[Reprinted from ANNALS OF THE INSTITUTE OF STATISTICAL MATHEMATICS, Vol. 53, No. 4, 835–852, 2001]

WEIGHTED KAPLAN-MEIER TESTS FOR UMBRELLA ALTERNATIVES

WEI-CHEN LEE AND YUH-ING CHEN

Ann. Inst. Statist. Math.
Vol. 53, No. 4, 835–852 (2001)

WEIGHTED KAPLAN-MEIER TESTS FOR UMBRELLA ALTERNATIVES

WEI-CHUN LEE AND YUH-ING CHEN

Institute of Statistics, National Central University, Chung-Li 32054, Taiwan, R.O.C.,
e-mail: ychen@stat.ncu.edu.tw

(Received January 14, 2000; revised May 17, 2000)

Abstract. The problem of testing for umbrella alternatives in a one-way layout with right-censored survival data is considered. Testing procedures based on the two-sample weighted Kaplan-Meier statistics suggested by Pepe and Fleming (1989, *Biometrics*, **45**, 497–507; 1991, *J. Roy. Statist. Soc. Ser. B*, **53**, 341–352) are suggested for both cases when the peak of the umbrella is known or unknown. The asymptotic relative efficiency of the weighted Kaplan-Meier test and the weighted logrank test proposed by Chen and Wolfe (2000, *Statist. Sinica*, **10**, 595–612) is computed for the umbrella peak-known setting where the piecewise exponential survival distributions have the proportional or crossing hazards, or the related hazards differ at early or late times. Moreover, the results of a Monte Carlo study are presented to investigate the level and power performances of the umbrella tests. Finally, application of the proposed procedures to an appropriated data set is illustrated.

Key words and phrases: Asymptotic relative efficiency, Monte Carlo study, umbrella alternative, weighted Kaplan-Meier statistic, weighted logrank statistic.

1. Introduction

In animal experiments or clinical trials, dose-response studies are frequently used to assess the relative treatment effects of increasing dose levels of a substance where the response of interest is, for instance, time to tumor occurence or the prolonged survival time of patients with a particular disease. These studies may lead to randomly right-censored data, since they may be terminated at preassigned time, subjects who randomly enter a study may be lost to follow-up randomly, or death may be due to a competing risk which is not of interest.

For the i-th sample ($i = 1, \dots, k$), let $T_{i1}, \dots, T_{in_i}$ be independent and identically distributed (iid) random variables each with a continuous distribution function F_i, and let $D_{i1}, \dots, D_{in_i}$ be iid random variables each with a continuous distribution function G_i, where D_{ij} is the censoring time associated with the survival time T_{ij}. Suppose that the k sample are mutually independent and that the T_{ij} and D_{ij} are also mutually independent. In such a setting, we actually observe only $X_{ij} = \min\{T_{ij}, D_{ij}\}$ and the indicator of censorship $\delta_{ij} = I\{T_{ij} \le D_{ij}\}, j = 1, \dots, n_i, i = 1, \dots, k$. Let $S_i = 1 - F_i$ and $C_i = 1 - G_i$, $i = 1, \dots, k$. When increasing dose levels may lead to a larger or at least equal efficacy, Liu *et al.* (1993) proposed a generalization of the Jonckheere (1954)-Terpstra (1952) test for $H_0 : (S_1 = \cdots = S_k)$ against the ordered alternatives $H_{1O} : (S_1 \le \cdots \le S_k$, with at least one strict inequality) (Barlow *et al.* (1972)) based on two-sample weighted logrank statistics (see, for example, Gill (1980)). When, however, an increasing dose-response relationship with a downturn in response at high doses is anticipated, Chen and Wolfe (2000) suggested testing procedures based on two-sample weighted logrank statistics against the peak-known umbrella alternative H_{1U}^p : ($S_1 \leq \cdots \leq S_p \geq \cdots \geq S_k$, with at least one strict inequality) or the peak-unknown umbrella alternative H_{1U} : ($S_1 \leq \cdots \leq S_p \geq \cdots \geq S_k$, for some p, $1 \leq p \leq k$, with at least one strict inequality) (Mack and Wolfe (1981)).

However, the two-sample weighted logrank statistic, as a function of the difference between two cumulative hazards, might not be sensitive to testing against the alternative about the survival distributions. Therefore, Pepe and Fleming (1989) developed a two-sample test based on the weighted Kaplan-Meier statistic. Pepe and Fleming (1989, 1991) further indicated that the two-sample weighted Kaplan-Meier (WKM) test is competitive to the two-sample logrank (LR) test for the proportional hazards model and the WKM test is even more powerful than the LR test when the two hazard functions are crossing. Therefore, in this paper, we consider tests based on two-sample weighted Kaplan-Meier statistics against the umbrella alternatives, since the hypotheses H_{1U}^p and H_{1U} only involve survival functions. The proposed weighted Kaplan-Meier umbrella tests then provide competitive alternatives to the weighted logrank umbrella tests suggested in Chen and Wolfe (2000) for right-censored survival data.

In Section 2 we propose a generalization of the Mack-Wolfe (1981) test for peak-known umbrella alternative based on two-sample weighted Kaplan-Meier statistics. In Section 3 the Pitman efficacy of the peak-known weighted Kalpan-Meier umbrella test under Lehmann alternatives is derived. The asymptotic relative efficiency of the proposed test and the competing weighted logrank umbrella test is then evaluated for different piecewise exponential survival distributions with a certain uniform censoring distribution. In Section 4 we generalize the Chen-Wolfe (1990) test based on weighted Kaplan-Meier statistics for the peak-unknown umbrella alternative with right-censored survival data. Section 5 contains a numerical example and Section 6 presents the results of a Monte Carlo study investigation of the level and power performances of the proposed tests for a variety of umbrella pattern treatment effects configurations. Conclusion and discussion are finally given in Section 7.

2. Peak-known umbrella test

To generalize the Mack-Wolfe (1981) peak(p)-known umbrella test for right-censored survival data, we consider the following statistic:

$$WKM_p = \sum_{1\leq i<}\sum_{j\leq p} \sqrt{\frac{n_i n_j}{N}} WKM_{ij} + \sum_{p\leq i<}\sum_{j\leq k} \sqrt{\frac{n_i n_j}{N}} WKM_{ji}, \tag{2.1}$$

where

$$WKM_{ij} = \int_0^{T_c} \hat{w}_{ij}(t)\{\hat{S}_j(t) - \hat{S}_i(t)\}dt$$

is the two-sample weighted Kaplan-Meier statistic comparing the j-th group with the i-th group, with

$$\hat{w}_{ij}(t) = \frac{\hat{C}_i(t-)\hat{C}_j(t-)}{\hat{p}_i\hat{C}_i(t-) + \hat{p}_j\hat{C}_j(t-)},$$

the weighted function downweights the difference between $\hat{S}_j(t)$ and $\hat{S}_i(t)$, where $\hat{S}_i(t)$ is the Kaplan-Meier (1958) survival estimator in group i, $\hat{C}_i(t)$ is the Kaplan-Meier

(1958) estimator of the censoring survival function $C_i(t)$, $\hat{p}_i = n_i/N$, $i = 1, \ldots, k$, and $T_c = \sup[t : \min\{\hat{C}_1(t), \ldots, \hat{C}_k(t), \hat{S}_1(t), \ldots, \hat{S}_k(t)\} > 0]$. Note that, for the uncensored data case with $\hat{w}_{ij}(\cdot) = 1$, the statistic WKM_p reduces to

$$\sum_{1 \le i <} \sum_{j \le p} \sqrt{\frac{n_i n_j}{N}} (\bar{X}_j - \bar{X}_i) + \sum_{p \le i <} \sum_{j \le k} \sqrt{\frac{n_i n_j}{N}} (\bar{X}_i - \bar{X}_j),$$

where $\bar{X}_i$ is the sample mean in group i. Therefore, the test based on WKM_p can be regarded as a generalization of the mean-based test for right-censored data.

To construct a test based on the statistic WKM_p, we obtain the null (H_0) asymptotic distribution of WKM_p in the following:

THEOREM 2.1. *Suppose that* $\hat{p}_i = n_i/N \to p_i$, $0 < p_i < 1$, $i = 1, \ldots, k$. *Assume that*

$$\frac{Y_i(t)}{n_i} \xrightarrow{p} \pi_i(t) = C_i(t)S(t)$$

uniformly in $t \in [0, \tau]$, *where* $Y_i(t) = \sum_{j=1}^{n_i} I_{[X_{ij} \ge t]}$ *is the at-risk numbers in group* i, $i = 1, \ldots, k$, $\tau = \sup[t : \min\{S(t), C_1(t), \ldots, C_k(t)\} > 0]$ *and* $S(t)$ *is the survival function under* H_0. *Then the null asymptotic distribution of the statistic* WKM_p *is a normal distribution with mean zero and variance,* $N(0, \sigma_p^2)$, *where*

$$(2.2) \quad \sigma_p^2 = \sum_{j=1}^{k} \int_0^\tau h_j(t)\{1 - \Delta\Lambda(t)\} d\Lambda(t),$$

$$(2.3) \quad h_j(t) = \begin{cases} \left[\sum_{\substack{i \ne j \\ 1 \le i \le p}} \sqrt{p_i} \operatorname{sgn}(j-i) \left\{ \int_t^\tau w_{ij}(u)S(u)du \right\} \right]^2 \Big/ \pi_j(t) & \text{if } 1 \le j \le p-1; \\ \left[\sum_{\substack{i \ne p \\ 1 \le i \le k}} \sqrt{p_i} \left\{ \int_t^\tau w_{ip}(u)S(u)du \right\} \right]^2 \Big/ \pi_p(t) & \text{if } j = p; \\ \left[\sum_{\substack{i \ne j \\ p \le i \le k}} \sqrt{p_i} \operatorname{sgn}(i-j) \left\{ \int_t^\tau w_{ji}(u)S(u)du \right\} \right]^2 \Big/ \pi_j(t) & \text{if } p+1 \le j \le k, \end{cases}$$

and

$$w_{ij}(t) = \frac{C_i(t-)C_j(t-)}{p_i C_i(t-) + p_j C_j(t-)},$$

$\operatorname{sgn}(x) = 1$ *if* $x > 0$ *and* -1 *for* $x < 0$, $\Lambda(\cdot)$ *is the cumulative hazard function under* H_0 *and* $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$.

PROOF. See Appendix A.1.

Let $\hat{S}(t)$ be the Kaplan-Meier estimator of the common survival function calculated from the combined k samples. Since the Kaplan-Meier estimator is consistent, the consistent estimator of $\pi_j(t)$ is obtained as $\hat{\pi}_j(t) = \hat{S}(t)\hat{C}_j(t-)$ and the consistent estimators of $d\Lambda(t)$ and $1-\Delta\Lambda(t)$ are given by $d\hat{\Lambda}(t) = -d\hat{S}(t)/\hat{S}(t-)$ and $1-\Delta\hat{\Lambda}(t) = \hat{S}(t)/\hat{S}(t-)$, respectively. Hence, the estimator

$$\widehat{\text{Var}}(WKM_p) = \sum_{j=1}^{k} \int_0^{T_c} \hat{h}_j(t) \frac{-\hat{S}(t)d\hat{S}(t)}{\hat{S}^2(t-)} \tag{2.4}$$

provides with a consistent estimator of σ_p^2, where

$$\hat{h}_j(t) = \begin{cases} \left[\sum_{\substack{i \neq j \\ 1 \le i \le p}} \sqrt{\hat{p}_i} \text{sgn}(j-i) \left\{ \int_t^{T_c} \hat{w}_{ij}(u)\hat{S}(u)du \right\} \right]^2 \Big/ \hat{S}(t)\hat{C}_j(t-) & \\ \qquad \text{if} \quad 1 \le j \le p-1; & \\ \left[\sum_{\substack{i \neq p \\ 1 \le i \le k}} \sqrt{\hat{p}_i} \left\{ \int_t^{T_c} \hat{w}_{ip}(u)\hat{S}(u)du \right\} \right]^2 \Big/ \hat{S}(t)\hat{C}_p(t-) & \\ \qquad \text{if} \quad j = p; & \\ \left[\sum_{\substack{i \neq j \\ p \le i \le k}} \sqrt{\hat{p}_i}\, \text{sgn}(i-j) \left\{ \int_t^{T_c} \hat{w}_{ji}(u)\hat{S}(u)du \right\} \right]^2 \Big/ \hat{S}(t)\hat{C}_j(t-) & \\ \qquad \text{if} \quad p+1 \le j \le k. & \end{cases} \tag{2.5}$$

Slutsky's theorem further implies that the null asymptotic distribution of the statistic

$$WKM_p^* = \frac{WKM_p}{\sqrt{\widehat{\text{Var}}(WKM_p)}} \tag{2.6}$$

is a standard normal. Therefore, we propose to reject H_0 in favor of H_{1U}^p if

$$WKM_p^* \ge z(\alpha),$$

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

3. Asymptotic relative efficiency for the peak-known setting

Note that, under a sequence of contiguous alternatives such that

$$\sqrt{N}\{S_j^N(t) - S_i^N(t)\} \to D_{ij}(t)$$

unifromly on $[0,\tau)$ for some bounded function $D_{ij}(\cdot)$, $i \neq j = 1,\ldots,k$, the mean of WKM_p is

$$\Delta = \sum_{1 \le i <} \sum_{j \le p} \sqrt{p_i p_j} \int_0^{\tau} w_{ij}(t)D_{ij}(t)dt + \sum_{p \le i <} \sum_{j \le k} \sqrt{p_i p_j} \int_0^{\tau} w_{ji}(t)D_{ji}(t)dt.$$

Moreover, the Martingale Central Limit Theorem implies that the asymptotic distribution of WKM_p is $N(\Delta, \sigma_p^2)$, where σ_p^2 is stated in (2.2). Note that the test based on the statistic WKM_p which has a higher power against H_{1U}^p should have a larger value of the parameter Δ. Therefore, the Pitman efficacy based on WKM_p is given by

$$\text{eff}(WKM_p) = \Bigg[\sum_{1\le i<}\sum_{j\le p}\sqrt{p_ip_j}\int_0^\tau w_{ij}(t)D_{ij}(t)dt + \sum_{p\le i<}\sum_{j\le k}\sqrt{p_ip_j}\int_0^\tau w_{ji}(t)D_{ji}(t)dt\Bigg]^2 \Bigg/ \sigma_p^2. \tag{3.1}$$

To evaluate the Pitman efficacy in equation (3.1), we consider the Lehmann alternatives:

$H_0 : (S_i = S \text{ for } i = 1, \ldots, k)$

and

$H_1 : (S_i^N = S^{1-\theta_i/\sqrt{N}}$ and $\theta_1 \le \cdots \le \theta_p \ge \cdots \ge \theta_k$ for some $p, 1 \le p \le k,$ with at least one strict inequality).

Since under the Lehmann alternatives, we observe $D_{ij}(t) \to (\theta_j - \theta_i)S(t)\ln S(t)$ for i, $j = 1, \ldots, k$. Hence,

$$\text{eff}(WKM_p) = \Bigg[\sum_{1\le i<}\sum_{j\le p}\sqrt{p_ip_j}\int_0^\tau w_{ij}(t)(\theta_i - \theta_j)S(t)\ln S(t)dt + \sum_{p\le i<}\sum_{j\le k}\sqrt{p_ip_j}\int_0^\tau w_{ji}(t)(\theta_j - \theta_i)S(t)\ln S(t)dt\Bigg]^2 \Bigg/ \sigma_p^2. \tag{3.2}$$

Note that the efficacy for the weighted Kaplan-Meier umbrella test under the assumption of equal censoring and equal sample sizes is given in Appendix A.2, and the related efficacy for the weighted logrank umbrella test can be found in Section 3 of Chen and Wolfe (2000).

Notice that the Lehmann alternatives with constants θ_i's correspond to the proportional hazards model. However, the Lehmann alternatives with $S(t) = \exp(-t)$ and some time-related θ_i's generate the alternatives which have piecewise exponential survival distributions but with different hazards at different time periods. For example, for the hazard functions

$$\lambda_{i1}I_1(t) + \lambda_{i2}\{1 - I_1(t)\},$$

where $I_1(t) = 1$ if $t < t_1$ and 0, otherwise, the corresponding time-related θ_i is

$$\lambda_{i1}I_1(t) + \{\lambda_{i2} + (\lambda_{i1} - \lambda_{i2})t_1/t\}\{1 - I_1(t)\}.$$

In this case, the hazard function with the same λ_{i2} (λ_{i1}) value but different λ_{i1} (λ_{i2}) values generate hazards that are different at early (late) times. Moreover, for the hazard functions

$$\lambda_{i1}I_1(t) + \lambda_{i2}I_2(t) + \lambda_{i3}I_3(t) + \lambda_{i4}\{1 - I_1(t) - I_2(t) - I_3(t)\},$$

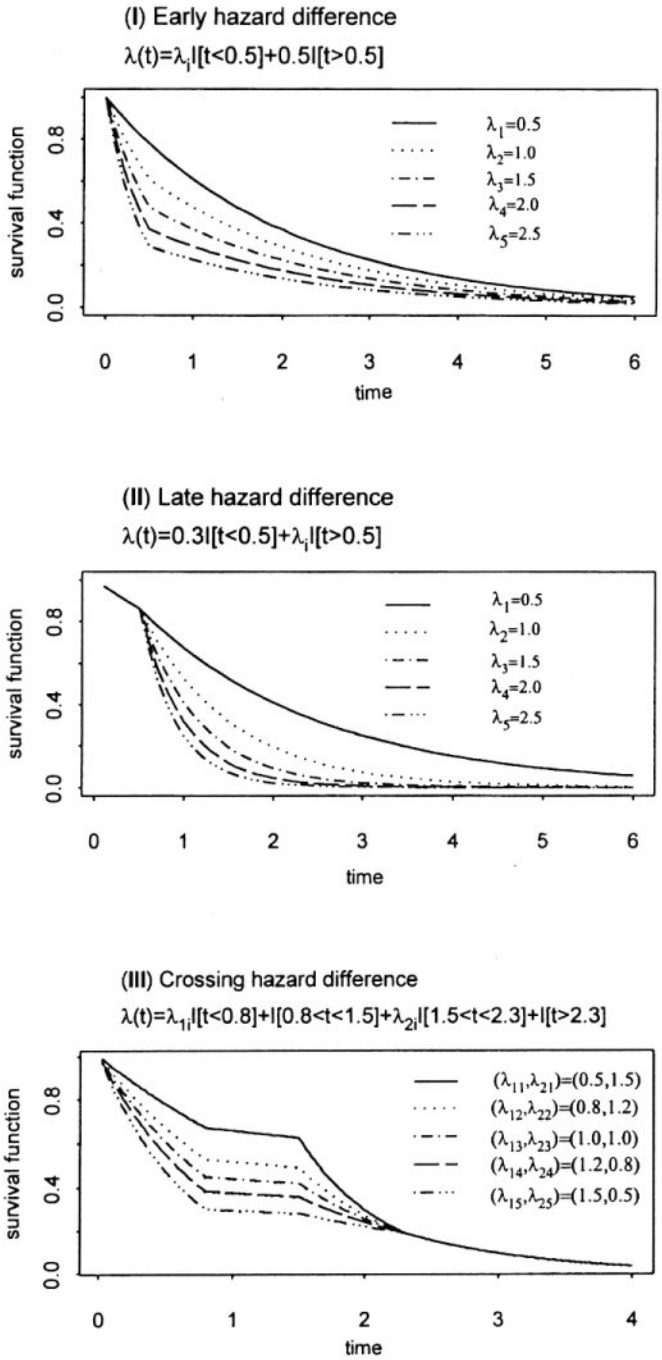

Fig. 1. Survival configuration for alternatives.

where $I_j(t) = 1$ if $t_{j-1} \leq t < t_j$ and 0, otherwise, with $t_0 = -\infty$, for $j = 1, 2, 3$, the corresponding θ_i is

$$\begin{aligned}\lambda_{i1} I_1(t) &+ \{\lambda_{i2} + (\lambda_{i1} - \lambda_{i2}) t_1/t\} I_2(t) + \{\lambda_{i3} + (\lambda_{i1} - \lambda_{i2}) t_1/t + (\lambda_{i2} - \lambda_{i3}) t_2/t\} \\ &\cdot I_3(t) + \{\lambda_{i4} + (\lambda_{i1} - \lambda_{i2}) t_1/t + (\lambda_{i2} - \lambda_{i3}) t_2/t + (\lambda_{i3} - \lambda_{i4}) t_3/t\} \\ &\cdot \{1 - I_1(t) - I_2(t) - I_3(t)\}.\end{aligned}$$

Note that the piecewise exponential distributions with different values of the λ_{ij} yield crossing hazards.

To compute the asymptotic relative efficiency (ARE) between the weighted logrank test and the weighted Kaplan-Meier test for peak-known (p) umbrella settings, we consider $k = 5$ with equal sample size and $p = k$. The survival distributions considered

Table 1. Asymptotic efficiencies of WKM tests relative to WLR tests for $k = 5$ and $p = 5$ with equal sample size and covariate $(\theta_1, \ldots, \theta_5)$ with censoring distribution $U(0, R)$.

θ_i	R	$\frac{WKM}{LR}$	$\frac{WKM}{PPW}$	$\frac{WKM}{WLRL}$
(a) Standard exponential distribution				
$(1,2,3,4,5)$	3.185	1.06	1.50	1.25
	9.901	1.08	1.48	1.39
$(1,1,1,1,2)$	3.185	1.63	2.34	1.95
	9.901	1.69	2.31	2.17
$(1,2,4,8,16)$	3.185	1.31	1.85	1.54
	9.901	1.33	1.82	1.72
(b) Piecewise exponential distribution[1]				
(I) Early hazard difference		1.31	0.97	27.4
(II) Late hazard difference		0.70	1.86	0.21
(III) Crossing hazard difference		1.22	1.03	4.97

1: $R = 2$ for piesewise exponential distribution

herein are exponentials with scale parameters for $i = 1, \ldots, k$ (1) $\theta_i = i$, (2) $\theta_i = 2^{(i-1)}$, and (3) $\theta_i = 1$, if $i \leq k-1$ and 2, if $i = k$, and a variety of piecewise exponentials with different values of the λ_{ij}'s. The piecewise exponential survival functions under consideration are presented in Fig. 1. The uniform distributions over 0 and R, namely, $U(0, R)$ with a variety of R values, are employed as censoring distributions. Note that the weight function in the weighted logrank (WLR) test statistic is generally taken as $\{S(t)\}^\rho\{(1 - S(t))\}^\gamma$, $\rho \geq 0$, $\gamma \geq 0$ (Fleming and Harrington (1991)). In this paper, for the comparison with the weighted Kaplan-Meier (WKM) test, we consider the WLR test based on the logrank (LR, $\rho = \gamma = 0$) statistic, Peto-Prentice-Wilcoxon (PPW, $\rho = 1$, $\gamma = 0$) statistic or the statistic with $\rho = 0$, $\gamma = 1$, denoted by WLRL, which is suitable for the late hazard differences. The values of the ARE between the WKM test and WLR test are then reported in Table 1.

We can see, from Table 1, that the WKM test is competitive to the LR test, which is known to be optimal for the proportional hazards model. The WKM test also provides with a competitor to the PPW test, which is believed to be appropriate for the early hazard differences. For the late hazard differences, however, the efficacy of the WKM test is less than that of the LR or WLRL test. This is not surprising, since the WKM test puts less weight on late time, thereby reducing its efficacy for detecting the late occuring hazard differences. Neverthless, for the survival functions with crossing hazards, the WKM test is superior to any WLR test under consideration.

4. Peak-unknown umbrella test

As noted in Chen and Wolfe (1990), if the peak of the umbrella is unknown, the alternative H_{1U} can be viewed as a union of k individual umbrella alternatives with the peak at group $1, \ldots, k$, respectively. This way of viewing H_{1U} leads to a natural extension for the peak-unknown setting to the test procedure which rejects H_0 for large values of

$$WKM^*_{\max} = \max(WKM^*_1, \ldots, WKM^*_k), \tag{4.1}$$

where WKM^*_p, $p = 1, \ldots, k$, are given in (2.6).

Notice that, for any constants $a_1, \ldots, a_k$, we have

$$\sum_{p=1}^{k} a_p WKM_p = \sum_{i=1}^{k} \int_0^{T_c} \left\{ \sum_{p=1}^{k} a_p H_i^{(p)} \right\} dM_i$$

where the $H_i^{(p)}$ are specified in (A.1). The Martingale Central Limit Theorem and the Cramér-Wald device then imply that, under H_0,

$$(WKM_1^*, \ldots, WKM_k^*) \xrightarrow{d} \boldsymbol{N}(0, \boldsymbol{R}), \qquad \text{as} \quad N \to \infty,$$

where

$$\boldsymbol{R} = \left(\frac{\mathrm{Cov}(WKM_p, WKM_q)}{\sqrt{\sigma_p^2 \sigma_q^2}} \right),$$

with σ_p^2, $p = 1, \ldots, k$, as stated in (2.2),

$$\mathrm{Cov}(WKM_p, WKM_q) = \sum_{i=1}^{k} \int_0^{\tau} h_i^{(pq)} \{1 - \Delta\Lambda\} d\Lambda$$

and $h_i^{(pq)}$ is the limit of $H_i^{(p)} H_i^{(q)} Y_i$, $p \neq q = 1, \ldots, k$. Since, for $p \neq q = 1, \ldots, k$, consistent estimators for the $\mathrm{Cov}(WKM_p, WKM_q)$ are

$$\widehat{\mathrm{Cov}}(WKM_p, WKM_q) = \sum_{i=1}^{k} \int_0^{T_c} \hat{h}_i^{(pq)}(t) \frac{-\hat{S}(t) d\hat{S}(t)}{\hat{S}^2(t-)},$$

a consistent estimator for $\boldsymbol{R}$ is obtained as

$$\hat{\boldsymbol{R}} = \left(\frac{\widehat{\mathrm{Cov}}(WKM_p, WKM_q)}{\sqrt{\widehat{\mathrm{Var}}(WKM_p) \widehat{\mathrm{Var}}(WKM_q)}} \right),$$

where $\widehat{\mathrm{Var}}(WKM_p)$ is given in (2.4).

Let $(Z_1, \ldots, Z_k)$ be a random vector which has a k-variate normal distribution with zero mean vector and correlation matrix $\hat{\boldsymbol{R}}$, and let $z_{\max}(k, \alpha)$ be the upper α-th percentile of the distribution of $\max(Z_1, \ldots, Z_k)$. We then obtain an approximate level α test for the umbrella alternative H_{1U} by rejecting H_0 if

$$WKM_{\max}^* \geq z_{\max}(k, \alpha).$$

Note that the value of $z_{\max}(k, \alpha)$ can be obtained by employing the program in Schervish (1984) for $k \leq 7$.

On the other hand, under the assumption of a common censoring distribution, we obtain the $\mathrm{Cov}(WKM_p, WKM_q)$ values, for $p < q$, in the following:

$$\mathrm{Cov}(WKM_p, WKM_q) = \left\{ \sum_{j=1}^{p-1} \left(\sum_{\substack{i \neq j \\ 1 \leq i \leq p}} \frac{\sqrt{p_i}}{p_i + p_j} \mathrm{sgn}(j - i) \right) \left(\sum_{\substack{i \neq j \\ 1 \leq i \leq q}} \frac{\sqrt{p_i}}{p_i + p_j} \mathrm{sgn}(j - i) \right) \right.$$

Table 2. Values of $z_{\max}(k,\alpha)$ for common censoring distribution and equal sample sizes.

	k								
α	2	3	4	5	6	7	8	9	10
0.01	2.58	2.72	2.80	2.86	2.90	2.93	2.99	3.02	3.05
0.05	1.96	2.12	2.22	2.28	2.33	2.36	2.37	2.38	2.39
0.10	1.65	1.82	1.92	1.99	2.03	2.07	2.08	2.09	2.10

$$
\begin{aligned}
&+\left(\sum_{\substack{i\neq p\\1\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_p}\right)\left(\sum_{\substack{i\neq p\\1\le i\le q}}\frac{\sqrt{p_i}}{p_i+p_p}\operatorname{sgn}(p-i)\right)\\
&+\sum_{j=p+1}^{q-1}\left(\sum_{\substack{i\neq j\\p\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_j}\operatorname{sgn}(i-j)\right)\left(\sum_{\substack{i\neq j\\1\le i\le q}}\frac{\sqrt{p_i}}{p_i+p_j}\operatorname{sgn}(j-i)\right)\\
&+\left(\sum_{\substack{i\neq q\\p\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_q}\operatorname{sgn}(i-q)\right)\left(\sum_{\substack{i\neq q\\1\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_q}\right)\\
&+\sum_{j=q+1}^{k}\left(\sum_{\substack{i\neq j\\p\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_j}\operatorname{sgn}(i-j)\right)\left(\sum_{\substack{i\neq j\\q\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_j}\operatorname{sgn}(i-j)\right)\Bigg\}\\
&\cdot\int_0^\tau\left[\int_t^\tau C(u)S(u)du\right]^2\frac{-dS(t)}{C(t-)S(t-)S(t)}.
\end{aligned}
$$

The value of the σ_p^2 (2.2) can be simplified as

$$
\begin{aligned}
\sigma_p^2=&\left\{\sum_{j=1}^{p-1}\left(\sum_{\substack{i\neq j\\1\le i\le p}}\frac{\sqrt{p_i}}{p_i+p_j}\operatorname{sgn}(j-i)\right)^2+\sum_{j=p+1}^{k}\left(\sum_{\substack{i\neq j\\p\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_j}\operatorname{sgn}(i-j)\right)^2\right.\\
&\left.+\left(\sum_{\substack{i\neq p\\1\le i\le k}}\frac{\sqrt{p_i}}{p_i+p_p}\right)^2\right\}\int_0^\tau\left[\int_t^\tau C(u)S(u)du\right]^2\frac{-dS(t)}{C(t-)S(t-)S(t)}.
\end{aligned}
$$

For this setting, we derive the correlations for the case of equal sample size and compute ($k\le 7$) or simulate ($k>7$) the critical values $z_{\max}(k,\alpha)$ from the k-variate normal distribution with known correlation matrix $\boldsymbol{R}$. Note that the correlation coefficient between WKM_1 and WKM_k is -1. Therefore, the value of $z_{\max}(k,\alpha)$ is also the upper α-th percentile of the distribution of $\max(|Z_1|,Z_2,\ldots,Z_{k-1})$. The critical values $z_{\max}(k,\alpha)$, for the case of common censoring distribution, equal sample sizes, $k=2(1)10$ and $\alpha=0.01$, 0.05 and 0.1 are then reported in Table 2. We recommend use of these critical values for situations where the sample sizes are equal and the assumption of common censoring distribution is tenable. Otherwise, we can obtain the estimated correlation matrix

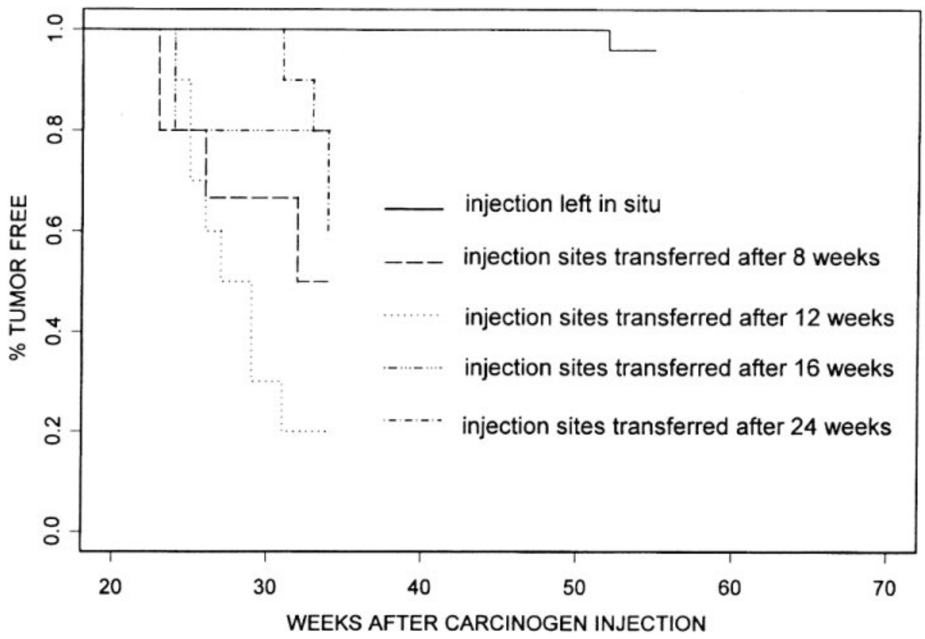

Fig. 2. The Kaplan-Meier estimates for the injection sites-transfer data.

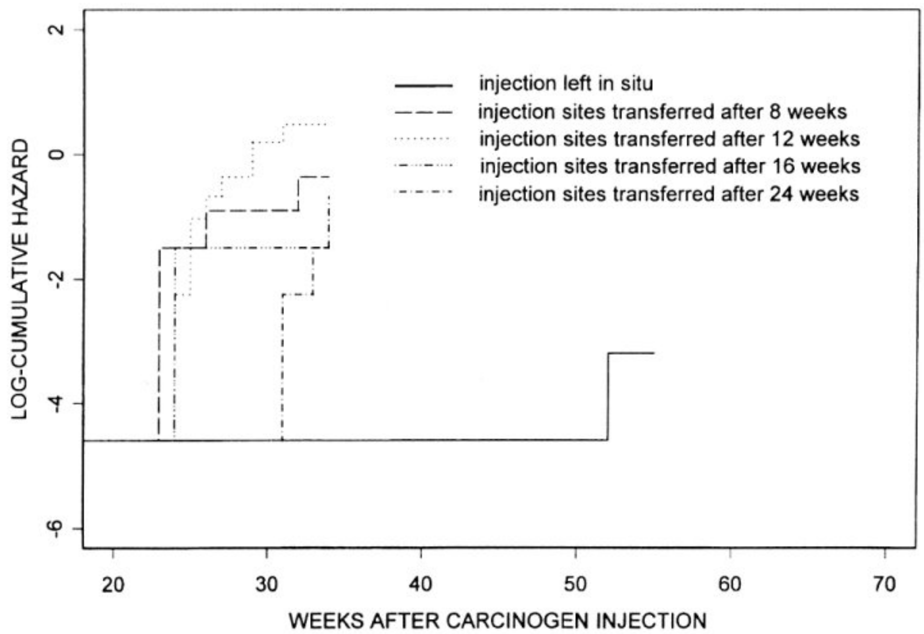

Fig. 3. The cumlative hazard functions for the injection sites-transfer data.

$\hat{\boldsymbol{R}}$ from the data, and then compute or simulate the critical value $z_{\max}(k, \alpha)$ from the appropriate multivariate normal distribution with the estimated correlation matrix $\hat{\boldsymbol{R}}$.

5. An example

We consider the numerical example analyzed in Chen and Wolfe (2000) which is the study presented in Homburger and Treger (1970) to investigate whether the carcinogenic effect of transplantation of combined injection sites from previous animal hosts receiving injections of a large dose (500 μg) of a weaker carcinogen, benz[α]-anthracene (BA) in

Table 3. Umbrella test statistics for the injection sites-transfer data.

Peak (p)	1	2	3	4	5
(a) Weighted Kaplan-Meier (WKM)					
WKM_p^*	1.016	2.995	3.824	1.324	−1.016
$\hat{\boldsymbol{R}}$	1.000	0.636	0.114	−0.455	−1.000
		1.000	0.481	−0.060	−0.630
			1.000	0.419	−0.112
				1.000	0.472
$p - value = 3 \times 10^{-4}$					
(b) Logrank (LR)[1]					
A_p^*	−2.040	2.922	5.377	3.884	5.536
$\hat{\boldsymbol{R}}$	1.000	0.300	0.152	−0.133	−0.754
		1.000	0.584	0.482	0.081
			1.000	0.681	0.327
				1.000	0.678
$p - value = 4 \times 10^{-7}$					
(c) Peto-Prentice-Wilcoxon (PPW)[1]					
A_p^*	−2.234	3.229	5.659	4.196	5.437
$\hat{\boldsymbol{R}}$	1.000	0.285	0.115	−0.172	−0.803
		1.000	0.663	0.549	0.126
			1.000	0.753	0.384
				1.000	0.712
$p - value = 1 \times 10^{-7}$					

1: A_p^* is the weighted logrank umbrella test statistic suggested in Chen and Wolfe (2000).

tricaprylin (glycerol triocanoate), has on tumor growth in the secondary recipients. Of interest, in particular, are possible differences in carcinogenic effects relative to the length of time elapsed between the original injections in the host and the site transferred to the transplant recipients. A group of 40 C57BL/6 J male mice at periods of 8, 12, 16, and 24 weeks after the original injection into the 40 donor mice. In addition, a control group of 50 animals were also injected directly with 500 μg of BA, which was left in situ. The measurement of record for each study group was the time (after the initial transplants or injection, in the case of the control group) at which a tumor was first palpated. For those animals which did not develop tumors, the time recorded is the number of weeks between the initial transplants (or injection for the control group) and the end of the study when animals were sacrificed and autopsied (or death for those which died without tumors). Thus, those animals with no incidence of tumors yield censored data for this study. The Kaplan-Meier estimates of the survival (tumor free) functions for the five studied groups of animals are presented in Fig. 2 and the relevant summary statistics for testing against the umbrella alternative are reported in Table 3.

We observe, from Table 3, that both the weighted Kaplan-Meier peak-unknown umbrella test and the Peto-Pentrice-Wilcoxon peak-unknown umbrella test claim that the survivals have an umbrella pattern with the peak, possibly, at the third group of injection sites transferred after 12 weeks, while the logrank peak-unknown umbrella test concludes that the survivals of the five groups follow an ordered pattern. Note that the

Table 5. Power estimates for $k = 4$, nominal level $\alpha = 0.05$, censoring distribution $U(0, R)$ and $n_1 = \cdots = n_4 = n$.

Parameters						Peak-known				Peak-unknown			
λ_1	λ_2	λ_3	λ_4	R	n	WKM	LR	$WLRL$	PPW	WKM	LR	$WLRL$	PPW
						(a) Exponential distribution							
1	1	1.5	2	9.901	20	0.631	0.619	0.509	0.544	0.420	0.386	0.292	0.300
					30	0.795	0.774	0.673	0.673	0.627	0.561	0.438	0.447
				3.185	20	0.580	0.507	0.392	0.458	0.355	0.284	0.205	0.242
					30	0.727	0.653	0.521	0.599	0.517	0.409	0.288	0.367
1	1.5	2	1	9.901	20	0.607	0.741	0.653	0.647	0.472	0.575	0.491	0.476
					30	0.787	0.882	0.792	0.806	0.671	0.765	0.649	0.654
				3.185	20	0.560	0.627	0.527	0.569	0.405	0.453	0.730	0.400
					30	0.720	0.782	0.692	0.728	0.572	0.639	0.528	0.569
						(b) Early hazard difference ($\lambda(t) = \lambda_i I_{[t\le 0.5]} + 0.5 I_{[t>0.5]}$)							
2	1.5	1	0.5	2	20	0.798	0.686	0.370	0.752	0.599	0.451	0.197	0.506
				2	30	0.922	0.838	0.501	0.887	0.788	0.645	0.290	0.724
2	1.5	0.5	1	2	20	0.698	0.598	0.351	0.640	0.591	0.496	0.236	0.557
				2	30	0.842	0.759	0.448	0.796	0.783	0.682	0.332	0.742
						(c) Late hazard difference ($\lambda(t) = 0.3 I_{[t\le 0.5]} + \lambda_i I_{[t>0.5]}$)							
2	1.5	1	0.3	2	20	0.600	0.713	0.749	0.564	0.379	0.481	0.512	0.317
				2	30	0.769	0.867	0.904	0.721	0.562	0.687	0.736	0.489
2	1.5	0.3	1	2	20	0.541	0.683	0.777	0.530	0.396	0.532	0.610	0.381
				2	30	0.693	0.828	0.914	0.671	0.579	0.722	0.808	0.547
						(d) Crossing hazard difference ($\lambda(t) = \lambda_{i1} I_{[t\le 0.8]} + 0.1 I_{[0.8<t\le 1.5]} + \lambda_{i2} I_{[1.5<t\le 2.3]} + I_{[t>2.3]}$)							
1.5	1	0.8	0.5[1]	2	20	0.712	0.634	0.403	0.659	0.496	0.390	0.213	0.414
0.5	1	1.2	1.5	2	30	0.865	0.798	0.534	0.828	0.702	0.585	0.322	0.617
1.5	0.8	0.5	1[1]	2	20	0.702	0.653	0.456	0.663	0.525	0.493	0.320	0.503
0.5	1.2	1.5	1	2	30	0.832	0.796	0.570	0.809	0.705	0.656	0.417	0.671

1: the first row is λ_{i1} and the second row is λ_{2i}.

tial distributions were considered. In the power study, we used exponential distributions with various scale parameters, denoted by λ_i, $i = 1, \ldots, k$. We also simulated the powers for the alternatives with piecewise exponential distributions corresponding to early, late or crossing hazards differences. Note that, for standard exponential distributions, the $U(0, R)$ distributions corresponding to probabilities of censorship 0.1 ($R = 9.901$) and 0.3 ($R = 3.185$), respectively, were considered in the level study. These uniform censoring distributions were then used in the power study. For the piecewise exponential distributions, the $U(0, 2)$ were employed as the censoring distribution in both the level and power studies. Therefore, in the level study, case (1)–(3) have probabilities of censorship 0.51, 0.56 and 0.51, respectively. However, the populations involved in the power study have different censoring probabilities due to different survival time distribution.

For each of these settings, we employed 5,000 replications in obtaining the level or

Table 6. Power estimates for $k = 4$, nominal level $\alpha = 0.05$, exponential and piecewise exponential survival distributions with uniform censoring distribution $U(0, R)$.

Parameters						$p = 2$		$p = 3$		peak-unknown	
λ_1	λ_2	λ_3	λ_4	R	n	WKM	LR	WKM	LR	WKM	LR
(a) Exponential distribution											
1	1.5	2	1	9.901	20	0.315	0.426	0.607	0.741	0.472	0.575
					30	0.418	0.573	0.787	0.882	0.671	0.765
				3.185	20	0.293	0.377	0.560	0.627	0.405	0.453
					30	0.393	0.512	0.720	0.782	0.572	0.639
1	1.5	2	1.5	9.901	20	0.104	0.160	0.461	0.529	0.355	0.421
					30	0.115	0.195	0.627	0.677	0.529	0.593
				3.185	20	0.102	0.148	0.419	0.424	0.367	0.328
					30	0.114	0.179	0.562	0.579	0.449	0.469
(b) Early hazard difference ($\lambda(t) = \lambda_i I_{[t\le 0.5]} + 0.5 I_{[t>0.5]}$)											
2	1	0.5	1.5	2	20	0.290	0.382	0.798	0.752	0.632	0.601
				2	30	0.367	0.498	0.920	0.890	0.805	0.765
1.5	1	0.5	2	2	20	0.382	0.436	0.734	0.690	0.567	0.531
				2	30	0.485	0.574	0.856	0.853	0.735	0.730
(c) Late hazard difference ($\lambda(t) = 0.3 I_{[t\le 0.5]} + \lambda_i I_{[t>0.5]}$)											
2	1	0.3	1.5	2	20	0.176	0.323	0.604	0.803	0.428	0.628
				2	30	0.226	0.418	0.784	0.927	0.613	0.822
1.5	1	0.3	2	2	20	0.225	0.347	0.558	0.757	0.383	0.579
				2	30	0.258	0.456	0.717	0.892	0.552	0.761
(d) Crossing hazard difference ($\lambda(t) = \lambda_{i1} I_{[t\le 0.8]} + 0.1 I_{[0.8<t\le 1.5]} + \lambda_{i2} I_{[1.5<t\le 2.3]} + I_{[t>2.3]}$)											
1.5	1	0.8	0.5[1]	2	20	0.053	0.090	0.712	0.634	0.496	0.390
0.5	1	1.2	1.5	2	30	0.048	0.103	0.865	0.798	0.702	0.585
1.5	0.8	0.5	1[1]	2	20	0.218	0.287	0.702	0.653	0.525	0.493
0.5	1.2	1.5	1	2	30	0.288	0.368	0.832	0.796	0.705	0.656

1: the first row is λ_{i1} and the second row is λ_{2i}.

power estimates under the nominal level $\alpha = 0.05$. Since the simulation results for $k = 4$ and $k = 5$ are quite similar, we only report the level and power estimates for $k = 4$ populations. The level estimates are presented in Table 4 and the power estimates are reported in Table 5. Note that the results in Table 6 provide information about how the peak-known umbrella test performs when it corresponds to the wrong peak.

We observe, from Table 4, that, for the peak-known setting, the weighted Kaplan-Meier (WKM) test and the weighted logrank (WLR) test maintain their levels reasonably well. For the peak-unknown setting, however, the level performance of the WKM test is better than the WLR test. In fact, the WLR test tends to be conservative in holding its level.

The results of the power study in comparisons between the WKM and WLR tests given in Table 5 are generally in a good agreement with those of the asymptotic relative efficiency presented in Table 1. Moreover, we observe that the WKM or WLR test has excellent power against umbrella pattern treatment effects when the peak is correctly chosen. However, the results in Table 6 reveal that the power of the peak-known umbrella test declines when the peak is incorrectly selected. In these cases, the peak-known WKM or, for example, the LR umbrella test with incorrect peak is even less powerful than the associated peak-unknown umbrella test.

7. Conclusion and Discussion

The asymptotic relative efficiencies and the simulation results both indicate that none of the competing umbrella tests is uniformly better than the other for the peak-known and peak-unknown settings, respectively. If the hazard difference is known or can be clearly visualled from the log-cumulative hazards plot to occur at early or late time, then the weighted logrank umbrella test can be used with appropriate weight function. If, however, the hazard functions are apparently crossing, then the weighted Kaplan-Meier umbrella test is suggested. Nevertheless, when the time period at which the hazard difference occurs is unknown or can not be identified clearly, the weighted Kaplan-Meier umbrella test may be used, since it would not be the worst one, although it might not be the best one.

It was noted in Chen and Wolfe (2000) that, when the peak of the umbrella is not certain, the peak-unknown weighted logrank umbrella test is even more powerful than the corresponding peak-known weighted logrank umbrella test with incorrectly selected umbrella peaks. The weighted Kaplan-Meier umbrella tests share the same merit. Therefore, if the peak-unknown weighted Kaplan-Meier umbrella test based on $WKM^*_{\max}$ (4.1) rejects H_0 and $WKM^*_{\max} = WKM^*_{\hat{p}}$, then we may estimate the unknown peak group to be at $\hat{p}$. Hence, we expect that the test based on $WKM^*_{\max}$ would not only provide a suitable test, but also give a reasonable estimation of the location of the peak group in those problems involving peak-unknown umbrella pattern treatment effects. In fact, a confidence set for the unknown umbrella peak would be of more interest in practice. Therefore, an extension of the work in Pan (1997) is needed. This topic will be addressed in a separate paper.

Appendix

A.1 *Proof of Theorem* 2.1

Let $I_{[A]}$ be an indicator function of event A. Following Theorem 3.2.3 in Fleming and Harrington (1991), the Kaplan-Meire estimator $\hat{S}_i(t)$ can be represented by

$$S(t) - S(t)\int_0^t \frac{\hat{S}_i(v)}{S(v)} I_{[Y_i(v)>0]} dM_i(v) + B_i(t),$$

where

$$B_i(t) = I_{[T_c<t]} \frac{\hat{S}_i(T_c)\{S(T_c) - S(t)\}}{S(T_c)}$$

and the $M_i(\cdot)$ are indenpent mean-zero martingales. Then equation (2.1) can be written as

$$WKM_P = \sum_{j=1}^{k}\int_0^{T_c} H_j(t)dM_j(t) - \sum_{j=1}^{k}\int_0^{T_c} V_j(t)B_j(t)dt \qquad \text{(A.1)}$$

where

$$H_j(t) = \begin{cases} \sum_{\substack{i\neq j \\ 1\leq i\leq p}} \sqrt{\frac{n_i n_j}{N}} \operatorname{sgn}(j-i) \left\{ \int_t^{T_c} \hat{w}_{ij}(u) S(u) du \right\} \frac{\hat{S}_j(t)}{S(t)} \frac{I_{[Y_j(t)>0]}}{Y_j(t)} & \text{if } 1 \leq j \leq p-1; \\ \sum_{\substack{i\neq p \\ 1\leq i\leq k}} \sqrt{\frac{n_i n_p}{N}} \left\{ \int_t^{T_c} \hat{w}_{ip}(u) S(u) du \right\} \frac{\hat{S}_p(t)}{S(t)} \frac{I_{[Y_p(t)>0]}}{Y_p(t)} & \text{if } j = p; \\ \sum_{\substack{i\neq j \\ p\leq i\leq k}} \sqrt{\frac{n_i n_j}{N}} \operatorname{sgn}(i-j) \left\{ \int_t^{T_c} \hat{w}_{ji}(u) S(u) du \right\} \frac{\hat{S}_j(t)}{S(t)} \frac{I_{[Y_j(t)>0]}}{Y_j(t)} & \text{if } p+1 \leq j \leq k \end{cases}$$

and

$$V_j(t) = \begin{cases} \sum_{\substack{i\neq j \\ 1\leq i\leq p}} \sqrt{\frac{n_i n_j}{N}} \operatorname{sgn}(j-i) \hat{w}_{ij}(t) & \text{if } 1 \leq j \leq p-1; \\ \sum_{\substack{i\neq p \\ 1\leq i\leq k}} \sqrt{\frac{n_i n_p}{N}} \hat{w}_{ip}(t) & \text{if } j = p; \\ \sum_{\substack{i\neq j \\ p\leq i\leq k}} \sqrt{\frac{n_i n_j}{N}} \operatorname{sgn}(i-j) \hat{w}_{ji}(t) & \text{if } p+1 \leq j \leq k. \end{cases}$$

The second term of (A.1) converges to zero and $H_j^2(t)Y_j(t) \to^p h_j(t)$, umiformly in t for $j = 1, \ldots, k$, where the $h_j(t)$ in (2.3) satisfy the conditions stated in Theorem 6.2.1 in Fleming and Harrington (1991) known as the Martingale Central Limit Theorem. Therefore, the theorem holds.

A.2 *Pitman efficency of WKM_p under equal censoring and equal sample sizes*

The simplified form of equation (3.2) assuming equal censoring is given as

$$\text{(A.2)}\ \text{eff}(WKM_p) = \left[\sum_{1\leq i<} \sum_{j\leq p} \frac{\sqrt{p_i p_j}}{p_i + p_j} (\theta_i - \theta_j) + \sum_{p\leq i<} \sum_{j\leq k} \frac{\sqrt{p_i p_j}}{p_i + p_j} (\theta_j - \theta_i) \right]^2 \cdot \left[\int_0^\tau C(t) S(t) \ln S(t) dt \right]^2 \Bigg/ \left\{ \left[\sum_{j=1}^{p-1} \left(\sum_{\substack{i\neq j \\ 1\leq i\leq p}} \frac{\sqrt{p_i}}{p_i + p_j} \operatorname{sgn}(j-i) \right)^2 + \left(\sum_{\substack{i\neq p \\ 1\leq i\leq k}} \frac{\sqrt{p_i}}{p_i + p_p} \right)^2 + \sum_{j=p+1}^{k} \left(\sum_{\substack{i\neq j \\ p\leq i\leq k}} \frac{\sqrt{p_i}}{p_i + p_j} \operatorname{sgn}(i-j) \right)^2 \right] \right.$$

$$\cdot \int_0^\tau \left(\int_t^\tau C(u)S(u)du\right)^2 \frac{-dS(t)}{C(t)S^2(t)} \Bigg\}.$$

Under the assumption of equal sample sizes, equation (A.2) can be further simplified as

$$\text{(A.3)}\ \text{eff}(WKM_p) = \left[\sum_{1\le i<}\sum_{j\le p}(\theta_i-\theta_j)+\sum_{p\le i<}\sum_{j\le k}(\theta_j-\theta_i)\right]^2 \left[\int_0^\tau C(t)S(t)\ln S(t)dt\right]^2$$

$$\Bigg/ \Bigg\{ k \left[\sum_{j=1}^{p-1} \left(\sum_{\substack{i\neq j\\ 1\le i\le p}} \text{sgn}(j-i) \right)^2 + (k-1)^2 \right.$$

$$\left. + \sum_{j=p+1}^{k} \left(\sum_{\substack{i\neq j\\ p\le i\le k}} \text{sgn}(i-j) \right)^2 \right]$$

$$\cdot \int_0^\tau \left(\int_t^\tau C(u)S(u)du\right)^2 \frac{-dS(t)}{C(t)S^2(t)} \Bigg\}.$$

References

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

Chen, Y. I. and Wolfe, D. A. (1990). A study of distribution-free tests for umbrella alternatives, *Biometrical J.*, **32**, 47–57.

Chen, Y. I. and Wolfe, D. A. (2000). Umbrella tests for right-censored survival data, *Statist. Sinica*, **10**, 595–612.

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

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

Homburger, F. and Treger, A. (1970). Transplantation technique for acceleration of carcinogenesis by benz[α]anthracene or 3,4,9,10-dibenzpyrene [benzo(rst)penta-phene], *Journal of the National Cancer Institute*, **44**, 357–360.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives, *Biometrika*, **41**, 133–145.

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

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.

Mack, M. A. and Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives, *J. Amer. Statist. Assoc.*, **76**, 175–181.

Pan, G. (1997). Confidence subset containing the unknown peaks of an umbrella ordering, *J. Amer. Statist. Assoc.*, **92**, 307–314.

Pepe, M. S. and Fleming, T. R. (1989). Weighted Kaplan-Meier statistics: A class of distance tests for censored survival data, *Biometrics*, **45**, 497–507.

Pepe, M. S. and Fleming, T. R. (1991). Weighted Kaplan-Meier statistics: Large sample and optimality considerations, *J. Roy. Statist. Soc. Ser. B*, **53**, 341–352.

Schervish, M. J. (1984). Multivariate normal probabilities with error bound, *Applied Statistics*, **33**, 81–94.

Terpstra, T. J. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking, *Indag. Math.*, **145**, 327–333.