Generalized Supremum Tests for the Equality of Cause Specific Hazard Rates

Subhash C. Kochar
K. F. Lam
and
Paul S.F. Yip

April 24, 2002
isid/ms/2002/11

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Generalized Supremum Tests for the Equality of Cause Specific Hazard Rates

Subhash C. Kochar

Indian Statistical Institute, 7, SJS Sansanwal Marg, New Delhi-110016, India

K. F. Lam[1] and Paul S.F. Yip

Department of Statistics & Actuarial Science, The University of Hong Kong, HK

Abstract. In this paper we propose two new classes of asymptotically distribution-free Renyi-type tests for testing the equality of two risks in a competing risk model with possible censoring. This work extends the work of Aly, Kochar and McKeague [1994, *Journal of American Statistical Association*, **89**, 994-999] and many of the existing tests for this problem belong to these newly proposed classes. The asymptotic properties of the proposed tests are investigated. Simulation studies are done to compare the performance with existing tests. A competing risks data set is analyzed to demonstrate the usefulness of the procedure.

Keywords: competing risks, counting processes, cumulative incidence function, martingales, Nelson-Aalen estimator of cumulative hazard, ordered alternatives

[1]Correspondence Address: Dr. K.F. Lam, Department of Statistics & Actuarial Science, The University of Hong Kong, Hong Kong. E-mail: hrntlkf@hku.hk, Fax: (852) 2858-9041

1. Introduction

Consider a competing risks model with two causes of failure. Let T denote the lifetime of a subject, assumed to be continuous, with distribution function F and survivor function S, and let δ denote the cause of failure, that is, $\{\delta = j\}$ is the event that the failure is due to risk j, $j = 1, 2$. In many practical situations it is important to know whether the various risks under consideration are equally serious or whether some risks are *more serious* than others, within the environment in which the 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), which is defined by

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t,\ \delta = j | T \geq t), \qquad j = 1, 2. \tag{1}$$

The overall hazard rate for time to failure satisfies the relation $\lambda_T(t) = \lambda_1(t) + \lambda_2(t)$. Cause specific hazard rates provide detailed information on the extent of each type of risk at each time point t. In models where the various causes of failure are independent, $\lambda_j(t)$ reduces to the (ordinary) hazard rate corresponding to the marginal distribution of failure from the j^{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. Censoring is possible arising from removal of subjects before failure from either cause 1 or cause 2 and it may be due to combination of other competing risks. Denote the censoring time by C and its survivor function by S_C. We assume that $S_C(t) > 0$ for all t and C is independent of T. We now identify three causes of failure, $\delta = 0, 1, 2$, where $\{\delta = 0\}$ is the event that the subject was censored. Under right censoring, we observe n independent, identically distributed copies (X_i, δ_i), $i = 1, \ldots, n$ of (X, δ), where $X = \min(T, C)$. More specifically, on the basis of these data, we formulate the problem of testing the null hypothesis,

$$H_0 : \lambda_1(t) = \lambda_2(t) \quad \text{for all } t, \tag{2}$$

against the alternative

$$H_a : \lambda_1(t) \leq \lambda_2(t) \quad \text{for all } t, \text{ with strict inequality for some } t. \tag{3}$$

In the literature such comparisons have also been made in terms of the *cumulative incidence functions* F_1 and F_2 (see Gray, 1988 and Luo & Turnbull, 1999), where

$$F_j(t) = P(T \leq t, \delta = j) = \int_0^t S(u)\lambda_j(u)\, du,\ j = 1, 2.$$

Note that the null hypothesis H_0 in (2) is equivalent to

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

and H_a in (3) implies

$$H_b : F_1(t) \leq F_2(t),\ t \geq 0 \text{ with strict inequality for some } t.$$

Several tests have been proposed in the literature for testing H_0 against various alternatives (see Kochar, 1995 and Carriere & Kochar, 2000). Most of the tests discussed in the literature can be expressed as functionals of weighted log-rank type statistics of the form

$$L_n(t) = \int_0^t w(u)\, d(\hat{\Lambda}_2 - \hat{\Lambda}_1)(u), \tag{4}$$

where $\Lambda_j(t) = \int_0^t \lambda_j(u)du$ is the cumulative cause specific hazard rate function for risk j, $j = 1, 2$ and the Nelson-Aalen estimator (see, e.g., Fleming and Harrington, 1991) of Λ_j is

$$\hat{\Lambda}_j(t) = \sum_{i:X_i \le t} I(\delta_i = j)/R_i$$

where $R_i = \#\{k : X_k \ge X_i\}$ is the size of the risk set at time X_i-. The weight function $w(u)$ reflects the importance attached to the difference between the CSHRs at time u. The tests proposed by Yip and Lam (1992) are based on studentized $L_n(\infty)$ statistics for various choices of w. Although these tests may be able to detect certain kinds of departure from H_0 with high power, they may not be consistent against general alternatives.

Aly, Kochar and McKeague (1994) proposed two Renyi-type tests for this problem. The first one, which is suitable for comparing cumulative incidence functions, is based on the statistic

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

where

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

and $\hat{S}_T$ and $\hat{S}_C$ are the product-limit estimators of S_T and S_C, respectively. Their second test which is suitable for testing against H_a is based on the statistic

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

The rationale behind these tests is that ϕ_n is an estimator of

$$\phi(t) = \int_0^t S_T(u-)S_C(u-)^{1/2}(\lambda_2(u) - \lambda_1(u))\, du$$

and H_a holds if and only if ϕ is increasing. Thus large positive values of D_{4n} give evidence of a departure from H_0 in the direction of H_a. This property will continue to hold if instead of $S_T(u-)S_C(u-)^{1/2}$, we use some other suitable *nonnegative* weight function. It was shown in Aly, Kochar and McKeague (1994) that the choice of the weight function $\hat{S}_T(u-)\hat{S}_C(u-)^{1/2}$ leads to asymptotically distribution-free tests when the data are censored and these tests are exactly distribution-free otherwise. An unpleasant property of these tests is that they are very conservative. This is probably due to the fact that the finite sample distributions of the statistics $n^{1/2}D_{3n}$ and $n^{1/2}D_{4n}$ cannot be approximated closely by their respective asymptotic distributions when n is not extremely large.

In Section 2, we propose two new classes of asymptotically distribution-free tests which are similar to the studentized versions of the D_{3n} and D_{4n} statistics, but with arbitrary nonnegative weight functions chosen from a flexible class of weight functions. In Section 3, we carry out an intensive simulation study to compare the performance of the various tests. It seems from this study that the studentized statistics using the estimated covariance functions appear to converge to the asymptotic null distribution much faster, which improves the small sample approximations significantly. Moreover, the proposed tests are highly flexible and this approach unifies the existing procedures. The proposed methods are illustrated by application to data from Hoel (1972) in Section 4. Section 5 contains some closing remarks and discussion.

2. The proposed classes of tests

In this section, we generalize the tests of Aly, Kochar and McKeague (1994) by taking different weight functions w. With suitable studentization, this yields a versatile family of tests. It is well known that under H_0, $n^{1/2}L_n(t)$ is a martingale with predictable variation process $\sigma^2(t)$ which, under some mild conditions, can be estimated consistently by

$$S_n^2(t) = \int_0^t \frac{w^2(u)}{\bar{Y}^2(u)/n}\, d\bar{N}(u), \tag{5}$$

where $\bar{Y}(u) = \sum_{i=1}^n I(X_i \geq u)$ is the total number of items at risk at $u-$, and $\bar{N}(u)$ is the total number of deaths up to time u. Using $L_n(t)$ as in (4), we propose the following three classes of test statistics for testing H_0:

$A_n(w) = \frac{L_n(\infty)}{S_n(\infty)}$,

$B_n(w) = \sup_{0\leq t<\infty} \frac{L_n(t)}{S_n(\infty)}$,

$C_n(w) = \sup_{0\leq s<t<\infty} \frac{L_n(t)-L_n(s)}{S_n(\infty)}$.

Large values of the statistic indicate statistical significance for tests of H_0. It follows from the results given in the Appendix that under H_0 and under some regularity conditions, $\{n^{1/2}\, L_n(t)/S_n(\infty)\}$ converges weakly to $\{W(t),\ t \geq 0\}$, a standard Brownian motion. As a consequence, under H_0,

$$n^{1/2}\, A_n(w) \to Z, \quad \text{a standard normal variable,} \tag{6}$$

$$P[n^{1/2}\, B_n(w) > b] \to P[\sup_{0<t<1} W(t) > b] = 2(1-\Phi(b)),\ b \geq 0, \tag{7}$$

$$n^{1/2}\, C_n(w) \to \sup_{0\leq t\leq 1} |W(t)|, \tag{8}$$

where Φ is the standard normal distribution function.

Consequently, for $c > 0$

$$P(n^{1/2}C_n \leq c) \to \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\{-\pi^2(2k+1)^2/8c^2\}. \tag{9}$$

Using (9) the asymptotic 0.90, 0.95 and 0.99 quantiles of $n^{1/2}C_n$ 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: $F_1(t) \neq F_2(t)$ for some t, which is equivalent to $\lambda_1(t) \neq \lambda_2(t)$ for some t. In that case it is natural to use the Kolmogorov–Smirnov type test statistic $B_n^* = \sup_{t\geq 0} |L_n(t)|/S_n(\infty)$. Using the same kind of arguments as in Aly, Kochar and McKeague (1994), it follows that under H_0, $n^{1/2}B_n^*$ converges in distribution to $\sup_{0\leq t\leq 1}|W(t)|$. This gives an omnibus test—consistent against *arbitrary* departures from H_0. The corresponding two-sided analogue of C_n is $C_n^* = \sup_{0\leq s<t<\infty}|L_n(t) - L_n(s)|/S_n(\infty)$ and its asymptotic null distribution is given in the Appendix.

The class A_n was proposed and studied by Yip and Lam (1992). In this class the choice of the weight function $w(u) = \bar{Y}(u)$ leads to the sign test whereas the choice $w(u) = \bar{Y}(u)\bar{N}(u-)$ gives a test which is equivalent to the one proposed by Bagai, Deshpandé and Kochar (1989a) for testing the equality of two hazard rates. On the other hand, the weight function $w(u) = \bar{Y}^2(u)$ results in the statistic proposed by Bagai, Deshpandé and Kochar (1989b) for testing against a stochastic ordering alternative. Previous studies show that the tests belonging to the class A_n have good power for testing against some specific alternatives, but they cannot be expected to be consistent against all alternatives to H_0. As will be seen later, the tests belonging to the classes B_n and C_n are sensitive to a wider range of alternatives and at the same time they maintain the full efficiency of the corresponding statistics belonging to the class A_n. In the uncensored case, the tests of Aly, Kochar and McKeague (1994) are extensions of the sign test to Renyi-type statistics and they are seen to be quite powerful for testing against alternatives where the cause specific hazard rates are proportional to each other and for this alternative the sign test is the UMP test. Similar observations were made by Gill (1980) and Fleming et al. (1987) in the case of classical two-sample problem when they extend the linear rank statistics to Renyi-type statistics using the same score function.

3. Simulations and power comparisons

To illustrate the flexibility of the proposed classes of tests, a large scale simulation study was conducted. The null hypothesis H_0 was tested against the alternatives

(i) $H_1 : \lambda_2(t) = (\beta + 1)\ \lambda_1(t)$;

(ii) $H_2 : \lambda_2(t) = \lambda_1(t)\ \{1 + \beta\ \Lambda_1(t)\}$; and

(iii) $H_3 : \Lambda_2(t) = \{\Lambda_1(t)\}^{\exp(\beta/2)}$.

The alternatives H_1 and H_2 belong to the class of order restricted alternatives H_a and one-sided tests were carried out for these. The alternative H_3 was considered by Lam (1998) where the two CSHRs cross and hence a 2-sided test was carried out. For simplicity, we let $\lambda_1(t) = 1$, the level of significance $\alpha = 0.05$, and β is set to be 0 and 1 at which $\beta = 0$ corresponds to the null hypothesis. For H_1, the failure times $T = \min(Y_1, Y_2)$ were generated from the absolutely continuous bivariate exponential distribution of Block and Basu (1974) with density

$$f(y_1, y_2) = \begin{cases} \frac{(\lambda_0+\lambda_1+\lambda_2)\lambda_1(\lambda_0+\lambda_2)}{\lambda_1+\lambda_2} e^{-\lambda_1 y_1-(\lambda_0+\lambda_2)y_2} & \text{if } y_1 < y_2, \\ \frac{(\lambda_0+\lambda_1+\lambda_2)\lambda_2(\lambda_0+\lambda_1)}{\lambda_1+\lambda_2} e^{-\lambda_2 y_2-(\lambda_0+\lambda_1)y_1} & \text{if } y_1 > y_2 \end{cases}$$

where λ_0 is the dependence parameter and $\lambda_0 = 0$ corresponds to the independence of the two risks. In this case, the cause specific hazard rates are proportional to each other and are given by

$$\lambda_j(t) = \frac{\lambda_j(\lambda_0 + \lambda_1 + \lambda_2)}{\lambda_1 + \lambda_2} \qquad j = 1, 2.$$

We set $\lambda_0 = 0$ and 1 in the study under H_1. For H_2 and H_3, we simply assumed the two risks to be independent of each other. In all the three cases, the censoring variable C was taken to be independently exponentially distributed. Three levels of censoring, namely no censoring, moderate and heavy censoring were considered to study the effect due to censoring. For each combination of the alternative hypothesis and the set of parameters assumed, 10000 data sets, each with a sample size of $n = 100$ were generated.

The weight functions used are

(a) $w_1(u) = \bar{Y}(u)$;

(b) $w_2(u) = \bar{Y}(u)\ \hat{\Lambda}(u-)$;

(c) $w_3(u) = \bar{Y}^2(u)$;

(d) $w_4(u) = \bar{Y}(u)\ \bar{N}(u-)$;

(e) $w_5(u) = \hat{S}_T(u-)\ \hat{S}_C(u-)^{1/2}$.

where the weight functions w_1 and w_2 are the optimal weight functions for the class of tests A_n, which give rise to asymptotically locally most powerful tests for H_1 and H_2, respectively (Yip and Lam, 1993). The tests generated by these five weight functions are compared with the tests $n^{1/2}D_{3n}$ and $n^{1/2}D_{4n}$ of Aly, Kochar and McKeague (1994), denoted by (f) under the classes B_n and C_n, respectively. The empirical type I error rates and the empirical powers of the tests with $\beta = 1$ are given in Tables 1 to 3.

Under H_0 ($\beta = 0$), the tests of Aly, Kochar and McKeague (1994) are more conservative in the sense that their empirical type I error probabilities are much smaller than the nominal level of significance 0.05, particularly when the censoring proportion is large. However, the tests proposed in this paper perform much better as their empirical type I error rates are quite close to the nominal level, and are not much affected by the magnitude of the censoring proportion. This indicates that the studentized technique has improved the rate of convergence of the proposed statistics to their asymptotic values which gives rise to more accurate inferential procedures.

The simulation study also demonstrates the importance of the weight function used. When testing against order restricted alternatives H_a and H_b, all tests with weight functions considered above perform quite well. The powers of the tests highly depend on the choice of the weight function. The three classes of tests, with optimal weight function generated from A_n, give good power for all values of β, and not just for local alternatives. In particular, under the usual order restricted alternatives, the test based on A_n is, in general, more powerful than the tests based on B_n and C_n for any arbitrary nonnegative weight function $w(u)$. In the cases with crossing CSHRs, the Renyi-type of tests based on B_n^* and C_n^* are generally more sensitive and more powerful than that of A_n^*. It is observed that the proposed two classes of tests are more versatile in the sense that they are power robust. The Renyi-type tests are generally more sensitive to departure from null hypothesis as is illustrated by the following example.

4. An example

The three classes of tests were applied to a set of mortality data given in Hoel (1972) which has been studied by many researchers in the field of competing risks analysis. The data were obtained from a laboratory experiment on RFM strain male mice which had received a radiation dose of 300 roentgens at ages of 5 to 6 weeks and were kept in a conventional laboratory environment. Causes of death were classified into three distinct groups, namely thymic lymphoma, reticulum cell sarcoma, and *other causes*. In this application, the deaths due to *other causes* are treated as censored observations and are assumed to be independent of the two types of cancer. The estimates of the cumulative hazard of dying from thymic lymphoma and reticulum cell sarcoma are given in Figure 1. Aly, Kochar and McKeague (1994), based on the plots of the smoothed estimates of the CSHRs, suggested that the CSHRs of the two types of cancer cross at about 500 days.

Hence, we only considered a 2-sided alternative using the complete data set. The weight functions (a) to (e) of Section 3 were used, and were compared with the tests of Aly, Kochar and McKeague (1994). The test statistics and the corresponding p-values (in parentheses) given in Table 4 suggest that the result is highly significant. By comparing the p-values of the tests, it is noticed the tests based on the C_n^* are more robust while the tests based on A_n^* and B_n^* may be more sensitive to the weight functions adopted. The proposed classes of tests are highly flexible and when we do not have any idea of the order of crossings of the two CSHRs, the tests based on C_n^* are recommended as they tend to give more robust results.

5. Discussion

The non-studentized tests of Aly et al. (1994) are asymptotically distribution-free only when the weight function $\hat{S}_T(u-)\ \hat{S}_C(u-)^{1/2}$ is used. Flexibility can be gained when different weight functions are adopted, but studentization is necessary in order to retain the asymptotic distribution-free properties. Simulation studies show that the studentized test statistics have better performance than the non-studentized statistics of Aly et al. (1994) in the sense that the finite sample distributions of the studentized statistics can be closely approximated by their respective asymptotic distributions under the null hypothesis. Empirically the studentized tests are almost unbiased even for moderate sample sizes, irrespective of the choice of the weight functions and censoring proportion. Choices of weight function have been proposed and discussed widely in the literature. However, the choice of weight function should be based on the investigator's desire to emphasize either early or late departures between the CSHRs as the data from different clinical trials may have different characteristics. For example unexpected early or late occurrences of the event may not be very informative and hence a weight function with lighter weight at both ends would be adopted by the investigator. The supremum version of the tests, namely B_n and C_n would be more sensitive to the cases where two CSHRs differ substantially for some range of t but not necessarily elsewhere. Nevertheless, tests based on weight function $w_1(u) = \bar{Y}(u)$ has reasonable power in practice in most situations. Together with the classes of tests C_n or C_n^*, which are less sensitive to the choice of weight functions, would be good tests to start with in general when we have no information about the characteristics of the data.

Acknowledgments

The authors acknowledge the valuable comments and suggestions of the Associate Editor and two referees. This research was supported by a RGC grant and the research funding of the Faculty of Social Sciences, the University of Hong Kong.

References

E. A. A. Aly, S. C. Kochar, and I. W. McKeague, "Some tests for comparing cumulative incidence functions and cause- specific hazard rates," *J. Amer. Statist. Assoc.*, 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," *Commun. Statist. 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.

H. W. Block and A. P. Basu, "A continuous bivariate exponential extension," *J. Amer. Statist. Assoc.*, vol. 69, pp. 1031-1037, 1974.

K. C. Carriere and S. C. Kochar, "Comparing sub- survival functions in a competing risks model," *Lifetime Data Analysis*, vol. 6, pp. 85-97, 2000.

W. Feller, "The asymptotic distribution of the range of sums of independent random variables," *Annals of Mathematical Statistics*, vol. 22, pp. 427-432, 1951.

T. R. Fleming, D. P. Harrington and M. O'Sullivan, "Supremum versions of the Log-rank and generalized Wilcoxon statistics," *J. Amer. Statist. Assoc.*, vol. 82, pp. 312-320, 1987.

T. R. Fleming and D. P. Harrington, *Counting Processes and Survival Analysis.* Wiley, New York, 1991.

R. D. Gill, *Censoring and Stochastic Integrals.* Mathematical Centre Tracts, vol. 124, *Mathematisch Centrum*, Amsterdam, (1980).

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

D. G. Hoel, "A representation of mortality data by competing risks," *Biometrics*, vol. 28, pp. 475- 488, 1972.

S. C. Kochar, "A review of some distribution-free tests for the equality of cause specific hazard rates, " *Analysis of Censored Data - the IMS -LNMS*, vol. 27, pp. 117-162, eds. J. V. Deshpandé and Hira Koul, 1995.

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.

X. Luo and B. W. Turnbull, "Comparing two treatments with multiple competing risks endpoints," *Statistica Sinica*, vol. 9, pp. 985-997, 1999.

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.

P. Yip and K. F. Lam, "A class of non-parametric tests for the equality of failure rates in a competing risks model," *Commun. Statist. 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,” *Commun. Statist. Theory Meth.*, vol. 22, pp. 3199-3222, 1993.

Appendix

The proof of the following theorem follows from Aly, Kochar and McKeague (1994).

Theorem : Let w be a locally bounded predictable non-negative weight function such that $nw^2(u)/\bar{Y}(u) \rightarrow K(u)$ in probability for each u and $\int_0^\infty K(u)d(\Lambda_1 + \Lambda_2)(u) < \infty$. Then under H_0

$$n^{1/2}L_n(t) \xrightarrow{\mathcal{D}} W(\sigma(t))$$

where $\{W(t), t \geq 0\}$ is a standard Brownian motion and $\sigma^2(t) = \int_0^\infty K(u)d(\Lambda_1 + \Lambda_2)(u)$ which can be estimated consistently by $S_n^2(t)$ of (5).

It follows from this and from Gill (pp. 80-81, 1980) that under the conditions of the above theorem and under H_0,

$$\frac{\sqrt{n}L_n(t)}{S_n(\infty)} \xrightarrow{\mathcal{D}} W(t).$$

The asymptotic null distributions as given by (6), (7) and (8) now follow easily from this and the details given in Aly, Kochar and McKeague (1994).

Now we consider the asymptotic null distribution of the statistic $\sqrt{n}C_n^* = \sqrt{n}\ \sup_{0 \leq s < t < \infty} |L_n(t) - L_n(s)|/S_n(\infty)$. Since the statistic $n^{1/2}C_n^*$ converges in distribution to $C^{**} = \sup_{0 \leq s < t \leq 1} |W(t) - W(s)|$ with W being a standard Brownian motion. It is easy to see that C^{**} has the same distribution as the range of the standard Brownian motion $(||W^-|| + ||W^+||)$ where $||W^-|| = \min(0, \inf W(t))$, $||W^+|| = \max(0, \sup W(t))$. The range of the standard Brownian motion was studied by Feller (1951) with density function given by (Eq. (3.6) of Feller (1951) by setting $t = 1$)

$$h(x) = 8 \sum_{k=1}^{\infty} (-1)^{k-1} k^2\ \phi(kx)$$

where ϕ is the density function of a standard normal variable Z. The 95% and 99% quantiles are found to be 2.497 and 3.023, respectively.

Figure 1: Nelson-Aalen estimates of the CSHRs for the two types of cancer.

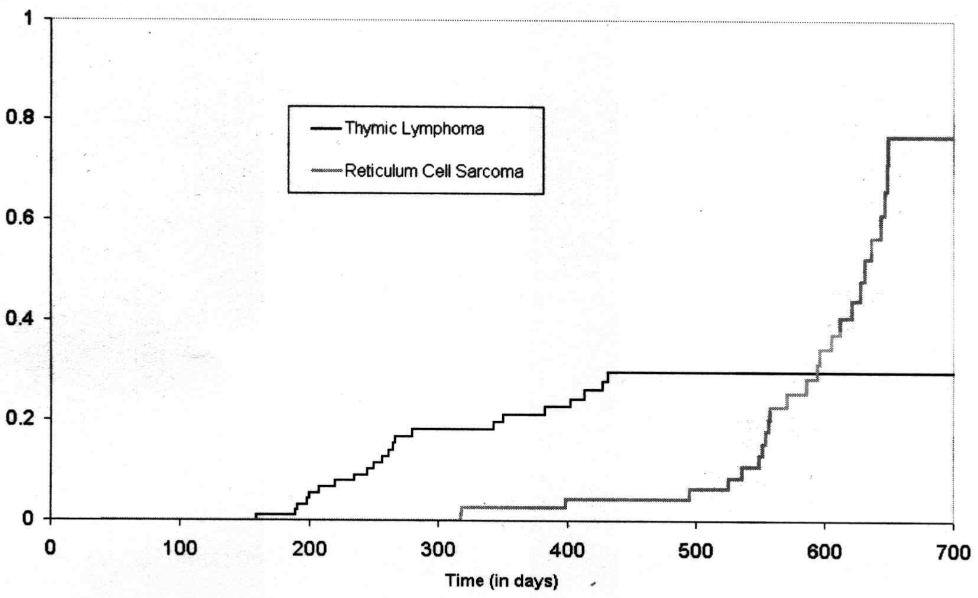

Table 1: Empirical type I error rates and powers of the tests under H_1

	No censoring			18 - 35% censoring			45 - 60% censoring		
Empirical type I error rates ($\beta = 0$)									
$\lambda_0 = 0.0$	A_n	B_n	C_n	A_n	B_n	C_n	A_n	B_n	C_n
(a)	0.0458	0.0460	0.0396	0.0515	0.0447	0.0365	0.0517	0.0418	0.0332
(b)	0.0532	0.0376	0.0314	0.0505	0.0317	0.0269	0.0537	0.0302	0.0216
(c)	0.0530	0.0470	0.0367	0.0515	0.0465	0.0316	0.0495	0.0413	0.0284
(d)	0.0497	0.0416	0.0372	0.0510	0.0419	0.0351	0.0548	0.0423	0.0332
(e)	0.0525	0.0439	0.0404	0.0509	0.0430	0.0368	0.0533	0.0410	0.0347
(f)	-	0.0359	0.0307	-	0.0271	0.0216	-	0.0138	0.0086
$\lambda_0 = 1.0$	A_n	B_n	C_n	A_n	B_n	C_n	A_n	B_n	C_n
(a)	0.0441	0.0480	0.0363	0.0520	0.0443	0.0344	0.0491	0.0408	0.0321
(b)	0.0493	0.0336	0.0292	0.0500	0.0323	0.0268	0.0511	0.0319	0.0230
(c)	0.0478	0.0465	0.0345	0.0531	0.0498	0.0325	0.0493	0.0409	0.0284
(d)	0.0477	0.0395	0.0360	0.0506	0.0402	0.0357	0.0526	0.0398	0.0336
(e)	0.0501	0.0456	0.0395	0.0510	0.0448	0.0362	0.0500	0.0398	0.0344
(f)	-	0.0378	0.0286	-	0.0305	0.0219	-	0.0188	0.0126
Empirical powers ($\beta = 1.0$)									
$\lambda_0 = 0.0$	A_n	B_n	C_n	A_n	B_n	C_n	A_n	B_n	C_n
(a)	0.9850	0.9496	0.9266	0.9034	0.8737	0.8330	0.7714	0.7264	0.6607
(b)	0.7862	0.7894	0.6869	0.6755	0.6587	0.5343	0.5290	0.4865	0.3604
(c)	0.9030	0.8478	0.8184	0.8203	0.7462	0.7024	0.6687	0.5738	0.5108
(d)	0.9072	0.8981	0.8380	0.8165	0.7963	0.7099	0.6743	0.6361	0.5429
(e)	0.9609	0.9428	0.9210	0.8991	0.8732	0.8299	0.7618	0.7240	0.6537
(f)	-	0.9292	0.9000	-	0.8348	0.7755	-	0.6053	0.5076
$\lambda_0 = 1.0$	A_n	B_n	C_n	A_n	B_n	C_n	A_n	B_n	C_n
(a)	0.9536	0.9453	0.9224	0.9219	0.8920	0.8588	0.8167	0.7740	0.7185
(b)	0.7771	0.7850	0.6804	0.6979	0.6870	0.5666	0.5716	0.5318	0.4098
(c)	0.9034	0.8477	0.8196	0.8415	0.7665	0.7239	0.7163	0.6264	0.5666
(d)	0.9020	0.8868	0.8280	0.8415	0.8203	0.7417	0.7228	0.6851	0.5909
(e)	0.9571	0.9377	0.9175	0.9163	0.8907	0.8541	0.8088	0.7709	0.7110
(f)	-	0.9219	0.8946	-	0.8599	0.8136	-	0.6872	0.5986

Table 2: Empirical type I error rates and powers of the tests under H_2

	No censoring			25 - 35% censoring			55 - 60% censoring		
Empirical type I error rates ($\beta = 0$)									
	A_n	B_n	C_n	A_n	B_n	C_n	A_n	B_n	C_n
(a)	0.0422	0.0435	0.0374	0.0506	0.0424	0.0322	0.0472	0.0396	0.0295
(b)	0.0494	0.0329	0.0277	0.0471	0.0295	0.0239	0.0507	0.0282	0.0196
(c)	0.0476	0.0447	0.0331	0.0510	0.0454	0.0310	0.0462	0.0401	0.0262
(d)	0.0477	0.0398	0.0358	0.0473	0.0364	0.0301	0.0489	0.0367	0.0287
(e)	0.0481	0.0425	0.0386	0.0509	0.0423	0.0332	0.0487	0.0380	0.0287
(f)	-	0.0352	0.0284	-	0.0271	0.0181	-	0.0120	0.0066
Empirical powers ($\beta = 1.0$)									
	A_n	B_n	C_n	A_n	B_n	C_n	A_n	B_n	C_n
(a)	0.4509	0.3914	0.4038	0.2731	0.2019	0.2025	0.1322	0.0958	0.0896
(b)	0.5931	0.5285	0.4678	0.3599	0.2829	0.2322	0.1721	0.1155	0.0861
(c)	0.2162	0.1485	0.1395	0.1258	0.0876	0.0720	0.0816	0.0593	0.0425
(d)	0.6023	0.5190	0.4969	0.3578	0.2829	0.2622	0.1690	0.1235	0.1106
(e)	0.4323	0.3466	0.3665	0.2739	0.2059	0.2102	0.1446	0.1084	0.1004
(f)	-	0.3047	0.3141	-	0.1600	0.1472	-	0.0512	0.0366

Table 3: Empirical type I error rates and powers of the tests under H_3

	No censoring			30 - 45% censoring			45 - 70% censoring		
Empirical type I error rates ($\beta = 0$)									
	A_n^*	B_n^*	C_n^*	A_n^*	B_n^*	C_n^*	A_n^*	B_n^*	C_n^*
(a)	0.0553	0.0410	0.0424	0.0540	0.0460	0.0350	0.0471	0.0376	0.0276
(b)	0.0480	0.0332	0.0275	0.0491	0.0298	0.0235	0.0456	0.0217	0.0140
(c)	0.0480	0.0434	0.0305	0.0541	0.0465	0.0285	0.0490	0.0423	0.0239
(d)	0.0466	0.0401	0.0349	0.0482	0.0396	0.0331	0.0473	0.0337	0.0271
(e)	0.0482	0.0432	0.0369	0.0515	0.0453	0.0349	0.0472	0.0336	0.0268
(f)	-	0.0329	0.0274	-	0.0243	0.0175	-	0.0081	0.0032
Empirical powers ($\beta = 1.0$)									
	A_n^*	B_n^*	C_n^*	A_n^*	B_n^*	C_n^*	A_n^*	B_n^*	C_n^*
(a)	0.1105	0.3204	0.2363	0.2559	0.4389	0.2917	0.4157	0.4894	0.3582
(b)	0.2228	0.1639	0.2092	0.0687	0.0386	0.0538	0.0494	0.0313	0.0174
(c)	0.5132	0.6619	0.4854	0.6491	0.6956	0.5398	0.6702	0.6485	0.4978
(d)	0.1333	0.0957	0.1774	0.0516	0.0575	0.0607	0.1156	0.1280	0.0793
(e)	0.1317	0.3752	0.2190	0.2110	0.3677	0.2353	0.2900	0.3547	0.2349
(f)	-	0.3169	0.1750	-	0.2657	0.1485	-	0.1507	0.0755

Table 4: Test statistics (p-values) using different weight functions for the rats data

	$n^{1/2}A_n^*$	$n^{1/2}B_n^*$	$n^{1/2}C_n^*$
(a)	2.0656 (0.0388671)	2.4529 (0.0283428)	4.5185 (0.0000249)
(b)	4.6612 (0.0000031)	4.6612 (0.0000063)	5.0209 (0.0000021)
(c)	-1.4994 (0.1337626)	3.5473 (0.0007782)	3.5473 (0.0015564)
(d)	4.7409 (0.0000021)	4.7409 (0.0000043)	5.5396 (0.0000001)
(e)	2.6433 (0.0082097)	2.6433 (0.0164195)	4.8243 (0.0000056)
(f)	-	2.4316 (0.0300615)	4.4380 (0.0000363)