Comparing Sub-Survival Functions in a Competing Risks Model

K. C. CARRIERE kc.carriere@ualberta.ca
Department of Mathematical Sciences, University of Alberta, Edmonton, T6G 2G1, CANADA

SUBHASH C. KOCHAR
Indian Statistical Institute, 7, SJS Sansanwal Marg, New Delhi-110016, INDIA

Received November 17, 1998; Revised May 20, 1999; Accepted July 14, 1999

Abstract. In the competing risks literature, one usually compares whether two risks are equal or whether one is "more serious." In this paper, we propose tests for the equality of two competing risks against an ordered alternative specified by their sub-survival functions. These tests are naturally developed as extensions of those based on hazard rates and cumulative incidence functions. We note that the interpretation of the new test results is more direct compared to the situation when the hypotheses are framed in terms of their cumulative incidence functions. The proposed tests are of the Kolmogrov–Smimov type, based on maximum differences between sub-survival functions. Our simulation studies indicate that they are excellent competitors of the existing tests, that are based mainly on differences between cumulative incidence functions. A numerical example will demonstrate the advantages of the proposed tests.

Keywords: Competing risks, cumulative incidence function, cause specific hazard rate, ordered alternatives, Nelson–Aalen estimator of cumulative hazard

1. Introduction

The competing risks problem involves subjects or experimental units exposed to multiple risks, but where the actual failure (or death) is attributable to only one cause. In this paper, we examine the case of exactly two risks. Let the notional lifetimes of a unit under these two risks be denoted by random variables X and Y, respectively, that are nonnegative. In general, X and Y need not be independent. This structure yields observables $(T,\ \delta)$ where $T = \min(X, Y)$ is the time of failure and $\delta = 2 - I(X \leq Y)$ is the cause of failure. The $I(A)$ denotes the indicator function of the event A. We assume that $P(X = Y) = 0$ and that lifetimes are continuous type random variables. Thus, for each subject we observe (T, δ).

The *cause specific hazard rate* corresponding to the i^{th} cause is defined by

$$h_i(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P[t \leq T < t + \Delta t,\ \delta = i \mid T \geq t]$$

$i = 1, 2$. The overall hazard rate is given by $h_T(t) = h_1(t) + h_2(t)$. When the causes of failure are independent, $h_i(t)$ is the ordinary hazard rate corresponding to the marginal distribution of failure times from the i^{th} cause. Prentice et al. (1978) emphasize that only those quantities which can be expressed in terms of cause specific hazard rates are estimable and can be estimated from the competing risks data even if the risks are dependent. In this paper, all the quantities we are interested in are functionals of cause specific hazard rates and hence identifiability will not be a problem.

For the competing risks problem, it is often of interest to distinguish between the following alternatives: (i) the two risks are equal, and (ii) one risk is greater than the other. In the literature, such comparisons have been made in terms of cause specific hazard rates and *cumulative incidence* (sub-distribution) functions, $F_i(t) = P[T \leq t, \delta = i]$, $i = 1, 2$. In this paper, we advocate the use of *sub-survival functions*, $\tilde{F}_i(t) = P[T > t, \delta = i]$, $i = 1, 2$ for making such comparisons. Using sub-survival functions allows for a more direct interpretations of hypotheses than using cumulative incidence functions does.

Let S_T denote the survival function of T. Then the cumulative incidence functions and the sub-survival functions can be expressed in terms of the cause specific hazard rates by the relations,

$$F_i(t) = \int_0^t h_i(u) S_T(u)\, du, \quad \tilde{F}_i(t) = \int_t^\infty h_i(u) S_T(u)\, du, \tag{1}$$

for $i = 1, 2$.

Based on a random sample (T_j, δ_j), $j = 1, \ldots, n$ on (T, δ), we consider the problem of testing the null hypothesis given by,

$$H_0\colon \tilde{F}_1(t) = \tilde{F}_2(t), \quad t \geq 0, \tag{2}$$

against the alternative,

$$H_1\colon \tilde{F}_1(t) \leq \tilde{F}_2(t), \quad t \geq 0, \tag{3}$$

with strict inequality for some t.

The alternative H_1 can also be expressed as

$$P[\delta = 1 \mid T > t] \leq P[\delta = 2 \mid T > t], \quad t \geq 0.$$

In this form it has the following interpretation: Given that a unit has survived up to time t, the conditional probability of its failing in the future from cause 2 is *uniformly* greater than that from cause 1. Thus H_1 indicates risk Y being "more serious" than risk X in some sense. Also we note that even though the sub-survival functions $\tilde{F}_1$ and $\tilde{F}_2$ may not be expected *a priori* to be equal, except under some special situations (Aly, Kochar and McKeague, 1994), it is the natural choice of null hypothesis for the ordered alternative H_1. Similarly, the null hypothesis is that of equality in the two-sample survival analysis problem for testing the alternative whether survival in one group is better than survival in another group.

Also note that H_0 is equivalent to H_0': $h_1(t) = h_2(t)$ for all t as well as to H_0'': $F_1(t) = F_2(t)$ for all t. H_1 is implied by the more stringent alternative

$$H_A\colon h_1(t) \leq h_2(t), \quad t \geq 0,$$

with strict inequality for some t. However, it is possible that the cause specific hazard rates cross each other, but their survival functions are ordered. An example of this is when X and Y are independent with X having exponential distribution with hazard rate 0.5 and Y having Weibull distribution with shape parameter 2 and scale parameter one.

Aly, Kochar and McKeague (1994) and Sun and Tiwari (1995) considered the problem of testing the null hypothesis H_0 against the alternative

$$H_2\colon\ F_1(t) \le F_2(t),\ t \ge 0,$$

and with strict inequality for some t.

While H_2 is also implied by H_A, H_1 and H_2 are not necessarily equivalent. The alternative H_2 does not seem to have the same natural interpretation that the alternative H_1 has. It follows from the discussion below that the alternative H_1 is also somewhat more stringent than H_2.

Consider the case when X and Y are *independent* with distribution functions F and G, respectively. Let $\overline{F}$ and $\overline{G}$ be the corresponding survival functions. In this case the cause specific hazard rates are identical to the hazard rates corresponding to the marginal distributions of the random variables X and Y; and the cumulative incidence functions and the sub-survival functions can be expressed as

$$\begin{aligned} F_1(x) &= \int_0^x \overline{G}(u)\, dF(u), \quad F_2(x) = \int_0^x \overline{F}(u)\, dG(u); \\ \tilde{F}_1(x) &= \int_x^\infty \overline{G}(u)\, dF(u), \quad \tilde{F}_2(x) = \int_x^\infty \overline{F}(u)\, dG(u). \end{aligned}$$

Now suppose that X is stochastically greater than Y ($X \overset{st}{\ge} Y$). That is, $\overline{F}(x) \ge \overline{G}(x)$ for all $x \ge 0$. This implies

$$F_1(x) = \int_0^x \overline{G}(u)\, dF(u) \le \int_0^x \overline{F}(u)\, dF(u) \le \int_0^x \overline{F}(u)\, dG(u) = F_2(x),$$

for $x \ge 0$. The last inequality follows because the function $\overline{F}(u)\, I(0 \le u \le x)$ is nonincreasing in u and $X \overset{st}{\ge} Y$. Thus in this case $X \overset{st}{\ge} Y$ implies H_2. But, in general, the alternative H_1 may not be implied by this constraint as the function $\overline{F}(u)\, I(x \le u \le \infty)$ is not monotone in u. However, as discussed above, a sufficient condition for H_1 to hold is that the hazard rate of X is smaller than that of Y.

Clearly there are situations where analysis based on sub-survival functions is more meaningful and revealing than analysis based on cumulative incidence functions. The two approaches, however, address different aspects of the competing risks problem. It is plausible that in some cases the cumulative incidence functions cross but their sub-survival functions are ordered (as in the case of the data set considered later in this paper) and vice versa.

Several tests are available in the literature for testing the equality of competing risks and the relevant references can be found in the review paper by Kochar (1995). Gray (1988) proposed a class of c-sample tests for comparing the cumulative incidence function of one risk over c different populations. The corresponding two-sample problem has been studied by Lin (1997). But to the best of our knowledge, we are not aware of any test designed specifically for testing H_0 against H_1 in terms of sub-survival functions. In this paper we propose new Kolmogrov-Smirnov type tests for this problem based on maximum differences between the two estimated sub-survival functions. These tests are similar in

spirit to the tests of Aly, Kochar and McKeague (1994) which were based on the maximum differences between the empirical cumulative incidence functions.

In the next section we introduce the test statistic for the case of uncensored data and give its exact and asymptotic distributions. In Section 3, we extend our test to the case when the data are randomly censored on the right. Section 4 is devoted to power comparisons. In Section 5, we analyze a data set using the procedures developed in this paper. Concluding remarks are given in the last section.

2. Tests for Uncensored Data

We test the null hypothesis H_0 against the alternative H_1 on the *uncensored* competing risks data $\{(T_j, \delta_j),\ j = 1, \ldots, n\}$ for n independent and identical units.

Let

$$\gamma(t) = \tilde{F}_2(t) - \tilde{F}_1(t). \tag{4}$$

Note that $\gamma(t) \equiv 0$ under the null hypothesis, but under H_1, $\gamma(t) \geq 0$, for $t \geq 0$ and with a strict inequality for some t. We base our test on the statistic

$$D^{\star}_{1n} = \sup_{0 \leq t < \infty} \gamma_n(t),$$

where $\gamma_n(t) = \tilde{F}_{2n}(t) - \tilde{F}_{1n}(t)$ and $\tilde{F}_{in}(t) = n^{-1} \sum_{j=1}^{N} i\{\delta_j = i, T_j > t\}$ is the empirical sub-survival function for cause i, $i = 1, 2$. Positive large values of $D^{\star}_{1n}$ provide evidence in favor of H_1. The statistic $D^{\star}_{1n}$ is similar to the statistic

$$D_{1n} = \sup_{0 \leq t < \infty} [F_{2n}(t) - F_{1n}(t)]$$

proposed by Aly, Kochar and McKeague (1994) for testing against the alternative H_2. Here $F_{in}(t)$ is the empirical sub-distribution function of the i^{th} cause.

Let $T_{(1)} \leq \ldots \leq T_{(n)}$ be the ordered failure times and let $\delta_{[j]}$, $j = 1, \ldots, n$ be the corresponding causes of failures. Then the statistic $D^{\star}_{1n}$ can be expressed as

$$D^{\star}_{1n} = \max_{0 \leq j \leq n} \frac{1}{n} \left\{ (n - j) - 2 \sum_{k=j+1}^{N} V_k \right\} = \max_{0 \leq j \leq n} Z_j / n$$

where

$$V_k = \begin{cases} 1 & \text{if } \delta \text{ corresponding to } T_{(k)} \text{ is } 1, \\ 0 & \text{otherwise;} \end{cases}$$

$Z_j = \eta_{j+1} + \cdots + \eta_n$, $Z_n = 0$, and $\eta_k = 1 - 2V_k$, for $k = 1, \ldots, n$. As under H_0, T and δ are independent (cf. Kochar and Proschan, 1991) and Z_j is a symmetric random walk. It follows that the null distribution of $D^{\star}_{1n}$ is the same as D_{1n} and is given by

$$P[nD^{\star}_{1n} = k] = \frac{1}{2^n} \binom{n}{[\frac{n-k}{2}]},\ k = 0, 1, \ldots, n.$$

Its asymptotic null distribution is

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

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

An alternative approach would be to base a test on an estimator of the following average value of $\gamma(.)$,

$$\gamma_A = \int_0^{\infty} \gamma(t)\, dF_T(t) = P[\delta_1 = 2, T_1 > T_2] - \frac{1}{2},$$

where (T_j, δ_j), $j = 1, 2$ are two independent copies of (T, δ) and F_T is the distribution function of T. A U-statistic estimator of this parameter γ_A leads to the test statistic

$$U_1 = \sum_{j=1}^{n} (R_j - 1) I\{\delta_j = 2\},$$

where R_j is the rank of T_j among $T_1, \ldots, T_n$. This statistic U_1 was initially proposed and studied by Bagai, Deshpandé and Kochar (1989a) for comparing the hazard rates of two competing risks. They also proposed another statistic $U_2 = \sum_{j=1}^{n} (n - R_j + 1) I(\delta_j = 2)$ for testing against stochastic ordering between two independent competing risks (1989b). It can be shown that U_2 is equivalent to the U-statistic estimator of

$$\int_0^{\infty} [F_2(x) - F_1(x)] dF_T(x) = P[\delta_1 = 2, T_1 \leq T_2] - \frac{1}{2},$$

an average difference between the sub-distribution functions.

Yip and Lam (1992, 1993) proposed a class of asymptotically distribution-free tests for testing the equality of the hazard rates of *independent* competing risks. Lam (1998) subsequently proved that the asymptotic null distributions of these statistics remain unchanged even when the risks are dependent. They used the counting processes approach to study the asymptotic properties of their tests. Their class includes the asymptotically equivalent versions of the U_1 and U_2 tests as special cases.

When an ordered alternative is unsuitable, it can be of interest to test H_0 against the general alternative: $\bar{F}_1(t) \neq \bar{F}_2(t)$ for some t. In that case it is natural to use the Kolmogrov-Smirnov test statistic

$$\bar{D}_n = \sup_{t \geq 0} |\gamma_n(t)|.$$

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.

3. Tests for Censored Data

We now modify the new test for the censored data. An item is censored, if its actual observation is unavailable 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 i.i.d. copies, $(\tilde{T}_j, \tilde{\delta}_j)$, $j = 1, \ldots, n$, of $\tilde{T} = \min(T, C)$ and $\tilde{\delta} = \delta\varepsilon$ where $\varepsilon = I(T \leq C)$.

The product limit (PL) estimator of S_T is given by

$$\hat{S}_T(t) = \prod_{i=1}^{n} \left(1 - \frac{\varepsilon_{[i]}}{n-i+1}\right)^{I\{\tilde{T}_{(i)} \leq t\}}$$

where $\tilde{T}_{(1)} \leq \cdots \leq \tilde{T}_{(n)}$ are the ordered $\tilde{T}_i$ values and $\varepsilon_{[i]}$ is the concomitant of the ith order statistic, that is, $\varepsilon_{[i]} = \varepsilon_j$ if $\tilde{T}_{(i)} = \tilde{T}_j$. If the largest observation is uncensored, the PL estimator at that point equals zero. If the largest observation is censored, the PL estimator can never equal zero and is undefined beyond the largest observation.

We see from $Eq.(1)$ that in the case of censored data, a natural estimator of $\tilde{F}_i$, is

$$\hat{\tilde{F}}_i(t) = \int_t^\infty \hat{S}_T(u-)\, d(\hat{\Lambda}_i(u)),$$

where $\hat{S}_T$ is the product-limit estimator of S_T and $\hat{\Lambda}_j$ is the Aalen estimator (1978) of the cumulative cause specific hazard rate function $\Lambda_i(t) = \int_0^t h_i(u)\, du$, given by

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

Here $R_j = \#\{k:\ \tilde{T}_k \geq \tilde{T}_j\}$ is the size of the risk set at time $\tilde{T}_j -$.

A suitable modification of the function $\gamma(t) = \tilde{F}_2(t) - \tilde{F}_1(t)$ for the censored case is

$$\begin{aligned} \gamma^*(t) &= \int_t^\infty S_C(u-)^{1/2} d(\tilde{F}_2 - \tilde{F}_1)(u) \\ &= \int_t^\infty S_T(u-) S_C(u-)^{1/2} d(\Lambda_2 - \Lambda_1)(u), \end{aligned}$$

which coincides with γ when there is no censoring. For a justification of this, see Aly, Kochar and McKeague (1994). An obvious choice of γ_n^*, an estimator of $\gamma^*(t)$ is

$$\gamma_n^*(t) = \int_t^\infty \hat{S}_T(u-) \hat{S}_C(u-)^{1/2} d(\hat{\Lambda}_2 - \hat{\Lambda}_1)(u),$$

where $\hat{S}_C$ is the PL estimator of S_C. Note that the quantity $S_T(u) S_C(u)^{1/2}$ vanishes at $\tilde{T}_{(n)}$, the largest observation. Positive large values of the test statistic

$$D_{3n}^\star = \sup_{0 \leq t < \infty} \gamma_n^*(t), \tag{5}$$

are significant for testing H_0 against H_1.

The $D_{3n}^\star$ test is asymptotically distribution-free with the same limiting null distribution as in the uncensored case. The proof follows on the lines of Aly, Kochar and McKeague (1994) and is omitted.

Table 1. Estimated sizes and powers of the tests for the LIFR distribution at asymptotic levels of 5%.

Test	θ				
	0.0	0.5	1.5	2.5	3.5
			Uncensored		
D_1	4.67	19.81	58.23	82.61	92.27
D_1^*	4.43	25.78	71.04	90.61	96.89
U_1	4.92	31.36	79.37	94.35	98.36
			Lightly censored (26–31%)		
D_3	3.85	11.74	36.82	60.15	76.46
D_3^*	4.13	15.67	47.05	71.94	85.80
U_1^*	4.37	17.94	51.22	75.66	88.80
			Heavily censored (54–59%)		
D_3	3.36	5.68	15.73	28.32	42.50
D_3^*	4.10	7.53	20.41	36.44	51.37
U_1^*	3.24	5.46	17.96	32.43	47.69

The statistic U_1 can also be easily modified to handle the case of randomly censored data as $U_{1n}^* = \int_0^\infty \gamma_n^*(t)\, d(1 - \hat{S}_T(t))$. The asymptotic null distribution of $\sqrt{n}U_{1n}^*$ is $N(0,\ 1/3)$ and its large values are significant for testing against H_1.

It may be noted that in case the censoring distribution has support on a finite interval $[0, \tau]$, then the null and alternatives are really regarding the quantity $\int_t^\tau h_i(u)S_T(u)du = \tilde{F}_i(t) - \tilde{F}_i(\tau)$ rather than the sub-survival functions. While this would not affect the validity of the tests, caution should be exercised in the interpretation of the results.

4. Efficiency and Power Comparisons

The alternative H_A implies both H_1 and H_2. However, we are not aware of any tests designed specifically for comparing the sub-survival functions of competing risks in H_1, while H_2 has many tests available. For some alternatives belonging to H_A (and hence to H_1 and H_2), we performed a simulation study to compare the powers of the D_1^*, D_3^* and U_1^* tests with the D_1, D_2 and U_1 tests.

In the first study we consider the case when X and Y are independent with X having standard exponential distribution and Y having linearly increasing failure rate (LIFR) distribution with hazard rate $h_\theta(x) = (1 + \theta x)$. The case $\theta = 0$ corresponds to the null hypothesis H_0 and values of $\theta > 0$ correspond to the alternative H_A (and hence H_1 and H_2).

The censoring was taken to be exponential with parameters 1 and 3, corresponding to "light" and "heavy" censoring (about 28% and 56% censored). We used asymptotic critical values at 5% level. Table 1 gives the estimated powers of the various tests based on 10,000 samples, each of size 100 for this alternative. Although not reported in the Table, the estimated powers of the U_2 and the sign tests in the *uncensored* case and at the above θ values were found to be rather non-competitive. These values for the two tests are 4.93, 11.92, 29.72, 55.85, 62.60 and 4.39, 22.27, 63.52, 86.33, 94.18, respectively.

At $n = 100$, the empirical sizes of all tests under consideration were a little too conservative compared to the nominal 5% with an exception of D_1 and U_1 for the uncensored case. The U_1 ($U_1^\star$) test that seemed to excel in the uncensored (light censored) case did not do as well in the heavily censored case. The general finding in this case is that U_1 ($U_1^\star$) is the best in uncensored (lightly censored) case, while $D_3^\star$ improves the empirical size and power considerably in the heavily censored case. But the results using the new tests $D_1^\star$ and $D_3^\star$ were generally improved from the previous tests D_1 and D_3 in all cases. Power improvement was rather substantial in the uncensored and the lightly censored cases, especially.

These simulations results are not surprising in the light of the findings of Yip and Lam (1992, 1993). They observed that in this case, the differences in the upper tails of the distributions are more relevant. The newly proposed tests put more weight on late failures.

Next we consider the case when (X, Y) follows the absolutely continuous bivariate exponential (ACBVE) distribution of Block and Basu (1974) with density

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

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

In this case the cause specific hazard rates $h_j(t) = \lambda_j\lambda/(\lambda_1 + \lambda_2)$, $j = 1, 2$ are proportional, and the alternative hypotheses H_A, 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. Again the censoring was taken to be exponential with parameters 1 and 3, corresponding to "light" and "heavy" censoring (about 22% and 45% censored). We used asymptotic critical values at 5%. Again the simulation results reported in Table 2 are based on 10,000 samples each of size 100. Only the results for $\lambda_0 = 1$ are reported as there is only marginal effect of this parameter on the power functions of the tests. Note that the case $\lambda_2 = 1.0$ corresponds to the null hypothesis. We used the exact null mean and variance in the asymptotic normal approximation of U_1. Improvements using the new test in empirical sizes and powers are not evident. With an exception of the U_1 test, which was notably poor, the other tests performed quite similarly.

It is evident from these studies that for the LIFR alternative, the $D_1^\star$ ($D_3^\star$) test performs better than the D_1 (D_3) test, while for the ACBVE distribution, they perform almost equally well. This shows that the $D_1^\star$ ($D_3^\star$) and U_1 ($U_1^\star$) tests are good competitors of the D_1 (D_3) and U_2 tests. Whereas the tests based on cumulative distribution functions give more importance to early failures, the newly proposed tests are more suitable when the differences in the upper tails of the distributions are more prominent.

5. Hoel's Data Revisited

We revisit the mortality data set given in Hoel (1972) and also analyzed earlier by Aly, Kochar and McKeague (1994). A radiation dose of 300 rads were given to 99 RMF strain male mice at 5–6 weeks of age and they were kept in a conventional laboratory environment.

Table 2. Estimated sizes and powers of tests for the ACBVE distribution with $\lambda_0 = 1$ at asymptotic levels of 5%.

Test	λ_2			
	1.0	1.5	2.0	2.5
	Uncensored			
D_1	4.85	59.59	95.08	99.66
D_1^*	4.42	58.07	94.36	99.56
U_1	4.84	52.65	90.47	98.51
	Lightly censored (18–25%)			
D_3	4.02	50.16	89.66	98.84
D_3^*	4.19	47.94	88.23	98.51
U_1^*	4.36	39.98	78.04	94.60
	Heavily censored (40–50%)			
D_3	3.07	35.35	76.34	94.51
D_3^*	3.43	32.96	72.07	92.20
U_1^*	3.05	22.38	52.69	75.96

The cause of death was recorded as one of thymic lymphoma, reticulum cell sarcoma, and other causes. Similarly in Aly, Kochar and McKeague (1994), we took "other causes" as censoring (39% were in this category), and used the two types of cancer mortality as the two causes of failure for comparison purposes. Thus $\tilde{F}_1$ and $\tilde{F}_2$ denote the sub-survival functions of the risks lymphoma and sarcoma, respectively. (Note that Aly, Kochar and McKeague (1994) label the two risks the other way around). We assume that the two diseases are lethal and independent of other causes of death . We do not need to assume that the two diseases are independent of each other. In Figure 1 and Figure 2, we plot the cumulative incidence functions and the sub-survival functions of the two risks. It is clear from these figures that for this data set the two cumulative incidence functions cross at about 500 days, but their sub-survival functions are ordered.

Figure 3 plots the function $\sqrt{n}\gamma_n^*$. Observe that $\sqrt{n}\gamma_n^*(t)$ first increases and then decreases, but it remains *nonnegative*. This plot indicates that the two cause specific hazard rates are not ordered; otherwise, it should have been monotone. The graph of $\sqrt{n}\gamma_n^*(t)$ appears to favor the alternative H_1 that the two sub-survival functions are ordered. For testing H_0 against H_1 we obtained a highly significant value of $\sqrt{n}D_{3n}^* = 4.8058$ with a p-value of $.1541 \times 10^{-5}$ (the value of U_1^* for this data is 2.6252 with a corresponding p-value of .0086). Both D_{3n}^* and U_1 tests reject H_0 in favor of the alternative H_1 that the sub-survival function of the risk of death from sarcoma is greater than that from lymphoma.

Aly, Kochar and McKeague (1994) reported that $\sqrt{n}D_{3n} = 3.69$ with $p < .0003$. It was concluded that the cumulative incidence for lymphoma was larger than that for sarcoma before 500 days; this was reversed after 500 days. On the other hand, our test indicates that the sub-survival function for sarcoma is always larger than that for lymphoma, supporting H_1.

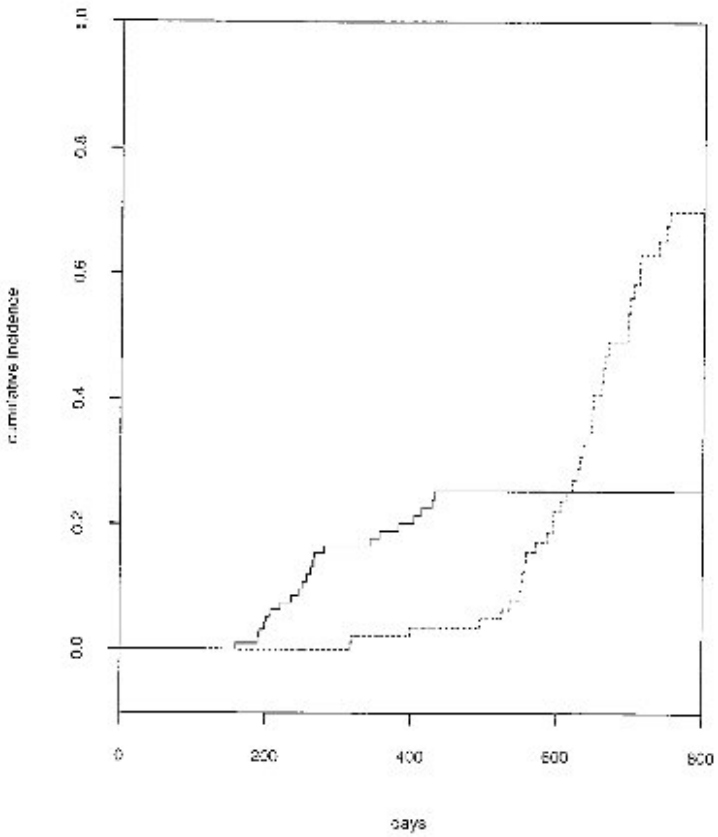

Figure 1. Cumulative incidence of lymphoma (___) and sarcoma (- - -).

6. Concluding Remarks

In this paper we have considered the competing risks problem and have shown that it is easier to interpret hypotheses expressed in terms of sub-survival functions rather than in terms of cumulative incidence functions. We have proposed new Kolmogrov-Smirnov type tests for the problem of testing the equality of two competing risks against the alternative that their sub-survival functions are ordered. These tests are similar to the ones proposed by Aly, Kochar and McKeague (1994) for comparing the cumulative incidence functions of two competing risks. The tests proposed in this paper give more weight to late failures which typically occur in the upper tails of the distributions and where the differences in the distributions are more important. We conclude that the tests proposed in this paper are good competitors of the existing ones. The new tests are expected to perform better in situations where the differences in the sub-survival functions are more prominent. The two approaches to the competing risks problem based on the differences of the cumulative

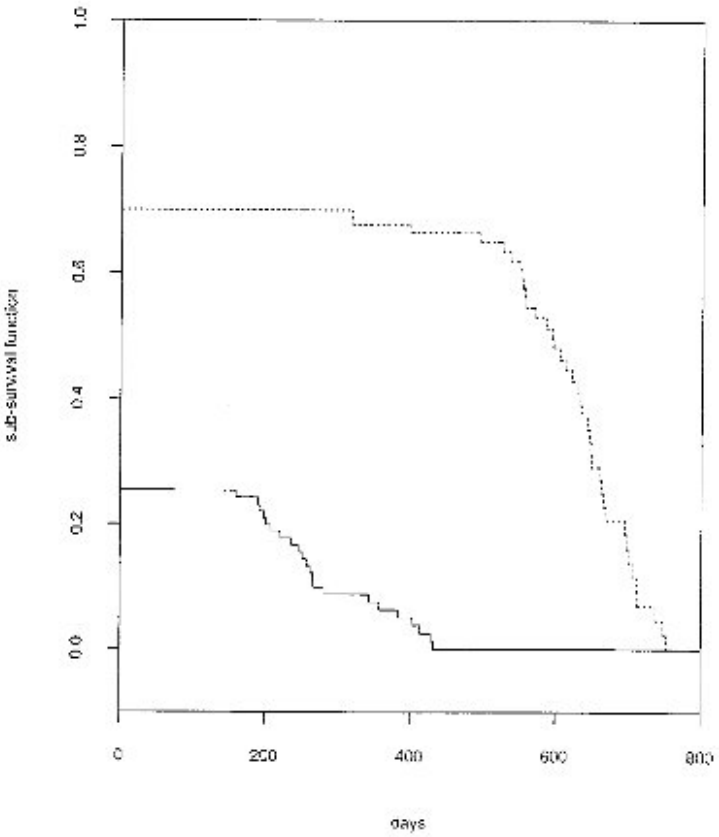

Figure 2. Sub-survival functions of lymphoma (____) and sarcoma (- - -).

incidence functions and the differences of the sub-survival functions are complementary to each other, addressing to different aspects of the problem.

Acknowledgments

The authors are grateful to the two referees for their constructive comments and helpful suggestions which led to improved presentation of the paper.

K. C. Carriere is supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada and a Scholar Award (6609-2120-48) with the National Health Research Development Program of Canada. Additional funding was provided by a Senior Scholar Award and Establishment Grant from the Alberta Heritage Foundation for Medical Research.

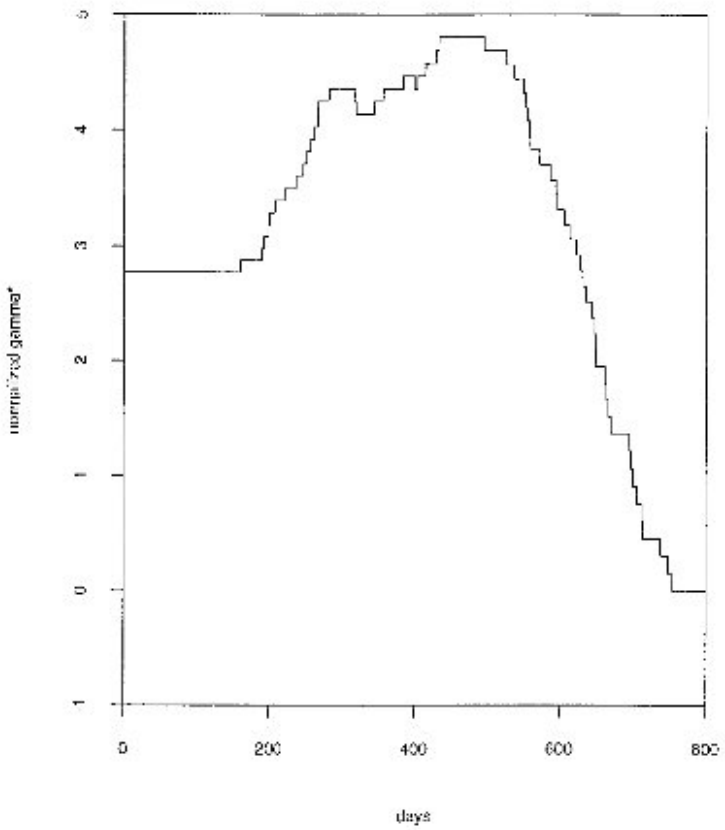

Figure 3. Plot of normalized gamma*.

This work partly done while Subhash C. Kochar was visiting the University of Alberta, Canada.

References

E. A. A. Aly, S. C. Kochar, and I. W. McKeague, "Some tests for comparing cumulative incidence functions and cause- specific hazard rates," *Journal of the American Statistical Association* vol. 89 pp. 994–999, 1994.

I. Bagai, J. V. Deshpandé, and S. C. Kochar, "A distribution-free test for the equality of failure rates due to two competing risks," *Communications in Statistics Part A—Theory Meth.* vol. 18 pp. 107–120, 1989a.

I. Bagai, J. V. Deshpandé, and S. C. Kochar, "Distribution-free tests for stochastic ordering among two independent risks," *Biometrika* vol. 76 pp. 775–778, 1989b.

R. J. Gray, "A class of *k*-sample tests for comparing the cumulative incidence of a competing risk," *Annals of Statistics* vol. 16 pp. 1141–1154, 1988.

S. C. Kochar, "A review of some distribution-free tests for the equality of cause specific hazard rates," in J. V. Desphandé and H. L. Koul (eds.), *Analysis of Censored Data*, IMS-LNMS vol. 27 pp. 147–162, 1995.

S. C. Kochar and F. Proschan, "Independence of time a nd cause of failure in the dependent competing risks model," *Statistica Sinica* vol. 1 pp. 295–299, 1991.

K. F. Lam, "A class of tests for the equality of k cause-specific hazard rates in a competing risks model," *Biometrika* vol. 85 pp. 179–188, 1998.

D. Y. Lin, "Non-parametric inference for cumulative incidence functions in competing risks studies," *Statistics in Medicine* vol. 16 pp. 901–910, 1997.

R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson, N. Flourney, V. T. Farewell, and N. E. Breslow, "The analysis of failure times in the presence of competing risks," *Biometrics* vol. 34 pp. 541–554, 1978.

Y. Sun and R. C. Tiwari, "Comparing cause-specific hazard rates of a competing risks model with censored data," in J. V. Desphandé and H. L. Koul (eds.), *Analysis of Censored Data*, IMS-LNMS vol. 27 pp. 255–270, 1995.

P. Yip and K. F. Lam, "A class of non-parametric tests for the equality of failure rates in a competing risks model," *Communications in Statistics Part A—Theory Meth.* vol. 21 pp. 2541–2556, 1992.

P. Yip and K. F. Lam, "A multivariate nonparametric test for the equality of failure rates in a competing risks model," *Communications in Statistics Part A—Theory Meth.* vol. 22 pp. 3199–3222, 1993.