SOME TESTS FOR COMPARING CAUSE-SPECIFIC HAZARD RATES

BY EMAD-ELDIN A.A. ALY, SUBHASH C. KOCHAR AND IAN W. MCKEAGUE

University of Alberta, Indian Statistical Institute and Florida State University

July, 1992

FSU Technical Report No. M-870
USARO Technical Report No. D-127
AFOSR Technical Report No. 91-272
SOME TESTS FOR COMPARING CAUSE-SPECIFIC HAZARD RATES

Emad-Eldin A.A. Aly\textsuperscript{1}

\textit{The University of Alberta, Edmonton, Canada}

Subhash C. Kochar\textsuperscript{2}

\textit{Indian Statistical Institute, New Delhi, India}

Ian W. McKeague\textsuperscript{3}

\textit{Florida State University, Tallahassee, Florida}

\textbf{Abstract.} 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 deciding the most appropriate course of treatment for a patient) it is important to know whether the forces of two given risks are equal or whether one is "more serious" than the other. We propose some distribution-free tests for comparing their cause-specific hazard rates and cumulative incidence functions 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.


\textit{Key words and phrases.} Competing risks, ordered alternatives, cumulative incidence function, distribution-free tests, right-censored data, counting processes.

\textsuperscript{1} Supported by an NSERC Canada grant at the University of Alberta.
\textsuperscript{2} Part of this research was done while visiting the University of Alberta supported by the NSERC Canada grant of the first author.
\textsuperscript{3} Partially supported by US Army Research Office Grant DAAL03-90-G-0103 and US Air Force Office of Scientific Research Grant AFOSR91-0048.
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 the notional (or latent) lifetimes of the unit under these two risks be denoted by $X$ and $Y$, respectively. In general, $X$ and $Y$ are dependent. Also, being lifetimes, they are nonnegative. We only observe $(T, \delta)$ 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 $A$.

On the basis of the competing risks data it is often useful to distinguish between the following alternatives: (i) the forces of the two risks are equal, and (ii) the force of one risk is greater than that of the other, within the environment in which the two risks are acting simultaneously. Such comparisons can be made in terms of cumulative incidence functions and cause-specific hazard rates, defined as follows. The cumulative incidence function corresponding to cause $j$ is

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

and the cause-specific hazard rate (CSHR) for cause $j$ is

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

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 $T$. In the case when $X$ 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) show that in general only probabilities expressible as functions of $g_1$ and $g_2$ may be estimated from the observable data $(T, \delta)$.

We mention two practical examples in which it is important to be able to make comparisons between cause-specific hazard rates (as well as cumulative incidence functions). The first example arises in reliability testing. Suppose that either of two components in a series system can be replaced to improve overall system reliability. One would replace the second component in preference to the first if $g_1 < g_2$ or $F_1 < F_2$. The second example, this one biomedical, comes from a paper of Benichou and Gail (1990). It concerns time to cancer recurrence in patients
with surgically resected cancer. In deciding whether to give a toxic therapy in the hope of preventing cancer recurrence it is appropriate to compare the cumulative incidence function for cancer recurrence (Benichou and Gail call this the absolute risk of recurrence), \( F_1 \), and the cumulative incidence function for other risks (e.g. other causes of death), \( F_2 \). A physician would be reluctant to recommend a toxic cancer treatment in an elderly patient in whom \( F_1 < F_2 \) or \( g_1 < g_2 \). Benichou and Gail go on to discuss the importance of the concept of absolute risk in evaluating public health measures to prevent disease. Gray (1988) also draws attention to the usefulness of comparing cause-specific incidence for different types of failure.

In this article we propose some methods for comparing CSHR’s and cumulative incidence functions. Our tests are less subjective than graphical procedures based on inspections of estimates of the CSHR’s themselves. We are interested in testing the hypotheses

\[
H_1 : g_1(t) \leq g_2(t), \ t \geq 0 \\
H_2 : F_1(t) \leq F_2(t), \ t \geq 0
\]

with strict inequalities for some \( t \). These hypotheses represent different ways of saying that risk \( Y \) is “more serious” than risk \( X \). Clearly \( H_1 \) implies \( H_2 \). We regard \( H_1 \) and \( H_2 \) as alternatives to the null hypothesis of equal risks:

\[
H_0 : g_1(t) = g_2(t), \ t \geq 0.
\]

Note that there is often no reason to expect \textit{a priori} that the cause specific risks \( g_1 \) and \( g_2 \) are equal (except, say, when \( g_1 \) and \( g_2 \) represent two identical components in a series system), but this is the natural choice of null hypothesis against which the ordered alternatives \( H_1 \) and \( H_2 \) should be tested. A similar choice of null hypothesis is made in the two-sample survival analysis problem when testing whether survival in one group is better than survival in another group, cf. Pepe and Fleming (1989).

In the case that \( X \) and \( Y \) are \textit{independent}, Bagai, Deshpandé and Kochar (1989 a,b) proposed distribution-free tests for testing the equality of two competing risks against stochastic ordering and failure rate ordering. Sen (1979) proposed nonparametric tests with maximum asymptotic relative efficiency for interchangeability of the competing risks against alternatives specified in terms of \( P[\delta = j | T = t] \). Aras and Deshpandé (1989) proposed locally most powerful rank tests for testing \( H_0 \) against various parametric alternatives expressed in terms of \( F_1 \) and \( F_2 \).
The remainder of the paper is organized as follows. In Section 2 we consider the problem of testing the null hypothesis of equality of CSHR's against the alternatives specified by $H_1$ and $H_2$. Our tests are of Kolmogorov–Smirnov type and are distribution-free. Formulae for the exact null distributions of the test statistics are provided. The asymptotic distributions are also derived. In Section 3 we consider an extension of our tests to deal with right-censored data. The results of a simulation study and an example are discussed in Section 4. In Section 5 we describe another extension of our tests to allow comparisons on any given time interval.

2. Testing for the equality of CSHR’s against ordered alternatives

In this section we introduce our tests for $H_1$ and $H_2$ in the uncensored case. The tests are based on the competing risk data $\{(T_i, \delta_i); i = 1, \ldots, n\}$ for $n$ independent and identical units.

2.1 Testing $H_0$ against $H_1$. A distribution-free test of $H_0$ vs. $H_1$ can be constructed using the fact that, under $H_1$, the function $\psi(t) = F_1(t) - F_2(t)$ is nonincreasing in $t$. This is a consequence of the identity $F_j(t) = \int_0^t g_j(u) S_T(u) du$ and provides a rationale for the test statistic

$$D_{1n}^+ = \sup_{0 \leq s < t < \infty} \{0, \psi_n(s) - \psi_n(t)\},$$

where $\psi_n(t) = F_{1n}(t) - F_{2n}(t)$. Here $F_{jn}(t) = n^{-1} \sum_{i=1}^n I\{\delta_i = j, T_i \leq t\}$ is the empirical estimator of $F_j$. Positive values of $D_{1n}^+$ provide evidence that $g_2$ is larger than $g_1$ in some interval. Note that

$$D_{1n}^+ = \max_{0 \leq i < j \leq n} \frac{1}{n} \left\{0, j - i - 2 \sum_{t=i+1}^j W_t \right\}$$

$$= \max_{0 \leq i < j \leq n} \frac{1}{n} \left\{0, \sum_{t=i+1}^j \eta_t \right\}$$

$$= \max_{0 \leq i < j \leq n} \frac{1}{n} \left\{0, Z_j - Z_i \right\},$$

(2.1)

where

$$W_i = \begin{cases} 1 & \text{if } \delta \text{ corresponding to } T_{(i)} \text{ (the } i\text{th ordered } T_i) \text{ is } 1 \\ 0 & \text{otherwise}, \end{cases}$$
\[ \eta_i = 1 - 2W_i, \ i = 1, 2, \ldots, n, \ Z_0 = 0 \text{ and } Z_k = \eta_1 + \ldots + \eta_k, \ \dot{k} = 1, 2, \ldots, n. \]

To obtain the exact null distribution of \( D_{1n}^+ \), we argue as follows. Kochar and Proschan (1991) proved that \( T \) and \( \delta \) are independent under \( H_0 \). Consequently, under \( H_0 \), \( W_1, \ldots, W_n \) are i.i.d. Bernoulli random variables with \( P(W_i = 0) = P(W_i = 1) = \frac{1}{2} \) and \( \eta_1, \ldots, \eta_n \) are i.i.d. with \( P(\eta_i = -1) = P(\eta_i = 1) = \frac{1}{2} \). It follows, using Simons (1983), that

\[
P\{nD_{1n}^+ < t\} = P\{-t < \min_{0 \leq j \leq n} Z_j < \max_{0 \leq j \leq n} Z_j < i + 1\}, \tag{2.2}\]

cf. Aly, Gombay and Kochar (1992, proof of Theorem 3.1 (i)). By applying Csáki (1986, (2.6)), this yields an exact formula for the null distribution function of \( D_{1n}^+ \). The asymptotic null distribution can be obtained from (2.1) and the discussion preceding (2.2) by using the invariance principle for partial sums (see, for example, Chapter 2 of Csörgő and Révész (1981)). We summarize this discussion in the following theorem.

**Theorem 2.1.** Under \( H_0 \)

\[
P\{nD_{1n}^+ < t\} = \frac{2}{2t + 1} \sum_{j=0}^{2t} \left\{ \cos \frac{j\pi}{2t+1} \right\} \sin \left\{ \frac{j\pi(t + 1)}{2t+1} \right\} \left\{ 1 + \sin \frac{j\pi}{2t+1} \right\} \tag{2.3}\]

\[
\times \left\{ \frac{1 - (-1)^j}{2} \right\} / \sin \frac{j\pi}{2t+1}
\]

for \( t = 1, \ldots, n + 1 \), and

\[
\sqrt{n}D_{1n}^+ \xrightarrow{D} \sup_{0 \leq x \leq 1} |W(x)|
\]

where \( \{W(t), t \geq 0\} \) is a standard Brownian motion. Consequently, for \( c > 0 \)

\[
P\{\sqrt{n}D_{1n}^+ \leq c\} \rightarrow \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\{ -\pi^2(2k+1)^2/8c^2 \}. \tag{2.4}\]

The exact formula (2.3) can be used to generate a table of critical values, see Aly, Gombay and Kochar (1992). Using (2.4) the asymptotic 0.90, 0.95 and 0.99 quantiles of \( \sqrt{n}D_{1n}^+ \) are found to be 1.96, 2.241 and 2.807, respectively.

**2.2 Testing \( H_0 \) against \( H_2 \).** Since under \( H_2 \), \( \psi(t) \) is nonpositive, a natural test statistic for testing \( H_0 \) against \( H_2 \) is given by

\[
D_{2n}^+ = \sup_{0 \leq t < \infty} \{0, -\psi_n(t)\}.
\]
We reject $H_0$ for large values of $D_{2n}^+$. Note that $D_{2n}^+$ can be written as $\max_{0 \leq j \leq n} Z_j$, where $Z_0, Z_1, \ldots, Z_n$ are as in (2.1). Thus, by Lemma 4.8.1 of Rényi (1970),

$$P\{n D_{2n}^+ = k\} = \frac{1}{2^n} \binom{n}{\lfloor \frac{n-k}{2} \rfloor}, \ k = 0, 1, 2, \ldots, n$$

under $H_0$. This gives the exact null distribution of $D_{2n}^+$. The asymptotic null distribution is obtained using the invariance principle for partial sums: under $H_0$

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

where $\Phi$ is the standard normal distribution function.

3. Censored data

In this section we consider an extension of our tests to allow for the possibility of right-censoring. The underlying censoring mechanism will be represented by a censoring time $C$ which is assumed independent of the latent failure times $X$ and $Y$. Denote the survival function of $C$ by $S_C$ and assume that $S_C(t) > 0$ for all $t$.

Under right-censoring we observe $n$ iid copies, $(\bar{T}_i, \bar{\delta}_i), \ i = 1, \ldots, n$, of $\bar{T} = \min(T, C)$ and $\bar{\delta} = \delta I(T \leq C)$. Our approach is to look for a suitable modification of the function $\psi$. Recall that $\psi(t) = \int_0^t S_T(u-) (g_1(u) - g_2(u)) \, du$. In order to obtain distribution-free tests of $H_1$ and $H_2$ in the censored case, look at the function

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

The factor $S_C(u-)^{1/2}$ turns out to be precisely what is needed to compensate for the censoring. We have that $\phi(t)$ nonincreasing under $H_1$ and identically zero under $H_0$, so to test $H_0$ against $H_1$ it is natural to use the test statistic

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

where $\phi_n$ is a suitable estimator of $\phi$. Similarly,

$$D_{4n}^+ = \sup_{0 \leq t < \infty} \{0, -\phi_n(t)\},$$
can be used to test $H_0$ against $H_2$. An obvious choice of $\phi_n$ 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 $\Lambda_j(t) = \int_0^t g_j(u) \, du$:

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

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

We note that $\hat{\Lambda}_j$ is a special case of an estimator discussed by Aalen and Johansen (1978) in connection with estimation of the transition probabilities of a non-homogeneous Markov chain with finitely many states. Indeed, we are dealing with a three-state non-homogeneous Markov chain having two absorbing states corresponding to the two types of failure.

The following result, proved in the Appendix, shows that $D_{3n}^+$ and $D_{4n}^+$ are asymptotically distribution-free and have the same limiting distributions they do in the uncensored case.

**Theorem 3.1.** Under $H_0$

$$\sqrt{n} D_{3n}^+ \overset{d}{\rightarrow} \sup_{0 \leq x \leq 1} |W(x)| \quad \text{and} \quad \sqrt{n} D_{4n}^+ \overset{d}{\rightarrow} \sup_{0 \leq x \leq 1} W(x).$$

Our approach easily 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 needs to be imposed on the dependency between the multiple risks, although the corresponding latent failure times need to be independent of the censoring, as before. Let $T$ be the minimum of a finite collection of latent failure times which include $X$ and $Y$, and let $\delta$ denote the corresponding cause of failure. Extensions of $D_{3n}^+$ and $D_{4n}^+$ that preserve the above 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 $P[\delta = 1 \text{ or } 2]$, see the Appendix.
4. Simulation results and an example

Our test procedures are consistent against their respective alternatives, \( H_1 \) or \( H_2 \). However, we would like to know whether they are powerful enough for practical applications. For that purpose, we carried out a simulation study, and the results show that our tests are readily able to detect these ordered alternatives.

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

\[
f(x, y) = \begin{cases} 
\frac{\lambda_1 \lambda (\lambda_2 + \lambda_0)}{\lambda_1 + \lambda_2} e^{-\lambda_1 x} - (\lambda_2 + \lambda_0) y & \text{if } x < y \\
\frac{\lambda_2 \lambda (\lambda_1 + \lambda_0)}{\lambda_1 + \lambda_2} e^{-\lambda_2 y} - (\lambda_1 + \lambda_0) x & \text{if } x < y 
\end{cases}
\]

where \((\lambda_0, \lambda_1, \lambda_2)\) are parameters and \( \lambda = \lambda_0 + \lambda_1 + \lambda_2 \). The CSHR’s are given by

\[
g_j(t) = \frac{\lambda_j \lambda}{\lambda_1 + \lambda_2},
\]

so \( H_1 \) holds if and only if \( \lambda_1 < \lambda_2 \). Under this model \( H_1 \) and \( H_2 \) are equivalent. The parameter \( \lambda_0 \) controls the degree of dependence between \( X \) and \( Y \); they are independent if and only if \( \lambda_0 = 0 \). We set \( \lambda_1 = 1 \) and considered various higher values of \( \lambda_2 \) corresponding to larger and larger 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, respectively). For the sake of comparison we included results for the uncensored case as well. We used asymptotic critical levels of 5%.

Inspection of Table 1 shows that use of the asymptotic critical levels gives somewhat conservative tests, and this effect increases as the censoring becomes more severe. The test based on \( D_{3n}^+ \) appears to be more conservative than the one based on \( D_{4n}^+ \). However, the tests become less conservative as the sample size increases (in fact we have found that the levels of the tests are close to their nominal 5% values for sample sizes over 500, even 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 \( \lambda_0 = 1 \).

As an application 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 which had received a radiation dose of 300 rads at 5–6 weeks of age and were kept
in a conventional laboratory environment. The cause of death was classified into thymic lymphoma, reticulum cell sarcoma, and other causes. For us, “other causes” represents censoring (39% were censored), and the two types of cancer mortality are taken to be the two causes of failure that we wish to compare, i.e. $g_1$ and $g_2$ are the CSHR's from lymphoma and sarcoma respectively. Our analysis depends on the assumption that the two diseases are lethal and independent of other causes of death, but we do not need to assume that they are independent of one another.

[Insert Figures 1 and 2 about here]

**Figure 1.** Aalen estimates of cumulative CSHR's for lymphoma (dashed line) and sarcoma (solid line).

**Figure 2.** Plot of $\sqrt{n}\phi_n(t)$ (solid line) and the asymptotic 5% critical levels for $D_{4n}^+$ (dashed lines).

We obtained $D_{3n}^+ = 4.81$, which gives a $P$-value of less than .01 for testing $H_0$ against $H_1$. Also, $D_{4n}^+ = 2.77$, which gives a $P$-value of .0056 for testing $H_0$ against $H_2$. When the roles of lymphoma and sarcoma were reversed, we obtained $D_{3n}^+ = D_{4n}^+ = 2.03$, so the $P$-values for the two tests are close to .1 and .05 respectively.

Our conclusion is that the two cause-specific hazard rates are unequal. Note that we cannot conclude that the CSHR for sarcoma is uniformly larger than the CSHR for lymphoma; the large value $D_{3n}^+ = 4.81$ only indicates that the sarcoma CSHR is larger than the lymphoma CSHR in some age interval. Indeed, inspection of a plot of the two cumulative CSHR estimates (Figure 1) suggests that up to 500 days there is moderate risk of lymphoma, yet negligible risk of sarcoma. After 500 days the situation reverses: there is negligible risk of lymphoma but high risk of sarcoma, and it is this large difference that the test statistic $D_{3n}^+$ is picking up. This is also reflected in the plot of $\sqrt{n}\phi_n(t)$ in Figure 2. Such plots are useful in avoiding misinterpretation of the test statistics. Plots of estimates of the CSHR's themselves are also useful; these can be made by finding smoothed derivatives of the cumulative CSHR estimates, see Ramlau-Hansen (1983), and are somewhat easier to interpret than plots of the cumulative CSHR estimates. However, our tests offer a less subjective comparison of CSHR's than can be made from a simple visual inspection of such plots.
5. Comparing CSHR’s in \([t_1, t_2]\)

It is often useful to compare CSHR’s (or cumulative incidence functions) in a
given time interval \([t_1, t_2]\), rather than at all times. For instance, an examination
of plots of the cumulative CSHR estimates for Hoel’s data strongly suggests that
the CSHR for sarcoma is much larger than the CSHR for lymphoma after 500 days.
In the second example discussed in the Introduction, Benichou and Gail (1990, p.
820) are interested in comparing the CSHR for cancer recurrence with the CSHR
for other risks at times between one and five years following surgical treatment.

It is straightforward to generalize our tests in Section 3 to deal with such cases.
We want a test of

\[ H_0^* : g_1(t) = g_2(t), \quad t_1 \leq t < t_2 \]

against the alternative

\[ H_1^* : g_1(t) \leq g_2(t), \quad t_1 \leq t < t_2 \]

with strict inequality for some \( t \in [t_1, t_2) \). We replace \( \phi \) by the function

\[ \phi^*(t) = (S_T(t_1) - S_T(t_2))^{-1/2} \int_{t_1}^{t} S_T(u-)S_C(u-)^{1/2}(g_1(u) - g_2(u)) \, du. \]

Clearly \( H_1^* \) is equivalent to \( \phi^* \) nonincreasing on \([t_1, t_2]\). As before, we suggest the
test statistic

\[ D_{5n}^+ = \sup_{t_1 \leq s < t \leq t_2} \{0, \phi^*_n(s) - \phi^*_n(t)\}, \]

where \( \phi^*_n(t) \) is obtained by substituting \( \hat{S}_T \) etc. into \( \phi^* \). It can be shown by routine
modifications of the proof of Theorem 3.1 that \( \sqrt{n}D_{5n}^+ \) converges in distribution to
\( \sup_{0 \leq x \leq 1} |W(x)| \) under \( H_0^* \).

When this test was applied to Hoel’s data, we obtained the highly significant
values of \( D_{5n}^+ = 5.56 \) (resp. 3.69) when testing whether the CSHR for sarcoma is
larger (resp. smaller) than the CSHR for lymphoma after (resp. before) 500 days.
This confirms our earlier conjectures arising from examination of Figures 1 and 2.

APPENDIX

Proof of Theorem 3.1. Suppose we can show that

\[ \sqrt{n}\phi_n \xrightarrow{D} W(F_T(\cdot)). \]  

(A.1)
Then, the second part of the theorem is clear. Using the continuous mapping theorem,

\[ \sqrt{n} \sup_{0 \leq s < t < \infty} \{ \phi_n(s) - \phi_n(t) \} \xrightarrow{D} \sup_{0 \leq s < t < \infty} \{ W(F_T(s)) - W(F_T(t)) \} \]

\[ \equiv \sup_{0 \leq u < v \leq 1} \{ W(u) - W(v) \} \]

\[ \equiv \sup_{0 \leq u \leq 1} \{ W(u) - \inf_{0 \leq v \leq u} W(v) \}. \]

The statement of the first part of the theorem now follows by

\[ W(u) - \inf_{0 \leq v \leq u} W(v) \xrightarrow{D} \sup_{0 \leq v \leq u} W(v) - W(u) \]

and the following well known result of Lévy (1948):

\[ \sup_{0 \leq v \leq u} W(v) - W(u) \equiv |W(u)|, \]

see Chung and Williams (1983). 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

\[ \hat{\Lambda}_j(t) = \int_0^t \frac{d\tilde{N}_j(u)}{\tilde{Y}(u)}, \]

where \( 1/0 \equiv 0 \),

\[ \tilde{Y} = \sum Y_i, \quad \tilde{N}_j = \sum N_{ij}, \]

\[ Y_i(u) = I(\tilde{T}_i \geq u), \quad N_{ij}(u) = I(\tilde{T}_i \leq u, \tilde{\delta}_i = j), \]

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 above processes. Let \( \bar{M}_j = \sum M_{ij} \). The predictable variation process of \( \bar{M}_j \) is \( \int_0^t \tilde{Y}(u) d\Lambda_j(u) \). By \( P(X = Y) = 0 \), the counting processes \( \tilde{N}_1 \) and \( \tilde{N}_2 \) 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 \( \bar{M}_1 - \bar{M}_2 \) is \( \int_0^t \tilde{Y}(u) d\Lambda_0(u) \), where \( \Lambda_0 = \Lambda_1 + \Lambda_2 \). Under \( H_0 \)

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

\[
\int_0^t \frac{\hat{S}_T(u-) \hat{S}_C(u-)}{\hat{Y}(u)/n} d\Lambda_0(u).
\]

By the Glivenko–Cantelli theorem, \( \hat{Y}(u)/n \) converges uniformly in \( u \) to \( P(\hat{T} \geq u) = S_T(u-)S_C(u-) \) almost surely. Hence, by the uniform consistency of the product-limit estimator on \( [0,t] \), the above 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 \( \Lambda_0 \); see Prentice et al. (1978). The appropriate Lindeberg condition is easily checked. (A.1) follows by Rebolledo’s (1980) martingale convergence theorem.

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

\[
\sqrt{\frac{n}{p_n}} \phi_n \overset{\mathcal{D}}{\rightarrow} W(F_{12}(\cdot)),
\]

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.

References


Table 1. Observed levels and powers of tests for equality of CSHR's based on $D_{3n}^\pm$ (resp. $D_{4n}^\pm$) at an asymptotic level of 5%. The underlying distribution of $(X, Y)$ is Block and Basu's (1974) ACBVE with $\lambda_1 = 1$.

(b) Uncensored

<table>
<thead>
<tr>
<th>$\lambda_2$</th>
<th>$n = 50$</th>
<th>n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_0 = 0$</td>
<td>$\lambda_0 = 1$</td>
</tr>
<tr>
<td>1.0</td>
<td>3.86 (4.90)</td>
<td>3.86 (4.90)</td>
</tr>
<tr>
<td>1.5</td>
<td>32.37 (39.46)</td>
<td>32.38 (39.46)</td>
</tr>
<tr>
<td>2.0</td>
<td>67.46 (74.95)</td>
<td>67.46 (74.95)</td>
</tr>
<tr>
<td>2.5</td>
<td>87.66 (91.96)</td>
<td>87.66 (91.96)</td>
</tr>
</tbody>
</table>

(b) Lightly censored (18%–33%)

<table>
<thead>
<tr>
<th>$\lambda_2$</th>
<th>$n = 50$</th>
<th>n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_0 = 0$</td>
<td>$\lambda_0 = 1$</td>
</tr>
<tr>
<td>1.0</td>
<td>2.89 (3.64)</td>
<td>2.98 (3.87)</td>
</tr>
<tr>
<td>1.5</td>
<td>21.95 (27.64)</td>
<td>24.19 (30.00)</td>
</tr>
<tr>
<td>2.0</td>
<td>51.95 (60.52)</td>
<td>55.40 (63.64)</td>
</tr>
<tr>
<td>2.5</td>
<td>76.35 (82.91)</td>
<td>78.49 (84.80)</td>
</tr>
</tbody>
</table>

(c) Heavily censored (40%–60%)

<table>
<thead>
<tr>
<th>$\lambda_2$</th>
<th>$n = 50$</th>
<th>n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_0 = 0$</td>
<td>$\lambda_0 = 1$</td>
</tr>
<tr>
<td>1.0</td>
<td>1.49 (2.29)</td>
<td>2.16 (2.82)</td>
</tr>
<tr>
<td>1.5</td>
<td>11.09 (16.02)</td>
<td>14.76 (19.79)</td>
</tr>
<tr>
<td>2.0</td>
<td>30.38 (39.76)</td>
<td>37.26 (46.75)</td>
</tr>
<tr>
<td>2.5</td>
<td>53.11 (63.73)</td>
<td>60.73 (70.27)</td>
</tr>
</tbody>
</table>

Note: The data were created using the uniform random number generator of Marsaglia, Zaman and Tsang (1990) and an algorithm of Friday and Patil (1977, Corollary 3.3). 10000 samples were used to obtain each entry in the table.
**Title:** SOME TESTS FOR COMPARING CAUSE-SPECIFIC HAZARD RATES

**Abstract:** 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 deciding the most appropriate course of treatment for a patient) it is important to know whether the forces of two given risks are equal or whether one is “more serious” than the other. We propose some distribution-free tests for comparing their cause-specific hazard rates and cumulative incidence functions 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.
**SOME TESTS FOR COMPARING CAUSE-SPECIFIC HAZARD RATES**

**AUTHOR(S)**

Emaâd-Eldin A.A. Aly, Subhash C. Kochar and Ian W. McKeague

**PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Department of Statistics  
Florida State University  
Tallahassee, FL 32306-3033

**SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Research Office  
P.O. Box 12211  
Research Triangle Park, NC 27709-2211

**SUPPLEMENTARY NOTES**

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

**ABSTRACT (Maximum 200 words)**

ABSTRACT. 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 deciding the most appropriate course of treatment for a patient) it is important to know whether the forces of two given risks are equal or whether one is "more serious" than the other. We propose some distribution-free tests for comparing their cause-specific hazard rates and cumulative incidence functions 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.