Analysis of Censored Data
IMS Lecture Notes – Monograph Series (1995) Volume 27

A Review of Some Distribution-free Tests for the Equality of Cause Specific Hazard Rates

Subhash C. Kochar
Indian Statistical Institute
7, SJS Sansanwal Marg
New Delhi-110016, India

Abstract

We consider the competing risks problem when the available data is in the form of times and causes of failure. In many practical situations it is important to know whether the various risks under consideration are equally fatal or whether some risks are *more serious* than others in terms of their cause specific hazard rates and the cumulative incidence functions. In this paper we review some of the recently proposed distribution-free tests for the above problem. Since the data are invariably censored in such problems, the results from the powerful theory of counting processes and maringales are very useful in studying the asymptotic properties of such procedures.

Key words : Competing risks, cumulative incidence function, counting processes, martingales, Nelson - Aalen estimator of cumulative hazard, ordered alternatives, random walk.

1 Introduction

In the standard competing risks model, an experimental unit or subject is exposed to several risks but the actual failure (or death) is attributed to exactly one cause. Let us assume that a unit is 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, 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 $P(X = Y) = 0$. Thus, the observed data is in the form of (T, δ) for each observed item.

It is well known that the joint and the marginal probability distributions of X and Y are not identifiable on the basis of the observable random variables (T, δ) unless X and Y are independent. This aspect of nonidentifiability has been discussed by Cox (1959), Tsiatis (1975) and Kalbfleisch and Prentice (1980), among others.

If X and Y are independent then it is meaningful to compare the marginal probability distributions of the respective risks when the other risk has been eliminated from the enviornmnet. Because of the nonidentifiability inherent in the competing risks model, the independence of X and Y cannot be tested on the basis of the (T, δ) data and must be assumed a prior on the basis of the physical or biological process leading to the failure of the unit. In many practical situations it is not realistic to assume the independence of X and Y. In such situations a comparison of the marginal distributions of X and Y is not meaningful as these functions may not represent the probability distributions of X and Y in any practical situation. For detailed discussions on this point, see Gail (1975), Prentice *et al.* (1978), Slud *et al.* (1988) and Prentice and Kalbfleisch (1988).

Yet on the basis of the competing risks data it is often useful to distinguish between the following alternatives: (i) the two risks are equal, and (ii) one risk is greater than the other, within the environment in which the two risks are acting simultaneously.

To quantify this, the concept of (ordinary) hazard rate has been generalized in the competing risks model to the notion of *cause specific hazard rates* (CSHR). In the continuous case the i^{th} cause specific hazard rate is defined by

$$h_i(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \, pr[t \leq T < t + \Delta t, \; \delta = i | T \geq t] \tag{1.1}$$

$i = 1, 2$. If T is discrete, the i^{th} cause specific hazard rate is given by $pr[T = t, \delta = i | \; T \geq t]$. In either case the overall hazard rate for time to failure is then given by $h(t) = h_1(t) + h_2(t)$. Cause specific hazard rates provide detailed information on the extent of each type of risk at each time t. In models where the various causes of failure are independent, $h_i(t)$ reduces to the (ordinary) hazard rate corresponding to the marginal distribution of failure from the i^{th} cause. Prentice et al. (1978) emphasize that only those quantities which are expressible 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, our hypotheses are phrased in terms of cause specific hazard rates and hence identifiability is not a problem.

Based on a random sample (T_i, δ_i) $i = 1, 2, \ldots, n$ on (T, δ), we consider the problem of testing the null hypothesis,

$$H_0 : h_1(t) = h_2(t) \quad \text{for all } t \;, \tag{1.2}$$

against the alternative

$$H_1 : h_1(t) \leq h_2(t) \quad \text{for all } t \;, \tag{1.3}$$

with strict inequality for some t.

Such comparisons can also be made in terms of *cumulative incidence function*

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

corresponding to each cause j. Observe that

$$F_j(t) = \int_0^t h_j(u) S_T(u)\, du\ , \tag{1.4}$$

where $S_T(u) = pr[T > t] = 1 - F_1(t) - F_2(t)$ is the survival function of T. The null hypothesis H_0 is equivalent to

$$H_0 : F_1(t) = F_2(t),\ t \geq 0.$$

Another interesting alternative is

$$H_2 : F_1(t) \leq F_2(t),\ t \geq 0 \tag{1.5}$$

and with strict inequality for some t. H_1 also implies

$$H_3 : \tilde{F}_1(t) \leq \tilde{F}_2(t),\ t \geq 0 \tag{1.6}$$

and with strict inequality for some t, where

$$\begin{aligned} \tilde{F}_j(t) &= pr[T > t, \delta = j] \\ &= \int_t^\infty h_j(u) S_T(u)\, du. \end{aligned} \tag{1.7}$$

All the above three alternatives H_1, H_2 and H_3 say that risk Y is *more serious* than risk X in some sense. It follows from (1.4) and (1.7) that H_1 implies H_2 as well as H_3. For various probabilistic interpretations of the above alternatives and their implications, see Aly and Kochar (1993).

Note that there may be no reason to expect *a priori* that the cause specific hazard rates are equal (except, say, when they represent identical components in a series system), but this is a natural choice of null hypothesis for the ordered alternatives H_1, H_2 and H_3.

Besides applications in the health sciences, these procedures have applications in industrial accelerated life tests (cf. Froda, 1987). When comparing the quality of k different brands of a component, several components may be tested in series. The components are functioning in the same environment and their times to failure are generally dependent. The system fails as soon as one of the components fails. This experimental design identifies weak components early in the experiment thus saving valuable time. On the basis of such data, one might like to test whether components supplied by different suppliers are of the same quality against ordered alternatives. This type of testing gives rise to the above type of data.

The following result by Kochar and Proschan (1991) is very useful in studying the distributional properties of various test statistics. It extends the famous characterization result on proportional hazards of independent competing risks to the dependent case (cf. Armitage (1959), Allen (1963) and Sethuraman (1965)).

LEMMA 1.1 *Let*

$$T = min(X_1, X_2, \ldots, X_n) \text{ and } \delta = i \text{ if } T = X_i \quad (i = 1, 2, \ldots, n).$$

Then T *and* δ *are independent if and only if the the cause specific hazard rates* $h_1(t), h_2(t), \ldots, h_n(t)$ *of* $X_1, X_2, \ldots, X_n$ *are proportional.*

Assuming that the underlying risks are independent and the lifetimes are continuous, various authors have proposed nonparametric tests for testing the equality of two or more hazard rates against ordered alternatives. Using the competing risks data, Froda (1987) has proposed locally most powerful rank tests for testing the equality of two risks against scale alternatives. Bagai, Deshpandé and Kochar (1989 a,b) developed distribution-free rank tests for testing the equality of two hazard rates against stochastic ordering and hazard rate ordering alternatives. Neuhaus (1991) has proposed asymptotically optimal rank tests for comparing several independent competing risks differing in their location or scale parameters. Yip and Lam (1992 and 1993) suggested a class of weighted logrank type statistics. Gray (1988), generalizing the approach of Harrington and Fleming (1982), has proposed a class of c-sample tests for comparing the crude incidence function, $F_1(t) = pr[T \leq t, \delta = 1]$ of the first risk over c different populations. The case of two *dependent risks* has been considerd by Sen (1979), Deshpandé (1989), Aras and Deshpandé (1992), Aly, Kochar and McKeague (1994) and by Sun and Tiwari (1995). Sen (1979) proposed nonparametric tests with maximum asymptotic relative efficiency for testing the interchangeability of two competing risks against alternatives expressed in terms of $\pi(t) = pr[\delta = 1 | T = t]$. Where as Aras and Deshpandé (1992) derived locally most powerful rank tests of H_0 against various parametric alternatives, Aly, Kochar and McKeague (1994) have proposed Kolmogrov-Smirnov type tests for testing the equality of two competing risks. In the case of discrete (grouped) data, Dykstra, Kochar and Robertson (1995) have studied the likelihood ratio test for testing the equality of k CSHRs against ordered alternatives of the type H_2.

In the next section we briefly describe the various tests proposed in the literature for comparing *independent* competing risks. In the third section we discuss the case of dependent competing risks. Section 4 is devoted to asymptotic relative efficiency and power comparisons.

2 Tests for comparing independent competing risks

Assuming that X and Y are independent with respective distribution functions $F(t)$ and $G(t) = F(\sigma t)$, Froda (1987) proposed locally most powerful rank tests for testing the null hypothesis $\sigma = 1$ against the alternative $\sigma < 1$. The test statistics are of the type $S_n = \sum_{i=1}^{n} a(R_i, F)\delta_i^*$, where $\delta_i^* = \delta_i - 1$ and R_i is the rank of T_i among $T_1, T_2, \ldots, T_n$. In general, the score function $a(i, F)$ depends on F. In the case of negative exponential distribution, the locally most powerful rank test against the scale alternative is based on large values of the *sign statistic* $\sum \delta_i^*$. Bagai, Deshpandé and Kochar (1989 b) extended this approach to obtain locally most powerful rank tests for more general alternatives. Neuhaus (1991) proposed asymptotically optimal rank tests for comparing $k\,(\geq 2)$ independent competing risks differing in their location or scale parameters.

Using Gehan (1965) type arguments, Bagai, Deshpandé and Kochar (1989 a) proposed the following U-statistic for comparing the hazard rates of F and G.

$$U_1 = \left[\binom{n}{2}\right]^{-1} \sum\sum_{1\leq i<j\leq n} \phi\{(T_i, \delta_i^*), (T_j, \delta_j^*)\}, \tag{2.1}$$

where

$$\phi_1\{(T_i, \delta_i^*), (T_j, \delta_j^*)\} = \begin{cases} 1 & \text{if } \delta_i^* = 1 \text{ and } T_i > T_j \\ & \text{or } \delta_j^* = 1 \text{ and } T_j > T_i \text{ ;} \\ 0 & \text{otherwise.} \end{cases}$$

The statistic U_1 can also be expressed as

$$W_1 = \binom{n}{2} U_1 = \sum_{i=1}^{n} (R_i - 1)\delta_i^*, \tag{2.2}$$

which is a Wilcoxon signed rank type statistic. It has the same null distribution as the Wilcoxon signed rank statistic with n replaced by $n + 1$.

The statistic U_1 can also be used for testing H_0 against H_3. Under H_3,

$$\Delta_1(t) = \tilde{F}_2(t) - \tilde{F}_1(t) \geq 0, \quad \forall t \geq 0.$$

It can be shown that

$$\int_0^\infty \Delta_1(t)\, dF_T(t) = pr[\delta_1^* = 1, T_1 > T_2] - \frac{1}{2}, \tag{2.3}$$

where (T_i, δ_i^*) $i = 1, 2$ are two independent copies of (T, δ^*) and F_T denotes the distribution function of T. A U-statistic estimator of this parameter leads to statistic U_1. A test based on large values of U_1 is significant for testing for testing H_0 against H_3. As pointed out by Kochar and Proschan (1991), its null distribution remains valid under H_0 even when the risks are dependent. Deshpandé (1989) has also independently proposed and studied the properties of this test for comparing dependent risks.

Similarly, for testing H_0 against H_2, one can consider the measure of deviation $\psi(t) = F_2(t) - F_1(t)$. It can be seen that

$$\int_0^\infty \psi(t)\, dF_T(t) = pr[\delta_1^* = 1, T_1 \leq T_2] - \frac{1}{2}. \tag{2.4}$$

A U-statistic estimator of this parameter is

$$U_2 = \left[\binom{n}{2}\right]^{-1} \sum\sum_{1 \leq i < j \leq n} \phi_2\{(T_i, \delta_i^*), (T_j, \delta_j^*)\}, \tag{2.5}$$

where

$$\phi_2\{(T_i, \delta_i^*), (T_j, \delta_j^*)\} = \begin{cases} 1 & \text{if } \delta_i^* = 1 \text{ and } T_i \leq T_j \\ & \text{or } \delta_j^* = 1 \text{ and } T_j \leq T_i \text{ ;} \\ 0 & \text{otherwise} \end{cases}$$

which is equivalent to the statistic

$$W_2 = \binom{n}{2} U_2 = \sum_{i=1}^{n} (n - R_i + 1)\delta_i^*. \tag{2.6}$$

Bagai, Deshpandé and Kochar (1989 b) proposed this test on heuristic grounds, again using Gehan type arguments. In the light of Lemma 1.1, it follows that U_2 also has the same null distribution as the Wilcoxon signed rank statistic.

Yip and Lam (1992) have proposed a class of non-parametric tests for testing the equality of hazard rates of two independent competing risks. Their class includes the tests U_1 and U_2 as special cases. They have used the counting processes approach to study the asymptotic properties of their tests.

Let $N_i(t)$ denote the number of failures from cause i during the interval $(0, t]$ for $i = 1, 2$; and let $Y(t) = n - N_1(t-) - N_2(t-)$ be the number of survivors just before time t. Let $w(u)$ be a locally bounded predictable process. Yip and Lam (1992) proposed a class of tests based on studentized versions of the following statistics, with their asymptotic null variances estimated from the data itself,

$$L_w = \int_0^\infty w(u)[d\hat{\Lambda}_1(u) - d\hat{\Lambda}_2(u)], \tag{2.7}$$

where $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are the Nelson estimators of the cumulative hazard functions for the independent competing risks and are given by

$$\hat{\Lambda}_j(t) = \sum_{i:T_i \le t} \frac{I(\delta_i = j)}{Y(T_i)}.$$

The weight $w(\cdot)$ reflects the relative importance attached to the difference between the CSHR's at time u. Different choices of $w(u)$, lead to different test statistics. In particular, the choice $w(u) = Y(u)N(u-)$ will lead to a statistic equivalent to U_1 whereas the choice $w(u) = \{Y(u)\}^2$ will yield U_2. As pointed out by Yip and Lam (1992), where as the first choice of the weight function puts more weight on the later part of the experiment, the second choice puts more weight on the earlier part of the experiment. They have also studied tests corresponding to other choices of weight functions.

In a follow up paper, Yip and Lam (1993) extended their two-risks procedures to compare k (> 2) independent competing risks.

3 Tests for comparing dependent risks

Sen (1979) considers fixed sample as well as sequential tests for the null hypothesis of bivariate symmetry of the joint distribution of (X, Y), the notional lifetimes under the two risks. The alternatives are expressed in terms of $\pi(t) = pr[\delta = 1|T = t]$, the conditional probability that the first risk is the cause of failure given that a failure occurs at time t. Then he obtains tests with maximum Pitman and Bahadur asymptotic relative efficiencies. Aras and Deshpandé (1992) adopt a different approach. They model this problem in terms of cumulative incidence functions F_1 and F_2. By assuming different parametric as well as Lehmann type alternatives between them, they obtain locally most powerful rank tests for the above problem. It remains to be seen whether such formulations are realistic in practice.

In practice, it may not be easy to model the competing risks problem in term of the function $\pi(t)$ of Sen (1979) or specify the exact relationship between the cumulative incidence functions as advocated by Aras and Deshpandé (1992). Moreover, none of the above mentioned procedures discuss the situation when the data is censored. Recently Aly, Kochar and McKeague (1994) have proposed Kolmogrov-Smirnov type tests for testing the equality of two competing risks against alternatives H_1 and H_2. They considered both the censored and the uncensored cases. In the rest of this section, we focus on these procedures.

3.1 The case of uncensored data

Since

$$\begin{aligned}\psi(t) &= \int_0^T S_T(u)(h_2(u) - h_1(u))\,du \\ &= F_2(t) - F_1(t),\end{aligned}$$

it follows that under H_0, $\psi(t) \equiv 0$ and H_1 holds if and only if $\psi(t)$ is nondeccreasing in t. Let

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

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 large values of D_{1n} provide evidence in favor of H_1.

The exact null distribution of D_{1n} is given by:

$$pr[nD_{1n} < t] = \frac{2}{2t+1} \sum_{j=0}^{2t} \left\{\cos \tfrac{j\pi}{2t+1}\right\} \sin\left\{\tfrac{j\pi(t+1)}{2t+1}\right\} \left\{1 + \cos \tfrac{j\pi}{2t+1}\right\} \times \left\{\tfrac{1-(-1)^j}{2}\right\} \Big/ \sin \tfrac{j\pi}{2t+1}, \tag{3.1}$$

for $t = 1, \ldots, n+1$.

The asymptotic null distribution of D_{1n} is obtained using the invariance principle for partial sums. Under H_0,

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

where $\{W(t),\ t \ge 0\}$ is a standard Bronian motion. Consequently, for $c > 0$

$$pr\{\sqrt{n} D_{1n} \le c\} \to \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\{-\pi^2 (2k+1)^2 / 8c^2\}. \tag{3.2}$$

The exact formula (3.1) can easily be used to generate a table of critical values. Using (3.5) 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.

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 $h_1(t) \ne h_2(t)$ for some t. In that case it is natural to use the Kolmogorov–Smirnov test statistic

$$\bar{D}_n = \sup_{t \ge 0} |\psi_n(t)|. \tag{3.3}$$

Under H_0, $\sqrt{n}\bar{D}_n$ converges in distribution to $\sup_{0 \le x \le 1} |W(x)|$. This gives an omnibus test–consistent against *arbitrary* departures from H_0.

For testing H_0 against H_2, Aly, Kochar and McKeague (1994) proposed the statistic

$$D_{2n} = \sup_{0 \leq t < \infty} \psi_n(t). \tag{3.4}$$

Again large values of D_{2n} being significant for testing H_0 against H_2.

Under H_0

$$pr\{nD_{2n} = k\} = \frac{1}{2^n}\binom{n}{[\frac{n-k}{2}]}, \; k = 0, 1, \ldots, n.$$

The asymptotic null distribution is obtained using the invariance principle for partial sums (see, e.g., Chapter 2 of Csörgő and Révész (1981)). Under H_0

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

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

3.2 The case of 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

$$\begin{aligned}\phi(t) &= \int_0^T S_C(u-)^{1/2} d(F_2 - F_1)(u) \\ &= \int_0^T S_T(u-) S_C(u-)^{1/2} (h_2(u) - h_1(u))\, du,\end{aligned} \tag{3.5}$$

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 censoring in order that our test statistics remain (asymptotically) distribution-free. H_0 is equivalent to $\phi(t) = 0$ for all $t \geq 0$, but H_1 holds if and only if ϕ is increasing.

Thus large positive values of

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

give evidence of a departure from H_0 in the direction of H_1. An obvious choice of ϕ_n is

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

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 (1978) of the cumulative CSHR function $\Lambda_j(t) = \int_0^t h_j(u)\, du$:

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

where $R_i = \#\{k : \tilde{T}_k \ge \tilde{T}_i\}$ 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. The problem at hand concerns a three-state chain with two absorbing states corresponding to the two types of failure.

Since $\phi(t) > 0$ for some t under H_2, positive large values of the test statistic

$$D_{4n} = \sup_{0 \le t < \infty} \phi_n(t), \tag{3.6}$$

give evidence of a departure from H_0 in the direction of H_2.

The estimate $\phi_n(t)$ is similar in spirit to a weighted logrank statistic, being of the form

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

where w is a locally bounded, predictable weight function. The weight $w(u)$ reflects the relative importance attached to the difference between the CSHRs at time u. Our choice of w, which essentially controls instability in the tails, is designed to give an asymptotically distribution-free test. As discussed in the previous section, Yip and Lam (1992) have suggested statistics of the type $L_n(\infty)$, for various choices of w, as test statistics for H_0 (in the case of uncensored data and independent X and Y).

As stated in the next theorem that D_{3n} and D_{4n} are asymptotically distribution-free with the same limiting distributions that obtained in the uncensored case.

THEOREM **3.1** *Under* H_0

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

Recently Sun and Tiwari (1995) have proposed a new test for testing H_0 against H_1 when the risks are possibly dependent and the data is randomly right censored. Their test is an extension of the W_2 test of Bagai, Deshpandé and Kochar (1989 b) to the censored case.

3.3 Extensions to more than 2 risks

Aly, Kochar and McKeague (1994) show that their approach can be extended 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. They also discuss the appropriate modifications to be made to their tests when one is interested only in the dominance of one risk over the other in a specified interval. It is not clear whether the other available tests discussed in this paper are amenable to this modification. Neuhaus (1991) has proposed asymptotically optimal rank tests for comparing k independent competing risks differing in their location or scale parameters.

Dykstra, Kochar and Robertson (1995) have studied the problem of comparing the cause specific hazard rates of k competing risks when the data is discrete (or grouped). No assumptions are made about the independence of the risks and they allow observations to be right censored but do assume that the censoring distribution is independent of actual time to failure. First they obtain nonparametric maximum likelihood estimates of the various cause specific hazard functions under the null as well as the alternative hypotheses in preparation towards constructing the likelihood ratio statistic for testing H_0 against H_1. The maximum likelihood estimates of the cause specific hazard rates under the order restriction imposed by H_1 are also of independent interest. The likelihood ratio test is a similar one and its asymptotic null distribution is of the chi-bar square type.

It can be shown that the various tests discussed in this section are consistent against their intended alternatives.

4 Efficiency and power comparisons

First we consider the case when the risks are independent and the data is not censored. For comparison purposes, we use the sign test $S = \sum_{i=1}^{n} \delta_i^*$, which is the locally most powerful rank test of H_0 against proportional CSHRs in the absence of censoring, see Aras and Deshpandé (1992). We consider the following alternatives for Pitman asymptotic relative efficiency comparisons.

(i)

$$h_\theta(x) = (1+\theta)h_0(x),$$

the proportional hazards model.

(ii)

$$h_\theta(x) = 1 + \theta(1 - e^{-x}),$$

the Makeham distribution.

(iii)

$$h_\theta(x) = [1 - \theta \ln \overline{F}(x)]h_0(x).$$

(linearly increasing failure rate when F is exponential)

The case $\theta = 0$ corresponds to the null hypothesis and $\theta > 0$ corresponds to H_1. The asymptotic relative efficiency of a test U with respect to the sign test S for the alternative (i) above is denoted by $e_i(U, S)$. These values are reported below.

$$e_1(U_1, S) = 0.75, \quad e_1(U_2, S) = 0.75$$
$$e_2(U_1, S) = 1.47, \quad e_2(U_2, S) = 0.27$$
$$e_3(U_1, S) = 1.69, \quad e_3(U_2, S) = 0.19 .$$

Since the Kolmogrov- Smirnov type tests do not satisfy the Noether's (1955) conditions for Pitman asymptotic relative efficiencies, we performed a simulation study to comapre the powers of the various tests under consideration.

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

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

where $(\lambda_0, \lambda_1, \lambda_2)$ are parameters and $\lambda = \lambda_0 + \lambda_1 + \lambda_2$. In this case, the CSHRs

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

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 $\lambda_0 = 0$. Since T and δ are independent whenever the CSHRs are proportional, it follows that the sign test S is the locally most powerful rank test in this case.

The following table gives the estimated powers of the different tests based on 10,000 samples, each of size 100 generated from the ACBVE distribution with $\lambda_0 = 0$, $\lambda_1 = 1.0$; and $\lambda_2 = 1.0, 1.4, 1.8$ and 2.2, respectively. Note that the case $\lambda_2 = 1.0$ corresponds to the null hypothesis. 5% asymptotic critical levels were used in this study. We used the exact null mean and variance in the asymptotic normal approximation of U_1 (U_2).

Table 4.1
Estimated Powers of the D_1, D_2, U_1, U_2 and the Sign Tests
at Asymptotic Levels of 5%

Test	λ_2			
	1.0	1.4	1.8	2.2
D_1	3.76	41.98	82.53	96.83
D_2	4.85	47.71	86.98	98.14
U_1	4.79	43.06	80.54	95.42
U_2	4.99	42.67	80.81	95.77
Sign	4.39	49.50	88.29	98.66

This table shows that there is atmost a 2% loss of power in using D_2 instead of the sign test when there is no censoring. For D_1, this loss is somewhat more. For this alternative, the U_1 and U_2 tests perform equally well, reconfirming the observation made by Yip and Lam (1992).

Next we consider the case when X and Y are independent with X having exponential distribution with unit hazard rate and Y having linearly increasing failure rate distribution with hazard rate $h_\theta(x) = (1 + \theta x)$. Table 4.2 gives the estimated powers of the various tests based on 10,000 samples, each of size 100 for this alternative. Again 5% asymptotic critical levels were used in this study.

Table 4.2
Estimated Powers of the D_1, D_2, U_1, U_2 and the Sign Tests
at Asymptotic Levels of 5%

Test	θ			
	0.5	1.5	2.5	3.5
D_1	18.21	58.25	83.62	92.86
D_2	19.81	58.23	82.61	92.27
U_1	31.36	79.37	97.31	98.36
U_2	11.92	29.72	55.85	62.60
Sign	22.27	63.52	86.33	94.18

For this alternative, the U_1 test outperforms the others, followed by the Sign test; and the D_1, D_2 tests. The U_2 test is less powerful for this alternative. The U_1 statistic puts more weight on the later part of the experiment, where as U_2 gives more importance to early failures. U_1 will be a very good choice in the competing risks set up since survival distributions are restricted to be non-negative, significant differences will, in general, become more obvious as times goes by. As shown in Bagai, Deshpandé and Kochar (1989 b) the U_2 test is more suitable for location type alternatives. The above studies show that the Kolmogrov- Smirnov type tests D_1 and D_2 are quite powerful in the absence of specific knowledge about the alternatives. They are easy to implement and the tables for their exact null distributions are available when the data is uncensored. In case, the data is censored, asymptotic critical values can be used though the resulting tests will be somewhat conservative.

Aly, Kochar and McKeague (1994) also consider the case when the data is censored and the variables X and Y are dependent. Based on their simulation study on ACBVE distribution, they made the following remarks.

(i) The use of the asymptotic critical levels gives somewhat conservative tests, and this effect increases as the censoring becomes more severe. However, the test based on D_{4n} appears to be less conservative (and more powerful) than the one based on D_{3n}, and both tests become less conservative as the sample size increases.

(ii) The levels of the tests are close to their nominal 5% values for sample size 500, except under heavy censoring.

(iii) There is no apparent adverse effect on the levels or the power due to lack of independence of X and Y.

Acknowledgements

The author is thankful to the referee for making useful suggestions on an earlier draft of this paper.

References

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

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5**, 141–150.

Allen, W. R. (1963). A note on the conditional probability of failure when hazards are proportional. *Operations Research* **11**, 658-659.

Armitage, P. (1959). The comparison of survival curves. *J. Roy. Statist. Soc. Ser. A* **122**, 279-300.

Aly, E. A. A. and Kochar, S. (1993). On hazard rate ordering of dependent variables. *Adv. Appl. Prob.* **25**, 477-482.

Aly, E. A. A., Kochar, S. C. and McKeague, I. W. (1994). Some tests for comparing cumulative incidence functions and cause- specific hazard rates.*J. Amer. Statist. Assoc.* **89**, 994-999.

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. *Commun. Statist. Theory Meth.* **18**, 107-120.

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

Cox, D. R.(1959). The analysis of exponentially distributed lifetimes with two types of failures. *J.R.S.S. Ser. B.* **21**, 411–421.

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

Deshpandé, J. V. (1989). A test for bivariate symmetry of dependent competing risks. Technical Report.

Dykstra, R., Kochar, S. and Robertson, T. (1995). Likelihood based inference for cause specific hazard rates under order restrictions.*J. Mult. Anal.* **54**, 263-274.

Froda, S. (1987). A signed rank test for censored paired lifetimes. *Commun. Statist. Theory Meth.* **16**, 3497–3517.

Gail, M. (1982). *Competing risks*, Encyclopedia in Statist. Sci. (Edited by S. Kotz and N.L. Johnson), 75–81.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Statist.* **16**, 1141–1154.

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

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Stat. Assoc.* **53**, 457–481.

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.

Neuhaus, G. (1991). Some linear and nonlinear rank tests for competing risks models. *Commun. Statist. Theory Meth.*, **20** 667–701.

Noether, G.E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26**, 64-68.

Prentice, R. L., Kalbfleisch, J. D. (1988). Reply to Slud, Byar and Schatzkin, *Biometrics* **44**, 1205.

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

Sen, P. K. (1979). Nonparametric tests for interchangeability under competing risks. In *Contributions to Statistics, Jaroslav Hajek Memorial Volume* (J. Jureckova, ed.), 211–228. Reidel, Dordrecht.

Sethuraman, J. (1965). On a characterization of the three limiting types of the extreme. *Sankhya Ser A* **27**, 357-364.

Slud, E. V., Byar, D. P. and Schatzkin, A. (1988). Dependent competing risks and the latent failure model. *Biometrics* **44**, 1203–1204.

Sun, Y. and Tiwari, R. C. (1995). Comparing cause-specific hazard rates of a competing risks model with censored data. Paper presented in the International Workshop on Censored Data held at Poona University, India, Dec. 28, 1994- Jan. 1, 1995.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci.* **72**, 20–22.

Yip, P. and Lam, K. F. (1992). A class of non-parametric tests for the equality of failure rates in a competing risks model. *Commun. Statist. Theory Meth.* **21**, 2541–2556.

Yip, P. and Lam, K. F. (1993). A multivariate nonparametric test for the equality of failure rates in a competing risks model. *Commun. Statist. Theory Meth.* **22**, 3199-3222.