MA 02142. USA Vol 25: 637–647, 1998

Testing for the Proportionality of Hazards in Two Samples Against the Increasing Cumulative Hazard Ratio Alternative

D. SENGUPTA
Indian Statistical Institute, Calcutta

A. BHATTACHARJEE
Reserve Bank of India, Mumbai

B. RAJEEV
Indian Statistical Institute, Calcutta

ABSTRACT. A number of tests of the proportional hazards hypothesis have been proposed in the past. In recent years, researchers have proposed tests geared specially for the alternative hypothesis of "increasing hazard ratio", keeping in mind the case of crossing hazards. This alternative may be too restrictive in many situations. In this paper we develop a test of the proportional hazards model for the weaker "increasing *cumulative* hazard ratio" alternative. The work is motivated by a data analytic example given by Gill & Schumacher (1987) where their test fails to reject the null hypothesis even though the faster ageing of one group is quite apparent from a plot. The normalized test statistic proposed here has an asymptotically normal distribution under either hypothesis. We also present two graphical methods related to our analytical test.

Key words: counting process, graphical method, hazard ratio, martingale convergence

1. Introduction

The proportional hazards (PH) model has played an instrumental part of data analysis in such areas as survival analysis, reliability, economics, demography and environmental studies. The validity of the PH assumption in a two-sample problem may be checked through one of the traditional graphical methods proposed by Cox (1972), Kay (1977), Andersen (1982), Arjas (1988) etc. (see Sengupta (1995) for a review). Several analytical tests are also available, see Schoenfeld (1980), Andersen *et al.* (1982), Wei (1984), Nagelkerke *et al.* (1984), Breslow *et al.* (1984) and Ciampi & Etezadi-Amoli (1985). Gill & Schumacher (1987) and Deshpande & Sengupta (1995) proposed analytical tests of the PH hypothesis against the alternative of "increasing hazard ratio", which may account for the "crossing hazards" phenomenon.

If F_1 and F_2 are two life distributions on the positive real line with hazard rates λ_1 and λ_2 and cumulative hazard functions Λ_1 and Λ_2, respectively, then the condition λ_1/λ_2 increasing is equivalent to the composition $\Lambda_1 \circ \Lambda_2^{-1}$ being convex on $[0, \infty)$. Using this equivalence, Lee & Pirie (1981) suggested the plotting of an estimator of Λ_1 (e.g. the Nelson–Aalen estimator) against that of Λ_2. It is expected that the graph would be approximately convex when the hazard ratio is increasing, and a straight line when the ratio is constant.

The "increasing hazard ratio" alternative may be too strong in some cases. Consider the situation where the hazard rate λ_2 has jump discontinuities. The ratio λ_1/λ_2 cannot be increasing unless λ_1 also has a jump of adequate size at every point of discontinuity of λ_2. On the other hand, the consistency of an "omnibus" test is not guaranteed. It would be nice to have a test which is consistent for a weaker alternative hypothesis.

A weaker form of relative ageing is represented by the condition "$\Lambda_1 \circ \Lambda_2^{-1}$ is star-shaped", that is, $\Lambda_1 \circ \Lambda_2^{-1}$ intersects any straight line passing through the origin at most once and from below. Convexity is a special case of star-shapedness. Sengupta & Deshpande (1994) showed that the above condition holds if and only if the cumulative hazard ratio (CHR) Λ_1/Λ_2 is an increasing function. Thus, the plot of Λ_1 against Λ_2 is star-shaped if and only if Λ_1/Λ_2 is increasing. The empirical plot of Lee & Pirie (1981) should also be approximately star-shaped when the CHR for the two groups is increasing. Such a phenomenon is indeed observed in the case of the Veterans' Administration data (Detre *et al.*, 1977). The plot given by Gill & Schumacher (1987) (with the coordinates interchanged) is star-shaped, but not convex. Hence, it is not surprising that the analytical tests proposed by Gill & Schumacher (1987) failed to reject the PH hypothesis in favour of the increasing hazard ratio alternative. Perhaps a test designed for the increasing CHR alternative would have been able to reject the PH hypothesis.

In this paper we propose a family of tests for the null and alternative hypotheses

$$\mathscr{H}_0\colon \Lambda_1(t)/\Lambda_2(t) = a \qquad \text{for all } t>0, \text{ for some } a>0,$$

$$\mathscr{H}_1\colon \Lambda_1(t)/\Lambda_2(t) \text{ is a non-constant increasing function of } t \text{ over } [0, \infty).$$

(The word "increasing" would mean "non-decreasing" throughout this paper). The family of statistics presented here are consistent for testing $\mathscr{H}_0$ vs $\mathscr{H}_1$. The asymptotic distribution of a suitably normalized form of the test statistic is standard normal both under $\mathscr{H}_0$ and $\mathscr{H}_1$. While the results are obtained in the general context of comparing two counting processes, the case of censored survival data is given special consideration.

2. Development of the test statistic

Let $N_i(t)$ for $i = 1, 2$ and $t \in [0, \infty)$ represent two components of a bivariate counting process. Let the Doob–Meier decomposition of the processes be of the form

$$dM_i(t) = dN_i(t) - Y_i(t)\,d\Lambda_i(t), \qquad i = 1, 2$$

where $\Lambda_i(\cdot)$, $i = 1, 2$ are deterministic functions on $[0, \infty)$ and $Y_i(\cdot)$, $i = 1, 2$ are non-negative processes which are predictable with respect to the filtration on which the martingales on the left hand side are defined. The above coincides with the "multiplicative intensity" model of the compensator process (see Aalen, 1978). When $N_i(t)$ corresponds to the number of failures or deaths up to time t in the ith group consisting of individuals with i.i.d. life distributions, $\Lambda_i(t)$ is the cumulative hazard rate corresponding to this distribution. In general, $N_i(t)$ may be the number of type i transitions in a Markov chain, $Y_i(t)$ the number at risk for type i transition and $\Lambda_i(t)$ the integrated transition rate.

Under $\mathscr{H}_1$, it is expected that $\Lambda_1(y)\Lambda_2(x) - \Lambda_1(x)\Lambda_2(y)$ would be non-negative for all $x < y$ and positive for some $x < y$. If the ratio Λ_1/Λ_2 is a fast increasing function, the above difference would be generally large. This fact may be used to define a measure of non-proportionality of the cumulative hazard functions,

$$q(w) = \iint_{0<x<y<\tau} w(x, y)[\Lambda_1(y)\Lambda_2(x) - \Lambda_1(x)\Lambda_2(y)]\,dx\,dy,$$

where $w(x, y)$ is a positive weight function and τ is a large positive number such that $\Lambda_j(\tau) < \infty$ for $j = 1, 2$. The idea is similar to that of Deshpande & Sengupta (1995), who considered a measure of non-proportionality of the hazard rates. The double integral may be reduced to products of single integrals by choosing the weight function $w(x, y) = k_1(y)k_2(x) - k_1(x)k_2(y)$, $k_1(\cdot)$ and $k_2(\cdot)$ being positive weight functions with an increasing ratio. With this choice, the above measure simplifies to

$$q(k_1, k_2) = t_{11}t_{22} - t_{12}t_{21}, \tag{2.1}$$

where

$$t_{ij} = \int_0^{\tau} k_i(s)\Lambda_j(s)\,ds, \qquad i = 1, 2, j = 1, 2.$$

Clearly, $q(k_1, k_2)$ is positive under $\mathscr{H}_1$ and zero under $\mathscr{H}_0$. Therefore a consistent estimator of this difference can serve as a test statistic for the problem at hand. Suppose for $j = 1, 2$, $\hat{\Lambda}_j(t)$ be the Nelson–Aalen estimator of $\Lambda_j(t)$ given by $\int_0^t dN_j(s)/Y_j(s)$, where the reciprocal of $Y_j(s)$ is defined to be 0 whenever $Y_j(s)$ is 0. Let $K_i(\cdot)$, $i = 1, 2$ be right-continuous functions with left limits (rcll) converging in probability to $k_i(\cdot)$, $i = 1, 2$, respectively. We define the test statistic as

$$Q_{K_1K_2} = T_{11}T_{22} - T_{12}T_{21},$$

where $T_{ij} = \int_0^{\tau} K_i(s)\hat{\Lambda}_j(s)\,ds$, $i = 1, 2$, $j = 1, 2$. It is shown in the appendix that a consistent estimator of the variance of the test statistic under the null hypothesis is

$$\widehat{\mathrm{var}}(Q_{K_1K_2}) = T_{21}T_{22}V_{11} - T_{21}T_{12}V_{12} - T_{11}T_{22}V_{12} + T_{11}T_{12}V_{22}, \tag{2.2}$$

where

$$V_{ij} = \int_0^{\tau}\int_0^{\tau} K_i(t)K_j(s)V(s \wedge t)\,ds\,dt, \qquad i = 1, 2, j = 1, 2,$$

and

$$V(t) = \int_0^t \frac{dN_1(s) + dN_2(s)}{Y_1(s)Y_2(s)}.$$

Note that the form of $Q_{K_1K_2}$ is similar to the statistic proposed by Gill & Schumacher (1987). In fact, if the cumulative hazard functions are replaced by the corresponding hazard rates, $q(k_1, k_2)$ becomes a measure of non-proportionality of the hazard rates. The family of statistics given by Gill & Schumacher (1987) may be motivated by this measure, although they did not mention it. An important difference between these two families is that the tests proposed here are not functions of the ranks alone; the actual lengths of time between successive jumps are made use of.

The weight functions $K_1(t)$ and $K_2(t)$ may be chosen so that $K_1(t)/K_2(t)$ is an increasing function, in order to make sure that $k_1(t)/k_2(t)$ is increasing. Gill & Schumacher (1987) have indicated several choices of weight functions for their family of statistics. Some of the choices are suitably normalized versions of

$$K_a(t) = Y_1(t)Y_2(t)$$

$$K_b(t) = Y_1(t)Y_2(t)[Y_1(t) + Y_2(t)]^{-1}$$

$$K_c(t) = Y_1(t)Y_2(t)[Y_1(t) + Y_2(t)]^{-1}\hat{S}(t)$$

$$K_d(t) = Y_1(t)Y_2(t)[Y_1(t) + Y_2(t)]^{-1}[\hat{S}(t)]^{1/2}$$

where $\hat{S}(t)$ is the Kaplan–Meier estimator computed from the combined sample. One may choose any pair of weight functions from the above that have an increasing ratio. All these weight functions are predictable, and hence satisfy the conditions of Gill & Schumacher (1987). Being rcll, these may also be used in the test statistic proposed here. In fact, the usable class of weight functions is larger here, because predictability is not required. For instance, one may replace the Kaplan–Meier estimator in the expression of $K_c(t)$ or $K_d(t)$ by a smoothed estimator.

3. Consistency and asymptotic normality

The form of the test statistic $Q_{K_1K_2}$ is similar to that of Gill & Schumacher (1987). However, here T_{ij} is not a stochastic integral but rather an ordinary Stieljes integral of a stochastic process. Therefore we take the following route to obtain the convergence results: (a) we show the convergence of the integral T_{ij} from that of the corresponding integrand (obtained from standard martingale convergence results); (b) subsequently we obtain the convergence of the test statistic by arguing that it is a constant function of the T_{ij}s.

The first step comes from the following theorem.

Theorem 3.1

Let $\mathbf{K}_n$ and $\mathbf{X}_n$ be vector stochastic processes with sample paths in $D[0,\infty)^p$ and $D[0,\infty)^q$ such that $\mathbf{K}_n \xrightarrow{\mathscr{P}} \mathbf{k}$ and $\mathbf{X}_n \xrightarrow{\mathscr{D}} \mathbf{X}$, where $\mathbf{k}$ is a deterministic function in $C[0,\infty)^p$ and $\mathbf{X}$ is a stochastic process with sample paths in $D[0,\infty)^q$. Then for every positive constant τ,

$$\int_0^\tau \mathbf{K}_n(t) \otimes \mathbf{X}_n(t)\,dt \xrightarrow{\mathscr{D}} \int_0^\tau \mathbf{k}(t) \otimes \mathbf{X}(t)\,dt.$$

(In the above, "$\otimes$" indicates the Kronecker product.)

Proof. See the appendix.

In order to study the convergence of T_{ij}, $i = 1, 2, j = 1, 2$, we replace $\mathbf{K}_n(t)$ and $\mathbf{X}_n(t)$ in the above theorem by $(K_1(t){:}\ K_2(t))'$ and a suitably normalized version of $(\hat{\Lambda}_1(t) - \Lambda_1(t){:}\ \hat{\Lambda}_2(t) - \Lambda_2(t))'$, respectively. (Here the prime (′) denotes the transpose of the vector in question.) The latter process can be written as

$$\begin{pmatrix} \hat{\Lambda}_1(t) - \Lambda_1(t) \\ \hat{\Lambda}_2(t) - \Lambda_2(t) \end{pmatrix} = \begin{pmatrix} \int_0^t Y_1^{-1}(s)\,dM_1(s) \\ \int_0^t Y_2^{-1}(s)\,dM_2(s) \end{pmatrix}.$$

We denote this vector martingale by $\mathbf{M}(t)$. Further, let

$$\mathbf{K}(\cdot) = \begin{pmatrix} K_1(\cdot) \\ K_2(\cdot) \end{pmatrix}, \quad \mathbf{k}(\cdot) = \begin{pmatrix} k_1(\cdot) \\ k_2(\cdot) \end{pmatrix}, \quad \boldsymbol{\Lambda}(\cdot) = \begin{pmatrix} \Lambda_1(\cdot) \\ \Lambda_2(\cdot) \end{pmatrix}, \quad \hat{\boldsymbol{\Lambda}}(\cdot) = \begin{pmatrix} \hat{\Lambda}_1(\cdot) \\ \hat{\Lambda}_2(\cdot) \end{pmatrix}$$

where $K_i(\cdot)$, $k_i(\cdot)\Lambda_i(\cdot)$ and $\hat{\Lambda}_i(\cdot)$ for $i = 1, 2$ are as defined in section 2. Finally, let $\mathbf{T} = (T_{11}T_{12}T_{21}T_{22})'$ and $\mathbf{t} = (t_{11}t_{12}t_{21}t_{22})'$. Notice that the dependence of each of these quantities on n is suppressed here for notational simplicity. The convergence of the integral takes place as indicated below.

Corollary 3.2

Suppose there is a positive sequence $\{a_n\}$, approaching infinity as n goes to ∞, such that the following three conditions hold for $j = 1, 2$:

$$a_n \int_0^s \frac{d\Lambda_j(u)}{Y_j(u)} \xrightarrow{\mathscr{P}} \int_0^s \frac{d\Lambda_j(u)}{y_j(u)} \qquad \forall s \in [0, \tau], \tag{3.1}$$

$$a_n \int_0^\tau Y_j^{-1}(u) I\left(\left|\frac{a_n}{Y_j(u)}\right| > \epsilon\right) d\Lambda_j(u) \xrightarrow{\mathscr{P}} 0 \qquad \forall \epsilon > 0, \tag{3.2}$$

$$\sqrt{a_n} \int_0^\tau I(Y_j(u) = 0)\,d\Lambda_j(u) \xrightarrow{\mathscr{P}} 0, \tag{3.3}$$

where y_1^{-1} and y_2^{-1} are bounded on $[0, \tau]$. Then

$$\mathbf{T} = \int_0^\tau \mathbf{K}(t) \cdot \mathbf{\Lambda}(t)\, dt \rightarrow \mathbf{t}, \tag{3.4}$$

$$\hat{\mathbf{T}} = \int_0^\tau \mathbf{K}(t) \cdot \hat{\mathbf{\Lambda}}(t)\, dt \rightarrow \mathbf{t}, \tag{3.5}$$

$$\sqrt{a_n}(\hat{\mathbf{T}} - \mathbf{T}) = \sqrt{a_n}\int_0^\tau \mathbf{K}(t) \cdot \mathbf{M}(t)\, dt \xrightarrow{\mathcal{D}} \int_0^\tau \mathbf{k}(t) \otimes \mathbf{W}(t)\, dt, \tag{3.6}$$

where $\mathbf{W}(\cdot)$ is a vector of two independent Gaussian processes $W_1(\cdot)$ and $W_2(\cdot)$ with zero mean, independent increments and variance function $\int_0^\tau y_i^{-1} d\Lambda_i(s)$, $i = 1, 2$, respectively.

Proof. The definition of $\mathbf{M}(\cdot)$ implies that its components are orthogonal martingales with variation processes $\int_0^t Y_j^{-1}(s)\, d\Lambda_j(s)$, $j = 1, 2$. Therefore the conditions (3.1)–(3.3) ensure, by a version of Rebolledo's martingale central limit theorem (see th. IV.1.2 of Andersen *et al.*, 1992), that

$$\sqrt{a_n}\mathbf{M}(t) \rightarrow \mathbf{W}(t).$$

The results (3.4), (3.5) and (3.6) follow from theorem 3.1 by replacing $\mathbf{X}_n(t)$ with $\mathbf{\Lambda}(t)$, $\hat{\mathbf{\Lambda}}(t)$ and $\sqrt{a_n}\,\mathbf{M}(t)$, respectively.

Remark. The stronger condition

$$\sup_{0 \le t \le \tau}\left|\frac{Y_j(t)}{a_n} - y_j(t)\right| \xrightarrow{p} 0 \text{ as } n \rightarrow \infty, \qquad j = 1, 2 \tag{3.7}$$

implies the conditions (3.1)–(3.3).

The second step in the asymptotic argument is similar to that of Gill & Schumacher (1987). The results (3.5)–(3.6), coupled with the version of the delta-method given by Gill & Schumacher (1985) imply that

$$Q_{K_1K_2} \xrightarrow{p} q(k_1, k_2),$$

$$\sqrt{a_n}(Q_{K_1K_2} - q(K_1, K_2)) \xrightarrow{\mathcal{D}} \int_0^\tau [c(t)W_1(t) - d(t)W_2(t)]\, dt,$$

where

$$c(t) = t_{22}k_1(t) - t_{12}k_2(t),$$

$$d(t) = t_{21}k_1(t) - t_{11}k_2(t).$$

The limiting distribution is therefore Gaussian with zero mean, while the variance is given by

$$\int_0^\tau\int_0^\tau [c(t)c(s)V_1(s \wedge t) + d(t)d(s)V_2(s \wedge t)]\, ds\, dt,$$

where

$$V_i(t) = \int_0^t \frac{d\Lambda_i(s)}{y_i(s)}, \qquad i = 1, 2.$$

Under the null hypothesis, the ratio $\Lambda_2(\cdot)/\Lambda_1(\cdot)$ is a constant θ, which can also be called the hazard ratio. Further, $c(\cdot)/d(\cdot)$ is also equal to θ under $\mathscr{H}_0$. Thus an alternative expression for the asymptotic null variance is

$$\operatorname{var}(\sqrt{a_n}Q_{K_1K_2}) = \int_0^\tau \int_0^\tau c(t)d(s)[\theta V_1(s\wedge t) + \theta^{-1}V_2(s\wedge t)]\,ds\,dt$$

$$= \int_0^\tau \int_0^\tau c(t)d(s) \int_0^{s\wedge t} \left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)}\right] ds\,dt$$

$$= t_{21}t_{22}v_{11} - t_{21}t_{12}v_{12} - t_{11}t_{22}v_{12} + t_{11}t_{12}v_{22},$$

where

$$v_{ij} = \int_0^\tau \int_0^\tau k_i(t)k_j(s) \int_0^{s\wedge t} \left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)}\right] ds\,dt, \qquad i = 1, 2,\ j = 1, 2.$$

This variance is estimated consistently by a_n times the expression given in (2.2), as shown in the appendix, provided $Y_i(t)/a_n \xrightarrow{\mathscr{P}} y_i(t)$ pointwise on $[0, \tau]$. Since $q(K_1, K_2) = 0$ is zero under $\mathscr{H}_0$ and positive under $\mathscr{H}_1$, the normalized statistic can be used for a one-sided test.

4. Graphical methods

The following three graphical procedures are of special interest here:

(a) the plot of $\hat{\Lambda}_1(t)$ vs $\hat{\Lambda}_2(t)$, proposed by Lee & Pirie (1981),
(b) the plot of $(\hat{\Lambda}_1(t) - \hat{\Lambda}_2(t))/\hat{\Lambda}_2(t)$ vs t, due to Dabrowska *et al.* (1989) and
(c) the plot of the log cumulative hazard difference $\log(\hat{\Lambda}_1(t)) - \log(\hat{\Lambda}_2(t))$ against t, suggested by Dabrowska *et al.* (1992).

A monotone trend in any of the last two plots suggests a monotone CHR of the two samples, while no trend corresponds to the PH model. Plot (a) is expected to be close to a straight line in the PH case and star-shaped when the CHR is (monotone) increasing. Thus, all the three plots are expected to bring out monotone CHR-type departures from the PH model, although they have so far been used to look for monotone hazard ratio.

The above plots can be quite unstable. Plots (b) and (c) can have wild fluctuations for small values of t (see Dabrowska *et al.*, 1989), while plot (a) may lack precision for large values of t. GS suggested a modification of plot (a), replacing $\hat{\Lambda}_i(t)$ with $\hat{\Lambda}_i^K(t) = \int_0^t K(s)\,d\hat{\Lambda}_i(s)$, $i = 1, 2$, where $K(\cdot)$ is a predictable weight function (see section 2). This modification can also be used in plots (b) and (c). The modified plots have the same characteristic features when the hazard ratio is constant or monotone, but such a feature no longer exists for monotone CHR.

To overcome this problem, we propose two graphical tests based on the estimated functions $T_i^K(t) = \int_0^t K(s)\hat{\Lambda}_i(s)\,ds$, $i = 1, 2$, where $K(\cdot)$ is a rcll weight function. The plot of $T_1^K(t)/T_2^K(t)$ against t is expected to be like a horizontal straight line when the PH model holds. On the other hand, a monotone ratio of the cumulative hazards of the two populations is expected to produce a monotone trend in the plot, irrespective of the choice of the weight function. Since $\hat{\theta}_K = T_1^K(\tau)/T_2^K(\tau)$ is a consistent estimator of the hazard ratio in the PH case, the horizontal straight line passing through the right end-point of the graph serves as a reference corresponding to the PH hypothesis.

The other suggested plot is that of $T_1^K(\cdot)$ against $T_2^K(\cdot)$. This graph is expected to be close to a straight line when the PH model holds and approximately convex or concave when the CHR is monotone. The straight line joining the origin with the end-point of the graph $(T_2^K(\tau), T_1^K(\tau))$, may serve as a reference for the PH hypothesis. The two suggested plots are expected to be smoother and more stable than their unweighted counterparts.

5. Data analysis

The analytic and graphical procedures proposed in sections 2 and 4 were used to analyse the ovarian cancer data set reported by Fleming *et al.* (1980), which describes the number of days from treatment to progression of disease. Here, groups 1 and 2 consist of 20 patients with high-grade tumor (stage IIA) and 15 patients with low-grade tumor (stage II), respectively. The statistic Q_{K_b,K_a} (after normalization) is 2.258. The corresponding two-sided p-value is 0.024, suggesting an increasing trend of the ratio $\Lambda_1(t)/\Lambda_2(t)$. This supports the findings of Gill & Schumacher (1987) and Deshpande & Sengupta (1995) that the hazard ratio is increasing.

The plot of $\Lambda_1(t)/\Lambda_2(t)$ vs. t, shown in Fig. 1 has by and large an increasing trend, but the fluctuations are substantial. Figure 2 shows the plot of $T_1^{K_b}(t)/T_2^{K_b}(t)$ against t which was suggested in section 4. This graph is smoother and more clearly suggestive of an increasing trend of the CHR.

The plot of $\Lambda_1^{K_b}(t)$ vs. $\Lambda_2^{K_b}(t)$ shown in Fig. 3 is approximately convex, indicating an increasing hazard ratio. However, the plot of $T_1^{K_b}(t)$ against $T_2^{K_b}(t)$ shown in Fig. 4 is smoother and clearly convex, suggesting an increasing CHR.

6. Concluding remarks

The role of the weight functions in the family of tests proposed here is crucial. An interesting question that can be posed in this connection is: "Can the weight functions be chosen 'optimally' according to some chosen criterion?" We have no clear answer to this question as yet. If a sequence of alternative hypotheses converging to $\mathscr{H}_0$ at a suitable rate is considered, it can be shown that the asymptotic relative efficiency is of the form $[\int_0^\tau l(t)g(t)\,dt]^2 / \int_0^\tau \int_0^\tau l(s)W(t,s)\,ds\,dt$, where $l(t)$ is the probability limit of the ratio of the weight functions, $g(t)$ is a function determined by $\Lambda_1(t)$ and $\Lambda_2(t)$, and $W(t, S)$ is a positive definite function of two variables, also determined by $\Lambda_1(t)$ and $\Lambda_2(t)$. A function

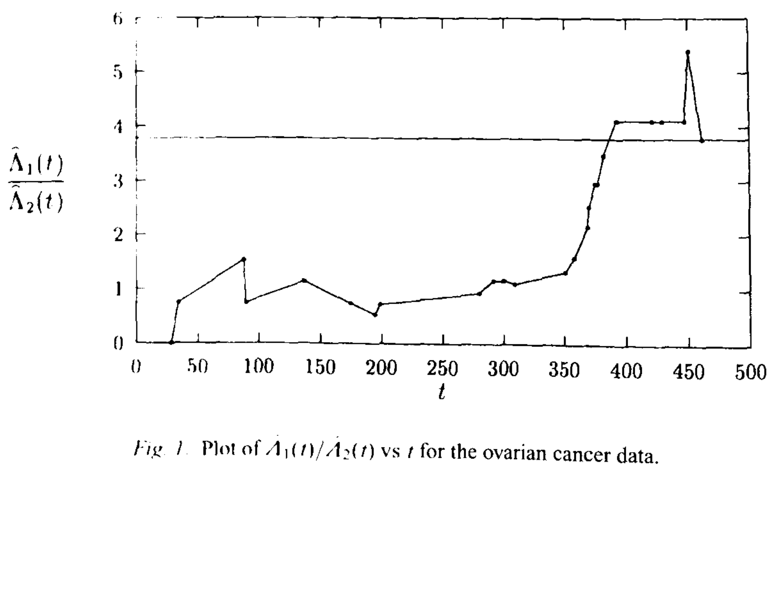

Fig. 1. Plot of $\Lambda_1(t)/\Lambda_2(t)$ vs t for the ovarian cancer data.

Under the null hypothesis, the ratio $\Lambda_2(\cdot)/\Lambda_1(\cdot)$ is a constant θ, which can also be called the hazard ratio. Further, $c(\cdot)/d(\cdot)$ is also equal to θ under $\mathscr{H}_0$. Thus an alternative expression for the asymptotic null variance is

$$\operatorname{var}(\sqrt{a_n}Q_{K_1K_2}) = \int_0^\tau\int_0^\tau c(t)d(s)[\theta V_1(s\wedge t) + \theta^{-1}V_2(s\wedge t)]\,ds\,dt$$

$$= \int_0^\tau\int_0^\tau c(t)d(s)\int_0^{s\wedge t}\left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)}\right]ds\,dt$$

$$= t_{21}t_{22}v_{11} - t_{21}t_{12}v_{12} - t_{11}t_{22}v_{12} + t_{11}t_{12}v_{22},$$

where

$$v_{ij} = \int_0^\tau\int_0^\tau k_i(t)k_j(s)\int_0^{s\wedge t}\left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)}\right]ds\,dt, \qquad i = 1, 2, j = 1, 2.$$

This variance is estimated consistently by a_n times the expression given in (2.2), as shown in the appendix, provided $Y_i(t)/a_n \xrightarrow{\mathscr{P}} y_i(t)$ pointwise on $[0, \tau]$. Since $q(K_1, K_2) = 0$ is zero under $\mathscr{H}_0$ and positive under $\mathscr{H}_1$, the normalized statistic can be used for a one-sided test.

4. Graphical methods

The following three graphical procedures are of special interest here:

(a) the plot of $\hat{\Lambda}_1(t)$ vs $\hat{\Lambda}_2(t)$, proposed by Lee & Pirie (1981),
(b) the plot of $(\hat{\Lambda}_1(t) - \hat{\Lambda}_2(t))/\hat{\Lambda}_2(t)$ vs t, due to Dabrowska *et al.* (1989) and
(c) the plot of the log cumulative hazard difference $\log(\hat{\Lambda}_1(t)) - \log(\hat{\Lambda}_2(t))$ against t, suggested by Dabrowska *et al.* (1992).

A monotone trend in any of the last two plots suggests a monotone CHR of the two samples, while no trend corresponds to the PH model. Plot (a) is expected to be close to a straight line in the PH case and star-shaped when the CHR is (monotone) increasing. Thus, all the three plots are expected to bring out monotone CHR-type departures from the PH model, although they have so far been used to look for monotone hazard ratio.

The above plots can be quite unstable. Plots (b) and (c) can have wild fluctuations for small values of t (see Dabrowska *et al.*, 1989), while plot (a) may lack precision for large values of t. GS suggested a modification of plot (a), replacing $\hat{\Lambda}_i(t)$ with $\hat{\Lambda}_i^K(t) = \int_0^t K(s)\,d\hat{\Lambda}_i(s)$, $i = 1, 2$, where $K(\cdot)$ is a predictable weight function (see section 2). This modification can also be used in plots (b) and (c). The modified plots have the same characteristic features when the hazard ratio is constant or monotone, but such a feature no longer exists for monotone CHR.

To overcome this problem, we propose two graphical tests based on the estimated functions $T_i^K(t) = \int_0^t K(s)\hat{\Lambda}_i(s)\,ds$, $i = 1, 2$, where $K(\cdot)$ is a rcll weight function. The plot of $T_1^K(t)/T_2^K(t)$ against t is expected to be like a horizontal straight line when the PH model holds. On the other hand, a monotone ratio of the cumulative hazards of the two populations is expected to produce a monotone trend in the plot, irrespective of the choice of the weight function. Since $\hat{\theta}_K = T_1^K(\tau)/T_2^K(\tau)$ is a consistent estimator of the hazard ratio in the PH case, the horizontal straight line passing through the right end-point of the graph serves as a reference corresponding to the PH hypothesis.

The other suggested plot is that of $T_1^K(\cdot)$ against $T_2^K(\cdot)$. This graph is expected to be close to a straight line when the PH model holds and approximately convex or concave when the CHR is monotone. The straight line joining the origin with the end-point of the graph $(T_2^K(\tau), T_1^K(\tau))$, may serve as a reference for the PH hypothesis. The two suggested plots are expected to be smoother and more stable than their unweighted counterparts.

5. Data analysis

The analytic and graphical procedures proposed in sections 2 and 4 were used to analyse the ovarian cancer data set reported by Fleming *et al.* (1980), which describes the number of days from treatment to progression of disease. Here, groups 1 and 2 consist of 20 patients with high-grade tumor (stage IIA) and 15 patients with low-grade tumor (stage II), respectively. The statistic Q_{K_h,K_a} (after normalization) is 2.258. The corresponding two-sided p-value is 0.024, suggesting an increasing trend of the ratio $\Lambda_1(t)/\Lambda_2(t)$. This supports the findings of Gill & Schumacher (1987) and Deshpande & Sengupta (1995) that the hazard ratio is increasing.

The plot of $\hat{\Lambda}_1(t)/\hat{\Lambda}_2(t)$ vs. t, shown in Fig. 1 has by and large an increasing trend, but the fluctuations are substantial. Figure 2 shows the plot of $T_1^{K_h}(t)/T_2^{K_h}(t)$ against t which was suggested in section 4. This graph is smoother and more clearly suggestive of an increasing trend of the CHR.

The plot of $\hat{\Lambda}_1^{K_h}(t)$ vs. $\hat{\Lambda}_2^{K_h}(t)$ shown in Fig. 3 is approximately convex, indicating an increasing hazard ratio. However, the plot of $T_1^{K_h}(t)$ against $T_2^{K_h}(t)$ shown in Fig. 4 is smoother and clearly convex, suggesting an increasing CHR.

6. Concluding remarks

The role of the weight functions in the family of tests proposed here is crucial. An interesting question that can be posed in this connection is: "Can the weight functions be chosen 'optimally' according to some chosen criterion?" We have no clear answer to this question as yet. If a sequence of alternative hypotheses converging to $\mathscr{H}_0$ at a suitable rate is considered, it can be shown that the asymptotic relative efficiency is of the form $[\int_0^\tau l(t)g(t)\,dt]^2/\int_0^\tau\int_0^\tau l(s)W(t,s)\,ds\,dt$, where $l(t)$ is the probability limit of the ratio of the weight functions, $g(t)$ is a function determined by $\Lambda_1(t)$ and $\Lambda_2(t)$, and $W(t, S)$ is a positive definite function of two variables, also determined by $\Lambda_1(t)$ and $\Lambda_2(t)$. A function

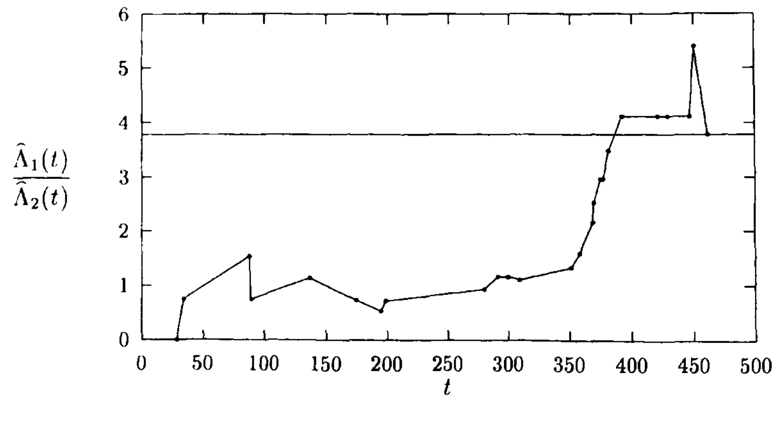

Fig. 1. Plot of $\hat{\Lambda}_1(t)/\hat{\Lambda}_2(t)$ vs t for the ovarian cancer data.

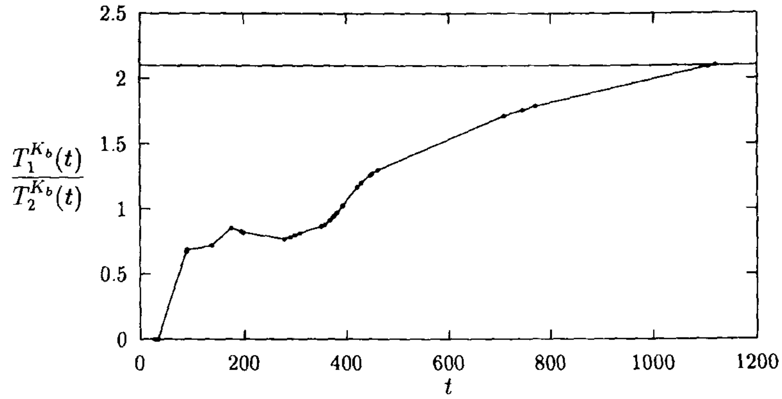

Fig. 2. Plot of $T_1^{K_b}(t)/T_2^{K_b}(t)$ vs t for the ovarian cancer data.

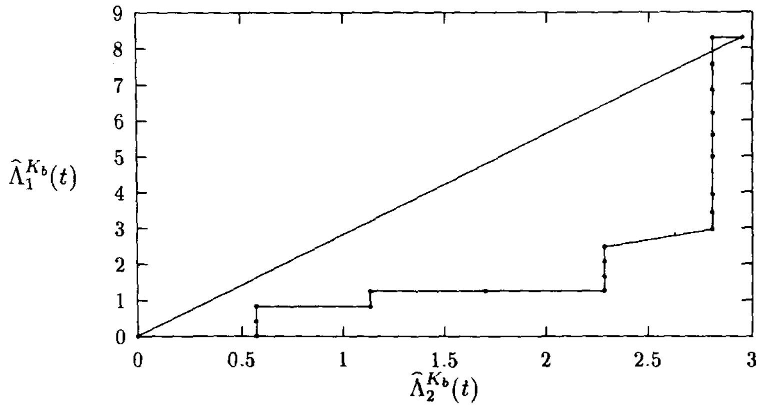

Fig. 3. Plot of $\hat{\Lambda}_1^{K_b}(t)$ vs $\hat{\Lambda}_2^{K_b}(t)$ for the ovarian cancer data.

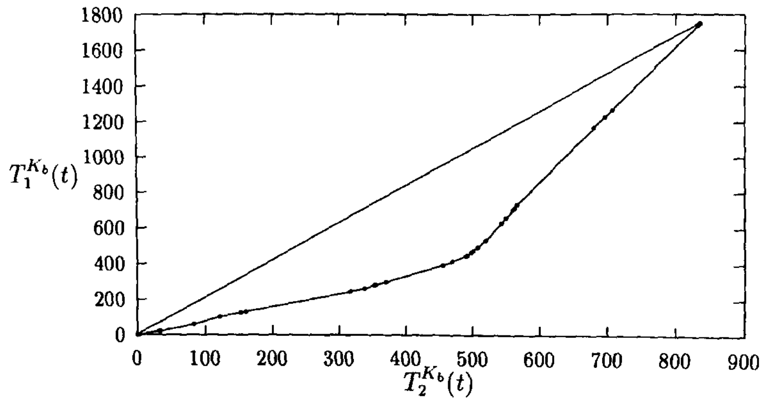

Fig. 4. Plot of $T_1^{K_b}(t)$ vs $T_2^{K_b}(t)$ for the ovarian cancer data.

$l(t)$ that maximizes this expression would lead to a suitable weight function. Unfortunately a closed form solution to this problem is not available. This is in contrast to the similar problem handled by Gill & Schumacher (1987), where the "optimal" solution could be obtained in closed form through the Cauchy–Schwartz inequality.

A small-scale simulation was performed to explore the role of the weight functions in the two-sample testing problem. The two samples were generated from an exponential distribution and a piecewise exponential distribution, respectively. Several combinations of weight functions were tried out. Out of these, the combination $Y_1(t)Y_2(t)\exp[-t/T_n]$ and $Y_1(t)Y_2(t)$, where T_n is the total time on test statistic for the combined sample, yielded the highest power. The former weight function could not have been used for the family of tests proposed by Gill & Schumacher (1987), since it is not predictable. This underscores the wide scope of the class of rcll weight functions considered here.

The analytical test proposed in section 2 can easily be adapted to the competing risks situation, where the hazard rates under consideration are the cause-specific hazard rates for two risks. The presence of other risks can also be accommodated.

The test can be generalized in two ways. First, the effect of covariates can be taken into consideration in a manner similar to Breslow (1974) and Dabrowska *et al.* (1992). The null hypothesis would then be equivalent to checking the proportionality of the effect of a binary covariate (such as a group indicator or a discretized covariate), assuming the other covariate effects to be proportional. (An extension to the Cox regression model with continuous covariates along the lines of Lin (1991) may not be possible). The other generalization may involve the cumulative γ-rate functions considered by Dabrowska *et al.* (1989), which includes as a special case the cumulative hazard function and the odds ratio function.

A nice feature of the graphical methods suggested here is that they produce smooth plots, even for small sample sizes. Thus the user need not be wary of reading too much from the shape of the plot.

References

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

Andersen, P. K. (1982). Testing for the goodness-of-fit of Cox's regression model. *Biometrics* **38**, 67–77.

Andersen, P. K., Borgan, O., Gill, R. & Kieding, N. (1982). Linear nonparametric tests for comparison of counting processes, with application to censored data. *Internat. Statist. Rev.* **50**, 219–258.

Andersen, P. K., Borgan, O., Gill, R. & Kieding, N. (1992). *Statistical models based on counting processes.* Springer-Verlag, New York.

Arjas, E. (1988). A graphical method for assessing goodness of fit of Cox's proportional hazard model. *J. Amer. Statist. Assoc.* **83**, 204–212.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Breslow, N. E., Edler, L. & Berger, J. (1984). A two-sample censored data rank test for acceleration. *Biometrics* **40**, 1049–1062.

Ciampi, A. & Etezadi-Amoli, J. (1985). A general model for testing the proportional hazards and accelerated failure time hypothesis in the analysis of censored survival data with covariates. *Comm. Statist. Theory Methods* **14**, 651–676.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.

Dabrowska, D. M., Doksum, K. A. & Song, J.-K. (1989) Graphical comparison of cumulative hazards for two populations. *Biometrika* **76**, 763–773.

Dabrowska, D. M., Doksum, K. A., Feduska, N. J., Husing, R. & Neville, P. (1992). Methods for comparing cumulative hazard functions in a semi-proportional hazard model. *Statist. Med.* **11**, 1465–1476.

Deshpande, J. V. & Sengupta, D. (1995). Testing for the hypothesis of proportional hazards in two populations. *Biometrika* **82**, 251–261.

Detre, K. M., Hultgren, H. & Takaro, T. (1977). Veterans' Administration cooperative study of surgery for

coronary arterial occlusive disease, III: methods and baseline characteristics including experience with medical treatment. *Amer. J. Cardiol.* **40**, 212–225.
Fleming, T. R., O'Fallon, J. R., O'Brien, P. C. & Harrington, D. P. (1980). Modified Kolmogorov–Smirnov test procedures with application to arbitrarily censored data. *Biometrics* **36**, 607–625.
Gill, R. & Schumacher, M. (1985). A simple test of the proportional hazards assumption. *Tech. Rep. No. MS-R8504*, Department of Mathematics and Statistics, Centrum voor Wiskunde en Informatica, Amsterdam.
Gill, R. & Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika* **74**, 289–300.
Jacod, J. & Shiryayev, A. N. (1980). *Limit theorems for stochastic processes*. Springer-Verlag, New York.
Kay, R. (1977). Proportional hazards regression models and the analysis of censored survival data. *Appl. Statist.* **26**, 227–237.
Lee, L. & Pirie, W. E. (1981). A graphical method for comparing trends in series of events. *Comm. Statist. Theory Methods* **10**, 827–848.
Lin, D. Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *J. Amer Statist. Assoc.* **86**, 725–728.
Nagelkerke, N. J., Oosting, J. & Hart, A. A. M. (1984). A simple test for goodness-of-fit of Cox's proportional hazards model. *Biometrics* **40**, 483–486.
Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* **67**, 145–153.
Sengupta, D. (1995). Graphical tools for censored survival data. In *Analysis of censored data* (eds H. L. Kaal & J. V. Deshpande) 193–217. Institute of Mathematical Statistics, Hayward, CA.
Sengupta, D. & Deshpande, J. V. (1994). Some results on the relative aging of life distributions. *J. Appl. Probab.* **31**, 991–1003.
Wei, L. J. (1984). Testing goodness-of-fit for proportional hazards model with censored observations. *J. Amer. Statist. Assoc.* **79**, 649–652.

Appendix

Proof of theorem 3.1

Consider the function h: $D[0, \infty)^p \times D[0, \infty)^q \to D[0, \infty)^{pq}$ defined by $h(\mathbf{k}, \mathbf{x})(t) = \mathbf{k}(t) \otimes \mathbf{x}(t)$. It is easy to show that h is continuous at all points $(\mathbf{k}, \mathbf{x})$ such that $\mathbf{k}$ is rcll and $\mathbf{x}$ is continuous. The probability that $(\mathbf{k}, \mathbf{X})$ does not belong to the continuity set of h is the same as the probability that $\mathbf{X}$ does not belong to $D[0, \infty)^q - C[0, \infty)^p$. The assumptions of the theorem ensures that this probability is zero. Therefore $\mathbf{K}_n(\cdot) \otimes \mathbf{X}_n(\cdot) \xrightarrow{\mathscr{D}} \mathbf{k}(\cdot) \otimes \mathbf{X}(\cdot)$ by virtue of the continuous mapping theorem.

Now consider the function f: $D[0, \infty)^{pq} \to \mathbb{R}^{pq}$ defined by $f(\mathbf{x}) = \int_0^\tau \mathbf{x}(t)\, dt$. To show that f is continuous, let $\mathbf{x}_n \to \mathbf{x}$ in $D[0, \infty)^{pq}$ and notice that every component of $f(\mathbf{x}_n)$ converges to the corresponding component of $f(\mathbf{x})$ by the dominated convergence theorem. Since the domain and range of f are spaces equipped with product topologies, this implies that $f(\mathbf{x}_n)$ converges to $f(\mathbf{x})$. Therefore f is continuous and the result of the theorem follows from the continuous mapping theorem.

Consistency of the variance estimator (2.2)

Assuming that $Y_i(t)/a_n \xrightarrow{\mathscr{P}} y_i(t)$ for $i = 1, 2$ pointwise on $[0, \tau]$, we have $a_n V(t) \xrightarrow{\mathscr{P}} v(t)$ in $D[0, \infty)$ under the usual Rebolledo conditions, where

$$v(t) = \int_0^t \left[\frac{dA_2(u)}{y_1(u)} + \frac{dA_1(u)}{y_2(u)}\right].$$

Let us also assume that $K_i \xrightarrow{\mathscr{P}} k_i$ for $i = 1, 2$, and that each of the functions v, k_1 and k_2 is continuous. In view of (3.5), we only have to show that $a_n V_{ij} \xrightarrow{\mathscr{P}} v_{ij}$, $i = 1, 2$, $j = 1, 2$.

We write v_{ij} as $\psi(k_i, \phi(k_j, v))$, where ψ and ϕ are functions from $D[0, \infty) \times D[0, \infty)$ to $\mathbb{R}$ and $D[0, \infty)$, respectively, defined as

$$\psi(k, l) = \int_0^\tau k(s) l(s)\, ds,$$

$$\phi(k, l)(t) = \int_0^\tau k(s) l(s \wedge t)\, ds.$$

In such a case $a_n V_{ij} = \psi(K_i, \phi(K_j, a_n V))$. The convergence of $a_n V_{ij}$ to v_{ij} in distribution is proved by showing that $\phi(K_j, a_n V) \rightarrow \phi(k_j, v)$. Since the limit of convergence in either step is deterministic, we can invoke the continuous mapping theorem and show that the functions ϕ and ψ are continuous at the limit points. To show the continuity of ϕ, let (k_{jn}, v_n) be a sequence in $D[0, \infty) \times D[0, \infty)$ converging to (k_j, v). Thus $k_{jn} \rightarrow k_j$ and $v_n \rightarrow v$ in $D[0, \infty)$. Since k_j and v are assumed to be continuous, prop. 1.17(b) of Jacod & Shiryayev (1980, p. 292) ensures that for each t $\sup_{s \leq t} |k_{jn}(s) - k_j(s)| \rightarrow 0$ and $\sup_{s \leq t} |v_n(s) - v(s)| \rightarrow 0$. Note that $\phi(k, v) \in C[0, \infty)$. It follows that for $s \in [0, \tau]$,

$$\begin{aligned} &|\phi(k_{jn}, v_n)(s) - \phi(k_j, v)(s)| \\ &= \left| \int_0^\tau [k_{jn}(t)(v_n(s \wedge t) - v(s \wedge t)) + v(s \wedge t)(k_{jn}(t) - k_j(t))]\, dt \right| \\ &\leq \sup_{s \in [0,\tau]} |v_n(s) - v(s)| \cdot \int_0^\tau |k_{jn}(t)| dt + \tau \cdot \sup_{s \in [0,\tau]} |v(s)| \cdot \sup_{s \in [0,\tau]} |k_{jn}(s) - k_j(s)|. \end{aligned}$$

Thus $\phi(k_{jn}, v_n)$ converges to $\phi(k_j, v)$ locally uniformly. Therefore $\phi(k_{jn}, v_n)$ converges to $\phi(k_j, v)$ in $D[0, \infty)$, and ϕ is continuous at (k_j, v). The continuity of ψ at $(k_i, \phi(k_j, v))$ is proved in a similar manner.

Received June 1996, in final form January 1998

Debasis Sengupta, Applied Statistics Unit, Indian Statistical Institute, 203, B.T. Road, Calcutta 700 035, India.