Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 3213–3222
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Identification of the minimum effective dose for right-censored survival data

Yuh-Ing Chen*, Yu-Mei Chang

Institute of Statistics, National Central University, Jhongli 32054, Taiwan, ROC

Received 17 May 2005; received in revised form 9 November 2006; accepted 15 November 2006
Available online 8 December 2006

Abstract

In this paper, we consider identifying the minimum effective dose (MED) in a dose–response study when survival data are subject to random right-censorship, where the MED is defined to be the smallest dose level under study that has survival advantage over the zero-dose control. To this end, we suggest single-step-down testing procedures based on three different types of weighted logrank statistics, respectively. The 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 testing procedures for identifying the MED is finally illustrated by using a numerical example of prostate cancer data.

Keywords: Closed test; Dose–response study; Minimum effective dose; Monte Carlo study; Right-censored data

1. Introduction

Dose–response studies are frequently conducted to evaluate the treatment effects of a drug in animal experiments or clinical trials for drug development, where subjects or patients are randomly allocated to groups to receive several increasing dose levels of the drug and a zero-dose control. 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 (Ruberg, 1995). Under the assumption that responses are distributed according to normal distributions with a common variance, Ruberg (1989) proposed single-step multiple tests based on different contrasts of sample means to identify the MED. Tamhane et al. (1996) then suggested use of more powerful step-down closed testing procedures based on some contrasts of sample means for the MED identification. On the other hand, when the assumption of normal distribution is not tenable, Chen (1999) considered the step-down closed testing procedure as suggested in Tamhane et al. (1996), but based on the Mann–Whitney statistics for identifying the MED. Jan and Shieh (2004) further proposed some contrasts of average ranks incorporated with the step-down closed testing scheme for the MED identification.

However, in animal experiments or clinical trials, it occurs quite often that the response corresponding to a certain dose level of a drug is time to tumor occurrence or the prolonged survival time of patients with a particular disease. In such studies, the survival data are usually subject to random right-censorship since these studies may be terminated at

* Corresponding author. Tel.: +886 3 4267227; fax: +886 3 4258602.
E-mail address: ychen@stat.ncu.edu.tw (Y.-I. Chen).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.11.017

preassigned times or subjects who randomly enter a study may be lost to follow-up randomly. Therefore, to identify the smallest dose level for randomly right-censored survival data that has survival advantage over the zero-dose control, we need new testing procedures.

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 each with a continuous survivor function S_i, and $C_{i1}, \ldots, C_{in_i}$ be i.i.d. random variables each with a continuous survivor function F_i, where C_{iu} is the censoring time associated with the survival time X_{iu}. Suppose that the zero population ($i = 0$) is the control and the other k populations correspond to the increasing dose levels. Furthermore, assume that the $k + 1$ samples are independent of each other and the censoring time C_{iu} is distributed independently of the lifetime X_{iu}. In such a setting, we often observe the bivariate vectors $\{\min(X_{iu}, C_{iu}), \delta_{iu}\}$, where $\delta_{iu} = 1$, if $\min(X_{iu}, C_{iu}) = X_{iu}$, and 0, otherwise. Let $S_0 < S_i$ denote $S_0(t) \leqslant S_i(t)$ for all t and $S_0(t) < S_i(t)$ for some t. In this paper, specifically, we consider to identify the MED

$$\text{MED} = \min\{i : S_0 < S_i,\ i = 1, \ldots, k\},$$

for the situation where the survival data are randomly right-censored.

In Section 2, we consider single-step-down multiple tests based on three types of weighted logrank (WLR) statistics, respectively, for the MED identification. In Section 3, the results of a Monte Carlo study investigation of the relative error rate, power and bias performances of the competing procedures are presented. Finally, in Section 4, the use of these testing procedures is illustrated with the numerical example involving patients with stage III prostate cancer (Byar and Corle, 1977) (Table 5).

2. Proposed testing procedures

For identifying the MED with randomly right-censored survival data, we consider testing for the null hypotheses $\text{H}_{0i} : (S_0 = S_1 = \cdots = S_i)$, $i = 1, 2, \ldots, k$. Since the family of null hypotheses $\text{H} = \{\text{H}_{0i}\}$ is closed under intersection in the sense that $\text{H}_{0i} \in \text{H}$ and $\text{H}_{0j} \in \text{H}$ imply $\text{H}_{0i} \cap \text{H}_{0j} \in \text{H}$ (Marcus et al., 1976), we use a level α single-step-down closed testing scheme for identifying the MED based on adjusted p values (Wright, 1992). Suppose that $T_1, \ldots, T_k$ are the statistics involved. The single-step-down closed testing procedure starts at the first step testing for H_{0k} based on the statistic T_k and computing the associated p value $p_1 = P\{T_k \geqslant t_k | H_{0k}\}$, where t_k is the observed value of T_k. If the test based on T_k fails to reject H_{0k}, then the MED is identified to be $\hat{MED} = k + 1$. In this case, the MED is declared to be a higher dose that is not under study. Otherwise, the test proceeds to next step testing for $\text{H}_{0(k-1)}$. In general, at step i involving with the zero-dose control and the first $k_i = k - i + 1$ non-zero dose groups, the statistic testing for H_{0k_i} is 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 then given by $p_i^* = \max(p_1, \ldots, p_i)$, where $p_j = P\{T_{k_j} \geqslant t_{k_j} | \text{H}_{0k_j}\}$ is the p value of the individual test based on T_{k_j}, $j = 1, 2, \ldots, i$. Therefore, if $p_i^* \leqslant \alpha$ and $i > 1$, then the test proceeds with $k_{i+1} = k_i - 1$, but if $p_i^* > \alpha$ or $i = 1$, the test stops and the MED is identified to be $\hat{MED} = k_i + 1$ or $\hat{MED} = 1$. 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 the identified MED.

In this paper, to identify the MED with randomly right-censored survival data based on the single-step-down closed testing scheme stated above, appropriate test statistics and the associated p values are needed. To this end, we consider in the following three types of WLR statistics. One called pairwise WLR statistic compares a certain dose-group with the zero-dose control, another referred to as the combined-groups WLR statistic contrasts a certain dose-group with all the lower dose groups combined, and the other termed step-type WLR statistic makes a comparison between the higher dose groups and the lower dose groups.

2.1. Pairwise weighted logrank statistics

For $i = 0, 1, \ldots, k$, let $D_i(t) = \#\{u : X_{iu} \leqslant t, \delta_{iu} = 1, u = 1, 2, \ldots, n_i\}$ be the number of patients in group i who have been dead by time t and let $Y_i(t) = \#\{u : X_{iu} \geqslant t, u = 1, \ldots, n_i\}$ be the number of patients in groups i who are still alive and uncensored at time t. Then, the pairwise WLR statistic comparing the ith dose level with the zero-dose control is given by

$$U_i = \int_0^{t_{0i}} W_{0i}(t) \frac{Y_0(t) Y_i(t)}{Y_{0i}(t)} \left\{ \frac{\mathrm{d}D_0(t)}{Y_0(t)} - \frac{\mathrm{d}D_i(t)}{Y_i(t)} \right\}, \quad i = 1, 2, \ldots, k,$$

where $t_{0i} = \sup\{t : Y_0(t)Y_i(t) > 0\}$, $Y_{0i}(t) = Y_0(t) + Y_i(t)$, and, as suggested in Fleming and Harrington (1991), $W_{0i}(t) = \{\hat{S}_{0i}(t-)\}^{\rho}\{1 - \hat{S}_{0i}(t-)\}^{\gamma}$, with $\hat{S}_{0i}(t)$ the Kaplan–Meier (Kaplan and Meier, 1958) estimator computed from the pooled groups 0 and i. Notice that setting $\rho = 0$, $\gamma = 0$ produces the logrank (LR) statistic (Mantel, 1966) which is optimal for the proportional hazards model, while setting $\rho = 1$, $\gamma = 0$ yields the Peto–Prentice–Wilcoxon (PPW) statistic (Peto and Peto, 1972; Prentice, 1978) which is appropriate for hazards different at early times.

Let, for $i \neq j = 1, 2, \ldots, k$, $t_{0ij} = \sup\{t : Y_0(t)Y_i(t)Y_j(t) > 0\}$, $D_{0i}(t) = D_0(t) + D_i(t)$, $D_{0ij}(t) = D_0(t) + D_i(t) + D_j(t)$ and $Y_{0ij}(t) = Y_0(t) + Y_i(t) + Y_j(t)$. Set $N = \sum_{u=0}^{k} n_u$. Under the null hypothesis H_{0i}, the variance of U_i and covariance between U_i and U_j can be estimated by

$$s_{ii}^{U} = \int_0^{t_{0i}} W_{0i}^2(t)\frac{Y_0(t)Y_i(t)}{Y_{0i}(t)}\left\{1 - \frac{\Delta D_{0i}(t) - 1}{Y_{0i}(t) - 1}\right\}\frac{\mathrm{d}D_{0i}(t)}{Y_{0i}(t)}$$

and

$$s_{ij}^{U} = \int_0^{t_{0ij}} W_{0i}(t)W_{0j}(t)\frac{Y_0(t)Y_i(t)Y_j(t)}{Y_{0i}(t)Y_{0j}(t)}\left\{1 - \frac{\Delta D_{0ij}(t) - 1}{Y_{0ij}(t) - 1}\right\}\frac{\mathrm{d}D_{0ij}(t)}{Y_{0ij}(t)},$$

with $\Delta D_{\cdot}(t) = D_{\cdot}(t) - D_{\cdot}(t-1)$. Let $U_i^* = U_i/\sqrt{s_{ii}^{U}}$, $i = 1, 2, \ldots, k$. The results in Chen (1998) imply that the asymptotic null (H_{0k}) distribution of the random vector $(U_1^*, \ldots, U_k^*)$ is a k-dimensional normal distribution with mean vector zero and the associated covariance matrix can be consistently estimated by $\mathbf{R}_k^{U} = (r_{ij}^{U})$, with $r_{ii}^{U} = 1$, and $r_{ij}^{U} = s_{ij}^{U}/\sqrt{s_{ii}^{U}s_{jj}^{U}}$, $i \neq j = 1, 2, \ldots, k$.

At step i in the closed testing procedure, the statistic involved is

$$U_{(k_i)}^* = \max(U_1^*, \ldots, U_{k_i}^*). \tag{1}$$

Suppose that $u_{(k_i)}$ is the observed value of $U_{(k_i)}^*$, then the associated adjusted p value is $p_i^* = \max(p_1, \ldots, p_i)$, where, for $j = 1, 2, \ldots, i$,

$$\begin{aligned} p_j = P\{U_{(k_j)}^* \geqslant u_{(k_j)} | \mathrm{H}_{0k_j}\} &\approx P\{\max(Z_1, \ldots, Z_{k_j}) \geqslant u_{(k_j)} | \mathrm{H}_{0k_j}\} \\ &= 1 - P\{Z_l < u_{(k_j)}, l = 1, \ldots, k_j\} \end{aligned}$$

and $(Z_1, \ldots, Z_{k_j})$ is a normal random vector with mean vector zero and covariance matrix $\mathbf{R}_{k_j}^{U}$. Notice that the approximate p value stated above can be computed by using the program in Gassmann et al. (2002).

2.2. Combined-groups weighted logrank statistics

The second type of WLR statistics compares the ith dose group with all the lower dose groups combined, including the zero-dose control. Therefore, the combined-groups WLR statistics under consideration are

$$G_i = \int_0^{t_i} W_{0+i}(t)\frac{Y_{0+(i-1)}(t)Y_i(t)}{Y_{0+i}(t)}\left\{\frac{\mathrm{d}D_{0+(i-1)}(t)}{Y_{0+(i-1)}(t)} - \frac{\mathrm{d}D_i(t)}{Y_i(t)}\right\}, \quad i = 1, 2, \ldots, k,$$

where $t_i = \sup\{t : Y_0(t)Y_1(t)\ldots Y_i(t) > 0\}$, $D_{0+i}(t) = \sum_{j=0}^{i} D_j(t)$ and $Y_{0+i}(t) = \sum_{j=0}^{i} Y_j(t)$, $i = 1, 2, \ldots, k$. The associated variances are then given by

$$s_{ii}^{G} = \int_0^{t_i} W_{0+i}^2(t)\frac{Y_{0+(i-1)}(t)Y_i(t)}{Y_{0+i}(t)}\left\{1 - \frac{\Delta D_{0+i}(t) - 1}{Y_{0+i}(t) - 1}\right\}\frac{\mathrm{d}D_{0+i}(t)}{Y_{0+i}(t)}, \quad i = 1, 2, \ldots, k.$$

Let $G_i^* = G_i/\sqrt{s_{ii}^{G}}$, $i = 1, 2, \ldots, k$. Notice that the statistics $G_1, \ldots, G_k$ are uncorrelated under the null hypothesis H_{0k} (Liu et al., 1993). Therefore, the asymptotic null (H_{0k}) distribution of the random vector $(G_1^*, G_2^*, \ldots, G_k^*)$ is a k-dimensional normal distribution with zero mean vector and identity covariance matrix.

At step i in the closed testing procedure, we suggest use of the statistic

$$G_{(k_i)}^* = \max(G_1^*, \ldots, G_{k_i}^*). \tag{2}$$

Suppose that the observed value of $G^*_{(k)_i}$ is $g_{(k_i)}$. The associated adjusted p value of the test is then given by $p^*_i = \max(p_1, \ldots, p_i)$, where $p_j = P\{G^*_{(k_j)} \geqslant g_{(k_j)} | \mathrm{H}_{0k_j}\} \approx 1 - \{\Phi(g_{(k_j)})\}^{k_j}$, $j = 1, \ldots, i$, and $\Phi(\cdot)$ is the distribution function of a standard normal distribution.

2.3. Step-type weighted logrank statistics

The step-type WLR statistic that compares the involved higher $(i - j + 1)$ dose levels with all the lower j dose levels, including the zero-dose control, is given by

$$V^{(i)}_j = \int_0^{t_i} W_{0+i}(t) \frac{Y_{0+(j-1)}(t) Y_{j+i}(t)}{Y_{0+i}(t)} \left\{ \frac{\mathrm{d}D_{0+(j-1)}(t)}{Y_{0+(j-1)}(t)} - \frac{\mathrm{d}D_{j+i}(t)}{Y_{j+i}(t)} \right\}, \quad 1 \leqslant j \leqslant i \leqslant k.$$

A standard argument on the asymptotic results about the two-sample WLR statistics implies that the related variances and covariances can be consistently estimated by, for $1 \leqslant j \neq m \leqslant i$,

$$s^{V^{(i)}}_{jj} = \int_0^{t_i} W^2_{0+i}(t) \frac{Y_{0+(j-1)}(t) Y_{j+i}(t)}{Y_{0+i}(t)} \left\{ 1 - \frac{\Delta D_{0+i}(t) - 1}{Y_{0+i}(t) - 1} \right\} \frac{\mathrm{d}D_{0+i}(t)}{Y_{0+i}(t)}$$

and

$$s^{V^{(i)}}_{jm} = \int_0^{t_i} W^2_{0+i}(t) \frac{Y_{0+(j-1)}(t) Y_{m+i}(t)}{Y_{0+i}(t)} \left\{ 1 - \frac{\Delta D_{0+i}(t) - 1}{Y_{0+i}(t) - 1} \right\} \frac{\mathrm{d}D_{0+i}(t)}{Y_{0+i}(t)}.$$

Let $V^{(i)^*}_j = V^{(i)}_j / \sqrt{s^{V^{(i)}}_{jj}}$, $1 \leqslant j \leqslant i$. Then, under the null hypothesis H_{0i}, the distribution of $(V^{(i)^*}_1, \ldots, V^{(i)^*}_i)$ can be approximated by an i-dimensional normal distribution with zero mean vector and covariance matrix $\mathbf{R}^V_i = (r^{V^{(i)}}_{jm})$, where $r^{V^{(i)}}_{jm} = 1$ and $r^{V^{(i)}}_{jm} = s^{V^{(i)}}_{jm} / \sqrt{s^{V^{(i)}}_{jj} s^{V^{(i)}}_{mm}}$, $1 \leqslant j \neq m \leqslant i$.

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

$$V^*_{(k_i)} = \max(V^{(k_i)^*}_1, \ldots, V^{(k_i)^*}_{k_i}). \tag{3}$$

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 $p^*_i = \max(p_1, \ldots, p_i)$, where for $j = 1, 2, \ldots, i$,

$$\begin{aligned} p_j &= P\{V^*_{(k_j)} \geqslant v_{(k_j)} | \mathrm{H}_{0k_j}\} \approx P\{\max(Z_1, \ldots, Z_{k_j}) \geqslant v_{(k_j)} | \mathrm{H}_{0k_j}\} \\ &= 1 - P\{Z_m < v_{(k_j)}, m = 1, \ldots, k_j\} \end{aligned}$$

and $(Z_1, \ldots, Z_{k_j})$ is a normal random vector with mean vector zero and covariance matrix $\mathbf{R}^V_{k_j}$. The associated approximate p values can be obtained, again, by using the program in Gassmann et al. (2002).

3. A simulation study

3.1. Description of the study

We conducted a Monte Carlo study to examine the relative error rate, power and bias performances of the proposed closed testing procedures based on the pairwise WLR statistics (1), combined-groups WLR statistics (2) and the step-type WLR test statistics (3), denoted by U, G and V, respectively. The error rate performances of the tests were 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 probability of correctly identifying the MED and $E(M\hat{E}D) - \mathrm{MED}$, respectively, were then used to assess the power and bias performances of the tests. We consider $k = 3$ and 4 treatments with sample sizes $n_0 = n_1 = \cdots = n_k = n = 10$, 30 and 50 in the EWE study and $n = 30$ and 50 in the FWE and power/bias study.

Notice that, under the Weibull distributions with the same shape parameters, the hazard functions are proportional. On the other hand, the hazard functions are different at early time under the lognormal distributions. Therefore, we

Table 1
Estimated experimentwise error rates for $\alpha = 0.05$, various survivor functions with uniform censoring distribution $U(0, r)$ and $n_0 = n_1 = \cdots = n_4 = n$

r	n	LR			PPW		
		U	G	V	U	G	V
Weibull							
4.410	10	0.062	0.045	0.061	0.051	0.036	0.053
	30	0.055	0.045	0.053	0.050	0.042	0.049
	50	0.050	0.044	0.049	0.046	0.043	0.046
1.745	10	0.059	0.039	0.057	0.051	0.030	0.052
	30	0.055	0.044	0.052	0.051	0.040	0.053
	50	0.049	0.043	0.049	0.048	0.040	0.049
Lognormal							
6.381	10	0.056	0.040	0.055	0.047	0.036	0.051
	30	0.055	0.046	0.054	0.050	0.044	0.049
	50	0.054	0.048	0.052	0.053	0.045	0.050
2.297	10	0.055	0.038	0.055	0.049	0.035	0.050
	30	0.055	0.043	0.054	0.051	0.045	0.053
	50	0.052	0.044	0.046	0.052	0.044	0.050

consider both the Weibull and lognormal distributions as survivor functions with proportional hazards and early hazard difference. We take the uniform distribution over $(0, r)$ as the censoring distribution in which patients are allowed to drop out the study at anytime with equal chance. The IMSL (International Mathematical and Statistical Library) routines RNUN, RNWIB and RNNOR were used to generate appropriate uniform, Weibull and normal variates, respectively (Morton and Popova, 2001). The exponential-transformed normal variate then gives necessary lognormal variate. In the EWE study, the survivor functions under study were Weibull with scale parameter 1 and shape parameter 2, and lognormal distribution with normal mean 0 and normal variance $\frac{1}{2}$. Various values of r, which correspond to the probability of censorship as 0.2 or 0.5 were considered in the EWE study. The corresponding uniform censoring distributions were then employed in the FWE and power/bias study. Notice that, in the FWE and power/bias study, the Weibull distributions under consideration have a common shape parameter 2 but various values of scale parameters θ_i, $i = 0, 1, \ldots, k$ with $\theta_0 = 1$, and the lognormal distributions have a common normal variance $\frac{1}{2}$, but different values of normal means θ_i, $i = 0, 1, \ldots, k$ with $\theta_0 = 0$. Also 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.

For each of these settings, 10,000 replicates were used to obtain the estimated EWE, FWE, power, and bias under the nominal level $\alpha = 0.05$. Therefore, the maximum standard error for the error rate and power estimators is about 0.005 ($\approx \sqrt{(0.5)(0.5)/10,000}$). In fact, the standard deviation of the error rate estimator is about 0.002 ($\approx \sqrt{(0.5)(0.95)/10,000}$). Since the relative EWE performance of the three testing procedures is similar for $k = 3$ and 4, we only presented the estimated EWE for $k = 4$ in Table 1. Again, the relative FWE and power/bias performances of the three testing procedures are similar for $k = 3$ and 4 and for $n = 30$ and 50. Therefore, we report, for simplicity, the estimates of FWE, power and bias for $k = 4$ and 50 in Tables 2–4, respectively. Notice that, when the true MED is 1, the FWE is zero since no type I error is involved. Therefore, we did not consider any treatment effect configurations with MED = 1 for the FWE study. Also notice that, in the bias study, we conventionally assign $\hat{MED} = 5$ for $k = 4$ if the MED is identified not to be any dose levels under study.

3.2. Description of the results

It is evident, upon examination of Table 1, that all the tests hold their EWE when sample size is 30 or larger. When sample size is 10, the pairwise and step-type PPW tests still maintain their EWE. However, for such a small sample size, both the combined-groups LR and PPW tests tend to be conservative in hold its EWE, while the pairwise and

Table 2
Estimated FWE for $\alpha = 0.05$, $n_0 = n_1 = \cdots = n_4 = 50$, and various survivor functions with uniform censoring distribution $U(0, r)$

θ_1	θ_2	θ_3	θ_4	True MED	LR			PPW		
					U	*G*	*V*	*U*	*G*	*V*
Weibull and $r = 1.745$										
1	1	1	1.5	4	0.051	0.043	0.048	0.051	0.040	0.048
1	1	1.5	1.5	3	0.051	0.044	0.051	0.049	0.042	0.050
1	1.5	1.5	1.5	2	0.048	0.043	0.049	0.047	0.041	0.048
1	1	1.3	1.5	3	0.048	0.037	0.049	0.045	0.035	0.046
1	1.3	1.3	1.5	2	0.044	0.030	0.046	0.042	0.028	0.044
1	1	1.5	1	3	0.051	0.043	0.047	0.048	0.041	0.044
1	1	1.5	1.3	3	0.051	0.043	0.051	0.049	0.041	0.049
1	1.3	1.5	1.3	2	0.045	0.033	0.046	0.043	0.031	0.045
1	1.5	1.3	1	2	0.048	0.041	0.046	0.046	0.038	0.043
Lognormal and $r = 2.297$										
0	0	0	0.5	4	0.052	0.041	0.049	0.052	0.042	0.053
0	0	0.5	0.5	3	0.051	0.045	0.053	0.050	0.044	0.050
0	0.5	0.5	0.5	2	0.050	0.045	0.051	0.048	0.043	0.049
0	0	0.3	0.5	3	0.047	0.035	0.051	0.046	0.035	0.048
0	0.3	0.3	0.5	2	0.044	0.029	0.047	0.043	0.029	0.045
0	0	0.5	0	3	0.051	0.044	0.050	0.050	0.043	0.047
0	0	0.5	0.3	3	0.051	0.044	0.053	0.050	0.043	0.050
0	0.3	0.5	0.3	2	0.045	0.032	0.048	0.044	0.033	0.046
0	0.5	0.3	0	2	0.050	0.043	0.047	0.048	0.042	0.045

step-type LR tests do not maintain their EWE. Moreover, as shown in Table 2, the FWE of all the testing procedures are well under controlled.

It is not surprising to find, from Tables 3 and 4 that, in terms of power and bias, the LR test performs better than the PPW test for Weibull distribution, while the PPW test is superior to LR test for lognormal distribution, since the Weibull distributions under study have the proportional hazards, but the hazards of lognormal distributions differ in early time. The step-type test outperforms in power and bias performances for ordered dose–response, but it performs poorly for umbrella patterned dose–response, especially, 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 true MED is higher than first dose level, however, the combined-groups test is superior to the pairwise test.

To sum up, when the sample size is 30 or larger for each group, both the experimentwise and familywise error rates of the three WLR tests are under controlled. In terms of power and bias performances, the step-type WLR test is recommended for ordered dose–response. For umbrella patterned dose–response, however, the pairwise or combined-groups WLR test is suggested. In fact, the pairwise WLR test is preferred if the true MED is relatively close to the first dose level; otherwise, the combined-groups WLR test is suggested.

4. An example

The data from a randomized and double-blinded clinical trial involving patients with stage III prostate cancer (Byar and Corle, 1977) were used for illustration. There were 292 patients with local extension of the disease, but without evidence of distant metastasis, where 75 took the placebo pill and served as the control group, 73 received 0.2 mg of diethylstilbestrol (DES) as dosage group 1, 73 were treated with 1.0 mg of DES as dosage group 2 and 71 were assigned to take 5.0 mg of DES as dosage group 3. Patients were followed according to a standard protocol at 6-month intervals or more frequently if required, and the survival times of the patients in the four groups since randomization 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 study so that the survivor function of the involved patients is better than that of the patients taking the placebo pill. The Kaplan–Meier

Table 3
Estimated power for $\alpha = 0.05$, $n_0 = n_1 = \cdots = n_4 = 50$, and various survivor functions with uniform censoring distribution $U(0, r)$

θ_1	θ_2	θ_3	θ_4	True MED	LR			PPW		
					U	G	V	U	G	V
Weibull and $r = 1.745$										
1	1	1	1.5	4	0.596	0.774	0.810	0.547	0.711	0.756
1	1	1.5	1.5	3	0.624	0.767	0.806	0.576	0.710	0.759
1	1.5	1.5	1.5	2	0.657	0.736	0.787	0.612	0.681	0.742
1.5	1.5	1.5	1.5	1	0.766	0.667	0.763	0.719	0.614	0.714
1	1	1.3	1.5	3	0.315	0.428	0.496	0.281	0.373	0.447
1	1.3	1.3	1.5	2	0.347	0.392	0.459	0.301	0.341	0.403
1.3	1.3	1.5	1.5	1	0.451	0.345	0.448	0.407	0.309	0.399
Average (ordered)					0.537	0.587	0.653	0.492	0.534	0.603
1	1	1.5	1	3	0.603	0.750	0.440	0.555	0.693	0.359
1	1	1.5	1.3	3	0.611	0.756	0.773	0.562	0.698	0.715
1	1.3	1.5	1.3	2	0.364	0.415	0.489	0.319	0.364	0.433
1	1.5	1.3	1	2	0.619	0.709	0.518	0.576	0.652	0.438
1.3	1.5	1.3	1	1	0.479	0.380	0.371	0.431	0.338	0.320
1.3	1.5	1	1	1	0.471	0.378	0.223	0.423	0.336	0.187
Average (umbrella)					0.525	0.565	0.469	0.478	0.514	0.409
Average (overall)					0.531	0.577	0.568	0.485	0.525	0.513
Lognormal and $r = 2.297$										
0	0	0	0.5	4	0.630	0.795	0.828	0.668	0.826	0.850
0	0	0.5	0.5	3	0.654	0.784	0.822	0.689	0.820	0.850
0	0.5	0.5	0.5	2	0.688	0.761	0.815	0.724	0.795	0.844
0.5	0.5	0.5	0.5	1	0.793	0.705	0.801	0.827	0.747	0.834
0	0	0.3	0.5	3	0.281	0.363	0.433	0.307	0.398	0.465
0	0.3	0.3	0.5	2	0.304	0.341	0.414	0.332	0.374	0.447
0.3	0.3	0.5	0.5	1	0.424	0.329	0.421	0.454	0.359	0.453
Average (ordered)					0.539	0.583	0.648	0.572	0.617	0.678
0	0	0.5	0	3	0.635	0.768	0.440	0.670	0.802	0.434
0	0	0.5	0.3	3	0.640	0.771	0.777	0.675	0.805	0.812
0	0.3	0.5	0.3	2	0.327	0.372	0.445	0.355	0.408	0.477
0	0.5	0.3	0	2	0.648	0.733	0.502	0.680	0.763	0.518
0.3	0.5	0.3	0	1	0.458	0.366	0.356	0.488	0.403	0.377
0.3	0.5	0	0	1	0.453	0.365	0.221	0.482	0.402	0.219
Average (umbrella)					0.527	0.563	0.457	0.558	0.597	0.473
Average (overall)					0.533	0.573	0.560	0.565	0.608	0.583

(Kaplan and Meier, 1958) estimates of the survivor functions for the four groups were shown in Fig. 1. The relevant summary statistics, including the statistics at each step, the identified MED, and the associated adjusted p values were then given in Table 5.

To apply the single-step-down close testing procedure based on the pairwise LR statistics, we compute $u_{(3)} = u_{(2)} = 2.37$ and $u_{(1)} = 0.44$, which gives $p_1 = P\{U^*_{(3)} \geqslant 2.37|\mathrm{H}_{(03)}\} \approx 0.025$, $p_2 = P\{U^*_{(2)} \geqslant 2.37|\mathrm{H}_{02}\} \approx 0.017$, and $p_3 = P\{U^*_{(1)} \geqslant 0.44|\mathrm{H}_{01}\} \approx 0.331$. Therefore, we identify the MED to be 1.0 mg of DES and the adjusted p value is $p^*_2 = 0.025$. The test based on the pairwise PPW statistics reaches the same conclusion as stated above, but has a different adjusted p value as 0.027.

The combined-groups LR statistics are obtained as $g_{(3)} = g_{(2)} = 2.26$ and $g_{(1)} = 0.44$, which produces $p_1 = P\{G^*_{(3)} \geqslant 2.26|\mathrm{H}_{03}\} \approx 0.035$, $p_2 = P\{G^*_{(2)} \geqslant 2.26|\mathrm{H}_{02}\} \approx 0.024$, and $p_3 = P\{G^*_{(1)} \geqslant 0.44|\mathrm{H}_{01}\} \approx 0.331$. Therefore, the LR statistics identifies the MED to be 1.0 mg of DES with an adjusted p value of 0.035. The test based on the combined-groups PPW statistics also identifies the 1.0 mg of DES to be the MED, but with a different adjusted p value of 0.034.

Table 4
Estimated power for $\alpha = 0.05$, $n_0 = n_1 = \cdots = n_4 = 50$, and various survivor functions with uniform censoring distribution $U(0, r)$

θ_1	θ_2	θ_3	θ_4	True MED	LR			PPW		
					U	*G*	*V*	*U*	*G*	*V*
Weibull and $r = 1.745$										
1	1	1	1.5	4	0.239	0.091	0.053	0.290	0.162	0.109
1	1	1.5	1.5	3	0.436	0.221	0.099	0.524	0.321	0.166
1	1.5	1.5	1.5	2	0.556	0.449	0.198	0.660	0.591	0.279
1.5	1.5	1.5	1.5	1	0.611	1.164	0.628	0.729	1.345	0.767
1	1	1.3	1.5	3	0.856	0.687	0.466	0.943	0.800	0.553
1	1.3	1.3	1.5	2	1.312	1.331	0.855	1.444	1.489	0.991
1.3	1.3	1.5	1.5	1	1.393	2.120	1.390	1.544	2.265	1.552
Average (ordered)					0.772	0.866	0.527	0.876	0.996	0.631
1	1	1.5	1	3	0.610	0.347	0.956	0.716	0.469	1.124
1	1	1.5	1.3	3	0.549	0.307	0.210	0.648	0.422	0.311
1	1.3	1.5	1.3	2	1.088	1.069	0.691	1.220	1.228	0.828
1	1.5	1.3	1	2	0.882	0.675	1.237	1.005	0.848	1.476
1.3	1.5	1.3	1	1	1.334	1.953	2.275	1.479	2.133	2.474
1.3	1.5	1	1	1	1.431	1.979	3.050	1.589	2.156	3.197
Average (umbrella)					0.982	1.055	1.403	1.110	1.209	1.568
Average (overall)					0.869	0.953	0.931	0.984	1.095	1.064
Lognormal and $r = 2.297$										
0	0	0	0.5	4	0.201	0.075	0.034	0.168	0.043	0.002
0	0	0.5	0.5	3	0.389	0.190	0.070	0.326	0.132	0.042
0	0.5	0.5	0.5	2	0.495	0.396	0.142	0.407	0.313	0.099
0.5	0.5	0.5	0.5	1	0.549	1.047	0.526	0.445	0.888	0.423
0	0	0.3	0.5	3	0.880	0.761	0.524	0.820	0.691	0.479
0	0.3	0.3	0.5	2	1.401	1.448	0.955	1.307	1.343	0.863
0.3	0.3	0.5	0.5	1	1.447	2.107	1.407	1.327	1.983	1.281
Average (ordered)					0.766	0.861	0.523	0.686	0.770	0.456
0	0	0.5	0	3	0.546	0.308	0.946	0.481	0.242	0.967
0	0	0.5	0.3	3	0.510	0.284	0.210	0.446	0.219	0.158
0	0.3	0.5	0.3	2	1.119	1.131	0.754	1.029	1.015	0.675
0	0.5	0.3	0	2	0.819	0.608	1.286	0.730	0.521	1.250
0.3	0.5	0.3	0	1	1.326	1.909	2.307	1.213	1.767	2.243
0.3	0.5	0	0	1	1.392	1.926	3.052	1.277	1.786	3.068
Average (umbrella)					0.952	1.028	1.426	0.863	0.925	1.394
Average (overall)					0.852	0.938	0.939	0.767	0.842	0.888

Finally, we use the single-step-down closed testing procedure based on step-type LR statistics to identify the MED. We observed $v_{(3)} = 2.12$ and computed $p_1 = P\{V^*_{(3)} \geqslant 2.12 | \mathrm{H}_{03}\} = 0.046$, which leads to the next step with $k_2 = 2$. Since $v_{(2)} = 2.26$ with $p_2 = P\{V^*_{(2)} \geqslant 2.26 | \mathrm{H}_{02}\} = 0.022$, we go to the next step with $k_3 = 1$. Now $p_3 = P\{V^*_{(1)} \geqslant 0.44 | \mathrm{H}_{01}\} \approx 0.331$, so we identify the MED to be 1.0 mg of DES and the adjusted p value is 0.046. Moreover, the test based on the step-type PPW statistics reaches the same conclusion as stated above, but the associated adjusted p value is 0.043.

Notice that, under significance level 0.05, all the three methods considered herein identify the 1.0 mg of DES as the MED at which the involved patients have a better chance to live longer than the patients taking the placebo pill. However, the survivor estimates in Fig. 1 show an apparent downturn in survivor function at the highest dose level, it is not surprising to see that the step-type test produces the largest adjusted p values. This result is, in fact, in a good agreement with the simulation results presented in previous section. Therefore, regarding to this particular application, either the combined-groups test or the pairwise test is better than the step-type test for MED identification.

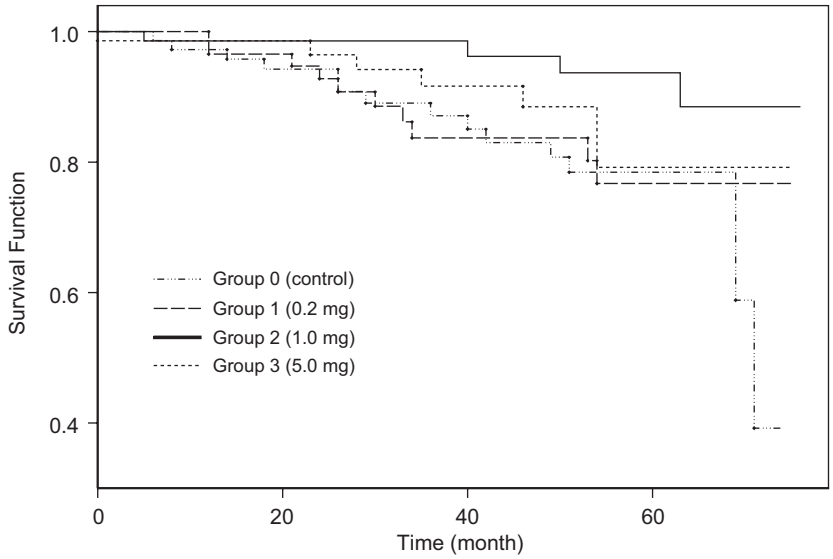

Fig. 1. Kaplan–Meier estimate of survivor functions for the prostate cancer data.

Table 5
Summary statistics for the prostate cancer data

	i	LR	PPW
$U_i^*(U_i, s_{ii}^U)$	1	0.44 (1.06, 5.89)	0.32 (0.69, 4.57)
	2	2.37 (5.00, 4.47)	2.33 (4.45, 3.63)
	3	1.15 (2.59, 5.10)	1.09 (2.19, 4.04)
$M\hat{E}D$		2	2
Adjusted p value		0.025	0.027
$G_i^*(G_i, s_{ii}^G)$	1	0.44 (1.06, 5.89)	0.32 (0.69, 4.57)
	2	2.26 (5.68, 6.30)	2.28 (5.18, 5.16)
	3	0.33 (0.80, 6.01)	0.33 (0.74, 4.97)
$M\hat{E}D$		2	2
Adjusted p value		0.035	0.034
$V_i^{(3)*}(V_i^{(3)}, s_{ii}^{V^{(3)}})$	1	1.67 (4.39, 6.92)	1.62 (3.88, 5.76)
	2	2.12 (6.25, 8.71)	2.14 (5.74, 7.21)
	3	0.33 (0.80, 6.01)	0.33 (0.74, 4.97)
$V_i^{(2)*}(V_i^{(2)}, s_{ii}^{V^{(2)}})$	1	1.64 (4.14, 6.35)	1.57 (3.59, 5.23)
	2	2.26 (5.68, 6.30)	2.28 (5.18, 5.16)
$V_i^{(1)*}(V_i^{(1)}, s_{ii}^{V^{(1)}})$	1	0.44 (1.06, 5.89)	0.32 (0.69, 4.57)
$M\hat{E}D$		2	2
Adjusted p value		0.046	0.043

Acknowledgments

This work was supported by the National Science Council of Taiwan under Grant NSC91-2118-008-001 and NSC92-2118-008-011.

References

Byar, D.P., Corle, D.K., 1977. Selecting optimal treatment in clinical trials using covariate information. J. Chronic Dis. 30, 445–459.
Chen, Y.I., 1998. Simple-tree weighted logrank tests for randomly right-censored data. Ann. Inst. Statist. Math. 50, 311–324.

Chen, Y.I., 1999. Nonparametric identification of the minimum effective dose. Biometrics 55, 1236–1240.
Fleming, T.R., Harrington, D.P., 1991. Counting Process and Survival Analysis. Wiley, New York.
Gassmann, H.I., Deák, I., Szántai, T., 2002. Computing multivariate normal probabilities: a new look. J. Comput. Graph. Statist. 11, 920–949.
Jan, S.L., Shieh, G., 2004. Nonparametric multiple test procedures for dose finding. Comm. Statist. Simulation Comput. 33, 1021–1037.
Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457–481.
Liu, P.Y., Green, S., Wolf, M., Crowley, J., 1993. Testing against ordered alternatives for censored survival data. J. Amer. Statist. Assoc. 88, 153–160.
Mantel, N., 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer. Chem. Rep. 50, 163–170.
Marcus, R., Peritz, E., Gabriel, K.R., 1976. On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63, 655–660.
Morton D.P., Popova, E., 2001. Monte Carlo simulations for stochastic optimization. In: Floudas, C.A., Pardalos, P.M. (Eds.), Encyclopedia of Optimization, vol. 3. Kluwer Academic Publishers, Dordrecht, pp. 439–447.
Peto, R., Peto, J., 1972. Asymptotically efficient rank invariant test procedures (with discussion). J. Roy. Statist. Soc. A 135, 185–206.
Prentice, R.L., 1978. Linear rank tests with right censored data. Biometrika 65, 167–179.
Ruberg, S.J., 1989. Contrasts for identifying the minimum effective dose. J. Amer. Statist. Assoc. 84, 816–822.
Ruberg, S.J., 1995. Dose response studies. I. Some design considerations. J. Biopharm. Statist. 5, 1–14.
Tamhane, A.C., Hochberg, Y., Dunnett, C.W., 1996. Multiple test procedures for dose finding. Biometrics 52, 21–37.
Wright, S.P., 1992. Adjusted p-values for simultaneous inference. Biometrics 48, 1005–1013.