Ann. Inst. Statist. Math.
Vol. 50, No. 2, 311–324 (1998)

SIMPLE-TREE WEIGHTED LOGRANK TESTS FOR RIGHT-CENSORED DATA

YUH-ING CHEN

Graduate Institute of Statistics, National Central University, Chung-Li, Taiwan 320, R.O.C.

(Received May 27, 1996; revised April 15, 1997)

Abstract. In this paper we are concerned with the problem of testing against the simple-tree alternative that there is at least one treatment more effective than the control when data are subject to random right-censorship. A class of tests based on linear combinations of two-sample weighted logrank statistics each comparing an individual treatment with the control is proposed. Asymptotic relative efficiencies of the simple-tree versions of Gehan-Wilcoxon, logrank and Peto-Prentice-Wilcoxon under Lehmann and scale alternatives are evaluated for various combinations of survival distributions and censoring probabilities. The results of a Monte Carlo level and power study are presented. An illustrated numerical example is also reported.

Key words and phrases: Asymptotic relative efficiency, right-censored data, simple-tree alternative.

1. Introduction

The problem of comparing several treatment groups with a control group occurs frequently in survival data analyses. For example, in comparative clinical trials, different therapies are often compared with a standard therapy or placebo in the prolongation of the survival time of the patient with a certain disease. In these cases, randomly right-censored data are often available, since subjects who randomly enter the study to take therapies under consideration may be lost to follow-up randomly or the study may be terminated at a preassigned time owing to time limitation.

For the i-th sample ($i = 0, 1, \ldots, k$), let $T_{i1}, \ldots, T_{in_i}$ be independent identically distributed (i.i.d.) random variables each with a continuous distribution function F_i, and $C_{i1}, \ldots, C_{in_i}$ be i.i.d. random variables each with a continuous distribution function G_i, where C_{iu} is the censoring time associated with the survival time T_{iu}. Suppose that the zero population ($i = 0$) is the control and the other k populations are treatments. Furthermore, assume that the $k + 1$ samples are independent of each other and the C_{iu} are distributed independently of the T_{iu}. In such a setting, we actually only observe the bivariate vectors (X_{iu}, δ_{iu}), where $X_{iu} = \min(T_{iu}, C_{iu})$, $\delta_{iu} = 1$, if $X_{iu} = T_{iu}$, and 0 otherwise.

Let $S_i = 1 - F_i$, $i = 0, 1, \ldots, k$. Suppose that the treatments are at least equivalent to the control and that a higher response corresponds to a better treatment effect. In this paper, specifically, we are concerned with testing the null hypothesis H_0: $(S_i = S_0,\ i = 1, 2, \ldots, k)$ against the simple-tree alternatives (Barlow *et al.* (1972)) H_1: $(S_i \geq S_0$ with strict inequality for at least one $i,\ i = 1, 2, \ldots, k)$ when the randomly right-censored data are involved. The problem of deciding treatments (if any) which are more effective than the control is also considered.

For the setting where data are subject to unequal patterns of censorship, Chakraborti and Desu (1991) based on the Gehan-Wilcoxon (Gehan (1965)) score and suggested a generalization of the Fligner-Wolfe (1982) test for the simple-tree alternative. They also proposed, according to Slepian's (1962) inequality, a multiple test based on two-sample Gehan-Wilcoxon statistics in determining which treatments are more effective than the control. However, the logrank statistic (Mantel (1966)) is probably the most commonly used two-sample test statistic and Gehan's generalized Wilcoxon statistic is a member of the general class of weighted logrank statistics (Tarone and Ware (1977)). Therefore, we consider in this paper generalizations of the Fligner-Wolfe test on the basis of weighted logrank statistics for the simple-tree alternatives. Multiple tests based on two-sample weighted logrank statistics are also suggested.

A class of tests based on linear combinations of two-sample weighted logrank statistics each comparing an individual treatment with the control is proposed. Three special simple-tree tests based on the Gehan-Wilcoxon, logrank and Peto-Prentice-Wilcoxon (Peto and Peto (1972), Prentice (1978)) statistics are investigated in detail. A numerical example (King *et al.* (1979)) studying the effect of diets on the development of tumors is illustrated. The Pitman efficacies of the simple-tree tests under Lehmann and scale alternatives are calculated. The optimal sample size allocation in the sense of maximizing the efficacies is then obtained. The asymptotic relative efficiencies (ARE) among these tests for Weibull and lognormal distributions are evaluated and the effect of censoring on ARE is explored. The results of a Monte Carlo simulation investigating the level and power performances of the simple-tree tests for small and moderate sample sizes are presented.

2. The proposed tests

For $i = 0, 1, \ldots, k$, let $D_i(t)$ be the number of patients in group i who have been observed to die by time t and $Y_i(t)$ the number of patients in group i who are still alive and uncensored at time t. In the area of martingale based analysis of censored data the two-sample weighted logrank statistic, for comparing the i-th treatment with the control, is written as

$$U_{0i} = \int_0^\infty K_{0i}(t) d\{\hat{\Lambda}_0(t) - \hat{\Lambda}_i(t)\}, \tag{2.1}$$

where $K_{0i}(t) = W_{0i}(t) Y_0(t) Y_i(t) / \{Y_0(t) + Y_i(t)\}$ and $\hat{\Lambda}_i(t) = \int_0^t dD_i(s)/Y_i(s)$ is Nelson's (1969) estimator of the cumulative hazard function of group i, $\Lambda_i(t)$. We consider in this paper three special cases of the weighted logrank statistics

which are of general interest: the Gehan-Wilcoxon (Gehan (1965), Prentice and Marek (1979)) statistic when $W_{0i}(t) = \{Y_0(t) + Y_i(t)\}/(n_0 + n_i)$, the logrank statistic (Mantel (1966)) when $W_{0i}(t) = 1$, and the Peto-Prentice-Wilcoxon (Peto and Peto (1972), Prentice (1978)) statistic when $W_{0i}(t) = \hat{S}_{0i}(t)$, where $\hat{S}_{0i}(t)$ is the Kaplan-Meier (1958) estimator in the combined samples of 0 and i. Note that the Chakraborti-Desu (1991) overall test considers only the Gehan-Wilcoxon two-sample statistic.

Remark 1. For the Gehan-Wilcoxon statistic, equation (2.1) is

$$U_{0i} = \frac{1}{n_0 + n_i} \left\{ \int_0^\infty Y_i(t) dD_0(t) - \int_0^\infty Y_0(t) dD_i(t) \right\}. \tag{2.2}$$

In comparing the i-th treatment group with the control group, we obtain (X_{iu}, δ_{iu}), $u = 1, \ldots, n_i$, and (X_{0v}, δ_{0v}), $v = 1, \ldots, n_0$. Therefore,

$$\int_0^\infty Y_i(t) dD_0(t) = \sum_{u=1}^{n_i} \sum_{v=1}^{n_0} I(X_{iu} > X_{0v}) I(\delta_{0v} = 1)$$

and, similarly,

$$\int_0^\infty Y_0(t) dD_i(t) = \sum_{u=1}^{n_i} \sum_{v=1}^{n_0} I(X_{0v} > X_{iu}) I(\delta_{iu} = 1),$$

where $I(s) = 1$, if statement s holds, and 0 otherwise. Hence, the counting process formulation of the Gehan-Wilcoxon statistic in (2.2) can be reduced to the more familiar expression (Gehan (1965)) as

$$U_{0i} = \frac{1}{n_0 + n_i} \sum_{u=1}^{n_i} \sum_{v=1}^{n_0} \phi(X_{iu}, \delta_{iu}; X_{0v}, \delta_{0v}), \tag{2.3}$$

where

$$\phi(X_{iu}, \delta_{iu}; X_{0v}, \delta_{0v}) = \begin{cases} +1 & \text{if } X_{iu} > X_{0v} \text{ and } \delta_{0v} = 1; \\ -1 & \text{if } X_{iu} < X_{0v} \text{ and } \delta_{iu} = 1; \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

For testing against the simple-tree alternative H_1: ($S_i \geq S_0$ with strict inequality for at least one i, $i = 1, 2, \ldots, k$), we propose to use the statistics of the form

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{k} \beta_i U_{0i},$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)$ is a vector of nonzero constants. Let $N = \sum_{i=0}^{k} n_i$. It can be shown, in Appendix A.1, that, under the null hypothesis H_0, the statistic

$N^{-1/2}U(\boldsymbol{\beta})/s$ has asymptotically a standard normal distribution, where $s = \sqrt{s^2}$ and s^2 is stated in equation (A.3). Therefore, the proposed test is to reject H_0 if

$$\hat{U}(\boldsymbol{\beta}) = N^{-1/2}U(\boldsymbol{\beta})/s \geq z(\alpha), \tag{2.5}$$

where $z(\alpha)$ is the upper α-th percentile of a standard normal distribution.

The choice of the constants $\beta's$ is open, see, for example, Chakraborti and Desu (1991) for a discussion on several possible choices of these constants. For simplicity, however, we suggest to employ $\beta_i = 1$, $i = 1, 2, \ldots, k$, namely, the simple-tree tests based on $\hat{U}(\mathbf{1})$ in the practical situations.

Remark 2. Andersen *et al.* (1993) proposed a trend test statistic which is a linear combination of the generalized Kruskal-Wallis (1952) rank-sum statistics for right-censored data. For Breslow's (1970) generalization of the Kruskal-Wallis statistic, in particular, this trend test is a linear combination of the statistics

$$V_i = \sum_{u=1}^{n_i} \sum_{j=0}^{k} \sum_{v=1}^{n_j} \phi(X_{iu}, \delta_{iu}; X_{jv}, \delta_{jv}), \qquad i = 0, 1, \ldots, k,$$

where $\phi(\cdot)$ is in (2.4). Note that the statistic V_i compares the i-th group with the combined groups from 0 to k, while the statistic U_{0i} in (2.3) compares the i-th treatment group with the control ($i = 0$) group.

If the proposed simple-tree test rejects the null hypothesis H_0, one would wish to determine which treatments are more effective than the control. According to Slepian's inequality, we then suggest, under an approximate experimentwise error rate α, to

$$\text{claim} \quad S_i > S_0 \quad \text{if} \quad \hat{U}_{0i} = N^{-1/2}U_{0i}/\sqrt{s_{ii}} \geq z(b) \quad \text{for} \quad i = 1, \ldots, k, \tag{2.6}$$

where s_{ii} is given in (A.1) and $\alpha = 1 - (1 - b)^k$. Note that the pairwise follow-up tests with the Gehan-Wilcoxon two-sample statistics was proposed in Chakraborti and Desu (1991).

3. An example

King *et al.* (1979) investigated the effect of diets on the development of tumors. Ninety rats of the same age and species and in similar physical condition were divided into three groups and were fed with low fat, saturated and unsaturated diets, respectively. The rats were observed for 200 days after an identical amount of tumor cells were injected into a foot pad of each rat and their tumor-free times were recorded and reported in Table 1. The tumor-free time of the rat without tumor at the end of the 200 days and the survival time of the rat dying accidently with no evidence of tumor are both regarded as censored times and underlines.

Table 1. Tumor-free time (days) of 90 rats on three different diets.

Low-fat	140	177	50	65	86	153	181	191	77	84
	87	56	66	73	119	140	200	200	200	200
	200	200	200	200	200	200	200	200	200	200
Saturated	124	58	56	68	79	89	107	86	142	110
	96	142	86	75	117	98	105	126	43	46
	81	133	165	170	200	200	200	200	200	200
Unsaturated	112	68	84	109	153	143	60	70	98	164
	63	63	77	91	91	66	70	77	63	66
	66	94	101	105	108	112	115	126	161	178

Source: King *et al.* (1979).

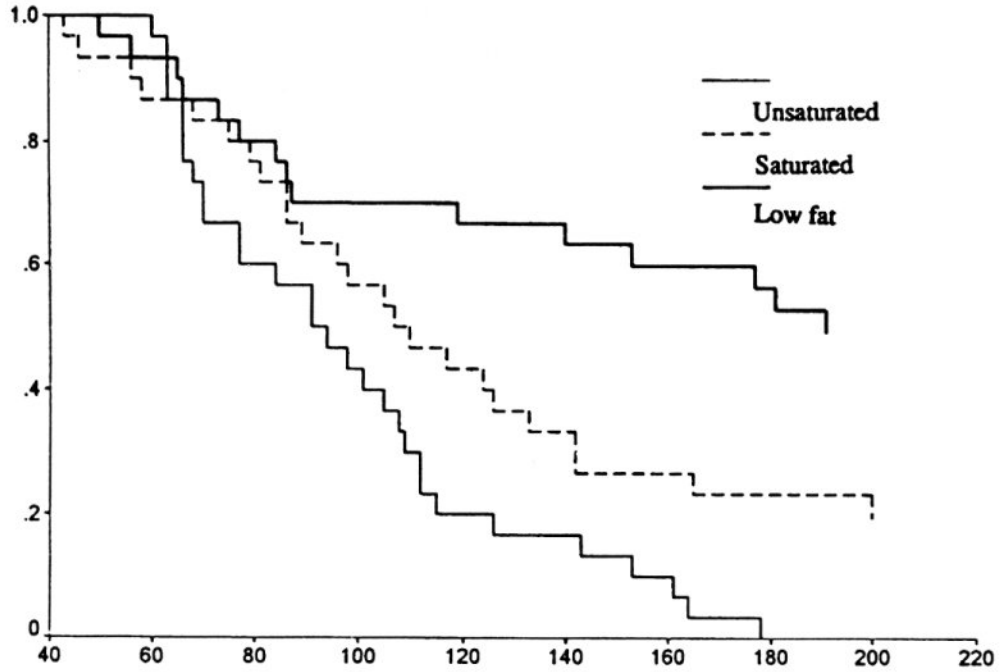

Fig. 1. Survival curves of rats in three diet groups.

The Kaplan-Meier survival function estimates for the three groups were shown in Fig. 1.

To compare the ability of the saturated (treatment 1) or unsaturated diet (treatment 2) relative to low fat diet (control) in keeping the rats tumor free, the two-sample weighted logrank statistics comparing the control group with the treatment groups are employed to construct the simple-tree tests. After some computations we have the relevant summary statistics in Table 2. According to these statistics, we found that the p-values for the logrank, Peto-Prentice-Wilcoxon and Gehan-Wilcoxon simple-tree tests are 8.9×10^{-5}, 1.7×10^{-3} and 1.3×10^{-3}, respectively. We then conclude that at least one of the saturated and unsaturated diets has shorter tumor free time than does the low fat diet. To determine which treatment diets that are unable to keep the rats tumor free compared to the control

Table 2. Summary statistics for the diet-tumor example.

Statistics	Logrank	Peto-Prentice-Wilcoxon	Gehan-Wilcoxon
U_{01}	6.568	4.053	4.200
U_{02}	14.010	7.292	7.667
$\sqrt{s_{11}}$	0.368	0.271	0.259
$\sqrt{s_{22}}$	0.323	0.223	0.230
s	0.579	0.409	0.415
$\hat{U}_{01}$	1.879	1.565	1.711
$\hat{U}_{02}$	4.566	3.455	3.513
$\hat{U}(\mathbf{1})$	3.749	2.926	3.013

diet, we found, at $\alpha = 0.05$, the critical value for the pairwise-wise comparisons in (2.6) is $z(0.0253) = 1.955$. Therefore, all three multiple tests based on two-sample logrank, Peto-Prentice-Wilcoxon and Gehan-Wilcoxon statistics lead to the conclusion that at $\alpha = 0.05$, only the unsaturated diet has shorter tumor free time compared to the low fat diet.

4. Asymptotic relative efficiency

Note that, under the simple-tree alternative H_1, we can express the statistic $N^{-1/2}U(\boldsymbol{\beta})$ as

$$\begin{aligned} N^{-1/2}U(\boldsymbol{\beta}) = & \int_0^\infty \frac{K_0(t)}{Y_0(t)} dM_0(t) - \sum_{i=1}^k \int_0^\infty \frac{K_i(t)}{Y_i(t)} dM_i(t) \\ & + \int_0^\infty K_0(t) \left\{ \frac{d\Lambda_0^N(t)}{d\Lambda(t)} - 1 \right\} d\Lambda(t) \\ & - \sum_{i=1}^k \int_0^\infty K_i(t) \left\{ \frac{d\Lambda_i^N(t)}{d\Lambda(t)} - 1 \right\} d\Lambda(t), \end{aligned}$$

where K_0 and K_i's are stated in Appendix A.1. Under the alternatives where the absolutely continuous distribution functions F_i^N can depend on N and $\sup_{0 \le t < \infty} |F_i^N(t) - F(t)| \to 0$ as $N \to \infty$, $i = 0, 1, \ldots, k$, for some absolutely continuous distribution function F, we obtain, along the lines of Section 7.4 in Fleming and Harrington (1991), the general formula for the Pitman efficacy of the simple-tree weighted logrank test in (2.5) is as

$$e = \frac{\{\int_0^\infty \kappa_0 \gamma_0 d\Lambda - \sum_{i=1}^k \int_0^\infty \kappa_i \gamma_i d\Lambda\}^2}{\sigma^2},$$

where

$$\kappa_0 = \lim_{N \to \infty} \sqrt{\frac{N}{n_0(N - n_0)}} K_0, \qquad \gamma_0 = \lim_{N \to \infty} \sqrt{\frac{n_0(N - n_0)}{N}} \{(d\Lambda_0^N / d\Lambda) - 1\},$$

$$\kappa_i = \lim_{N\to\infty} \sqrt{\frac{n_0+n_i}{n_0 n_i}} K_i, \qquad \gamma_i = \lim_{N\to\infty} \sqrt{\frac{n_0 n_i}{n_0+n_i}} \{(d\Lambda_i^N/d\Lambda) - 1\},$$
$$i = 1, 2, \ldots, k,$$

and

$$\sigma^2 = \int_0^\infty \left\{\sum_{i=1}^k h_i(t)\right\} \{1 - \Delta\Lambda(t)\} d\Lambda(t) \qquad \text{is stated in (A.2).}$$

Note that there are some situations where the $k+1$ groups of data are subject to the same pattern of censorship, see, for example, Chen (1994). Hence, we consider in this section the assumption of equal censoring; that is, $G_i = G$, $i = 0, 1, \ldots, k$. Let $\bar{G} = 1 - G$ and $S = 1 - F$. The efficacy of $\hat{U}(\boldsymbol{\beta})$ can be simplified to be

$$\frac{\left\{\sqrt{\frac{\lambda_0}{1-\lambda_0}} \int_0^\infty (\sum_{i=1}^k \lambda_{0i}\beta_i\omega_{0i})\gamma_0 \bar{G} dF - \sum_{i=1}^k \sqrt{\lambda_0\lambda_{0i}\beta_i} \int_0^\infty \omega_{0i}\gamma_i \bar{G} dF\right\}^2}{\lambda_0 \int_0^\infty \{(\sum_{i=1}^k \lambda_{0i}\beta_i\omega_{0i})^2 + \sum_{i=1}^k \lambda_{0i}(1-\lambda_{0i})\beta_i^2\omega_{0i}^2\}\bar{G} dF}.$$

To compute the efficacy discussed previously, two particular types of simple-tree contiguous alternatives are considered in the following:

I. Lehmann alternatives

$H_0 : S_i = S$ for $i = 0, 1, \ldots, k$

and

$H_1 : S_i = S^{1-b_i/\sqrt{N}}$ and $b_i \geq b_0$ with strict inequality for at least one i, $i = 1, 2, \ldots, k$, where S is an underlying survival function.

II. Scale alternatives

$H_0 : S_i = S$ for $i = 0, 1, \ldots, k$

and

$H_1 : S_i(t) = S(te^{-b_i/\sqrt{N}})$ and $b_i \geq b_0$ with strict inequality for at least one i, $i = 1, 2, \ldots, k$.

Note that the Lehmann alternatives correspond to the proportional hazards model, while the scale alternatives correspond to location shifts in log survival times. It can be seen that, for $i = 0, 1, \ldots, k$, $\sqrt{N}\{d\Lambda_i^N/d\Lambda - 1\} = -b_i$ for the Lehmann alternatives and $\lim_{N\to\infty} \sqrt{N}\{d\Lambda_i^N(t)/d\Lambda(t) - 1\} = -b_i\{\frac{tS'(t)}{S(t)} - \frac{tS''(t)}{S'(t)} - 1\}$ for the scale alternatives. We assume, without loss of generality, that $b_0 = 0$. The formulas for the efficacies of the simple-tree weighted logrank statistics under the assumption of equal censoring are then obtained, after some algebraic manipulations, as

$$\frac{\lambda_0\{\sum_{i=1}^k \sqrt{\lambda_{0i}} b_i\beta_i \int_0^\infty \omega_{0i}\bar{G} dF\}^2}{2\int_0^\infty \{(\sum_{i=1}^k \beta_i\omega_{0i})^2 + \sum_{i=1}^k \beta_i^2\omega_{0i}^2\}\bar{G} dF} \tag{4.1}$$

for the Lehmann alternatives and

$$\frac{\lambda_0\left\{\sum_{i=1}^{k}\sqrt{\lambda_{0i}}b_i\beta_i\int_0^\infty \omega_{0i}\left[\frac{tS'(t)}{S(t)}-\frac{tS''(t)}{S'(t)}-1\right]\bar{G}dF\right\}^2}{2\int_0^\infty\{(\sum_{i=1}^{k}\beta_i\omega_{0i})^2+\sum_{i=1}^{k}\beta_i^2\omega_{0i}^2\}\bar{G}dF} \tag{4.2}$$

for the scale alternatives.

Remark 3. The problem of allocating observations is usually of interest in practical situations. If se consider the setting of $\lambda_0 = a\lambda$ and $\lambda_1 = \lambda_2 = \cdots = \lambda_k = \lambda$, where $\lambda = (a+k)^{-1}$, then $\lambda_{0i} = (a+1)^{-1}$ for $i = 1, 2, \ldots, k$. Since, under such a sample size allocation, maximizing the efficacy in (4.1) or (4.2) is equivalent to maximizing $\lambda_0\lambda_{0i} = a[(a+1)(a+k)]^{-1}$, we find that the efficacy is maximized by taking $a = \sqrt{k}$. Hence, for equal censoring, the optimal design for the simple-tree test based on $\hat{U}(\boldsymbol{\beta})$ is the same as that for Dunnett's (1955) test.

Note that $\omega_{0i} = S\bar{G}$ gives the Gehan-Wilcoxon (G-W) statistic, $\omega_{0i} = 1$ yields the logrank (LR) statistic and $\omega_{0i} = S$ produces the Peto-Prentice-Wilcoxon (P-P-W) statistic. By replacing with appropriate weights, the efficacies for the LR, G-W and P-P-W statistics for equal censoring and sample sizes can be readily obtained as given in Appendix A.2. In fact, the efficacy of the Gehan-Wilcoxon simple-tree test is identical to that of the Chakraborti-Desu (1991) test for Lehmann alternatives when sample sizes are all equal. Moreover, for equal censoring and sample sizes, the asymptotic relative efficiencies among the tests considered here depend only on the survival and censoring distributions.

To evaluate the asymptotic relative efficiencies (ARE) among the simple-tree tests $\hat{U}(\boldsymbol{\beta})$, we consider the Weibull survival distribution with density function $f(t) = \eta t^{\eta-1}\exp(-t^\eta)$, $t > 0$, for the Lehmann alternatives and the lognormal survival distribution with density function $f(t) = \{\eta/(t\sqrt{2\pi})\}\exp\{-(\eta^2/2)(\log t)^2\}$, $t > 0$, for the scale alternatives, where $\eta = 0.5$, 1 and 2. We employ the uniform censoring distributions over $(0, R)$ with probabilities of censorship 0.1, 0.3 and 0.5. The values of the ARE's for the Lehmann and the scale alternatives with equal uniform censoring when sample sizes are all equal are reported in Table 3.

Table 3. Asymptotic relative efficiencies.

Alternatives	Censoring probability	G-W/LR			P-P-W/LR		
		$\eta = 0.5$	1.0	2.0	0.5	1.0	2.0
Lehmann	0.1	0.78	0.75	0.73	0.80	0.78	0.76
	0.3	0.83	0.78	0.75	0.89	0.85	0.80
	0.5	0.87	0.81	0.75	0.95	0.93	0.88
Scale	0.1	1.15	1.16	1.17	1.15	1.16	1.17
	0.3	1.11	1.13	1.15	1.12	1.14	1.16
	0.5	1.08	1.11	1.13	1.09	1.11	1.14

We can see, from Table 3, that the logrank simple-tree test is superior to either the Gehan-Wilcoxon simple-tree test proposed by Chakraborti and Desu (1991) or the Peto-Prentice-Wilcoxon simple-tree test for Lehmann alternatives, while the Gehan-Wilcoxon test and the Peto-Prentice-Wilcoxon test are both more efficient than the logrank test for scale alternatives. This is not surprising since Weibull distributions preserve the proportional hazards, but the hazards are far from being proportional for lognormal distributions. Note that the ARE's are also generally in agreement with the findings in Liu *et al.* (1993).

5. Monte Carlo study

To examine the relative level and power performances of the simple-tree tests based on $\hat{U}(\mathbf{1})$ in (2.5) for comparing several treatments with a control when observations are subject to random right-censorship and sample sizes are varied from small to moderate, we conducted a Monte Carlo study. We considered $k = 3$ treatments with sample sizes $n_0 = n_1 = \cdots = n_k = n = 10$, 20 and 30 in the level study and with $n = 20$ and 30 in the power study.

Exponential and lognormal distributions were considered as survival time distributions and the uniform distribution over $(0, R)$ was used as the censoring distribution. Appropriate uniform, normal and exponential variates were generated by using the IMSL routines DRNUN, DRNNOR and DRNEXP. The exponential-transformed normal variates then give necessary lognormal variates. In the level study, the standard exponential distribution and the lognormal distribution with zero normal mean and normal variance $\sigma^2 = 1/2$ were considered. In the power study, we used exponential distributions with various values of scale parameters θ_i's and lognormal distributions with normal variance $\sigma^2 = 1/2$ but different values of normal means θ_i's. Various values of R which correspond to the probability of censorship as 0.10, 0.30 and 0.50 were considered in the level study, the corresponding uniform distributions for probabilities of censorship 0.10 and 0.30 were then employed as censoring distributions in the power study. Note that the censoring probabilities were fixed for each population in the level study. However, in the power study, they might be varied for the four populations involved due to different survival time distributions.

For each of these settings, we used 1,000 replications to obtain the level or power estimates under the nominal level $\alpha = 0.05$. Therefore, the maximum standard error for the power estimate is about 0.016 ($\approx\sqrt{(0.5)(0.5)/1000}$). In fact, the standard error for the level estimate is less than 0.007 ($\approx\sqrt{(0.05)(0.95)/1000}$). The level and power estimates are presented in Tables 4 and 5.

It is evident, upon examination of Table 4, that the logrank, Gehan-Wilcoxon and Peto-Prentice-Wilcoxon simple-tree tests hold their levels reasonably when the common sample size is about 20. The power study presented in Table 5 shows that the unweighted logrank test is more powerful than the Gehan-Wilcoxon and Peto-Prentice-Wilcoxon tests for exponential distributions. However, the Gehan-Wilcoxon and Peto-Prentice-Wilcoxon tests are both superior to the logrank test for lognormal distributions. These results in fact coincide with the ones in comparing their asymptotic relative efficiencies presented in Table 3.

Table 4. Level estimates for $\alpha = 0.05$, uniform censoring and $n_0 = n_1 = n_2 = n_3 = n$.

n	Censoring probability	Exponential			Lognormal		
		LR	P-P-W	G-W	LR	P-P-W	G-W
10	0.1	0.071	0.057	0.065	0.082	0.063	0.068
	0.3	0.066	0.062	0.065	0.066	0.057	0.061
	0.5	0.058	0.054	0.067	0.058	0.059	0.061
20	0.1	0.062	0.054	0.057	0.049	0.051	0.053
	0.3	0.058	0.059	0.062	0.048	0.047	0.053
	0.5	0.055	0.054	0.057	0.056	0.051	0.054
30	0.1	0.058	0.052	0.052	0.052	0.045	0.045
	0.3	0.052	0.053	0.057	0.059	0.050	0.053
	0.5	0.056	0.058	0.056	0.054	0.053	0.056

Exponential: $f(t) = \exp(-t)$.
Lognormal: $f(t) = \{1/(t\sqrt{\pi})\}\exp\{-(\log t)^2\}$.

Table 5. Powers estimates for $\alpha = 0.05$, uniform censoring and $n_0 = n_1 = n_2 = n_3 = n$.

Survival distribution	n	θ_0	θ_1	θ_2	θ_3	Censoring probability					
						0.1			0.3		
						LR	P-P-W	G-W	LR	P-P-W	G-W
Exponential	20	1	1	1	2	0.187	0.175	0.176	0.164	0.161	0.151
		1	1	1.5	2	0.363	0.312	0.304	0.289	0.263	0.259
		1	1	2	2	0.469	0.417	0.416	0.384	0.359	0.344
		1	1.5	2	2	0.663	0.599	0.600	0.568	0.537	0.520
		1	2	2	2	0.788	0.729	0.731	0.673	0.624	0.625
	30	1	1	1	2	0.262	0.246	0.239	0.219	0.213	0.211
		1	1	1.5	2	0.502	0.423	0.424	0.398	0.378	0.360
		1	1	2	2	0.645	0.552	0.549	0.509	0.489	0.459
		1	1.5	2	2	0.837	0.773	0.772	0.729	0.686	0.663
		1	2	2	2	0.925	0.875	0.870	0.827	0.790	0.761
Lognormal	20	0	0	0	0.5	0.185	0.226	0.218	0.169	0.207	0.199
		0	0	0.2	0.5	0.306	0.359	0.349	0.280	0.316	0.310
		0	0	0.5	0.5	0.476	0.534	0.535	0.442	0.490	0.477
		0	0.2	0.5	0.5	0.620	0.680	0.670	0.562	0.618	0.611
		0	0.5	0.5	0.5	0.772	0.831	0.824	0.735	0.796	0.779
	30	0	0	0	0.5	0.246	0.277	0.268	0.227	0.238	0.238
		0	0	0.2	0.5	0.388	0.421	0.421	0.358	0.391	0.385
		0	0	0.5	0.5	0.604	0.653	0.651	0.577	0.604	0.604
		0	0.2	0.5	0.5	0.732	0.790	0.794	0.702	0.754	0.743
		0	0.5	0.5	0.5	0.893	0.931	0.929	0.864	0.890	0.891

Exponential: $f_i(t) = (1/\theta_i)\exp\{-t/\theta_i\}$.
Lognormal: $f_i(t) = \{1/(t\sqrt{\pi})\}\exp\{-(\log t - \theta_i)^2\}$.

6. Conclusion

A class of tests based on linear combinations of two-sample weighted logrank statistics is proposed for testing against the simple-tree alternatives when data are subject to random right-censorship. The asymptotic relative efficiencies and simulation results show that the unweighted logrank simple-tree test should be used when the assumption of proportional hazards is tenable and sample sizes are near 20. When the hazards are far from being proportional, both the Gehan-Wilcoxon and Peto-Prentice-Wilcoxon simple-tree tests are more powerful than the logrank test. However, as noted by Andersen *et al.* ((1993), 349–350), the weight function used in the Peto-Prentice-Wilcoxon test depends only on the survival experience, while the Gehan-Wilcoxon test uses a weight function that depends on survivals as well as censorings, we recommend to implement the Peto-Prentice-Wilcoxon simple-tree test for the non-proportional hazards model, especially, when the censoring patterns differ greatly in the populations under consideration.

In comparing several treatments with a control, experimenters are also interested in deciding which treatments (if any) are more effective than the control. In such cases, the pairwise follow-up tests based on two-sample weighted logrank tests are suggested. The choice of the weight function for the multiple test is, again, similar to the one for the overall simple-tree test.

Acknowledgements

This research was supported by the National Science Council of Taiwan under Grant NSC85-2121-M008-009. The author wishes to thank the referees for valuable comments and suggestions which substantially improved the presentation of this paper.

Appendix

A.1 *Asymptotic null distribution of* $N^{-1/2}U(\boldsymbol{\beta})$

Note that, when $S_i = S_0$, using the martingale framework, the statistic U_{0i} in (2.1) can be written as

$$U_{0i} = \int_0^\infty \frac{K_{0i}(t)}{Y_0(t)} dM_0(t) - \int_0^\infty \frac{K_{0i}(t)}{Y_i(t)} dM_i(t),$$

where the $M_i(t) = D_i(t) - \int_0^t Y_i(s)d\Lambda(s)$ are independent zero-mean martingales and $\Lambda(s)$ is the common cumulative hazard function. Suppose that $N \to \infty$ in such a way that $n_i/N \to \lambda_i$, $0 < \lambda_i < 1$, and, hence, $n_i/(n_0 + n_i) \to \lambda_i/(\lambda_i + \lambda_0) = \lambda_{0i}$, $i = 0, 1, \ldots, k$. If $Y_i(t)/n_i \xrightarrow{p} \pi_i(t)$, $i = 0, 1, \ldots, k$, uniformly as $N \to \infty$, then the three weight functions considered in this paper satisfy the property that $W_{0i}(t) \xrightarrow{p} \omega_{0i}(t)$ and, thus, $K_{0i}^2(t)[Y_0(t) + Y_i(t)]/[NY_0(t)Y_i(t)] \xrightarrow{p} h_{ii}(t)$ and $K_{0i}(t)K_{0j}(t)/[NY_0(t)] \xrightarrow{p} h_{ij}(t)$ uniformly as $N \to \infty$ for $i = 1, 2, \ldots, k$, where

$$h_{ii}(t) = \lambda_0\lambda_{0i}\omega_{0i}^2(t)\pi_0(t)\pi_i(t)/[(1-\lambda_{0i})\pi_0(t) + \lambda_{0i}\pi_i(t)]$$

$$h_{ij}(t) = \lambda_0\lambda_{0i}\lambda_{0j}\omega_{0i}(t)\omega_{0j}(t)\pi_0(t)\pi_i(t)\pi_j(t)/\{[(1-\lambda_{0i})\pi_0(t) + \lambda_{0i}\pi_i(t)] \cdot [(1-\lambda_{0j})\pi_0(t) + \lambda_{0j}\pi_j(t)]\},$$

$$i \neq j = 1, 2, \ldots, k.$$

Hence, the Martingale Central Limit Theorem (see, for example, Theorem 6.2.1 in Fleming and Harrington (1991)) implies that the asymptotic null (H_0) distribution of the random vector $N^{-1/2}(U_{01}, U_{02}, \ldots, U_{0k})$ is the k-dimensional normal with zero mean vector and covariance matrix $\Sigma = (\sigma_{ij})$, where

$$\sigma_{ij} = \int_0^\infty h_{ij}(t)[1 - \Delta\Lambda(t)]d\Lambda(t), \qquad i, j = 1, 2, \ldots, k,$$

with $\Lambda(t)$ the common cumulated hazard function and $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$. Note that unbiased and consistent estimators of σ_{ij} are then given by

$$\text{(A.1)} \qquad s_{ii} = \frac{1}{N}\int_0^\infty K_{0i}^2(t)\left[\frac{1}{Y_0(t)} + \frac{1}{Y_i(t)}\right]\left[1 - \frac{\Delta D(t) - 1}{Y(t) - 1}\right]\frac{dD(t)}{Y(t)}$$

and

$$s_{ij} = \frac{1}{N}\int_0^\infty \frac{K_{0i}(t)K_{0j}(t)}{Y_0(t)}\left[1 - \frac{\Delta D(t) - 1}{Y(t) - 1}\right]\frac{dD(t)}{Y(t)} \qquad \text{for} \quad i \neq j = 1, 2, \ldots, k,$$

where $D(t) = \sum_{i=0}^k D_i(t)$, $Y(t) = \sum_{i=0}^k Y_i(t)$ and $\Delta D(t) = D(t) - D(t-)$. Following the Cramer-Wold device, we observe that the asymptotic null distribution of $N^{-1/2}\sum_{i=1}^k \beta_i U_{0i}$ is normal with zero mean and variance $\sigma^2 = \sum_{i=1}^k \beta_i^2\sigma_{ii} + \Sigma\Sigma_{i\neq j}\,\beta_i\beta_j\sigma_{ij}$ for any nonzero constants β_i, $i = 1, 2, \ldots, k$.

Note that we can write $N^{-1/2}U(\boldsymbol{\beta})$, under H_0, as

$$N^{-1/2}U(\boldsymbol{\beta}) = \int_0^\infty \frac{K_0(t)}{Y_0(t)}dM_0(t) - \sum_{i=1}^k \int_0^\infty \frac{K_i(t)}{Y_i(t)}dM_i(t),$$

where $K_i(t) = N^{-1/2}\beta_i K_{0i}(t)$, $i = 1, 2, \ldots, k$, and $K_0(t) = \sum_{i=1}^k K_i(t)$. Moreover, we have $K_i^2(t)/Y_i(t) \xrightarrow{p} h_i(t)$, $i = 0, 1, \ldots, k$, where

$$h_0(t) = \lambda_0\pi_0(t)\left[\sum_{j=1}^k \lambda_{0j}\beta_j\omega_{0j}(t)\pi_j(t)/\{(1 - \lambda_{0j})\pi_0(t) + \lambda_{0j}\pi_j(t)\}\right]^2$$

and, for $i = 1, 2, \ldots, k$,

$$h_i(t) = \lambda_0\lambda_{0i}(1 - \lambda_{0i})\beta_i^2\omega_{0i}^2(t)\pi_0^2(t)\pi_i(t)/\{(1 - \lambda_{0i})\pi_0(t) + \lambda_{0i}\pi_i(t)\}^2.$$

Therefore, σ^2 can also be expressed as

$$\text{(A.2)} \qquad \sigma^2 = \int_0^\infty \sum_{i=0}^k h_i(t)[1 - \Delta\Lambda(t)]d\Lambda(t).$$

An unbiased and consistent estimator of σ^2 is then given by

$$\text{(A.3)} \qquad s^2 = \int_0^\infty \left\{\sum_{i=0}^k \frac{K_i^2(t)}{Y_i(t)}\right\}\left\{1 - \frac{\Delta D(t) - 1}{Y(t) - 1}\right\}\frac{dD(t)}{Y(t)}.$$

Hence, we obtain, by applying Slutsky's Theorem, that $N^{-1/2}U(\boldsymbol{\beta})/s$ is asymptotically standard normal, where $s = \sqrt{s^2}$.

A.2 *Pitman efficacies for equal censoring and sample sizes*

The Pitman efficacies of the Logrank (LR), Gehan-Wilcoxon (G-W) and Peto-Prentice-Wilcoxon (P-P-W) simple-tree tests for equal censoring and sample sizes derived directly from equations (4.1) and (4.2) are given in the following:

$$\text{eff(LR)} = \left\{\sum_{i=1}^{k} \beta_i b_i\right\}^2 \int_0^\infty \bar{G} dF \Bigg/ \left\{(k+1)\left[\left(\sum_{i=1}^{k} \beta_i\right)^2 + \sum_{i=1}^{k} \beta_i^2\right]\right\},$$

$$\text{eff(L-R)} = \left\{\sum_{i=1}^{k} \beta_i b_i\right\}^2 \left\{\int_0^\infty \bar{G}^2 S dF\right\}^2 \Bigg/ \left\{(k+1)\left[\left(\sum_{i=1}^{k} \beta_i\right)^2 + \sum_{i=1}^{k} \beta_i^2\right] \int_0^\infty \bar{G}^3 S^2 dF\right\}$$

and

$$\text{eff(P-P-W)} = \left\{\sum_{i=1}^{k} \beta_i b_i\right\}^2 \left\{\int_0^\infty \bar{G} S dF\right\}^2 \Bigg/ \left\{(k+1)\left[\left(\sum_{i=1}^{k} \beta_i\right)^2 + \sum_{i=1}^{k} \beta_i^2\right] \int_0^\infty \bar{G} S^2 dF\right\}$$

for the Lehmann alternatives, and

$$\text{eff(LR)} = \left\{\sum_{i=1}^{k} \beta_i b_i\right\}^2 \int_0^\infty \left(\frac{tS'(t)}{S(t)} - \frac{tS''(t)}{S'(t)} - 1\right) \bar{G} dF \Bigg/ \left\{(k+1)\left[\left(\sum_{i=1}^{k} \beta_i\right)^2 + \sum_{i=1}^{k} \beta_i^2\right]\right\},$$

$$\text{eff(G-W)} = \left\{\sum_{i=1}^{k} \beta_i b_i\right\}^2 \left\{\int_0^\infty \left(\frac{tS'(t)}{S(t)} - \frac{tS''(t)}{S'(t)} - 1\right) \bar{G}^2 S dF\right\}^2 \Bigg/ \left\{(k+1)\left[\left(\sum_{i=1}^{k} \beta_i\right)^2 + \sum_{i=1}^{k} \beta_i^2\right] \int_0^\infty \bar{G}^3 S^2 dF\right\},$$

and

$$\text{eff(P-P-W)} = \left\{\sum_{i=1}^{k} \beta_i b_i\right\}^2 \left\{\int_0^\infty \left(\frac{tS'(t)}{S(t)} - \frac{tS''(t)}{S'(t)} - 1\right) \bar{G} S dF\right\}^2 \Bigg/ \left\{(k+1)\left[\left(\sum_{i=1}^{k} \beta_i\right)^2 + \sum_{i=1}^{k} \beta_i^2\right] \int_0^\infty \bar{G} S^2 dF\right\}$$

for the scale alternatives.

REFERENCES

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, Wiley, New York.

Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship, *Biometrika*, **57**, 579–594.

Chakraborti, S. and Desu, M. M. (1991). Linear rank tests for comparing treatments with a control when data are subject to unequal patterns of censorship, *Statist. Neerlandica*, **45**, 227–254.

Chen, Y. I. (1994). A generalized steel procedure for comparing several treatments with a control under random right-censorship, *Comm. Statist. Simulation Comput.*, **23**, 1–16.

Dunnett, C. (1955). A multiple comparisons procedure for comparing several treatments with a control, *J. Amer. Statist. Assoc.*, **50**, 1096–1121.

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

Fligner, M. A. and Wolfe, D. A. (1982). Distribution-free tests for comparing several treatments with a control, *Statist. Neerlandica*, **36**, 119–127.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples, *Biometrika*, **52**, 203–223.

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

King, M., Bailey, D. M., Gibson, D. G., Pitha, J. V. and McCay, P. B. (1979). Incidence and growth of mammary tumors induced by 7, 12 dimethylbenz(a) antheacence as related to the dietary content of fat and antioxidant, *Journal of National Cancer Institute*, **63**, 656–664.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis, *J. Amer. Statist. Assoc.*, **47**, 583–621.

Liu, P. Y., Green, S., Wolf, M. and Crowley, J. (1993). Testing against ordered alternatives for censored survival data, *J. Amer. Statist. Assoc.*, **88**, 153–161.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemother. Rep.*, **50**, 163–170.

Nelson, W. (1969). Hazard plotting for incomplete failure data, *Journal of Quality Technology*, **1**, 27–52.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion), *J. Roy. Statist. Soc. Ser. A*, **135**, 185–206.

Prentice, R. L. (1978). Linear rank tests with right censored data, *Biometrika*, **65**, 165–179.

Prentice, R. L. and Marek, P. (1979). A qualitative discrepancy between censored data rank tests, *Biometrics*, **35**, 861–867.

Slepian, D. (1962). The one-sided barrier problem for Gaussian noise, *Bell System Technical Journal*, **41**, 463–501.

Taron, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions, *Biometrika*, **64**, 147–160.