \qquad t \geq 0,$$

* Emad-Eldin A. A. Aly is Professor, Department of Mathematical Sciences, The University of Alberta, Edmonton, Alberta, Canada T6G 2G1. Subhash C. Kochar is Associate Professor, Statistics-Mathematics Unit, Indian Statistical Institute, New Delhi, 110016, India. Ian W. McKeague is Professor, Department of Statistics, Florida State University, Tallahassee, FL 32306. Aly's work was supported by an NSERC Canada grant at the University of Alberta. Part of Kochar's research was done while visiting the University of Alberta, supported by Aly's NSERC Canada grant. McKeague's work was partially supported by U.S. Air Force Office of Scientific Research Grant AFOSR91-0048. The authors thank Ectit Gombay for helpful discussion. They also thank the associate editor and the referees for valuable suggestions that improved the presentation and focus of the paper.

with strict inequality for some t. H_2 is more restrictive than H_1, yet the common *parametric* models satisfy H_2 whenever they satisfy H_1. Such a test might be useful in reliability testing: In the foregoing example, suppose that the second component consists of another component of the same type as the first component and a third component in series. Then, because it is known a priori that $g_1(t) \le g_2(t)$ for all t, H_2 is a more natural choice of ordered alternative than H_1 (which includes the possibility that $g_2(t) < g_1(t)$ for some t).

Various authors have proposed tests of H_0 in the case that X and Y are *independent:* Bagai, Deshpandé, and Kochar (1989a,b) developed distribution-free rank tests against stochastic ordering and failure rate ordering alternatives; Neuhaus (1991) constructed asymptotically optimal rank tests against stochastic ordering; and Yip and Lam (1992) suggested a class of weighted logrank-type statistics. The case of *dependent* X, Y has been considered only recently: Aras and Deshpandé (1992) derived locally most powerful rank tests for H_0 against various parametric alternatives expressed in terms of F_1 and F_2. But none of these tests allow for the possibility of censoring, and they are sensitive to only a relatively small range of departures from H_0. The tests introduced in this article are asymptotically distribution-free, consistent against H_1, H_2, and applicable to right-censored data and dependent X, Y.

This article is organized as follows. In Section 2 we introduce our test statistics and give formulas for their exact null distributions. We also derive the asymptotic null distributions. In Section 3 we develop the extension of our approach to right-censored data and explain how to deal with multiple (rather than just two) competing risks. We also discuss comparisons on finite time intervals. Finally, we present the results of a simulation study and an example in Section 4.

2. UNCENSORED DATA

The tests of the null hypothesis H_0 introduced in this section are based on the uncensored competing risk data $\{(T_i, \delta_i), i = 1, \ldots, n\}$ for n independent and identical units.

First, consider a test of H_0 vs. H_1. Note that H_1 is equivalent to $\psi(t) \ge 0$ for all $t \ge 0$, with strict inequality for some t, where $\psi(t) = F_2(t) - F_1(t)$. Thus a natural test statistic for detecting the alternative H_1 is given by

$$D_{1n} = \sup_{0<t<\infty} \psi_n(t),$$

where $\psi_n(t) = F_{2n}(t) - F_{1n}(t)$ and $F_{jn}(t) = n^{-1} \sum_{i=1}^{n} I\{\delta_i = j, T_i \le t\}$ is the empirical cumulative incidence function for cause j. Positive values of D_{1n} provide evidence in favor of H_1. Note that

$$D_{1n} = \max_{0\le j\le n} \frac{1}{n}\left\{j - 2\sum_{i=1}^{j} W_i\right\} = \max_{0\le j\le n} Z_j/n,$$

where

$W_i = 1$ if δ corresponding to $T_{(i)}$ (the ith ordered T_i) is 1

$= 0$ otherwise,

$Z_k = \eta_1 + \cdots + \eta_k$, $Z_0 = 0$, and $\eta_i = 1 - 2W_i$. Kochar and Proschan (1991) proved that T and δ are independent under H_0. Consequently, under H_0, Z_j is a symmetric simple random walk starting at 0, and by lemma 4.8.1 of Rényi (1970),

$$P\{nD_{1n} = k\} = \frac{1}{2^n}\binom{n}{\left[\frac{n-k}{2}\right]}, \qquad k = 0, 1, \ldots, n.$$

This gives the exact null distribution of D_{1n}. The asymptotic null distribution is obtained using the invariance principle for partial sums (see, for example, Csörgő and Révész 1981, chap. 2): under H_0,

$$P\{\sqrt{n}D_{1n} > x\} \to P\{\sup_{0<t<1} W(t) > x\} = 2(1 - \Phi(x)), \qquad x \ge 0,$$

where $\{W(t), t \ge 0\}$ is a standard Brownian motion and Φ is the standard normal distribution function.

Next, consider testing H_0 versus H_2. The alternative H_2 is equivalent to ψ increasing (assume that the g_j are continuous). This is a consequence of the identity $\psi(t) = \int_0^t S_T(u)(g_2(u) - g_1(u))\,du$ and provides a rationale for the test statistic

$$D_{2n} = \sup_{0<s<t<\infty} \{\psi_n(t) - \psi_n(s)\}.$$

Positive values of D_{2n} provide evidence that $g_2(t)$ is larger than $g_1(t)$ for some t. The exact null distribution of D_{2n} is given by

$$P\{nD_{2n} < t\} = \frac{2}{2t+1}\sum_{j=0}^{2t}\left\{\cos\frac{j\pi}{2t+1}\right\}\sin\left\{\frac{j\pi(t+1)}{2t+1}\right\} \times\left\{1 + \cos\frac{j\pi}{2t+1}\right\}\left\{\frac{1-(-1)^j}{2}\right\}\Big/\sin\frac{j\pi}{2t+1} \quad (1)$$

for $t = 1, \ldots, n+1$. This follows from the identity

$$nD_{2n} = \max_{1\le j\le n}(Z_j - \min_{1\le i\le j} Z_i), \quad (2)$$

and equations (1), (4), and (5) of Page (1955).

The asymptotic null distribution can be obtained from (2) and the invariance principle for partial sums: under H_0,

$$\sqrt{n}D_{2n} \xrightarrow{\mathcal{D}} \sup_{0\le x\le 1} |W(x)|.$$

Consequently, for $c > 0$,

$$P\{\sqrt{n}D_{1n} \le c\} \to \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\{-\pi^2(2k+1)^2/8c^2\}. \quad (3)$$

The exact formula (1) can easily be used to generate a table of critical values. Using (3), the asymptotic .90, .95, and .99 quantiles of $\sqrt{n}D_{2n}$ are found to be 1.96, 2.241, and 2.807.

When an ordered alternative is unsuitable, it can be of interest to test H_0 against the general alternative, $F_1(t) \ne F_2(t)$ for some t, which is equivalent to $g_1(t) \ne g_2(t)$ for some t. In that case it is natural to use the Kolmogorov-

Smirnov test statistic $\bar{D}_n = \sup_{t\geq 0} |\psi_n(t)|$. Under H_0, $\sqrt{n}\bar{D}_n$ converges in distribution to $\sup_{0\leq x\leq 1} |W(x)|$. This gives an omnibus test—consistent against *arbitrary* departures from H_0 (see the proof of Theorem 3.1 in the Appendix).

3. CENSORED DATA AND OTHER EXTENSIONS

In this section we consider various extensions of our tests that will make them more widely applicable. We study the censored data case, additional causes of failure beyond those due to X and Y, and comparisons on subintervals.

3.1 Censored Data

Censoring arises when an item is removed from observation before failure due to X or Y. Denote the censoring time by C and its survival function by S_C. Assume that $S_C(t) > 0$ for all t and that C is independent of X and Y. Under right censoring, we observe n iid copies, $(\tilde{T}_i, \tilde{\delta}_i)$, $i = 1, \ldots, n$, of $\tilde{T} = \min(T, C)$ and $\tilde{\delta} = \delta I(T \leq C)$.

Our approach is to seek a suitable generalization of the function $\psi = F_2 - F_1$. Consider the function

$$\phi(t) = \int_0^t S_C(u-)^{1/2}\, d(F_2 - F_1)(u),$$

which coincides with ψ when there is no censoring. The integrand $S_C(u-)^{1/2}$ turns out to be precisely what is needed to compensate for the censoring for our test statistics to remain (asymptotically) distribution free. H_0 is equivalent to $\phi(t) = 0$ for all $t \geq 0$, but under H_1, $\phi(t) > 0$ for some t (see Lemma 1 in the Appendix). Thus positive values of the test statistic

$$D_{3n} = \sup_{0\leq t<\infty} \phi_n(t),$$

where ϕ_n is an estimator of ϕ, give evidence of a departure from H_0 in the direction of H_1. Because

$$\phi(t) = \int_0^t S_T(u-)S_C(u-)^{1/2}(g_2(u) - g_1(u))\, du,$$

H_2 is equivalent to ϕ increasing. Thus positive values of

$$D_{4n} = \sup_{0\leq s\leq t<\infty} \{\phi_n(t) - \phi_n(s)\}$$

give evidence of a departure from H_0 in the direction of H_2. An obvious choice of ϕ_n is

$$\phi_n(t) = \int_0^t \hat{S}_T(u-)\hat{S}_C(u-)^{1/2}\, d(\hat{\Lambda}_1 - \hat{\Lambda}_2)(u),$$

where $\hat{S}_T$ and $\hat{S}_C$ are the product-limit estimators of S_T and S_C and $\hat{\Lambda}_j$ is the Aalen estimator of the cumulative CSHR function $\Lambda_j(t) = \int_0^t g_j(u)\, du$:

$$\hat{\Lambda}_j(t) = \sum_{i:\tilde{T}_i\leq t} I(\tilde{\delta}_i = j)/R_i,$$

where $R_i = \#\{k: \tilde{T}_k \geq \tilde{T}_i\}$ is the size of the risk set at time $\tilde{T}_i-$.

The estimator $\hat{\Lambda}_j$ is a special case of an estimator discussed by Aalen and Johansen (1978) in connection with inference for the transition probabilities of a non–time-homogeneous Markov chain with finitely many states. Our approach could easily be generalized to deal with comparisons between such transition probabilities. The problem at hand concerns a three-state chain with two absorbing states corresponding to the two types of failure.

The estimate $\phi_n(t)$ is similar in spirit to a weighted log rank statistic of the form

$$L_n(t) = \int_0^t w(u)\, d(\hat{\Lambda}_1 - \hat{\Lambda}_2)(u),$$

where w is a locally bounded, predictable weight function. The weight $w(u)$ reflects the relative importance attached to the difference between the CSHR's at time u. Our choice of w, which essentially controls instability in the tails, is designed to give an asymptotically distribution-free test. Yip and Lam (1992) have suggested test statistics based on normalized $L_n(\infty)$ for various other choices of w. They considered only the case of uncensored data and independent X and Y, but their approach readily extends to the present setting.

The following result, proved in the Appendix, shows that D_{3n} and D_{4n} are asymptotically distribution free with the same limiting distributions as in the uncensored case.

Theorem 3.1. Under H_0,

$$\sqrt{n}D_{3n} \xrightarrow{\mathcal{D}} \sup_{0\leq x\leq 1} W(x) \quad \text{and} \quad \sqrt{n}D_{4n} \xrightarrow{\mathcal{D}} \sup_{0\leq x\leq 1} |W(x)|.$$

Moreover, the tests are consistent against their respective alternatives.

The omnibus test statistic $\bar{D}_n$ has a similar extension to the censored data setting.

3.2 Additional Competing Risks

Our approach further extends to the case of multiple (rather than just two) competing risks in which any two of the cause-specific risks are to be compared. No structure need be imposed on the dependency between the multiple risks, although the corresponding latent failure times must be independent of the censoring. Let T be the minimum of a finite collection of latent failure times that include X and Y (but not the censoring), and let δ denote the corresponding cause of failure. The cumulative incidence functions of X and Y and the various hypotheses are defined as before. Extensions of D_{3n} and D_{4n} that preserve the foregoing asymptotic distributions are obtained by using $\phi_n(t)/\sqrt{p_n}$ in place of $\phi_n(t)$, where

$$p_n = \int_0^\infty \hat{S}_T(u-)\, d(\hat{\Lambda}_1 + \hat{\Lambda}_2)(u)$$

is a consistent estimator of $P[\delta = 1 \text{ or } 2]$. Further details are given in the Appendix.

3.3 Comparisons on Subintervals

It is often useful to compare cumulative incidence functions (or CSHR's) in a given time interval, say $[t_1, t_2]$, rather than at all times. An example will be given in the next section.

It is straightforward to generalize our tests to deal with such comparisons. Defining $F_j^*(t) = P[t_1 < T \leq t, \delta = j]$, the null hypothesis is now $F_1^*(t) = F_2^*(t)$, $t_1 \leq t < t_2$, which is equivalent to $g_1(t) - g_2(t)$, $t_1 \leq t < t_2$.

Extensions of our earlier test statistion are obtained by using

$$\phi_n^*(t) = (\hat{S}_T(t_1) - \hat{S}_T(t_2))^{-1/2}(\phi_n(t) - \phi_n(t_1))$$

on the interval $[t_1, t_2)$ instead of $\phi_n(t)$ on $[0, \infty)$. Theorem 3.1 readily extends to this case.

4. SIMULATION STUDY AND AN EXAMPLE

Our test procedures are consistent against their respective alternatives; however, we would like to know whether they are powerful enough for practical purposes. In this section we report the results of a simulation study designed to address this question and apply our methods to a set of real data.

4.1 Simulation Results

For the distribution of (X, Y), we used Block and Basu's (1974) absolutely continuous bivariate exponential (ACBVE) distribution with density

$$f(x, y) = \frac{\lambda_1 \lambda(\lambda_2 + \lambda_0)}{\lambda_1 - \lambda_2} e^{-\lambda_1 x - (\lambda_2 - \lambda_0) y} \quad \text{if } x < y$$

$$= \frac{\lambda_2 \lambda(\lambda_1 + \lambda_0)}{\lambda_1 + \lambda_2} e^{-\lambda_2 y - (\lambda_1 + \lambda_0) x} \quad \text{if } x > y,$$

where $(\lambda_0, \lambda_1, \lambda_2)$ are parameters and $\lambda = \lambda_0 + \lambda_1 + \lambda_2$. The CSHR's

$$g_j(t) = \frac{\lambda_j \lambda}{\lambda_1 + \lambda_2}$$

Table 1. Observed Levels and Powers of Test for Equality of Cumulative Incidence Functions Based on D_{3n} at an Asymptotic Level of 5%

	n = 50		n = 100		n = 500	
λ_2	$\lambda_0 = 0$	$\lambda_0 = 1$	$\lambda_0 = 0$	$\lambda_0 = 1$	$\lambda_0 = 0$	$\lambda_0 = 1$
			Uncensored			
1.0	4.90	4.90	4.44	4.44	4.63	4.63
1.5	39.46	39.46	61.05	61.05	99.71	99.71
2.0	74.95	74.95	95.11	95.11	100	100
2.5	91.96	91.96	99.78	99.78	100	100
			Lightly censored (18–33%)			
1.0	3.64	3.87	4.16	4.06	4.71	4.64
1.5	27.64	30.00	47.97	51.22	97.94	98.52
2.0	60.52	63.64	87.64	89.76	100	100
2.5	82.91	84.80	98.57	98.75	100	100
			Heavily censored (40–60%)			
1.0	2.29	2.82	2.61	3.64	3.74	4.27
1.5	16.02	19.79	29.12	35.85	88.78	93.42
2.0	39.76	46.75	68.79	76.49	99.98	100
2.5	63.73	70.27	91.57	94.72	100	100

NOTE: The underlying distribution of (X, Y) is Block and Basu's (1974) ACBVE with $\lambda_1 = 1$. The data were created using the uniform random number generator of Marsaglia, Zaman and Tsang (1990) and an algorithm of Friday and Patil (1977, cor. 3.3). 10,000 samples were used to obtain each entry in the table.

Table 2. Observed Levels and Powers of Test for Equality of Cumulative Incidence Functions Based on D_{4n} at an Asymptotic Level of 5%

	n = 50		n = 100		n = 500	
λ_2	$\lambda_0 = 0$	$\lambda_0 = 1$	$\lambda_0 = 0$	$\lambda_0 = 1$	$\lambda_0 = 0$	$\lambda_0 = 1$
			Uncensored			
1.0	3.86	3.86	3.68	3.69	4.16	4.16
1.5	32.37	32.38	54.41	54.43	99.48	99.48
2.0	67.46	67.46	92.59	92.59	100	100
2.5	87.66	87.66	99.4	99.40	100	100
			Lightly censored (18–33%)			
1.0	2.89	2.98	3.45	3.37	4.28	4.53
1.5	21.95	24.19	41.32	44.86	96.73	97.79
2.0	51.95	55.40	83.31	85.57	100	100
2.5	76.35	78.49	97.47	97.97	100	100
			Heavily censored (40–60%)			
1.0	1.49	2.16	1.80	2.79	3.27	3.81
1.5	11.09	14.76	22.88	29.24	84.18	90.61
2.0	30.38	37.26	60.59	69.25	99.92	99.99
2.5	53.11	60.73	86.49	91.01	100	100

NOTE: The underlying distribution of (X, Y) is Block and Basu's (1974) ACBVE with $\lambda_1 = 1$. The data were created using the uniform random number generator of Marsaglia, Zaman and Tsang (1990) and an algorithm of Friday and Patil (1977, cor 3.3). 10,000 samples were used to obtain each entry in the table.

are proportional, and the alternative hypotheses H_1 and H_2 are equivalent to $\lambda_1 < \lambda_2$. The parameter λ_0 controls the degree of dependence between X and Y, with independence if and only if $\lambda_0 = 0$. We set $\lambda_1 = 1$ and considered various higher values of λ_2 corresponding to increasing departures from H_0. The censoring was taken to be exponential with parameter values 1 and 3, corresponding to "light" and "heavy" censoring (about 25% and 50% censored). For comparison, we included results for the uncensored case as well. We used asymptotic critical levels of 5%.

Inspection of Tables 1 and 2 shows that use of the asymptotic critical levels gives somewhat conservative tests, and that this effect increases as the censoring becomes more severe. But the test based on D_{3n} appears to be less conservative (and more powerful) than the one based on D_{4n}, and both tests become less conservative as the sample size increases. The levels of the tests are close to their nominal 5% values for sample size 500, except under heavy censoring. There is no apparent adverse effect on the levels or the power due to lack of independence of X and Y. (Pearson's correlation between X and Y is about .15 for the table entries corresponding to $\lambda_0 = 1$.)

Because T and δ are independent whenever the CSHR's are proportional, it follows that under the ACBVE distribution, the sign test (based on the proportion of failures from cause 1) is the locally most powerful rank test of H_0 against proportional CSHR's in the absence of censoring (see Aras and Deshpandé 1992). Our simulations indicated the power of the sign test to be 44%, 79%, and 94% for $n = 50$ and $\lambda_2 = 1.5, 2.0$, and 2.5 ($\lambda_0 = 0$ and 1, uncensored data). Comparing these figures with the second and third columns in Table 1 (uncensored), we find at most a 5% loss of power for our test compared to the sign test.

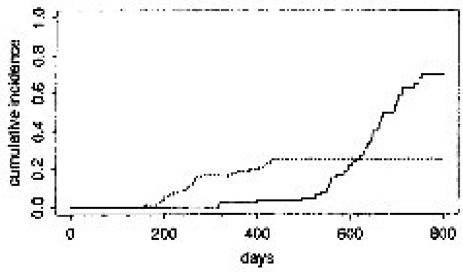

Figure 1. Cumulative Incidence for Lymphoma (---) and Sarcoma (——).

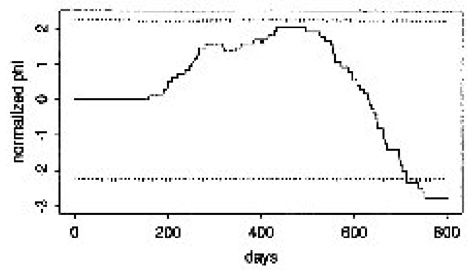

Figure 3. Plot of $\sqrt{n}\phi_n$ (Solid Line) and Corresponding Asymptotic 5% Critical Levels (Dashed Lines) for the Omnibus Test Based on $\bar{D}_n$.

4.2 Application to Real Data

We have analyzed a set of mortality data given in Hoel (1972). These data were obtained from a laboratory experiment on 99 RMF strain male mice that had received a radiation dose of 300 rads at 5–6 weeks of age and were kept in a conventional laboratory environment. Causes of death were classified into thymic lymphoma, reticulum cell sarcoma, and other causes. We shall treat "other causes" as censoring (39% were in this category), and take the two types of cancer mortality as the two causes of failure that we wish to compare; that is, F_1 and F_2 are the cumulative incidence functions for death from sarcoma and lymphoma in the absence of risk from other causes of death. Our analysis depends on the assumption that the two diseases are lethal and independent of other causes of death (which is biologically reasonable, according to Hoel). We do not need to assume that the two diseases are independent of one another. An alternative analysis of the data would be to treat other causes of death as a competing risk (cf. Sec. 3.2).

Plots of estimates of the cumulative incidence functions (Fig. 1) suggest that up to about 500 days, there is moderate probability of (death from) lymphoma and small probability of sarcoma. After 500 days, the situation reverses, with negligible probability of lymphoma but high probability of sarcoma. This is reflected in the plots of the smoothed CSHR estimates in Figure 2, which were obtained using an Epanechnikov kernel function and a bandwidth of 80 days applied to the cumulative CSHR estimates (cf. Ramlau-Hansen 1983).

Our tests offer a less subjective comparison than can be made from visual inspection of such plots. We obtained the highly significant $\sqrt{n}D_{3n} = 3.69$ (resp. 5.56) when testing whether the cumulative incidence for lymphoma is larger (resp. smaller) than the cumulative incidence for sarcoma before (resp. after) 500 days. (Both P values were less than .0003.) The tests based on D_{4n} gave similar results. But the omnibus test of $F_1 = F_2$ gave a considerably less significant result. We obtained $\sqrt{n}\bar{D}_n = 2.77$ (which is significant at the 5% level, but not at the 1% level); see Figure 3. This illustrates an advantage of testing against an appropriate *ordered* alternative on a suitable subinterval, as opposed to using the omnibus test.

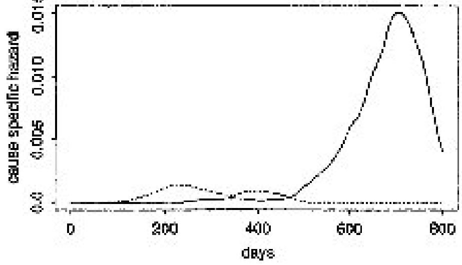

Figure 2. Cause Specific Hazard for Lymphoma (---) and Sarcoma (——).

APPENDIX: PROOFS

Proof of Theorem 3.1

Suppose we can show that

$$\sqrt{n}\,\phi_n \xrightarrow{\mathcal{D}} W(F_T(\cdot)). \tag{A.1}$$

Then the first part of the theorem is clear by the continuous mapping theorem. For the second part,

$$\sqrt{n}\sup_{0\le s<t<\infty}\{\phi_n(t)-\phi_n(s)\} \xrightarrow{\mathcal{D}} \sup_{0\le s<t<\infty}\{W(F_T(t))-W(F_T(s))\}$$
$$\stackrel{\mathcal{D}}{=} \sup_{0\le u<v\le 1}\{W(u)-W(v)\} = \sup_{0\le v\le 1} V(v),$$

where $V(v) = \sup_{0\le u\le v} W(u) - W(v)$. The second part now follows from the well-known result of Lévy (1948) that the processes $V(\cdot)$ and $|W(\cdot)|$ are identically distributed. It remains to prove (A.1), for which we use the counting process approach developed by Aalen (1978). Note that we can write $\hat{\Lambda}_j$ in the form

$$\hat{\Lambda}_j(t) = \int_0^t \frac{d\bar{N}_j(u)}{\bar{Y}(u)},$$

where $1/0 = 0$,

$$\bar{Y} = \sum Y_i, \qquad \bar{N}_j = \sum N_{ij},$$
$$Y_i(u) = I(\tilde{T}_i \ge u), \qquad N_{ij}(u) = I(\tilde{T}_i \le u, \tilde{\delta}_i = j),$$

for $j = 1, 2$, and the summations are over $i = 1, \ldots, n$. Let

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_i(u)\, d\Lambda_j(u).$$

Then M_{ij}, $i = 1, \ldots, n$ are orthogonal martingales under the natural filtration generated by the foregoing processes. Let $\bar{M}_j = \sum M_{ij}$. The predictable variation process of $\bar{M}_j$ is $\int_0^t \bar{Y}(u)\, d\Lambda_j(u)$. By $P(X = Y) = 0$, the counting processes $\bar{N}_1$ and $\bar{N}_2$ almost surely have no simultaneous jumps, so $\bar{M}_1$ and $\bar{M}_2$ are orthogonal martingales (this is a standard result from counting process theory). Thus the predictable variation process of $\bar{M}_2 - \bar{M}_1$ is $\int_0^t \bar{Y}(u)\, d\Lambda_0(u)$, where $\Lambda_0 = \Lambda_1 + \Lambda_2$. Under H_0,

$$\phi_n(t) = \int_0^t \frac{\hat{S}_T(u-)\hat{S}_C(u-)^{1/2}}{\bar{Y}(u)}\, d(\bar{M}_2 - \bar{M}_1)(u).$$

Because $\hat{S}_T(u-)$ and $\hat{S}_C(u-)$ are left continuous and adapted, they are predictable, so $\sqrt{n}\,\phi_n$ is a martingale with predictable variation process

$$\int_0^t \frac{\hat{S}_T(u-)^2\hat{S}_C(u-)}{\bar{Y}(u)/n}\, d\Lambda_0(u).$$

By the Glivenko–Cantelli theorem, $\bar{Y}(u)/n$ converges uniformly in u to $P(\tilde{T} \geq u) = S_T(u-)S_C(u-)$ almost surely. Hence, by the uniform consistency of the product-limit estimator on $[0, t]$, the foregoing variation process converges in probability to $\int_0^t S_T(u-)\, d\Lambda_0(u) = F_T(t)$. Here we have used the fact that the cumulative hazard function of T is Λ_0 (see Prentice et al. 1978). The appropriate Lindeberg condition is easily checked. (A.1) follows by Rebolledo's (1980) martingale convergence theorem.

We now turn to the proof that our tests are consistent against their respective alternatives. In general,

$$\sqrt{n}\,\phi_n(t) = \xi_n(t) + \sqrt{n}\bar{\phi}_n(t) + o_P(1) \tag{A.2}$$

uniformly in t, where $\xi_n \xrightarrow{\mathcal{D}} W(F_T(\cdot))$ as before and

$$\bar{\phi}_n(t) = \int_0^t \hat{S}_T(u-)\hat{S}_C(u-)^{1/2}\, d(\Lambda_2 - \Lambda_1)(u).$$

Now $\bar{\phi}_n$ converges in probability uniformly over bounded intervals to ϕ. Under H_1, $\phi(t) > 0$ for some t, by the lemma that follows this proof, so $\sqrt{n}D_{3n} \xrightarrow{P} \infty$ from (A.2). Under H_2, $\phi(t) - \phi(s) > 0$ for some $s < t$, so $\sqrt{n}D_{4n} \xrightarrow{P} \infty$ from (A.2).

Lemma 1. Under the alternative H_1, $\phi(t) \geq 0$ for all t, with strict inequality for some t.

Proof. The nonnegativity of ϕ follows from a result of Barlow and Proschan (1975, lemm. 7.1(b), p. 120), because $S_C^{1/2}$ is nonincreasing and $F_2 - F_1$ is nonnegative under H_1. For the strict inequality, if the nonnegative ϕ were to vanish *everywhere*, then so would $F_2 - F_1$ (we assumed that S_C never vanishes), but this possibility is excluded under H_1.

We conclude by indicating how to extend the proof of Theorem 3.1 to deal with more than two competing risks. In this setting the predictable variation process of $\sqrt{n}\,\phi_n$ converges in probability to $F_1 + F_2$. Because p_n is consistent for $P[\delta = 1 \text{ or } 2]$, it follows that

$$\sqrt{\frac{n}{p_n}}\,\phi_n \xrightarrow{\mathcal{D}} W(F_{12}(\cdot)),$$

where F_{12} is the conditional distribution function of $\min(X, Y)$ given that $\delta = 1$ or 2. This extends (A.1). The remaining steps of the proof are identical.

[Received July 1991. Revised May 1993.]

REFERENCES

Aalen, O. O. (1978), "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6, 701–726.

Aalen, O. O., and Johansen, S. (1978), "An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations," *Scandinavian Journal of Statistics*, 5, 141–150.

Aras, G., and Deshpandé, J. V. (1992), "Statistical Analysis of Dependent Competing Risks," *Statistics and Decisions*, 10, 323–336.

Bagai, I., Deshpandé, J. V., and Kochar, S. C. (1989a), "A Distribution-Free Test for the Equality of Failure Rates Due to Two Competing Risks," *Communications in Statistics, Part A—Theory and Methods*, 18, 107–120.

——— (1989b), "Distribution-Free Tests for Stochastic Ordering Among Two Independent Risks," *Biometrika*, 76, 775–778.

Barlow, R. E., and Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing Probability Models*, New York: Holt, Rinehart & Winston.

Benichou, J., and Gail, M. H. (1990), "Estimates of Absolute Cause-Specific Risk in Cohort Studies," *Biometrics*, 46, 813–826.

Block, H. W., and Basu, A. P. (1974), "A Continuous Bivariate Exponential Extension," *Journal of the American Statistical Association*, 69, 1031–1037.

Csörgő, M., and Révész, P. (1981), *Strong Approximations in Probability and Statistics*, New York: Academic Press.

Friday, D. S., and Patil, G. P. (1977), "A Bivariate Exponential Model With Applications to Reliability and Computer Generation of Random Variables," in *The Theory and Applications of Reliability*, Vol. 1, eds. C. P. Tsokos and I. N. Shimi, New York: Academic Press, pp. 527–549.

Froda, S. (1987), "A Signed Rank Test for Censored Paired Lifetimes," *Communications in Statistics, Part A—Theory and Methods*, 16, 3497–3517.

Gray, R. J. (1988), "A Class of *k*-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk," *The Annals of Statistics*, 16, 1141–1154.

Hoel, D. G. (1972), "A Representation of Mortality Data by Competing Risks," *Biometrics*, 28, 475–488.

Kochar, S. C., and Proschan, F. (1991), "Independence of Time and Cause of Failure in the Dependent Competing Risks Model," *Statistica Sinica*, 1, 295–299.

Lévy, P. (1948), *Processus Stochastiques et Mouvement Brownien*, Paris: Gauthier-Villars.

Marsaglia, G., Zaman, A., and Tsang, W. W. (1990), "Toward a Universal Random Number Generator," *Statistics and Probability Letters*, 8, 35–39.

Neuhaus, G. (1991), "Some Linear and Nonlinear Rank Tests for Competing Risks Models," *Communications in Statistics, Part A—Theory and Methods*, 20, 667–701.

Page, E. S. (1955), "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42, 523–527.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978), "The Analysis of Failure Times in the Presence of Competing Risks," *Biometrics*, 34, 541–554.

Ramlau-Hansen, H. (1983), "Smoothing Counting Process Intensities by Means of Kernel Functions," *The Annals of Statistics*, 11, 453–466.

Rebolledo, R. (1980), "Central Limit Theorems for Local Martingales," *Z. Wahrsch. verw. Gebiete*, 51, 269–286.

Rényi, A. (1970), *Foundations of Probability*, London: Holden-Day.

Yip, P., and Lam, K. F. (1992), "A Class of Non-Parametric Tests for the Equality of Failure Rates in a Competing Risks Model," *Communications in Statistics, Part A—Theory and Methods*, 21, 2541–2556.