JOURNAL OF BIOPHARMACEUTICAL STATISTICS
Vol. 14, No. 4, pp. 843–856, 2004

Finding the Maximum Safe Dose Level for Heteroscedastic Data

Ajit C. Tamhane[1,*] and Brent R. Logan[2]

[1]Departments of IE/MS and of Statistics, Northwestern University, Evanston, Illinois, USA
[2]Division of Biostatistics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

ABSTRACT

In this article we extend to the heteroscedastic setting the multiple stepwise test procedures proposed in Tamhane et al. [Tamhane. A. C., Dunnett, C. W., Green, J. W., Wetherington, J. F. (2001). Multiple test procedures for identifying a safe dose. *J. Am. Statist. Assoc.* 96:835–843] for finding the maximum safe dose. Toxicological data are often heteroscedastic; therefore, the extensions given herein should be highly useful in practice. Simulations are performed to study the Type I familywise error rate and power properties of the procedures. A real data example is given to illustrate the procedures.

Key Words: Dose–response; Multiple comparisons; Toxicology; Unequal variances.

Mathematics Subject Classification: 62J15; 62P10.

*Correspondence: Ajit C. Tamhane, Departments of IE/MS and of Statistics, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA; Fax: (847) 491-8005; E-mail: ajit@iems.northwestern-edu.

DOI: 10.1081/BIP-200035413
1054-3406 (Print); 1520-5711 (Online)

INTRODUCTION

In Tamhane et al. (2001), we developed three stepwise multiple-testing procedures (labeled SD1PC, SD2PC, and SD1HC) for finding the maximum safe dose (MAXSD) level of a compound using toxicological data. These procedures are applicable to homoscedastic data. However, toxicological data are often heteroscedastic (i.e., the data from different doses have different variances). In this article we extend the procedures in Tamhane et al. (2001) to the heteroscedastic setting. Extensive simulations are conducted, which show that the procedures applicable to homoscedastic data significantly exceed the nominal familywise Type I α level under certain heteroscedastic settings. On the other hand, the modified procedures for heteroscedastic data control the α level reasonably well under most heteroscedastic settings.

The outline of the article is as follows. Preliminaries and Notation gives the problem formulation and defines the notation. Test Procedures proposes the heteroscedastic extensions of the SD1PC, SD2PC, and SD1HC procedures. Simulation Results describes the simulation results on the Type I error rates and powers of homoscedastic and heteroscedastic versions of SD1PC, SD2PC, and SD1HC. Of these three procedures, SD2PC is generally the preferred procedure. A real data example is given in Example. Finally, conclusions are summarized in Concluding Remarks.

PRELIMINARIES AND NOTATION

Let $0, 1, \ldots, k$ denote increasing dose levels with 0 being the zero dose level (control). Assume that the measurements y_{ij} $(j = 1, 2, \ldots, n_i)$ on dose i are independent and normally distributed with mean μ_i and variance σ_i^2 (denoted by $y_{ij} \sim N(\mu_i, \sigma_i^2)$). Let $\bar{y}_i$ denote the sample mean and s_i^2 denote the sample variance based on $n_i - 1$ d.f. from dose i $(0 \leq i \leq k)$. A small μ_i (e.g., a lower yield of a crop contaminated by a herbicide) implies a more toxic response. The case where a large μ_i represents a more toxic response can be handled analogously.

We regard as *unsafe* a decrease in mean yield below a specified percentage (e.g., 10%) of the mean yield μ_0 at the zero dose level. In general, we specify $\lambda < 1$ (e.g., $\lambda = 0.90$ for a 10% decrease in the mean yield compared to μ_0). All doses with $\mu_i \leq \lambda\mu_0$ are regarded as unsafe and those with $\mu_i > \mu_0$ are regarded as safe. The MAXSD for specified λ is defined as

$$\text{MAXSD} = \max\{i : \mu_j > \lambda\mu_0 \ \forall j \leq i\}. \tag{1}$$

Define the minimum unsafe dose (MINUD) by MINUD = MAXSD + 1. If a larger μ_i represents a more toxic response, then $\lambda > 1$. In that case, doses with $\mu_i \geq \lambda\mu_0$ are regarded as unsafe and those with $\mu_i < \lambda\mu_0$ are regarded as safe. The MAXSD is then defined as $\text{MAXSD} = \max\{i : \mu_j < \lambda\mu_0 \ \forall j \leq i\}$.

We want to guarantee that the probability that *any* unsafe dose is declared safe is no more than a specified constant α. If $\widehat{\text{MAXSD}}$ denotes the estimated MAXSD, then this requirement translates to

$$P\{\widehat{\text{MAXSD}} > \text{MAXSD}\} \leq \alpha. \tag{2}$$

Now consider the family of hypothesis-testing problems

$$H_{0i} : \mu_i \leq \lambda\mu_0 \quad \text{vs.} \quad H_{1i} : \mu_i > \lambda\mu_0 \ \ (1 \leq i \leq k). \tag{3}$$

Here H_{0i} states that the ith dose is unsafe and H_{1i} states that the ith dose is safe. After testing the hypotheses, we set $\widehat{\text{MAXSD}} = \max\{i : H_{0i} \text{ is rejected}\}$. The error probability requirement Eq. (2) is satisfied if we control the type I familywise error rate (FWE) at level α for the family of hypotheses (Eq. (3)):

$$\text{FWE} = P\{\text{Any true } H_{0i} \text{ is rejected}\} \leq \alpha. \tag{4}$$

It was shown in Sec. 8 of Tamhane et al. (2001) that this requirement is satisfied if the null hypotheses in (Eq. (3)) are tested in a step-down (SD) manner beginning with H_{01}; if it is rejected, then test H_{02} and so on, each at level α, which is what SD2PC does. On the other hand, SD1PC and SD1HC test $\bigcap_{j=i}^{k} H_{0j} \subseteq H_{0i}$ in a step-down manner $(i = 1, 2, \ldots, k)$, each at level α. Under the assumption of monotonicity $\mu_0 \geq \mu_1 \geq \cdots \geq \mu_k$, we have $H_{0i} = \bigcap_{j=i}^{k} H_{0j}$; therefore, SD1PC and SD1HC procedure also control (Eq. (2)). However, if the means are not monotone, then only SD2PC controls this requirement as shown by Bauer (1997).

TEST PROCEDURES

Both SD1PC and SD2PC are based on pairwise contrasts and are described in SD1PC and SD2PC. But SD1HC is based on Helmert contrasts and is described in SD1HC procedure.

SD1PC and SD2PC

Define the t-statistics for pairwise contrasts as

$$t_i = \frac{\bar{y}_i - \lambda\bar{y}_0}{\sqrt{s_i^2/n_i + \lambda^2 s_0^2/n_0}} \quad (1 \leq i \leq k). \tag{5}$$

The SD1PC procedure tests H_{0i} using the statistic

$$t_{i,\max} = \max_{i \leq j \leq k} t_j, \tag{6}$$

whereas SD2PC uses ordinary t-tests based on the t_i. Each procedure tests H_{0i} if H_{0j} for $j = 1, \ldots, i-1$ have been rejected and rejects H_{0i} if the observed test statistic exceeds the upper α critical point of its null distribution. Equivalently, these tests can be applied in terms of the p-values. If p_i denotes the marginal p-value associated with t_i, then SD2PC rejects H_{0i} if $p_i < \alpha$. To apply SD1PC, one needs to calculate the adjusted p-value, denoted by $\tilde{p}_i$, which is the probability that under H_{0i}, the statistic $t_{i,\max}$ is at least at large as its observed value. To calculate $\tilde{p}_i$, we need to know the joint distribution of t_j $(i \leq j \leq k)$.

The exact joint distributions of the t_i are intractable and depend on the unknown variances σ_i^2. Therefore, we must use approximations to the critical points of these distributions. The marginal distribution of each t_i can be approximated by Student's t-distribution with degrees of freedom (d.f.) obtained by using the Welch–Satterthwaite approximation

$$\nu_i = \frac{(\lambda^2 s_0^2/n_0 + s_i^2/n_i)^2}{\lambda^4 s_0^4/n_0^2(n_0 - 1) + s_i^4/n_i^2(n_i - 1)}. \tag{7}$$

For SD1PC we need the upper α critical point of $t_{i,\max}$, which involves the joint distribution of $t_i, t_{i+1}, \ldots, t_k$. This distribution can be approximated by using the $(k - i + 1)$-variate t-distribution for $i = 1, 2, \ldots, k$. The correlation matrix of this distribution can be estimated by using

$$\rho_{jj'} = \frac{\lambda^2 s_0^2/n_0}{\sqrt{(\lambda^2 s_0^2/n_0 + s_j^2/n_j)(\lambda^2 s_0^2/n_0 + s_{j'}^2/n_{j'})}} \quad (i \le j < j' \le k). \tag{8}$$

These correlations have the product form $\rho_{jj'} = \gamma_j \gamma_{j'}$ where

$$\gamma_j = \sqrt{\frac{\lambda^2 s_0^2/n_0}{\lambda^2 s_0^2/n_0 + s_j^2/n_j}}. \tag{9}$$

So the computation of the desired critical point can be performed efficiently by using the iterated integral representation given in Eq. (1.1a) in Appendix 3 of Hochberg and Tamhane (1987). A Fortran program for this purpose based on Dunnett (1989) is available from http://lib.stat.cmu.edu/general. The average of the $\rho_{jj'}$ can be used to yield a good approximation; see Hochberg and Tamhane (1987, p. 146). For the d.f. of the multivariate t-distribution, the average of the ν_j $(i \le j \le k)$ can be used.

SD1HC Procedure

A Helmert contrast is defined as

$$C_{ij} = (\bar{y}_i + \cdots + \bar{y}_j) - (j - i + 1)\lambda \bar{y}_0 \quad (i \le j \le k). \tag{10}$$

The standard error of C_{ij} equals

$$s.e.(C_{ij}) = \sqrt{(j - i + 1)^2 \lambda^2 s_0^2/n_0 + \sum_{h=i}^{j} (s_h^2/n_h)}. \tag{11}$$

Define the statistic

$$t_{ij} = \frac{C_{ij}}{s.e.(C_{ij})}. \tag{12}$$

Then, in analogy with Eq. (6), the test statistic for testing H_{0i} is

$$t_{i,\max} = \max_{i \le j \le k} t_{ij}.$$

As in the case of SD1PC, H_{0i} is tested if H_{0j} $(j = 1, \ldots, i-1)$ are rejected and if the adjusted p-value associated with $t_{i,\max}$, denoted by $\tilde{p}_i$, is less then α. To calculate $\tilde{p}_i$, we need to know the joint distribution of the t_{ij} $(i \le j \le k)$.

The marginal distribution of t_{ij} can be approximated by Student's t-distribution with the Welch–Satterthwaite d.f.

$$\nu_{ij} = \frac{\{(j-i+1)^2\lambda^2 s_0^2/n_0 + \sum_{h=i}^{j}(s_h^2/n_h)\}^2}{\{(j-i+1)^4\lambda^4 s_0^4/n_0^2(n_0-1) + \sum_{h=i}^{j}[s_h^4/n_h^2(n_h-1)]\}}. \tag{13}$$

The correlation matrix of the t_{ij} for $i \le j < j' \le k$ can be estimated by using

$$\rho_{jj'}^{(i)} = \frac{\sum_{h=i}^{j}(s_h^2/n_h) + (j-i+1)(j'-i+1)\lambda^2 s_0^2/n_0}{\sqrt{\left[\sum_{h=i}^{j}(s_h^2/n_h) + (j-i+1)^2\lambda^2 s_0^2/n_0\right]\left[\sum_{h=i}^{j'}(s_h^2/n_h) + (j'-i+1)^2\lambda^2 s_0^2/n_0\right]}}. \tag{14}$$

These correlations do not have the product structure. The SAS-IML program based on Genz and Bretz (1999) with arbitrary correlation matrices available from http://www.bioinf.uni-hannover.de can be used in this case. The degrees of freedom for the multivariate t can be approximated by the average of the ν_{ij} $(i \le j \le k)$.

SIMULATION RESULTS

We performed Monte Carlo simulations to study the FWEs and powers of homoscedastic and heteroscedastic versions of SD1PC, SD2PC, and SD1HC. Here, power is defined as $p\{\widehat{\text{MAXSD}} = \text{MAXSD}\}$, i.e., the probability that the correct MAXSD is identified. We mainly focused our study on two types of configurations: (1) constant coefficient of variation (CV) in which the σ_i are proportional to the μ_i and (2) constant variance among the positive doses (which is different from the zero dose variance). The μ_i configurations were chosen to be the same as those in Table 5 of Tamhane et al. (2001) except that when studying the FWE, we set μ_{MINUD} equal to $\lambda\mu_0$ because it is the least favorable configuration at which the max FWE is attained.

We also studied a third configuration in which μ_{MAXSD} was close to μ_0, and $\sigma_0 = \sigma_1 = \cdots = \sigma_{\text{MAXSD}}$ while $\sigma_i > \sigma_0$ for $i >$ MAXSD. This configuration was chosen because the FWEs of the procedures designed for homoscedastic settings were most likely to be exceeded in this case. This occurs for two reasons. First, there is greater separation between the mean at the MAXSD and the mean at the MINUD, resulting in Type I error rate closer to its maximum. Second, the homoscedastic procedures use a biased pooled variance estimate, which results in a biased (but lower variance) estimate of the SE. This bias is more problematic for configuration (3) than for the first two configurations because the bias is in the liberal direction, resulting in greater Type I error rates for the homoscedastic procedures. In contrast, the first two

configurations generally have conservatively biased SE estimates. The results for the constant CV case are given in Table 1 (FWE) and Table 2 (Power). The results for the constant variance among the positive doses case are given in Table 3 (FWE) and Table 4 (Power). The FWE results for the third configuration are given in Table 5. All estimates are based on 10,000 simulation runs.

It is seen from Tables 1 and 3 that both the homoscedastic and heteroscedastic SD1PC, SD2PC, and SD1HC procedures control the FWE at or below $\alpha = 0.05$ for almost all the configurations studied. The FWE exceeds the nominal $\alpha = 0.05$ at the 5% significance level if the estimated FWE exceeds

$$0.05 + 1.96\sqrt{\frac{(0.05)(0.95)}{10000}} = 0.054.$$

Only in a few cases the estimated FWE is >0.054, and the excess is not practically significant in these cases. In many linear configurations, the FWEs of SD1PC and SD2PC are quite small and zero in several cases. This is because those configurations are not least favorable in terms of maximizing the FWE. However, quite a different picture emerges when we look at Table 5. Here we see that even the heteroscedastic procedures fail to control the FWE in about half the cases, but the maximum FWE is only 0.064. The performance of homoscedastic

Table 1. Simulated Type I error rates of SD1PC and SD2PC under constant CV for $\lambda = 0.85$, $\mu_0 = 20$, $\mu_{\text{MAXSD}} = 18$, $\sigma_0 = 1$, $n_0 = 16$, $n_1 = \cdots = n_5 = 8$, and $\alpha = 0.05$.

Function shape	MAXSD	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	Heteroscedastic procedures			Homoscedastic procedures		
			SD1PC	SD2PC	SD1HC	SD1PC	SD2PC	SD1HC
Step	1	(18, 17, 17, 17, 17)	0.053	0.045	0.054	0.048	0.050	0.054
	2	(18, 18, 17, 17, 17)	0.054	0.041	0.051	0.045	0.047	0.052
	2	(20, 18, 17, 17, 17)	0.048	0.048	0.053	0.040	0.042	0.044
	3	(18, 18, 18, 17, 17)	0.046	0.039	0.050	0.040	0.041	0.044
	3	(20, 18, 18, 17, 17)	0.046	0.042	0.053	0.044	0.041	0.045
	3	(20, 20, 18, 17, 17)	0.047	0.046	0.053	0.039	0.039	0.043
	4	(18, 18, 18, 18, 17)	0.044	0.036	0.048	0.026	0.037	0.033
	4	(20, 18, 18, 18, 17)	0.047	0.041	0.051	0.025	0.037	0.031
	4	(20, 20, 18, 18, 17)	0.047	0.045	0.050	0.019	0.033	0.025
	4	(20, 20, 20, 18, 17)	0.044	0.044	0.048	0.024	0.037	0.029
Linear	1	(18, 14, 10, 6, 2)	0.000	0.000	0.000	0.000	0.000	0.000
	1	(18, 17, 15, 13, 11)	0.014	0.044	0.028	0.018	0.062*	0.030
	2	(19, 18, 17, 15, 13)	0.018	0.047	0.031	0.019	0.052	0.029
	2	(19, 18, 14, 10, 6)	0.000	0.000	0.000	0.000	0.000	0.000
	3	(19, 19, 18, 17, 15)	0.027	0.044	0.038	0.025	0.048	0.032
	3	(19, 19, 18, 14, 10)	0.000	0.000	0.000	0.000	0.000	0.000
	4	(19, 19, 19, 18, 17)	0.045	0.044	0.046	0.025	0.039	0.031
	4	(19, 19, 19, 18, 14)	0.000	0.000	0.000	0.000	0.000	0.000

*Indicates that the corresponding estimated FWE significantly exceeds the nominal 5% level.

Table 2. Simulated powers of SD1PC and SD2PC under constant CV for $\lambda = 0.85$, $\mu_0 = 20$, $\mu_{\text{MAXSD}} = 18$, $\sigma_0 = 1$, $n_0 = 16$, $n_1 = \cdots = n_5 = 8$, and $\alpha = 0.05$.

Function shape	MAXSD	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	Heteroscedastic procedures			Homoscedastic procedures		
			SD1PC	SD2PC	SD1HC	SD1PC	SD2PC	SD1HC
Step	1	(18, 17, 17, 17, 17)	0.519	0.753*	0.653	0.584	0.785*	0.673
	2	(18, 18, 17, 17, 17)	0.530	0.615	0.657*	0.594	0.673*	0.671
	2	(20, 18, 17, 17, 17)	0.544	0.745*	0.673	0.599	0.785*	0.678
	3	(18, 18, 18, 17, 17)	0.563	0.516	0.661*	0.622	0.577	0.690*
	3	(20, 18, 18, 17, 17)	0.571	0.615	0.672*	0.618	0.666	0.683*
	3	(20, 20, 18, 17, 17)	0.593	0.750*	0.686	0.626	0.781*	0.689
	4	(18, 18, 18, 18, 17)	0.602	0.436	0.681*	0.672	0.517	0.717*
	4	(20, 18, 18, 18, 17)	0.616	0.509	0.686*	0.661	0.573	0.706*
	4	(20, 20, 18, 18, 17)	0.625	0.610	0.692*	0.680	0.657	0.719*
	4	(20, 20, 20, 18, 17)	0.649	0.744*	0.709	0.691	0.772*	0.722
	5	(18, 18, 18, 18, 18)	0.727	0.417	0.764*	0.719	0.485	0.766*
	5	(20, 18, 18, 18, 18)	0.740	0.490	0.766*	0.718	0.542	0.762*
	5	(20, 20, 18, 18, 18)	0.741	0.561	0.765*	0.713	0.596	0.748*
	5	(20, 20, 20, 18, 18)	0.757	0.657	0.780*	0.714	0.672	0.749*
	5	(20, 20, 20, 20, 18)	0.801*	0.800	0.797	0.725	0.796*	0.757
Linear	1	(18, 14, 10, 6, 2)	0.593	0.799*	0.717	0.775	0.901*	0.834
	1	(18, 17, 15, 13, 11)	0.555	0.748*	0.683	0.676	0.800*	0.748
	2	(19, 18, 17, 15, 13)	0.575	0.738*	0.692	0.662	0.794*	0.730
	2	(19, 18, 14, 10, 6)	0.615	0.793*	0.726	0.736	0.877*	0.799
	3	(19, 19, 18, 17, 15)	0.619	0.749*	0.702	0.659	0.779*	0.714
	3	(19, 19, 18, 14, 10)	0.651	0.798*	0.736	0.717	0.845*	0.769
	4	(19, 19, 19, 18, 17)	0.649	0.749*	0.705	0.701	0.780*	0.731
	4	(19, 19, 19, 18, 14)	0.694	0.791*	0.755	0.738	0.826*	0.769

*Indicates that the corresponding estimated power is the highest among the three procedures for the given configuration.

procedures is considerably worse with the maximum FWE as high as 0.142. This shows that the heteroscedastic procedures are much more robust, although not completely fool-proof.

We now turn to Tables 2 and 4 for the powers of SD1PC, SD2PC, and SD1HC (both their homoscedastic and heteroscedastic versions). The highest of the three powers for each version is marked with an asterisk in each case. We see that for step-shaped response functions SD1HC generally has the highest power, whereas for linear-shaped response functions SD2PC has the highest power. Therefore, SD1PC is not a contender. However, note that for step-shaped response functions, SD2PC has a higher power than SD1PC when MAXSD is low (doses 1–3), whereas the opposite is true when MAXSD is high (doses 4 and 5). These results are in agreement with those in Tamhane et al. (2001). What is surprising is that in almost all cases, the power of a homoscedastic procedure (SD1PC, SD2PC, or SD1HC) is higher than that of its heteroscedastic version. As mentioned above, this is because a pooled variance estimate is used, which is biased but has substantially less variance (and greater d.f.) for the small sample sizes used here. As long as the bias is negligible

Table 3. Simulated Type I error rates of SD1PC and SD2PC for $\lambda = 0.85$, $\mu_0 = 20$, $\mu_{MAXSD} = 18$, $\sigma_0 = 0.5$, $\sigma_1 = \cdots = \sigma_5 = 1$, $n_0 = 16$, $n_1 = \cdots = n_5 = 8$, and $\alpha = 0.05$.

Function shape	MAXSD	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	Heteroscedastic procedures			Homoscedastic procedures		
			SD1PC	SD2PC	SD1HC	SD1PC	SD2PC	SD1HC
Step	1	(18, 17, 17, 17, 17)	0.050	0.043	0.054	0.056*	0.050	0.045
	2	(18, 18, 17, 17, 17)	0.051	0.036	0.053	0.055*	0.044	0.048
	2	(20, 18, 17, 17, 17)	0.054	0.041	0.050	0.056*	0.049	0.048
	3	(18, 18, 18, 17, 17)	0.048	0.033	0.048	0.053	0.041	0.052
	3	(20, 18, 18, 17, 17)	0.050	0.036	0.056*	0.054	0.045	0.052
	3	(20, 20, 18, 17, 17)	0.046	0.044	0.055*	0.054	0.051	0.052
	4	(18, 18, 18, 18, 17)	0.039	0.025	0.053	0.032	0.037	0.042
	4	(20, 18, 18, 18, 17)	0.041	0.030	0.054	0.032	0.040	0.042
	4	(20, 20, 18, 18, 17)	0.042	0.033	0.047	0.032	0.044	0.042
	4	(20, 20, 20, 18, 17)	0.049	0.045	0.046	0.033	0.049	0.042
Linear	1	(18, 14, 10, 6, 2)	0.000	0.000	0.000	0.000	0.000	0.000
	1	(18, 17, 15, 13, 11)	0.014	0.040	0.030	0.015	0.050	0.026
	2	(19, 18, 17, 15, 13)	0.019	0.047	0.030	0.020	0.049	0.028
	2	(19, 18, 14, 10, 6)	0.000	0.000	0.000	0.000	0.000	0.000
	3	(19, 19, 18, 17, 15)	0.028	0.045	0.039	0.030	0.051	0.037
	3	(19, 19, 18, 14, 10)	0.000	0.000	0.000	0.000	0.000	0.000
	4	(19, 19, 19, 18, 17)	0.046	0.044	0.049	0.033	0.049	0.042
	4	(19, 19, 19, 18, 14)	0.000	0.000	0.000	0.000	0.000	0.000

*Indicates that the corresponding estimated FWE significantly exceeds the nominal 5% level.

or is in the conservative direction (SE estimate is too high), the Type I error will still be approximately controlled and the power will be substantially higher for homoscedastic procedures due to the reduced variance of the SE estimate and the resulting increase in d.f. This is the case for most of the configurations (1) and (2). However, in configuration (3), the pooled estimate is liberally biased (too low), and the homoscedastic procedures no longer control the FWE. This would seem to suggest that although the homoscedastic procedures appear to do better, they do not control the FWE for all situations, and so one should consider anticipated variance vs. dose mean patterns in deciding whether a homoscedastic procedures is appropriate. Also note that these results are likely to be sensitive to the samples sizes in the groups. Larger sample sizes will emphasize the importance of bias in the MSE much more, and so the pooled variance used in the homoscedastic procedures will be less beneficial.

EXAMPLE

A 90-day routine rat study was conducted to evaluate the toxicity of a crop protection compound. Test substance was added directly to the rodent diet and was thoroughly mixed to ensure homogeneous distribution. Three doses of the compound were compared with a zero dose control. The sample sizes in the four groups

Table 4. Simulated powers of SD1PC and SD2PC for $\lambda = 0.85$, $\mu_0 = 20$, $\mu_{MAXSD} = 18$, $\sigma_0 = 0.5$, $\sigma_1 = \cdots = \sigma_5 = 1$, $n_0 = 16$, $n_1 = \cdots = n_5 = 8$, and $\alpha = 0.05$.

Function shape	MAXSD	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	Heteroscedastic procedures			Homoscedastic procedures		
			SD1PC	SD2PC	SD1HC	SD1PC	SD2PC	SD1HC
Step	1	(18, 17, 17, 17, 17)	0.486	0.756*	0.682	0.614	0.807*	0.720
	2	(18, 18, 17, 17, 17)	0.497	0.602	0.682*	0.631	0.699	0.712*
	2	(20, 18, 17, 17, 17)	0.518	0.757*	0.688	0.641	0.810*	0.724
	3	(18, 18, 18, 17, 17)	0.539	0.482	0.693*	0.656	0.603	0.723*
	3	(20, 18, 18, 17, 17)	0.552	0.594	0.691*	0.662	0.694	0.725*
	3	(20, 20, 18, 17, 17)	0.579	0.750*	0.692	0.674	0.802*	0.735
	4	(18, 18, 18, 18, 17)	0.593	0.395	0.701*	0.725	0.532	0.761*
	4	(20, 18, 18, 18, 17)	0.609	0.494	0.717*	0.727	0.608	0.761*
	4	(20, 20, 18, 18, 17)	0.613	0.598	0.711*	0.733	0.697	0.763*
	4	(20, 20, 20, 18, 17)	0.642	0.746*	0.726	0.750	0.808*	0.775
	5	(18, 18, 18, 18, 18)	0.725	0.345	0.799*	0.779	0.499	0.808*
	5	(20, 18, 18, 18, 18)	0.733	0.421	0.785*	0.780	0.567	0.809*
	5	(20, 20, 18, 18, 18)	0.749	0.514	0.783*	0.784	0.643	0.809*
	5	(20, 20, 20, 18, 18)	0.765	0.647	0.794*	0.790	0.739	0.812*
	5	(20, 20, 20, 20, 18)	0.804*	0.801	0.804*	0.795	0.850*	0.821
Linear	1	(18, 14, 10, 6, 2)	0.531	0.797*	0.705	0.659	0.857*	0.752
	1	(18, 17, 15, 13, 11)	0.504	0.750*	0.697	0.647	0.807*	0.736
	2	(19, 18, 17, 15, 13)	0.551	0.756*	0.712	0.672	0.810*	0.740
	2	(19, 18, 14, 10, 6)	0.567	0.802*	0.718	0.688	0.859*	0.757
	3	(19, 19, 18, 17, 15)	0.594	0.752*	0.715	0.696	0.802*	0.746
	3	(19, 19, 18, 14, 10)	0.620	0.804*	0.724	0.721	0.852*	0.775
	4	(19, 19, 19, 18, 17)	0.643	0.751*	0.719	0.750	0.808*	0.775
	4	(19, 19, 19, 18, 14)	0.681	0.799*	0.744	0.778	0.857*	0.806

*Indicates that the corresponding estimated power is the highest among the three procedures for the given configuration.

were $n_0 = 18, n_1 = 20, n_2 = 19$, and $n_3 = 18$. The variable of interest was the kidney weight to the body weight ratio. A large value of this ratio is regarded as unsafe with a threshold of a 15% average increase over its value for the zero dose. Thus, we specify $\lambda = 1.15$.

The data are given in Table 6. The box plot of the data is shown in Fig. 1. The dose means seem to increase quadratically. The mean differences are highly significant ($F = 52.82$ with $p = 0.000$). The within group variances also appear to be different across the dose groups (Bartlett's $\chi^2 = 10.446$ with $p = 0.015$, Levene's $F = 3.115$ with $p = 0.030$). However, the normality of the data does not seem to be in serious doubt as can be seen from the normal plot of the residuals in Fig. 2. There appear to be a few outliers in this plot, but they are not identified as such by outlier tests; therefore, we did not delete them. Thus, the assumptions necessary to apply the proposed methods seem to be satisfied.

The inequalities in the hypotheses in Eq. (3) are reversed as follows:

$$H_{0i} : \mu_i \geq \lambda\mu_0 \quad \text{vs.} \quad H_{1i} : \mu_i < \lambda\mu_0 \quad (1 \leq i \leq k).$$

Table 5. Simulated Type I error rates of SD1PC and SD2PC for $\lambda = 0.85$, $\mu_0 = 20$, $\mu_{\text{MAXSD}} = 19$, $\sigma_i = 0.5$ for $0 \le i \le \text{MAXSD}$, $\sigma_i = 1$ for $i > \text{MAXSD}$, $n_0 = 16$, $n_1 = \cdots = n_5 = 8$, and $\alpha = 0.05$.

Function shape	MAXSD	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	Heteroscedastic procedures			Homoscedastic procedures		
			SD1PC	SD2PC	SD1HC	SD1PC	SD2PC	SD1HC
Step	0	(17, 17, 17, 17, 17)	0.064*	0.053	0.060*	0.062*	0.059*	0.047
	1	(19, 17, 17, 17, 17)	0.059*	0.052	0.056*	0.085*	0.069*	0.063*
	2	(19, 19, 17, 17, 17)	0.055*	0.056*	0.059*	0.114*	0.083*	0.085*
	3	(19, 19, 19, 17, 17)	0.056*	0.049	0.053	0.136*	0.110*	0.117*
	4	(19, 19, 19, 19, 17)	0.047	0.046	0.047	0.112*	0.142*	0.125*

*Indicates that the corresponding estimated FWE significantly exceeds the nominal 5% level.

Also, the sign of the test statistic (Eq. (5)) must be reversed. In other words, the pairwise contrast t-statistics are defined as

$$t_i = \frac{\lambda \bar{y}_0 - \bar{y}_i}{\sqrt{\lambda^2 s_0^2/n_0 + s_i^2/n_i}} \quad (1 \le i \le k).$$

Similarly, the signs on the Helmert contrasts defined in Eq. (10) must also be reversed. After making these minor modifications, we are now ready to apply the procedures.

SD1PC And SD2PC Procedures

Table 7 lists the t-statistics and their d.f. computed by using Eqs. (5) and (7), respectively, and their marginal p-values. We first apply SD2PC. Because $p_1 = 0.003 < \alpha = 0.05$, we reject H_{01} and conclude that dose 1 is safe. But $p_2 = 0.751 > \alpha = 0.05$, so H_{02} cannot be rejected and testing stops with $\widehat{\text{MAXSD}} = 1$.

To apply SD1PC, we need to calculate $\tilde{p}_1$ corresponding to $\max_{1 \le j \le 3} t_j = t_{1,\max} = 2.899$ for which we need the correlation matrix of t_1, t_2, t_3. This matrix is as follows [calculated by using Eq. (8)]:

$$\begin{bmatrix} 1 & 0.552 & 0.537 \\ & 1 & 0.520 \\ & & 1 \end{bmatrix}.$$

The average d.f. = 32.58, which is rounded down to 32. Using these values we get $\tilde{p}_1 = 0.009 < \alpha = 0.05$, so we reject H_{01} and conclude that dose 1 is safe. There is no need to calculate $\tilde{p}_2$ because it will be greater than $p_2 = 0.751$ and, hence, greater than $\alpha = 0.05$, so testing stops with $\widehat{\text{MAXSD}} = 1$.

Table 6. Kidney Wt./body Wt. $\times 10^3$.

	Dose			
	0	1	2	3
	6.593	7.062	7.006	9.569
	7.480	7.347	8.706	9.362
	6.930	7.733	7.257	10.911
	5.662	7.396	7.743	9.961
	6.789	8.173	7.026	9.497
	7.268	6.938	8.561	9.911
	6.647	6.988	7.674	8.544
	6.443	6.621	7.450	10.404
	6.713	7.508	8.188	10.421
	6.057	6.657	8.150	10.065
	6.253	7.787	7.619	9.670
	7.045	6.537	8.722	8.194
	6.552	7.369	7.387	8.989
	5.668	6.623	6.798	7.347
	6.354	6.456	7.617	7.260
	6.511	6.507	8.071	9.017
	7.111	6.154	7.020	8.847
	6.015	5.934	7.821	8.723
		6.909	7.063	
		7.252		
n_i	18	20	19	18
Mean	6.5606	6.9975	7.6778	9.2606
SD	0.5094	0.5755	0.5949	1.0052

SD1HC Procedure

We begin by testing H_{01}. The three Helmert contrast t-statistics with their d.f. are computed by using Eqs. (12) and (13):

$$t_{11} = 2.899, \quad t_{12} = 1.240, \quad t_{13} = -2.539,$$
$$\nu_{11} = 35.44, \quad \nu_{12} = 33.05, \quad \nu_{13} = 35.43.$$

The average d.f. $= 34.64$, which is rounded down to 34. The correlation matrix [computed using Eq. (14)] is

$$\begin{bmatrix} 1 & 0.868 & 0.762 \\ & 1 & 0.874 \\ & & 1 \end{bmatrix}.$$

Notice that these correlations are higher than those for pairwise contrasts. As a result, the adjusted p-value, $\tilde{p}_1$, corresponding to $\max_{1\le j\le 3} t_{1j} = t_{1,\max} = 2.899$ is

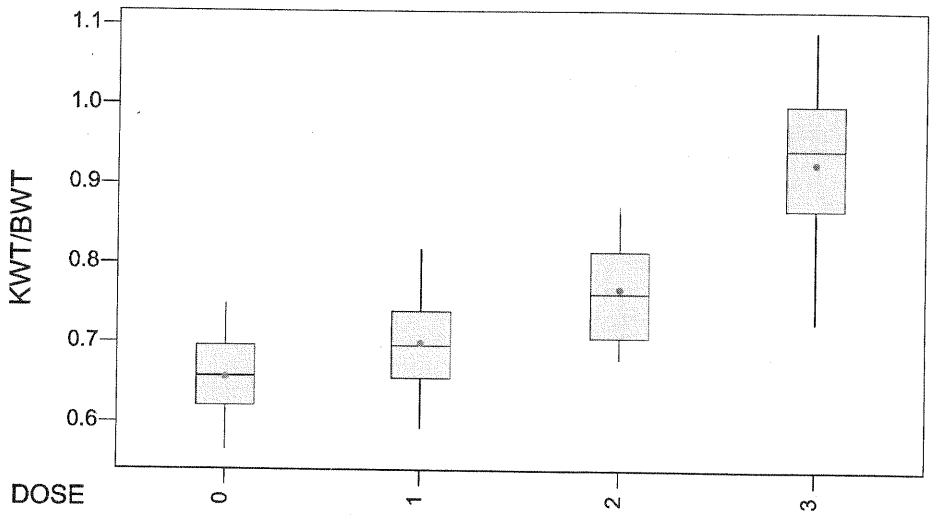

Figure 1. Box plot of kidney data.

smaller, namely, $\tilde{p}_1 = 0.007 < \alpha = 0.05$. Thus, H_{01} is rejected. The t-statistics and their d.f. for testing H_{02} are

$$t_{22} = -0.686, \quad t_{23} = -4.758; \quad \nu_{22} = 34.94, \quad \nu_{23} = 41.72.$$

Clearly, $t_{2,\max}$ is nonsignificant, so testing stops with $\widehat{\text{MAXSD}} = 1$.

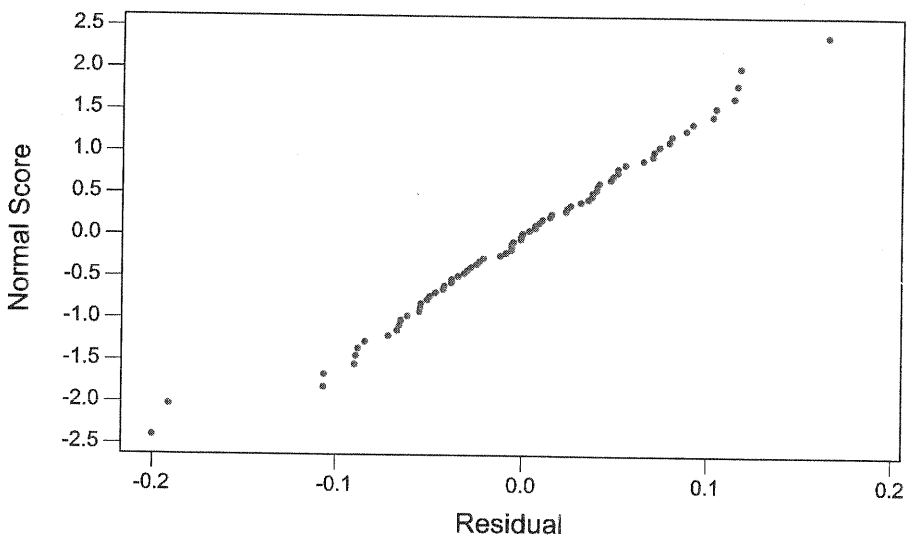

Figure 2. Normal plot of the residuals for kidney data.

Table 7. t-Statistics for pairwise contrasts, their degrees of freedom, and p-values.

	Dose		
i	1	2	3
t_i	2.899	−0.686	−6.257
ν_i	35.44	34.94	27.35
p_i	0.003	0.751	1.00

It is worth noting that if we use, $\lambda = 1.10$ then $p_1 = 0.121$ for SD2PC, $\tilde{p}_1 = 0.254$ for SD1PC and $\tilde{p}_1 = 0.196$ for SD1HC. Thus, dose 1 is not found to be safe by any of the procedures.

CONCLUDING REMARKS

We have given heteroscedastic extensions of SD1PC, SD2PC, and SD1HC for identifying MAXSD in dose–response studies for safety testing. All three procedures control the Type I error rates. However, in terms of power, SD1PC is dominated by either SD2PC or SD1HC as in the homoscedastic case considered in Tamhane et al. (2001). So either SD2PC or SD1HC is preferred, the former in the linear dose response case and the latter in the step dose response case. SD2PC has the advantage of being simpler and more generally applicable, but its power can be quite low for the step dose response case if MAXSD is a high dose.

ACKNOWLEDGMENTS

This research was partially supported by an educational grant from DuPont. We thank Dr. John Green of DuPont Safety, Health and Environmental Excellence Center, Haskell Laboratory for Toxicology and Industrial Medicine for providing the data for the problem and helpful discussions. We also thank two referees for useful suggestions.

REFERENCES

Bauer, P. (1997). A note on multiple testing procedure in dose finding. *Biometrics* 53:1125–1128.

Dunnett, C. W. (1989). Multivariate normal probability integrals with product correlation structure. *Appl. Statist.* 38:564–579. Algorithm AS251.

Genz, A., Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *J. Statist. Comput. Simul.* 63:361–378.

Hochberg, Y., Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York: Wiley.

Tamhane, A. C., Dunnett, C. W., Green, J. W., Wetherington, J. F. (2001). Multiple test procedures for identifying a safe dose. *J. Am. Statist. Assoc.* 96: 835–843.

Received May 2003
Accepted August 2003