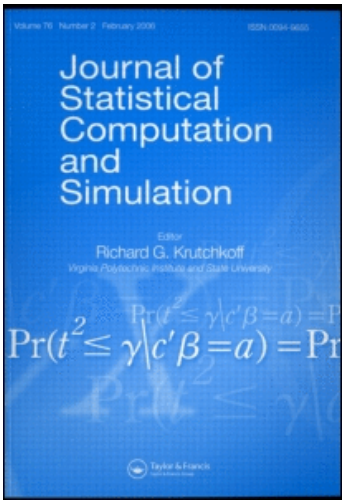

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713650378

Identification of the minimum effective dose based on weighted Kaplan-Meier statistics

Yu-Mei Chang[a]; Yuh-Ing Chen[b]
[a] Department of Statistics, Tunghai University, Taichung, Taiwan, Republic of China [b] Institute of Statistics, National Central University, Jhongli, Taiwan, Republic of China

First published on: 23 November 2010

To cite this Article Chang, Yu-Mei and Chen, Yuh-Ing(2010) 'Identification of the minimum effective dose based on weighted Kaplan-Meier statistics', Journal of Statistical Computation and Simulation,, First published on: 23 November 2010 (iFirst)
To link to this Article: DOI: 10.1080/00949650903451769
URL: http://dx.doi.org/10.1080/00949650903451769

Journal of Statistical Computation and Simulation
iFirst, 2010, 1–16

Taylor & Francis
Taylor & Francis Group

Identification of the minimum effective dose based on weighted Kaplan–Meier statistics

Yu-Mei Chang[a]* and Yuh-Ing Chen[b]

[a] *Department of Statistics, Tunghai University, Taichung 407, Taiwan, Republic of China;* [b] *Institute of Statistics, National Central University, Jhongli 320, Taiwan, Republic of China*

(*Received 22 December 2008; final version received 30 October 2009*)

Testing procedures are considered for identifying the minimum effective dose (MED) in a dose–response study with randomly right-censored survival data, where the MED is defined to be the smallest dose level under study that has survival advantage over the zero dose control. The proposed testing procedures are implemented in a step-down manner together with three different types of weighted Kaplan–Meier statistics. Comparative results of a Monte Carlo error rate and power/bias study for a variety of survival and censoring distributions are then presented and discussed. The application of the proposed procedures is finally illustrated for identifying the MED of the diethylstilbestrol in the treatment of prostate cancer.

Keywords: closed test; dose–response study; minimum effective dose; Monte Carlo study; right-censored data

1. Introduction

In animal experiments or clinical trials for drug development, dose–response studies are frequently conducted to evaluate the treatment effects of the drug under study. In such studies, subjects or patients are usually randomly allocated to different groups to receive either the placebo (or zero dose) or a variety of dose levels of the drug. One factor of interest in such studies is to identify the minimum effective dose (MED) of the drug, where the MED is defined to be the smallest dose level producing a clinically important response that can be declared statistically significantly more effective than the placebo response [1].

In animal experiments or clinical trials, randomly right-censored survival data are frequently observed, since the study may be terminated at a preassigned time due to time limitation, some events may be attributed to competing risks, which is not of interest in the present study, or subjects may be randomly lost to follow-up. For testing the equality of two survival distributions in the presence of randomly right censoring, the most commonly implemented test statistics are the weighted logrank (WLR) statistics [2]. Therefore, Chen and Chang [3] suggested step-down testing procedures based on different types of WLR statistics for identifying the MED. However,

*Correspondence author. Email: yumei0115@thu.edu.tw

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650903451769
http://www.informaworld.com

the WLR statistic is a function of the difference of the Nelson–Aalen [4,5] cumulative hazard functions, and it may not be sensitive against the stochastic ordering alternative, particularly when the hazard functions of two groups are crossing. As a result, Pepe and Fleming [6,7] developed a class of test statistics based on the sum of weighted differences in Kaplan–Meier [8] estimators of survival functions. Moreover, they showed that the weighted Kaplan–Meier (WKM) test is competitive with the logrank (LR) test under the proportional hazards alternative and may perform better under non-proportional hazards alternatives. Therefore, in this paper, we consider identifying the MED using the step-down closed testing procedure as in [3], but based on the contrast type of WKM statistics for randomly right-censored survival data.

In Section 2, we introduce the problem under study and review the step-down testing procedure in Tamhane *et al.* [9]. In Section 3, the step-down testing procedures based on three different types of WKM statistics, respectively, are then proposed for the MED identification. The results of a Monte Carlo study investigation of the relative error rate and power/bias performances of the competing procedures are presented and discussed in Section 4. The use of these testing procedures is further illustrated in Section 5 with the numerical example involving the prostate cancer data of [10]. Finally, Section 6 draws some conclusions and suggestions on the use of the proposed testing procedures.

2. Data setting, problem and step-down testing procedure

Suppose that the zero population ($i = 0$) is the zero dose control and the other k populations correspond to the increasing dose levels. For the ith sample ($i = 0, 1, \ldots, k$), let $X_{i1}, \ldots, X_{in_i}$ be independent and identically distributed (i.i.d.) random variables with a continuous survival function S_i, and $C_{i1}, \ldots, C_{in_i}$ be i.i.d. random variables with a continuous survival function F_i, where C_{ij} is the censoring time associated with the survival time X_{ij}. Furthermore, assume that the $k + 1$ samples are independent and the censoring time C_{ij} is independent of the lifetime X_{ij}. In this setting, we actually observed only $\{\min(X_{ij}, C_{ij})\}$ and the indicator of censorship $\delta_{ij} = I\{X_{ij} \le C_{ij}\}$, $j = 1, 2, \ldots, n_i$, $i = 0, 1, \ldots, k$. Let $S_i > S_0$ denote $S_i(t) \ge S_0(t)$ for all t and $S_i(t) > S_0(t)$ for some t. In this paper, we consider identifying the MED defined as

$$\text{MED} = \min\{i\colon S_i > S_0,\ i = 1, 2, \ldots, k\}.$$

To identify the MED with randomly right-censored survival data, we implement the step-down procedure [9] for testing the null hypotheses $H_{0i} : (S_0 = S_1 = \cdots = S_i)$, $i = 1, 2, \ldots, k$. The family of null hypotheses $H = \{H_{0i}\}$ is closed under intersection in the sense that $H_{0i} \in H$ and $H_{0j} \in H$ imply $H_{0i} \cap H_{0j} \in H$ [11]; hence, we use a level α step-down closed testing scheme for identifying the MED based on adjusted p-values [12]. Suppose that $T_1, \ldots, T_k$ are the statistics involved. The step-down closed testing procedure begins at the first step testing for H_{0k} based on the statistic $T_{(k)} = \max(T_1, \ldots, T_k)$ and computing the associated p-value $p_k = P\{T_{(k)} \ge t_{(k)} | H_{0k}\}$, where $t_{(k)}$ is the observed value of $T_{(k)}$. If the test based on fails to reject H_{0k}, then the MED is identified to be $\hat{\text{MED}} = k + 1$. In this case, the MED is declared to be beyond the dose levels under study. Otherwise, the test proceeds to the next step testing for $H_{0(k-1)}$. In general, at step i involving the zero dose control and the first $k_i = k - i + 1$ non-zero dose groups, the statistic for testing H_{0k_i} is $T_{(k_i)} = \max(T_1, \ldots, T_{k_i})$. Moreover, if the observed value of $T_{(k_i)}$ is $t_{(k_i)}$, the adjusted p-value of the closed test at this step is given by $p^*_{k_i} = \max(p_{k_1}, \ldots, p_{k_i})$, where $p_{k_i} = P\{T_{(k_i)} \ge t_{(k_i)} | H_{0k_i}\}$ is the p-value of the individual test based on . Therefore, if $p^*_{k_i} \le \alpha$ and $i<k$, then the test proceeds with $k_{i+1} = k_i - 1$, and if $p^*_{k_i} > \alpha$ or $i = k$, the test stops and the MED is identified to be $\hat{\text{MED}} = k_i + 1$ or $\hat{\text{MED}} = 1$. Finally, the associated adjusted p-value for the closed testing procedure is $p^*_{\hat{\text{MED}}}$, which provides with the evidence of data for supporting

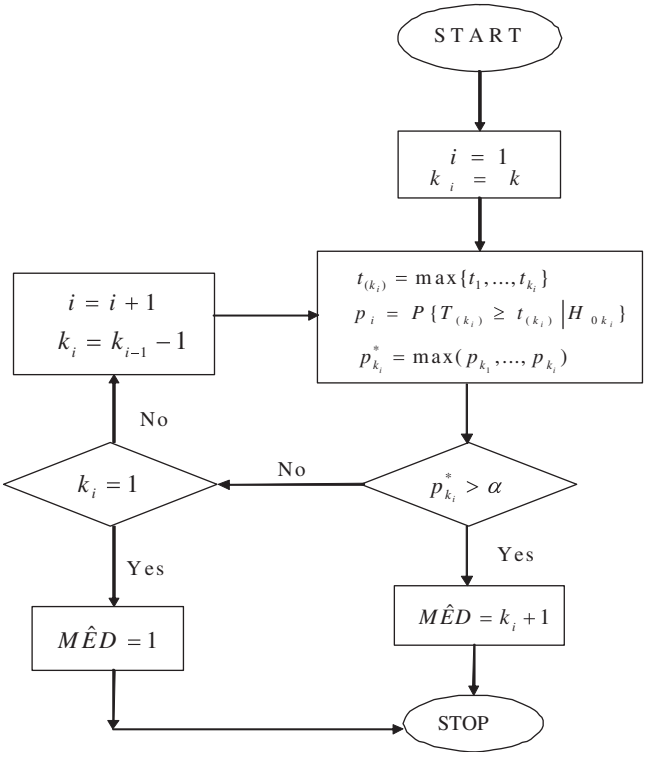

Figure 1. Flowchart for identifying the MED based on the adjusted p-values.

the identified MED. The flowchart for the closed testing scheme based on the adjusted p-value is shown in Figure 1.

3. Proposed testing procedures

3.1. *Pairwise WKM statistics*

The first type of test statistics under study is the two-sample WKM statistic [6,7] comparing the ith dose level and the zero dose control,

$$U_{0i} = \sqrt{\frac{n_0 n_i}{N}} \int_0^{t_{0i}} \hat{K}_{0i}(t)\{\hat{S}_i(t) - \hat{S}_0(t)\}\,\mathrm{d}t, \quad i = 1, 2, \ldots, k,$$

where $N = \sum_{i=0}^{k} n_i$, $t_{0i} = \sup\{t \mid \min_{j\in\{0,i\}}\{\hat{F}_j(t)\hat{S}_j(t)\} > 0\}$, $\hat{S}_i(t)$ and $\hat{F}_i(t)$ are the Kaplan–Meier [8] estimators of the survival distributions of the lifetime and censoring time in group i, respectively, and $\hat{K}_{0i}(t)$ is a random weight function given by

$$\hat{K}_{0i}(t) = \frac{\hat{F}_0^-(t)\hat{F}_i^-(t)}{\hat{\lambda}_0\hat{F}_0^-(t) + \hat{\lambda}_i\hat{F}_i^-(t)}, \quad \hat{\lambda}_i = \frac{n_i}{N}, \quad i = 0, 1, \ldots, k,$$

where $\hat{F}_i^-(t)$ is the estimated probability of not being censored before time t. Under the null hypothesis H_{0k}, a consistent estimator of the asymptotic ($N \to \infty$) variance of U_{0i},

$i = 1, 2, \ldots, k$, is obtained as

$$s_{ii}^U = -\int_0^{t_{0i}} \left(\int_t^{t_{0i}} \hat{K}_{0i}(v)\hat{S}_{0i}(v)\,\mathrm{d}v \right)^2 \frac{\hat{\lambda}_0 \hat{F}_0^-(t) + \hat{\lambda}_i \hat{F}_i^-(t)}{\hat{F}_0^-(t)\hat{F}_i^-(t)} \frac{\mathrm{d}\hat{S}_{0i}(t)}{\hat{S}_{0i}(t)\hat{S}_{0i}^-(t)}, \tag{1}$$

where $\hat{S}_{0i}$ is the Kaplan–Meier estimator computed from the combined samples of the control $i$and $\mathrm{d}\hat{S}_{0i}(t) = \hat{S}_{0i}(t) - \hat{S}_{0i}^-(t)$.

Let $U_{0i}^* = U_{0i}/\sqrt{s_{ii}^U}$, $i = 1, 2, \ldots, k$. It can be shown [13] that, under the null hypothesis H_{0k}, the asymptotic distribution of the random vector $(U_{01}^*, \ldots, U_{0k}^*)$ is the k-variate normal with mean vector zero and its associated variance–covariance matrix Γ_k^U can be consistently estimated by $\hat{\Gamma}_k^U = \{s_{ij}^U/\sqrt{s_{ii}^U s_{jj}^U}\}$, where s_{ii}^U is stated in Equation (1) and for $i \neq j$,

$$s_{ij}^U = -\sqrt{\lambda_i \lambda_j} \int_0^{t_{0ij}} \left(\int_t^{t_{0ij}} \hat{K}_{0i}(v)\hat{S}_{0ij}(v)\,\mathrm{d}v \right) \left(\int_t^{t_{0ij}} \hat{K}_{0j}(v)\hat{S}_{0ij}(v)\,\mathrm{d}v \right) \frac{\mathrm{d}\hat{S}_{0ij}(t)}{\hat{S}_{0ij}(t)\hat{S}_{0ij}^-(t)\hat{F}_0^-(t)},$$

where $t_{0ij} = \sup\{t \,|\, \min_{m\in\{0,i,j\}}\{\hat{F}_m(t)\hat{S}_m(t)\} > 0\}$, and $\hat{S}_{0ij}$ is the Kaplan–Meier [8] estimator computed from the combined samples of the control, i and j. Notice that, under equal sample sizes and equal censoring distributions, the covariance of U_{0i}^* and U_{0j}^* is equal to $1/2$ for all $i \neq j = 1, 2, \ldots, k$.

Suppose at the ith step in the closed testing scheme, the observed value of

$$U_{(k_i)}^* = \max(U_{01}^*, \ldots, U_{0k_i}^*) \tag{2}$$

is $u_{(k_i)}$, then the adjusted p-value of the test at the ith step is $p_{k_i}^* = \max(p_{k_1}, \ldots, p_{k_i})$, where for $j = 1, 2, \ldots, i$,

$$\begin{aligned} p_{k_j} &= P\{U_{(k_j)}^* \geq u_{(k_j)} | H_{0k_j}\} \approx P\{\max(Z_1, \ldots, Z_{k_j}) \geq u_{(k_j)} | H_{0k_j}\} \\ &= 1 - P\{Z_l < u_{(k_j)}, l = 1, 2, \ldots, k_j | H_{0k_j}\} \end{aligned}$$

and $(Z_1, \ldots, Z_{k_j})$ is a normal random vector with mean vector zero and variance–covariance matrix $\Gamma_{k_j}^U$.

3.2. *Combined-groups WKM statistics*

To compare the ith dose level with the combined all lower dose levels including the zero dose control, the second-type WKM statistic under consideration is

$$G_i = \sum_{m=0}^{i-1} U_{mi}, \quad i = 1, 2, \ldots, k,$$

where

$$U_{mi} = \sqrt{\frac{n_m n_i}{N}} \int_0^{t_i} \hat{K}_{mi}(t)\{\hat{S}_i(t) - \hat{S}_m(t)\}\,\mathrm{d}t, \quad \text{for } m < i,$$

and $t_i = \sup\{t \,|\, \min_{m\in\{0,1,\ldots,i\}}\{\hat{F}_m(t)\hat{S}_m(t)\} > 0\}$. Let $\sum_{i=1}^k G_i = \sum_{i=1}^k \sum_{m=0}^{i=1} U_{mi} = \sum\sum_{m<i} U_{mi}$. Since the asymptotic null distribution of the random vector $(U_{01}, \ldots, U_{0k})$ is a k-variate normal distribution, the asymptotic null distribution of $\sum_{i=1}^k G_i$ is a normal distribution. Therefore, the asymptotic null (H_{0k}) distribution of the random vector $(G_1, \ldots, G_k)$ is a k-variate normal with mean vector zero and the related variance–covariance matrix can be consistently

estimated, for $i = 1, 2, \ldots, k$, by

$$s_{il}^{G} = -\hat{\lambda}_i \sum_{m=0}^{i-1} \int_0^{t_i} \left(\int_t^{t_i} \hat{K}_{mi}(v)\hat{S}_{0+i}(v)\,\mathrm{d}v \right)^2 \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t)\hat{S}_{0+i}^{-}(t)\hat{F}_{m}^{-}(t)}$$

$$- \int_0^{t_i} \left(\sum_{m=0}^{i-1} \sqrt{\hat{\lambda}_m} \int_t^{t_i} \hat{K}_{mi}(v)\hat{S}_{0+i}(v)\,\mathrm{d}v \right)^2 \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t)\hat{S}_{0+i}^{-}(t)\hat{F}_{i}^{-}(t)},$$

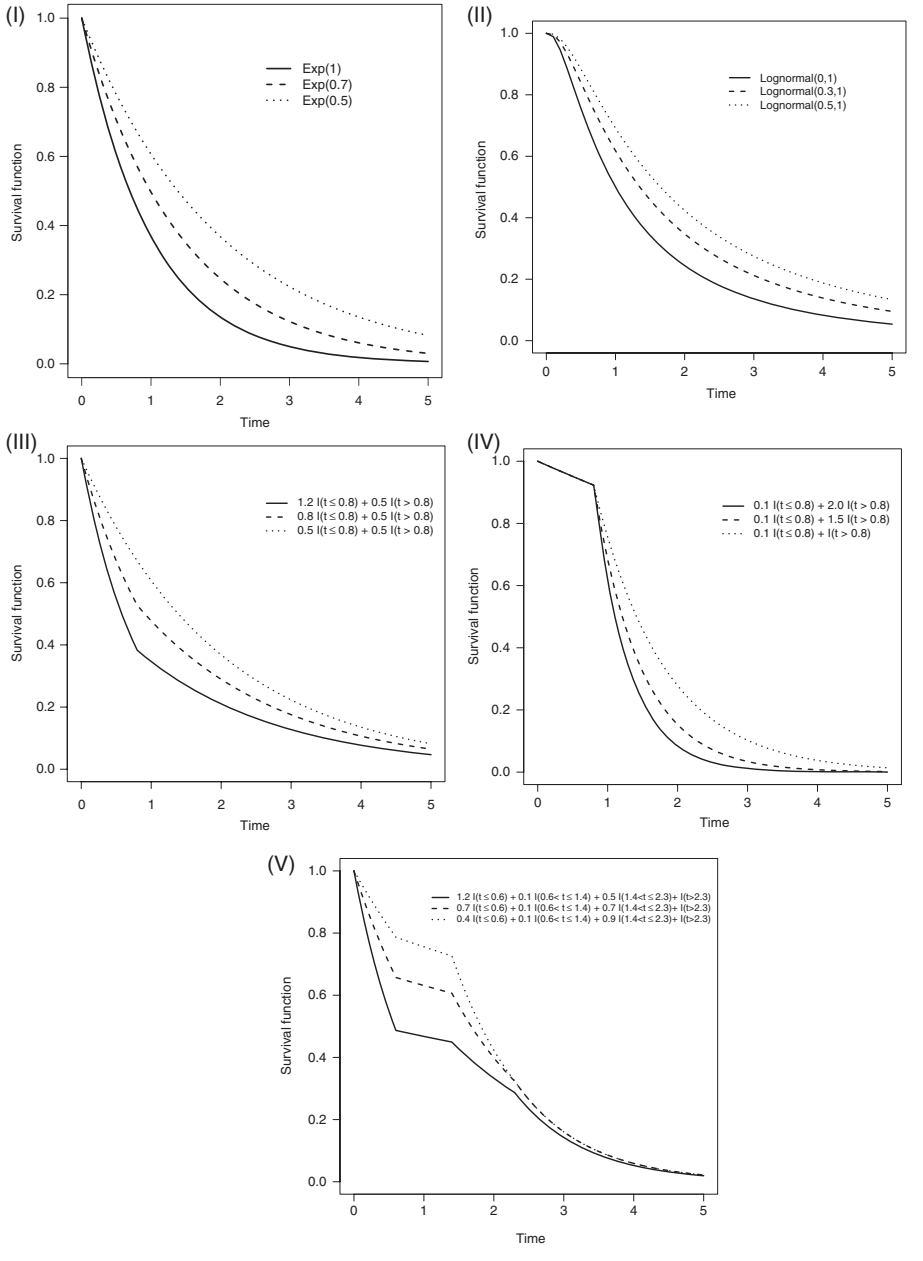

Figure 2. Survival function configurations for simulation study: (I) exponential survival functions; (II) Lognormal survival functions and (III)–(V) piecewise exponential survival functions corresponding to early, late and crossing hazard differences.

and for $i \neq j = 1, 2, \ldots, k$, by

$$s_{ij}^G = -\sqrt{\lambda_i \lambda_j} \sum_{m=0}^{i-1} \int_0^{t_j} \left(\int_t^{t_j} \hat{K}_{mi}(v) \hat{S}_{0+j}(v)\, \mathrm{d}v \right) \left(\int_t^{t_j} \hat{K}_{mj}(v) \hat{S}_{0+j}(v)\, \mathrm{d}v \right) \times \frac{\mathrm{d}\hat{S}_{0+j}(t)}{\hat{S}_{0+j}(t) \hat{S}_{0+j}^-(t) \hat{F}_m^-(t)} - \int_0^{t_j} \left(\sum_{m=0}^{i-1} \sqrt{\hat{\lambda}_m} \int_t^{t_i} \hat{K}_{mi}(v) \hat{S}_{0+j}(v)\, \mathrm{d}v \right) \times \left(\sqrt{\hat{\lambda}_j} \int_t^{t_j} \hat{K}_{ij}(v) \hat{S}_{0+j}(v)\, \mathrm{d}v \right) \frac{\mathrm{d}\hat{S}_{0+j}(t)}{\hat{S}_{0+j}(t) \hat{S}_{0+j}^-(t) \hat{F}_i^-(t)},$$

where $\hat{S}_{0+i}(t)$ is the Kaplan–Meier estimator based on the combined samples from 0 to i.

Table 1. Estimated EWEs for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3 = n$ with uniform censoring distribution $U(0, r)$.

		LR			WKM		
r	n	U^{L}	G^{L}	V^{L}	U	G	V
Exponential							
3.197	10	0.063	0.044	0.061	0.060	0.060	0.058
	30	0.052	0.044	0.050	0.053	0.052	0.054
	50	0.053	0.049	0.054	0.053	0.054	0.055
1.593	10	0.055	0.041	0.057	0.060	0.062	0.062
	30	0.051	0.044	0.054	0.052	0.054	0.056
	50	0.053	0.045	0.053	0.050	0.052	0.054
Lognormal							
4.879	10	0.059	0.045	0.057	0.056	0.059	0.058
	30	0.050	0.047	0.051	0.053	0.054	0.052
	50	0.050	0.042	0.047	0.050	0.050	0.053
2.401	10	0.056	0.042	0.054	0.056	0.061	0.061
	30	0.054	0.043	0.050	0.052	0.053	0.053
	50	0.048	0.044	0.049	0.053	0.052	0.051
Piecewise exponential with $\theta = 1.2I(t \leq 0.8) + 0.5I(t > 0.8)$							
3.655	10	0.058	0.043	0.059	0.057	0.058	0.057
	30	0.051	0.046	0.053	0.051	0.054	0.054
	50	0.052	0.046	0.051	0.051	0.052	0.053
1.457	10	0.056	0.042	0.056	0.061	0.064	0.062
	30	0.050	0.044	0.054	0.053	0.055	0.057
	50	0.051	0.046	0.054	0.050	0.051	0.054
Piecewise exponential with $\theta = 0.1I(t \leq 0.8) + 2I(t > 0.8)$							
4.095	10	0.062	0.045	0.061	0.053	0.051	0.052
	30	0.053	0.046	0.052	0.052	0.051	0.051
	50	0.055	0.049	0.054	0.055	0.053	0.055
2.423	10	0.062	0.044	0.063	0.052	0.051	0.053
	30	0.053	0.043	0.053	0.051	0.050	0.054
	50	0.054	0.047	0.057	0.053	0.054	0.056
Piecewise exponential with $\theta = 1.2I(t \leq 0.6) + 0.1I(0.6 < t \leq 1.4) + 0.5I(1.4 < t \leq 2.3) + I(t > 2.3)$							
4.618	10	0.060	0.044	0.060	0.056	0.056	0.056
	30	0.053	0.046	0.053	0.055	0.055	0.055
	50	0.054	0.049	0.051	0.054	0.054	0.054
2.182	10	0.054	0.042	0.058	0.057	0.060	0.060
	30	0.052	0.042	0.051	0.051	0.052	0.056
	50	0.051	0.047	0.043	0.052	0.055	0.054

Let $G_i^* = G_i/\sqrt{s_{ii}^G}$, $i = 1, 2, \ldots, k$. Therefore, under the null hypothesis H_{0k}, the distribution of $(G_1^*, \ldots, G_k^*)$ can be approximated by a k-variate normal with zero mean vector and the associated variance–covariance matrix Γ_k^G can be consistently estimated by $= \hat{\Gamma}_k^G = \{s_{ij}^G/\sqrt{s_{ii}^G s_{jj}^G}\}$, $i, j = 1, 2, \ldots, k$. Notice that, under equal censoring distributions, equal sample sizes and the null hypothesis H_{0k}, the statistics $(G_1, \ldots, G_k)$ are uncorrelated.

Suppose that, at the ith step of the closed testing scheme, the observed value of

$$G_{(k_i)}^* = \max(G_1^*, \ldots, G_{k_i}^*) \tag{3}$$

is $g_{(k_i)}$, and the associated adjusted p-value of the test is $p_{k_i}^* = \max(p_{k_1}, \ldots, p_{k_i})$, where, for $j = 1, 2, \ldots, i$,

$$\begin{aligned} p_{k_j} &= P\{G_{(k_j)}^* \geq g_{(k_j)} | H_{0k_j}\} \approx P\{\max(Z_1, \ldots, Z_{k_j}) \geq g_{(k_j)} | H_{0k_j}\} \\ &= 1 - P\{Z_l < g_{(k_j)}, l = 1, 2, \ldots, k_j | H_{0k_j}\} \end{aligned}$$

Table 2. Estimated FWE for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3$=50 with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^{L}	G^{L}	V^{L}	U	G	V
(I) *Exponential and* $r = 3.197$										
1	1	1	0.5	3	0.054	0.048	0.055	0.054	0.053	0.057
1	1	0.5	0.5	2	0.053	0.049	0.054	0.053	0.050	0.053
1	1	0.7	0.5	2	0.041	0.031	0.044	0.040	0.031	0.043
1	1	0.5	0.7	2	0.050	0.046	0.050	0.049	0.044	0.048
1	1	0.5	1	2	0.051	0.046	0.047	0.052	0.047	0.047
(II) *Lognormal with* $r = 4.879$										
0	0	0	0.5	3	0.049	0.042	0.050	0.049	0.048	0.052
0	0	0.5	0.5	2	0.050	0.041	0.051	0.048	0.042	0.049
0	0	0.3	0.5	2	0.045	0.029	0.046	0.043	0.030	0.044
0	0	0.5	0.3	2	0.047	0.038	0.048	0.051	0.041	0.051
0	0	0.5	0	2	0.044	0.035	0.039	0.047	0.038	0.040
(III) *Piecewise exponential and* $r = 3.655$										
1.2	1.2	1.2	0.5	3	0.053	0.048	0.053	0.051	0.051	0.053
1.2	1.2	0.5	0.5	2	0.051	0.043	0.052	0.051	0.045	0.052
1.2	1.2	0.8	0.5	2	0.043	0.030	0.045	0.043	0.031	0.045
1.2	1.2	0.5	0.8	2	0.053	0.045	0.054	0.053	0.046	0.052
1.2	1.2	0.5	1.2	2	0.044	0.036	0.040	0.046	0.039	0.040
(IV) *Piecewise exponential and* $r = 4.095$										
2	2	2	1	3	0.054	0.048	0.054	0.051	0.046	0.052
2	2	1	1	2	0.053	0.048	0.054	0.049	0.042	0.051
2	2	1.5	1	2	0.040	0.030	0.041	0.037	0.027	0.037
2	2	1	1.5	2	0.049	0.043	0.049	0.048	0.041	0.047
2	2	1	2	2	0.046	0.041	0.042	0.044	0.038	0.038
(V) *Piecewise exponential and* $r = 4.618$										
1.2	1.2	1.2	0.4[a]	3	0.046	0.040	0.051	0.048	0.048	0.051
0.5	0.5	0.5	0.9							
1.2	1.2	0.4	0.4[a]	2	0.049	0.035	0.052	0.050	0.044	0.051
0.5	0.5	0.9	0.9							
1.2	1.2	0.7	0.4[a]	2	0.041	0.026	0.043	0.045	0.034	0.047
0.5	0.5	0.7	0.9							
1.2	1.2	0.4	0.7[a]	2	0.047	0.032	0.049	0.052	0.043	0.052
0.5	0.5	0.9	0.7							
1.2	1.2	0.4	1.2[a]	2	0.048	0.034	0.042	0.049	0.043	0.043
0.5	0.5	0.9	0.5							

[a]The first row is θ_{1i} and the second row is θ_{2i}.

and $(Z_1, \ldots, Z_{k_j})$ is a normal random vector with mean vector zero and variance–covariance matrix $\Gamma^G_{k_j}$.

3.3. *Step-type WKM statistics*

For $1 \le j \le i \le k$, the third test considered is the step-type WKM statistic, which compares the involved higher $(i - j + 1)$ dose levels with all the lower j dose levels including the zero dose control and is given by

$$V_j^{(i)} = \sum_{m=0}^{j-1} \sum_{l=j}^{i} U_{ml}, \quad 1 \le j \le i \le k.$$

Notice that $V_j^{(i)} = G_i$ if $j = i$. Based on the asymptotic result [13] of the two-sample WKM statistic, the variance and covariance of $V_j^{(i)}$ can be consistently estimated, for $1 \le m \ne l \le i$, by

$$s_{mm}^{V^{(i)}} = -\sum_{j=0}^{m-1} \int_0^{t_i} \left(\sum_{h=m}^{i} \sqrt{\hat{\lambda}_h} \int_t^{t_i} \hat{K}_{jh}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right)^2 \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t) \hat{S}^-_{0+i}(t) \hat{F}^-_j(t)}$$
$$- \sum_{j=m}^{i} \int_0^{t_i} \left(\sum_{h=0}^{m-1} \sqrt{\hat{\lambda}_h} \int_t^{t_i} \hat{K}_{jh}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right)^2 \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t) \hat{S}^-_{0+i}(t) \hat{F}^-_j(t)}$$

Table 3. Estimated power for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3 = 50$ with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^{L}	G^{L}	V^{L}	U	G	V
(I) *Exponential and* $r = 3.197$										
1	1	1	0.5	3	0.669	0.814	0.834	0.653	0.816	0.837
1	1	0.5	0.5	2	0.695	0.786	0.816	0.680	0.788	0.811
1	0.5	0.5	0.5	1	0.806	0.744	0.800	0.792	0.735	0.782
1	1	0.7	0.5	2	0.264	0.322	0.374	0.259	0.336	0.380
1	0.7	0.7	0.5	1	0.339	0.281	0.325	0.335	0.279	0.311
Average (ordered)					0.555	0.589	0.630	0.544	0.591	0.624
1	1	0.5	0.7	2	0.680	0.779	0.745	0.665	0.780	0.735
1	1	0.5	1	2	0.673	0.774	0.440	0.657	0.773	0.429
1	0.5	0.5	0.7	1	0.792	0.734	0.734	0.780	0.723	0.706
1	0.5	0.7	1	1	0.728	0.698	0.370	0.716	0.679	0.344
1	0.5	1	1	1	0.724	0.702	0.188	0.705	0.682	0.172
Average (umbrella)					0.719	0.737	0.495	0.705	0.727	0.477
Average (overall)					0.637	0.663	0.563	0.624	0.659	0.551
(II) *Lognormal with* $r = 4.879$										
0	0	0	0.5	3	0.469	0.613	0.658	0.486	0.665	0.703
0	0	0.5	0.5	2	0.505	0.591	0.643	0.520	0.639	0.679
0	0.5	0.5	0.5	1	0.608	0.528	0.600	0.624	0.552	0.609
0	0	0.3	0.5	2	0.220	0.258	0.312	0.221	0.288	0.334
0	0.3	0.3	0.5	1	0.279	0.222	0.272	0.289	0.232	0.271
Average (ordered)					0.416	0.442	0.497	0.428	0.475	0.519
0	0	0.5	0.3	2	0.485	0.579	0.563	0.501	0.622	0.585
0	0	0.5	0	2	0.474	0.568	0.294	0.498	0.620	0.306
0	0.5	0.5	0.3	1	0.588	0.519	0.536	0.616	0.546	0.544
0	0.5	0.3	0	1	0.531	0.492	0.275	0.561	0.518	0.275
0	0.5	0	0	1	0.529	0.504	0.136	0.532	0.505	0.127
Average (umbrella)					0.521	0.532	0.361	0.542	0.562	0.367
Average (overall)					0.469	0.487	0.429	0.485	0.519	0.443

and

$$
\begin{aligned}
s_{ml}^{V^{(i)}} = &- \sum_{j=0}^{m-1} \int_0^{t_i} \left(\sum_{h=m}^{i} \sqrt{\hat{\lambda}_h} \int_t^{t_i} \hat{K}_{jh}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right) \left(\sum_{q=l}^{i} \sqrt{\hat{\lambda}_q} \int_t^{t_i} \hat{K}_{jq}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right) \\
&\times \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t) \hat{S}_{0+i}^{-}(t) \hat{F}_j^{-}(t)} + \sum_{j=m}^{l-1} \int_0^{t_i} \left(\sum_{h=0}^{m-1} \sqrt{\hat{\lambda}_h} \int_t^{t_i} \hat{K}_{jh}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right) \\
&\times \left(\sum_{q=l}^{i} \sqrt{\hat{\lambda}_q} \int_t^{t_i} \hat{K}_{jq}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right) \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t) \hat{S}_{0+i}^{-}(t) \hat{F}_j^{-}(t)} \\
&- \sum_{j=l}^{i} \int_0^{t_i} \left(\sum_{h=0}^{m-1} \sqrt{\hat{\lambda}_h} \int_t^{t_i} \hat{K}_{jh}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right) \left(\sum_{q=0}^{l-1} \sqrt{\hat{\lambda}_q} \int_t^{t_i} \hat{K}_{jq}(v) \hat{S}_{0+i}(v)\, \mathrm{d}v \right) \\
&\times \frac{\mathrm{d}\hat{S}_{0+i}(t)}{\hat{S}_{0+i}(t) \hat{S}_{0+i}^{-}(t) \hat{F}_j^{-}(t)},
\end{aligned}
$$

Table 4. Estimated power for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3 = 50$ with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^{L}	G^{L}	V^{L}	U	G	V
(III) *Piecewise exponential and* $r = 3.655$										
1.2	1.2	1.2	0.5	3	0.569	0.709	0.749	0.614	0.780	0.807
1.2	1.2	0.5	0.5	2	0.603	0.687	0.736	0.647	0.754	0.783
1.2	0.5	0.5	0.5	1	0.712	0.633	0.712	0.757	0.687	0.748
1.2	1.2	0.8	0.5	2	0.204	0.238	0.288	0.214	0.274	0.316
1.2	0.8	0.8	0.5	1	0.274	0.221	0.265	0.285	0.237	0.266
Average (ordered)					0.472	0.498	0.550	0.503	0.546	0.584
1.2	1.2	0.5	0.8	2	0.566	0.666	0.620	0.609	0.734	0.665
1.2	1.2	0.5	1.2	2	0.580	0.680	0.363	0.619	0.744	0.391
1.2	0.5	0.5	0.8	1	0.686	0.621	0.625	0.733	0.670	0.654
1.2	0.5	0.8	1.2	1	0.631	0.602	0.310	0.676	0.644	0.305
1.2	0.5	1.2	1.2	1	0.625	0.603	0.161	0.664	0.643	0.158
Average (umbrella)					0.618	0.634	0.416	0.660	0.687	0.435
Average (overall)					0.545	0.566	0.483	0.582	0.617	0.509
(IV) *Piecewise exponential and* $r = 4.095$										
2	2	2	1	3	0.588	0.751	0.777	0.503	0.682	0.714
2	2	1	1	2	0.617	0.725	0.754	0.534	0.653	0.680
2	1	1	1	1	0.731	0.655	0.722	0.646	0.567	0.610
2	2	1.5	1	2	0.154	0.191	0.230	0.128	0.172	0.200
2	1.5	1.5	1	1	0.206	0.165	0.188	0.172	0.138	0.144
Average (ordered)					0.459	0.497	0.534	0.397	0.442	0.470
2	2	1	1.5	2	0.583	0.698	0.613	0.496	0.631	0.523
2	2	1	2	2	0.594	0.714	0.415	0.509	0.638	0.355
2	1	1	1.5	1	0.716	0.647	0.619	0.631	0.555	0.509
2	1	1.5	2	1	0.654	0.619	0.287	0.564	0.531	0.225
2	1	1.5	1.5	1	0.644	0.620	0.167	0.555	0.531	0.135
Average (umbrella)					0.638	0.660	0.420	0.551	0.577	0.349
Average (overall)					0.549	0.579	0.477	0.474	0.510	0.410

where $\hat{S}_{0+i}(t)$ is the pooled Kaplan–Meier survival estimator based on the combined samples from 0 to i.

Let $V_j^{(i)*} = V_j^{(i)}/\sqrt{s_{jj}^{V(i)}}$, $1 \le j \le i$. Then, under the null hypothesis H_{0i}, the distribution of $(V_1^{(i)*}, \ldots, V_i^{(i)*})$ is asymptotically normal with zero mean vector and the corresponding variance–covariance matrix Γ_i^V can be consistently estimated by $\hat{\Gamma}_i^V = \{s_{ml}^{V^{(i)}}/\sqrt{s_{mm}^{V^{(i)}} s_{ll}^{V^{(i)}}}\}$, $1 \le m \ne l \le i$. Notice that, under equal sample sizes and equal censoring distributions, the covariance between $V_m^{(i)*}$ and $V_l^{(i)*}$ is reduced to $r_{ml}^{V^{(i)}} = \sqrt{m(i-l+1)/l(i-m+1)}$, for $1 \le m \ne l \le i$.

At step i of the step-down testing procedure, the test statistic involved is

$$V_{(k_i)}^* = \max(V_1^{(k_i)*}, \ldots, V_{k_i}^{(k_i)*}). \tag{4}$$

Suppose that the observed value of $V_{(k_i)}^*$ is $v_{(k_i)}$, then the associated adjusted p-value of the test is given by $p_{k_i}^* = \max(p_{k_1}, \ldots, p_{k_i})$, where for $j = 1, 2, \ldots, i$,

$$\begin{aligned} p_{k_j} &= P\{V_{(k_j)}^* \ge v_{(k_j)} | H_{0k_j}\} \approx P\{\max(Z_1, \ldots, Z_{k_j}) \ge v_{(k_j)} | H_{0k_j}\} \\ &= 1 - P\{Z_m < v_{(k_j)}, m = 1, 2, \ldots, k_j | H_{0k_j}\}, \end{aligned}$$

and $(Z_1, \ldots, Z_{k_j})$ is a normal random vector with mean vector zero and variance–covariance matrix $\Gamma_{k_j}^V$.

Hence, the MED can be identified using the step-down testing procedure described in Section 2 or the flowchart in Figure 1, based on the proposed adjusted p-values [12]. Notice that all the p-values of the proposed testing procedures stated above can be computed by using the programme in Gassmann *et al*. [14]. The *mvtnorm* package of Genz and Bretz [15] in R [16] software also provides the required integration technique for free and is easily accessible.

Table 5. Estimated power for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3 = 50$ with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^{L}	G^{L}	V^{L}	U	G	V
(V) *Piecewise exponential and* $r = 4.618$										
1.2	1.2	1.2	0.4[a]	3	0.322	0.411	0.473	0.548	0.727	0.763
0.5	0.5	0.5	0.9							
1.2	1.2	0.4	0.4[a]	2	0.344	0.391	0.460	0.574	0.687	0.729
0.5	0.5	0.9	0.9							
1.2	0.4	0.4	0.4[a]	1	0.432	0.361	0.428	0.684	0.618	0.678
0.5	0.9	0.9	0.9							
1.2	1.2	0.7	0.4[a]	2	0.158	0.173	0.229	0.233	0.303	0.351
0.5	0.5	0.7	0.9							
1.2	0.7	0.7	0.4[a]	1	0.216	0.168	0.210	0.303	0.252	0.291
0.5	0.7	0.7	0.9							
Average (ordered)					0.294	0.301	0.360	0.468	0.517	0.562
1.2	1.2	0.4	0.7[a]	2	0.329	0.388	0.389	0.561	0.677	0.636
0.5	0.5	0.9	0.7							
1.2	1.2	0.4	1.2[a]	2	0.323	0.378	0.201	0.546	0.667	0.314
0.5	0.5	0.9	0.5							
1.2	0.4	0.4	0.7[a]	1	0.424	0.358	0.384	0.665	0.601	0.597
0.5	0.9	0.9	0.7							
1.2	0.4	0.7	1.2[a]	1	0.381	0.343	0.206	0.616	0.579	0.284
0.5	0.9	0.7	0.5							
1.2	0.4	1.2	1.2[a]	1	0.362	0.341	0.077	0.597	0.573	0.106
0.5	0.9	0.9	0.5							
Average (umbrella)					0.364	0.362	0.251	0.597	0.619	0.387
Average (overall)					0.329	0.332	0.306	0.533	0.568	0.475

[a]The first row is θ_{1i} and the second row is θ_{2i}.

4. A simulation study

We conduct a simulation study to investigate the relative error rate and power/bias performances of the proposed testing procedures based on the pairwise WKM statistics in Equation (2), combined-groups WKM statistics in Equation (3) and the step-type WKM statistics in Equation (4), denoted by U, G and V, respectively, as well as the three LR statistics proposed in Chen and Chang [3], denoted by U^L, G^L and V^L. The error rate properties of the tests are evaluated by the experimentwise error rate (EWE, the probability of incorrectly declaring MED under the global null hypothesis H_{0k}) and familywise error rate (FWE, the probability of underestimating the MED). The power and bias performances of the tests are assessed by the probability of correctly identifying the MED and E(M$\hat{\text{E}}$D)–MED. In this study, we consider $k = 3$ dose levels (excluding control) with sample size $n_0 = n_1 = n_2 = n_3 = 10$, 30 and 50 in the EWE study and $n = 50$ in the FWE and power/bias study. The exponential, lognormal and piecewise exponential distributions are considered as survival time distributions and the uniform distribution over $(0, r)$ is used as the censoring distribution.

In the EWE study, the common survival distribution for each group is exponential distribution with scale parameter $\theta = 1$ and lognormal survival distribution with zero normal mean and standard deviation 1.0. Moreover, the piecewise exponential distributions with different hazards at different time periods are also investigated. Thus, various values of r, which correspond to

Table 6. Estimated bias for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3$=50 with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^{L}	G^{L}	V^{L}	U	G	V
(I) *Exponential and* $r = 3.197$										
1	1	1	0.5	3	0.189	0.063	0.033	0.206	0.051	0.028
1	1	0.5	0.5	2	0.341	0.212	0.113	0.365	0.200	0.126
1	0.5	0.5	0.5	1	0.396	0.660	0.416	0.432	0.673	0.468
1	1	0.7	0.5	2	0.915	0.843	0.653	0.935	0.811	0.654
1	0.7	0.7	0.5	1	1.448	1.745	1.461	1.475	1.726	1.505
Average (ordered)					0.658	0.705	0.535	0.683	0.692	0.556
1	1	0.5	0.7	2	0.473	0.298	0.332	0.507	0.296	0.360
1	1	0.5	1	2	0.502	0.313	0.977	0.531	0.314	0.999
1	0.5	0.5	0.7	1	0.474	0.709	0.709	0.505	0.732	0.803
1	0.5	0.7	1	1	0.798	0.901	1.888	0.834	0.958	1.965
1	0.5	1	1	1	0.827	0.895	2.435	0.886	0.953	2.483
Average (umbrella)					0.615	0.623	1.268	0.653	0.651	1.322
Average (overall)					0.636	0.664	0.902	0.668	0.671	0.939
(II) *Lognormal with* $r = 4.879$										
0	0	0	0.5	3	0.404	0.281	0.221	0.384	0.213	0.169
0	0	0.5	0.5	2	0.712	0.602	0.402	0.687	0.498	0.349
0	0.5	0.5	0.5	1	0.933	1.277	0.962	0.892	1.195	0.95
0	0	0.3	0.5	2	1.123	1.101	0.855	1.113	1.012	0.808
0	0.3	0.3	0.5	1	1.753	2.067	1.749	1.725	1.995	1.744
Average (ordered)					0.985	1.066	0.838	0.960	0.983	0.804
0	0	0.5	0.3	2	0.856	0.709	0.667	0.811	0.607	0.621
0	0	0.5	0	2	0.919	0.760	1.290	0.864	0.645	1.263
0	0.5	0.5	0.3	1	1.041	1.333	1.261	0.969	1.249	1.252
0	0.5	0.3	0	1	1.363	1.505	2.168	1.276	1.428	2.169
0	0.5	0	0	1	1.411	1.488	2.587	1.403	1.483	2.614
Average (umbrella)					1.118	1.159	1.595	1.065	1.082	1.584
Average (overall)					1.052	1.112	1.216	1.012	1.033	1.194

Note: The average bias is the mean absolute bias.

the probability of censorship as 0.3 or 0.5, are considered in the EWE study. As shown in Figure 2, the solid line in each panel represents the common survival function under the null hypothesis. Appropriate uniform, exponential and lognormal variates were generated by using the IMSL (International Mathematical and Statistical Library) routines RNUN, ENEXP and RNLNL, respectively [17].

In the FWE and power/bias study, we use the (I) exponential distributions with various values of scale parameters θ_i, $i = 0, 1, \ldots, k$ and the (II) lognormal distributions with same standard deviation 1.0 and various mean values θ_i, $i = 0, 1, \ldots, k$. Notice that exponential distributions have the proportional hazards, but the hazards of lognormal distributions differ in early time. In addition, a variety of piecewise exponential distributions with different values of θ_i, $i = 0, 1, \ldots, k$, at different time periods corresponding to early, late and crossing hazard differences alternatives, respectively, are also considered as follows:

(III) $\theta_i I(t \leq 0.8) + 0.5 I(t > 0.8)$,
(IV) $0.1 I(t \leq 0.8) + \theta_i I(t > 0.8)$,
(V) $\theta_{1i} I(t \leq 0.6) + 0.1 I(0.6 < t \leq 1.4) + \theta_{2i} I(1.4 < t \leq 2.3) + I(t > 2.3)$.

Notice that the dose–response configurations under study include both the ordered and umbrella patterns of treatment effects with either step or linear type of increments or decrements. Also notice that, under different survival distributions, the related value of r corresponding to the probability of censorship as 0.3 in the EWE study was used in the FWE and power/bias study.

Table 7. Estimated bias for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3 = 50$ with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^L	G^L	V^L	U	G	V
(III) *Piecewise exponential and* $r = 3.655$										
1.2	1.2	1.2	0.5	3	0.290	0.167	0.120	0.249	0.091	0.064
1.2	1.2	0.5	0.5	2	0.525	0.415	0.246	0.440	0.278	0.174
1.2	0.5	0.5	0.5	1	0.666	0.995	0.667	0.540	0.821	0.574
1.2	1.2	0.8	0.5	2	1.065	1.039	0.822	1.017	0.919	0.750
1.2	0.8	0.8	0.5	1	1.704	1.989	1.707	1.641	1.871	1.651
Average (ordered)					0.850	0.921	0.712	0.777	0.796	0.643
1.2	1.2	0.5	0.8	2	0.690	0.522	0.564	0.611	0.382	0.483
1.2	1.2	0.5	1.2	2	0.707	0.530	1.152	0.625	0.394	1.097
1.2	0.5	0.5	0.8	1	0.791	1.047	1.035	0.656	0.893	0.960
1.2	0.5	0.8	1.2	1	1.089	1.187	2.065	0.957	1.063	2.082
1.2	0.5	1.2	1.2	1	1.126	1.192	2.513	1.007	1.071	2.522
Average (umbrella)					0.881	0.896	1.466	0.771	0.761	1.429
Average (overall)					0.865	0.908	1.089	0.774	0.778	1.036
(IV) *Piecewise exponential and* $r = 4.095$										
2	2	2	1	3	0.270	0.126	0.090	0.366	0.202	0.163
2	2	1	1	2	0.474	0.321	0.205	0.632	0.457	0.336
2	1	1	1	1	0.566	0.901	0.611	0.771	1.128	0.915
2	2	1.5	1	2	1.111	1.030	0.868	1.222	1.107	0.971
2	1.5	1.5	1	1	1.831	2.029	1.826	1.979	2.130	1.998
Average (ordered)					0.850	0.881	0.720	0.994	1.005	0.877
2	2	1	1.5	2	0.671	0.467	0.602	0.844	0.605	0.780
2	2	1	2	2	0.673	0.450	1.041	0.850	0.607	1.171
2	1	1	1.5	1	0.666	0.950	1.049	0.897	1.207	1.376
2	1	1.5	2	1	1.024	1.139	2.134	1.281	1.399	2.322
2	1	1.5	1.5	1	1.066	1.139	2.494	1.332	1.407	2.590
Average (umbrella)					0.820	0.829	1.464	1.041	1.045	1.648
Average (overall)					0.835	0.855	1.092	1.017	1.025	1.262

Therefore, the $(k+1)$ groups under study may suffer different probabilities of censorship ranging from 0.3 to 0.5.

For each of these settings, 10,000 replicates are used to obtain the estimated EWE, FWE, power and bias under the nominal level $\alpha = 0.05$. Therefore, the standard deviation of the error rate estimator is about $0.002(\approx \sqrt{(0.05)(0.95)/10,000})$ and the maximum standard error for the power estimator is about $0.005(\approx \sqrt{(0.5)(0.5)/10,000})$. The estimated EWE for the group of LR tests and the group of WKM tests are presented in Table 1, and the estimates of FWE, power and bias for $k = 3$ and $n = 50$ are reported in Tables 2–8. Notice that, when the true MED is one, the FWE is zero since no type I error is involved. Therefore, we leave the FWE blank for the situation with MED $= 1$ in Table 2. Also notice that, in the power/bias study, we estimate the MED to be four if the MED is identified to be beyond $k = 3$.

Upon examination of the results in Table 1, it is evident that all the tests maintain or exceed their EWE only insubstantially when the sample size is 30 or larger. When the sample size is 10, the combined-groups LR tests tend to be conservative, while the pairwise LR, step-type LR and WKM tests do not maintain their EWE. Moreover, the results in Table 2 confirm that the FWE of all the tests are controlled.

The power and bias estimates presented in Tables 3–5 and 6–8, respectively, reveal that under exponential proportional hazards alternative, the WKM and LR tests are competitive, while the WKM test performs better than the LR test under early and crossing hazards difference alternatives but worse than the LR test under late hazards difference alternatives. Moreover, the step-type test outperforms the pairwise and combined-groups tests in power and bias performances for ordered dose–response, but it performs poorly for umbrella-patterned dose–response, especially for the case with a dramatic down turn. In general, when the true MED is the first dose level, the power or bias performance of the pairwise test is better than that of the combined-groups test. When the

Table 8. Estimated bias for $\alpha = 0.05$, $n_0 = n_1 = n_2 = n_3 = 50$ with uniform censoring distribution $U(0, r)$.

				True	LR			WKM		
θ_0	θ_1	θ_2	θ_3	MED	U^L	G^L	V^L	U	G	V
(V) *Piecewise exponential and* $r = 4.618$										
1.2	1.2	1.2	0.4[a]	3	0.322	0.411	0.473	0.326	0.153	0.113
0.5	0.5	0.5	0.9							
1.2	1.2	0.4	0.4[a]	2	0.556	0.487	0.401	0.586	0.410	0.269
0.5	0.5	0.9	0.9							
1.2	0.4	0.4	0.4[a]	1	1.060	1.033	0.759	0.742	1.020	0.762
0.5	0.9	0.9	0.9							
1.2	1.2	0.7	0.4[a]	2	1.483	1.803	1.505	1.059	0.942	0.750
0.5	0.5	0.7	0.9							
1.2	0.7	0.7	0.4[a]	1	1.357	1.388	1.113	1.658	1.918	1.663
0.5	0.7	0.7	0.9							
Average (ordered)					0.956	1.024	0.850	0.874	0.889	0.711
1.2	1.2	0.4	0.7[a]	2	1.166	1.107	1.005	0.700	0.501	0.531
0.5	0.5	0.9	0.7							
1.2	1.2	0.4	1.2[a]	2	1.210	1.141	1.466	0.759	0.539	1.243
0.5	0.5	0.9	0.5							
1.2	0.4	0.4	0.7[a]	1	1.558	1.831	1.727	0.855	1.096	1.120
0.5	0.9	0.9	0.7							
1.2	0.4	0.7	1.2[a]	1	1.814	1.952	2.367	1.129	1.251	2.145
0.5	0.9	0.7	0.5							
1.2	0.4	1.2	1.2[a]	1	1.912	1.976	2.763	1.208	1.279	2.679
0.5	0.9	0.9	0.5							
Average (umbrella)					1.532	1.601	1.866	0.930	0.933	1.544
Average (overall)					1.244	1.313	1.358	0.902	0.911	1.128

[a]The first row is θ_{1i} and the second row is θ_{2i}.

true MED is higher than the first dose level, however, the combined-groups test is superior to the pairwise test.

5. An example

The data from a randomized and double-blinded clinical trial involving patients aged greater than 60 with stage IV prostate cancer having evidence of distant metastasis [10] are used for illustration. Among the 176 patients, 47 took the placebo pill as the control group, 43 received 0.2 mg of diethylstilbestrol (DES), 40 were treated with 1.0 mg of DES and 46 took 5.0 mg of DES. All drugs were administered daily orally and the survival times until death were recorded. For patients who had not yet died or were dead due to causes other than the prostate cancer, the observed times were regarded as censored data. Herein, we are concerned with the minimum dosage of DES under which the survival rate of the involved patients is larger than that of the patients taking the placebo pill. The Kaplan–Meier [8] estimates for the survival functions of the four groups are shown in Figure 3. The relevant summary statistics, the identified MED and the associated adjusted p-values are then given in Table 9.

To apply the step-down closed testing procedure based on the WKM statistics, we have $k_1 = 3$ in this case. We first compute $u_{(3)} = 1.437$ and find its p-value as $p_3 = P\{U^*_{(3)} \geq 1.437|H_{03}\} = 0.171$. Since it is greater than the predetermined nominal level $\alpha = 0.05$, we identify the MED to be higher than 5.0 mg of DES. The combined-groups WKM statistic is obtained as $g_{(3)} = 1.620$ together with the p-value $p_3 = P\{G^*_{(3)} \geq 1.620|H_{03}\} = 0.149$. Hence, the combined-groups WKM statistics identify the MED to be larger than 5.0 mg of DES. Finally, to identify the MED based on the step-type WKM statistics, we observe $v_{(3)} = 2.115$ and $p_3 = P\{V^*_{(3)} \geq 2.115|H_{03}\} = 0.043$, which leads to the next step with $k_2 = 2$. Since $v_{(2)} = 1.440$ with $p_2 = P\{V^*_{(2)} \geq 1.440|H_{02}\} = 0.130$, we stop the procedure and identify the MED to be 5.0 mg of DES with the adjusted p-value as 0.043.

Notice that the adjusted p-values of the corresponding LR statistics are 0.251, 0.263 and 0.080, respectively. This would lead to the conclusion that under the significance level 0.05,

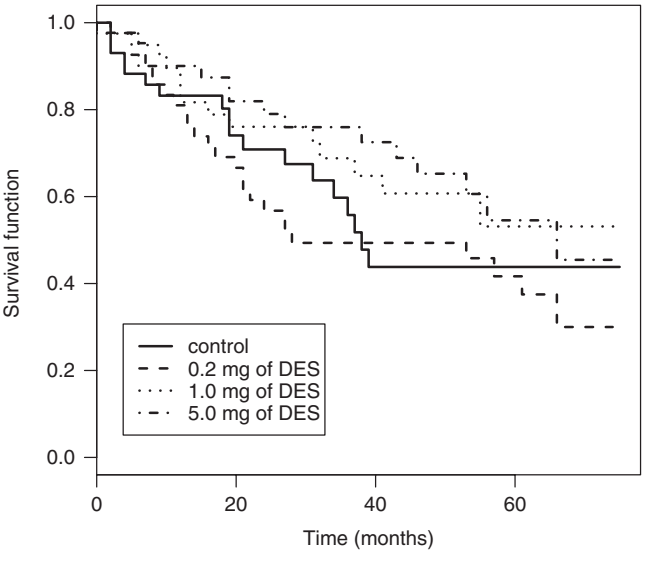

Figure 3. Kaplan–Meier estimate of survival functions for the prostate cancer data.

Table 9. Summary statistics for the stage IV prostate cancer data.

	i	LR	WKM
$U_i^*(U_i, s_{ii}^U)$	1	−0.667	−0.459
	2	0.932	0.966
	3	1.178	1.437
MÊD		4	4
Adjusted p-value		0.251	0.171
$G_i^*(G_i, s_{ii}^G)$	1	−0.667	−0.459
	2	1.300	1.440
	3	1.278	1.620
MÊD		4	4
Adjusted p-value		0.263	0.149
$V_i^{(3)*}(V_i^{(3)}, s_{ii}^{V^{(3)}})$	1	0.446	0.726
	2	1.823	2.115
	3	1.278	1.620
$V_i^{(2)*}(V_i^{(2)}, s_{ii}^{V^{(2)}})$	1	0.011	0.235
	2	1.300	1.440
$V_i^{(1)*}(V_i^{(1)}, s_{ii}^{V^{(1)}})$	1	−0.667	−0.459
MÊD		4	3
Adjusted p-value		0.080	0.043

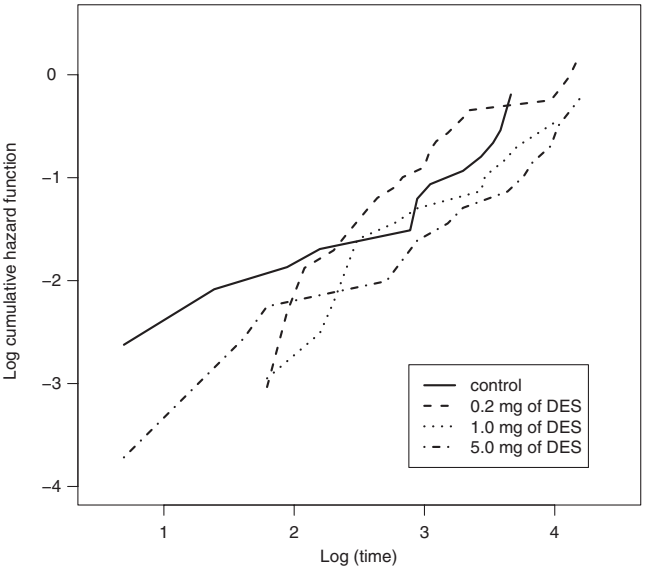

Figure 4. Log cumulative hazard functions for the prostate cancer data.

the MED of the DES in the treatment of the prostatic patients is beyond the largest dosage under study. In fact, the log cumulative hazard functions in Figure 4 cross each other. Therefore, it is not surprising that the LR tests produce the larger p-values than the associated WKM tests.

6. Conclusions

To sum up, both the group of WKM tests and the group of LR tests are recommended for identifying the MED in a dose–response study with proportional hazards. However, the WKM test is suggested when hazards are crossing or different at early times, while the LR test is suggested when hazards are different at late times. For the choice of types of testing procedures, the step-type test is recommended for ordered dose–responses. For umbrella pattern dose–responses, however, the pairwise or the combined-groups test is recommended. In this case, the pairwise test is preferred if the true MED is assumed to be relatively close to the first dose level, otherwise the combined-groups test is suggested.

Notice that, the weight function in this paper, $\hat{K}(t)$, in the WKM statistic is essentially used to down weigh the variance of the difference between two estimated survival functions over later time periods, especially, for heavy censoring. If some specific types of alternatives can be expected, different weight functions such as the ones in [18], $\hat{K}(t)\hat{S}_{(t)}^{-\rho}\{1-\hat{S}_{(t)}^{-}\}^{\gamma}$, can be employed for each individual comparison to gain better power. However, if these weight functions are not favoured, the difference between two restricted mean lifetimes [19] can also be considered.

Acknowledgements

We would like to thank the two reviewers for their constructive and detailed suggestions which greatly improved the presentation of the results in this paper. The work of the first author was supported by the National Science Council of Taiwan under Grant NSC96-2118-M-008-029-006 and the work of the second author was supported under NSC91-2118-M-008-001 and NSC92-2118-M-008-011.

References

[1] S.J. Ruberg, *Dose response studies. I. Some design considerations*, J. Biopharm. Stat. 5 (1995), pp. 1–14.
[2] T.R. Fleming and D.P. Harrington, *Counting Process and Survival Analysis*, Wiley, New York, 1991.
[3] Y.I. Chen and Y.M. Chang, *Identification of the minimum effective dose for right-censored survival data*, Comput. Stat. Data Anal. 51 (2007), pp. 3213–3222.
[4] W. Nelson, *Theory and applications of hazard plotting for censored failure data*, Technometrics 14 (1972), pp. 945–965.
[5] O.O. Aalen, *Nonparametric inference for a family of counting processes*, *Ann. Stat.* 6 (1978), pp. 701–726.
[6] M.S. Pepe and T.R. Fleming, *Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data*, Biometrics 45 (1989), pp. 497–507.
[7] M.S. Pepe and T.R. Fleming, *Weighted Kaplan–Meier statistics: A large sample and optimality considerations*, J. R. Stat. Soc. Ser. B 53 (1991), pp. 341–352.
[8] E.L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Am. Stat. Assoc. 53 (1958), pp. 457–481.
[9] A.C. Tamhane, Y. Hochberg, and C.W. Dunnett, *Multiple test procedures for dose finding*, Biometrics 52 (1996), pp. 21–37.
[10] D.P. Byar and D.K. Corle, *Selecting optimal treatment in clinical trials using covariate information*, Chronic Dis. 30 (1977), pp. 445–459.
[11] R. Marcus, E. Peritz, and K.R. Gabriel, *On closed testing procedures with special reference to ordered analysis of variance*, Biometrika 63 (1976), pp. 655–660.
[12] S.P. Wright, *Adjusted p-values for simultaneous inference*, Biometrics 48 (1992), pp. 1005–1013.
[13] Y. Chi, *Multiple testing procedures based on weighted Kaplan–Meier statistics for right censored survival data*, Stat. Med. 24 (2005), pp. 23–35.
[14] H.I. Gassmann, I. Deák, and T. Szántai, *Computing multivariate normal probabilities: A new look*, J. Comput. Graphical Stat. 11 (2002), pp. 920–949.
[15] A. Genz and F. Bretz, *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics, Vol. 195, Springer, New York, 2009, ISBN: 978-3-642-01688-2.
[16] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0, URL http://www.R-project.org.
[17] D.P. Morton and E. Popova, Monte Carlo simulations for stochastic optimization, in *Encyclopedia of Optimization*, C.A. Floudas and P.M. Pardalos, eds., Kluwer Academic Publishers, 2001, pp. 439–447.
[18] Y. Shen and J. Cai, *Maximum of the weighted Kaplan–Meier tests with application to cancer prevention and screening trials*, Biometrics 57 (2001), pp. 837–843.
[19] T. Karrison, *Restricted mean life with adjustment for covariates*, J. Am. Stat. Assoc. 82 (1987), pp. 1169–1176.