Biometrics 64, 877–885
September 2008

DOI: 10.1111/j.1541-0420.2007.00958.x

Stepwise Confidence Intervals for Monotone Dose–Response Studies

Jianan Peng,[1,*] Chu-In Charles Lee,[2] Karelyn A. Davis,[3] and Weizhen Wang[4]

[1]Department of Mathematics and Statistics, Acadia University, Wolfville, Nova Scotia, B4P 2R6 Canada
[2]Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Newfoundland, A1C 5S7 Canada
[3]School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6 Canada
[4]Department of Mathematics and Statistics, Wright State University, Dayton, Ohio 45435, U.S.A.
[*]*email:* jianan.peng@acadiau.ca

Summary. In dose–response studies, one of the most important issues is the identification of the minimum effective dose (MED), where the MED is defined as the lowest dose such that the mean response is better than the mean response of a zero-dose control by a clinically significant difference. Dose–response curves are sometimes monotonic in nature. To find the MED, various authors have proposed step-down test procedures based on contrasts among the sample means. In this article, we improve upon the method of Marcus and Peritz (1976, *Journal of the Royal Statistical Society, Series B* **38**, 157–165) and implement the dose–response method of Hsu and Berger (1999, *Journal of the American Statistical Association* **94**, 468–482) to construct the lower confidence bound for the difference between the mean response of any nonzero-dose level and that of the control under the monotonicity assumption to identify the MED. The proposed method is illustrated by numerical examples, and simulation studies on power comparisons are presented.

Key words: Dose–response study; Minimum effective dose; Multiple contrast test; Order-restricted inference.

1. Introduction

Dose–response studies are of central importance in pharmaceutical research and other areas such as behavioral analysis. For this purpose, a dose–response (DR) experiment is conducted in which several doses are administered to separate groups of experimental units. In most cases, a zero-dose control group, or placebo group, is included which serves as a standard against which the dose groups are to be compared with. The primary goal is to assess whether there is indeed a DR effect, and if a DR effect is found, then to identify the lowest dose level producing a desirable effect over that of the control. This lowest dose level is commonly referred to as the minimum effective dose (MED) (Ruberg, 1989). For example, Wöhr, Borta, and Schwarting (2005) presented summary data of a DR study of the rat, concerning the usefulness of immobility as a measure of fear conditioning. Four shock doses (0.2 mA $<$ 0.5 mA $<$ 0.8 mA $<$ 1.1 mA) were compared with the zero-dose control. It may be of interest to know the minimum effective (shock) dose, even though it is believed that the higher the shock dose, the more immobile are the animals.

Statistical determination of the MED may involve multiple testing procedures, regression methods (Tukey, Ciminera, and Heyse, 1985) or combination of both (Bretz, Pinheiro, and Branson, 2005). Many authors have proposed multiple test procedures, which test the appropriate hypotheses in a fixed sequence, to determine the MED, see Williams (1971, 1972), Ruberg (1989), Tamhane, Hochberg, and Dunnett (1996), Dunnett and Tamhane (1998), and Hellmich and Lehmacher (2005), among others. The present article improves upon previous methods by incorporating a monotonicity assumption of DR means and obtaining confidence intervals for use in a stepwise fashion.

For compatibility with existing literature, we consider a one-way layout $Y_{ij} = \mu_i + \epsilon_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, where ϵ_{ij} are i.i.d. normal with mean 0 and variance σ^2. The control group is indexed as 1 and the remaining $k-1$ treatment groups correspond to the $k-1$ increasing dose levels, with n_i subjects randomly assigned to group i, $i = 1, \ldots, k$. The statistic $S^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2/\nu$ is used as an estimator for σ^2, and it is independent of the sample means $\bar{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_k)$, where $\nu S^2/\sigma^2 \sim \chi^2_\nu$ and $\nu = \sum_{i=1}^{k} n_i - k > 0$. Suppose that the control group is a negative control group receiving a placebo. If the DR is monotonic, then defining MED as the minimum dose i such that $\mu_i > \mu_1 (i \geq 2)$ is meaningless as $MED = 2$ in this case. Accordingly, if the DR is continuous, then the MED should be defined as the minimum dose such that the mean response at that dose is clinically significantly better than the mean response of the control; that is

$$MED = \min\{i : \mu_i > \mu_1 + \delta\}, \tag{1}$$

where δ defines a clinically significant difference preassigned by an experimenter. Hsu and Berger (1999) provide further details of the above MED definition. The estimated MED ($\widehat{MED}$) is determined statistically from the observed DR relationship. For convenience, we write $MED = i$ and $\widehat{MED} = i$

to represent that the true MED or the estimated MED is the dose corresponding to μ_i.

For the MED problem, an error arises if the inferred MED, and any of the doses higher than the inferred MED, is in fact not efficacious. That is, the probability of declaring an ineffective dose to be effective should be controlled. Previous formulations of the MED problem, which have cast it as one of testing a family of null hypotheses of equalities against various alternatives, fail to control the aforementioned error rate, which is directional, as explained in Hsu and Berger (1999).

In some disease categories the DR is expected to be monotone, while in others the DR may be inverted-U or umbrella shaped, as stated in the International Conference on Harmonization (ICH) guideline E4. Bretz et al. (2005) considered 12 DR models in their simulation study, of which only the double-logistic model and quadratic model do not satisfy the monotonicity assumption. In this article, we assume the DR is monotonic, that is, $\mu_1 \leq \ldots \leq \mu_k$, and a larger μ_i indicates a better average outcome.

It is well known that interval estimation provides a visual perspective unmatched by a point estimate or a test statistic. The ICH guideline E9 "Statistical Principle for Clinical Trials" suggests that estimates of treatment effects should be accompanied by confidence intervals, whenever possible. More importantly, a lower confidence bounds method such as the DR method of Hsu and Berger (1999) automatically controls the probability of inferring any ineffective dose as effective (Hsu [1996] has pointed out that confidence intervals inference implies confident directions inference), as illustrated by the following theorem (suggested by the associate editor).

THEOREM 1.1. *Denote the lower confidence bound for* $\mu_i - \mu_1$ *by* $L_i(\bar{Y})(2 \leq i \leq k)$. *Assume that* $P\{\mu_{\widehat{MED}} - \mu_1 > L_{\widehat{MED}}(\bar{Y})\} \geq 1 - \alpha$ *and* $L_{\widehat{MED}}(\bar{Y}) \geq \delta$. *Then under the monotonicity assumption* $\mu_1 \leq \ldots \leq \mu_k$, *the probability of inferring any ineffective dose as effective is* $\leq \alpha$.

Proof. Under the assumption that $\mu_1 \leq \ldots \leq \mu_k, \mu_j - \mu_1 \geq \mu_{\widehat{MED}} - \mu_1$ for $\widehat{MED} \leq j \leq k$. Then

$$P\{\mu_{\widehat{MED}} - \mu_1 > L_{\widehat{MED}}(\bar{Y})\} = P\{\mu_j - \mu_1 > L_{\widehat{MED}}(\bar{Y}), \; j = \widehat{MED}, \ldots, k\} \geq 1 - \alpha.$$

Therefore, the result follows.

The SD2 methods of Dunnett and Tamhane (1998) cannot achieve the result in Theorem 1.1. Bauer (1997) and particularly Hsu and Berger (1999) indicate that it is dangerous to recklessly test equalities to detect MED.

The method in Hsu and Berger (1999) has wide applicability, nevertheless monotonicity is not assumed as it is based on pairwise t comparisons. However, such pairwise comparisons that do not fully use the natural ordering of doses have low power in some directions. In this article, we consider how to construct the lower confidence bound for $\mu_k - \mu_1$ under the monotonicity assumption and then adopt the DR method in Hsu and Berger (1999) to account for the monotone DR shape.

Marcus and Peritz (1976) discussed the construction of simultaneous lower confidence bounds under certain restricted normal model with known variances. They computed numerous partitions and then selected the maximum lower bound. Their process is very lengthy (see p. 160 of their paper). To improve computational efficiency, we propose a method analogous to Lee, Peng, and Liu (2006), which uses the Kuhn–Tucker (1951) equivalence theorem to formulate the necessary and sufficient condition for the partition.

This article is organized as follows. Constructing the lower confidence bound for $\mu_k - \mu_1$ is described in Section 2. The stepwise confidence interval procedure is introduced in Section 3. We apply our method to two data sets in Section 4 and report simulation results in Section 5. A brief discussion is given in Section 6 and proofs are provided in the Appendix on the website `http://www.biometrics.tibs.org`.

The Lower Confidence Bound for $\mu_k - \mu_1$

In this section we first consider how to test whether there is a DR effect by testing $H_{0k} : \mu_1 = \mu_2 = \cdots = \mu_k$ versus $H_{1k} : \mu_1 < \mu_k$ under the monotonicity assumption and outline how to construct a lower confidence bound for $\mu_k - \mu_1$. Although there are many tests available for assessing DR effects, we focus on the likelihood ratio test for its good power performance for a wide range of DR profiles and it is also related to the multiple contrast (MC) test statistic that we use to construct the lower confidence bound.

2.1 *The Likelihood Ratio Test*

Under the monotonicity assumption $\Omega_k = \{\mu : \mu_1 \leq \ldots \leq \mu_k\}$, the likelihood ratio test rejects H_{0k} in favor of H_{1k} for large values of

$$S_{01} = \sum_{i=1}^{k} n_i(\mu_i^* - \hat{\mu})^2 \Bigg/ \left\{ \sum_{i=1}^{k} n_i(\bar{Y}_i - \mu_i^*)^2/\nu + S^2 \right\},$$

where $\hat{\mu} = \sum_{i=1}^{k} n_i \bar{Y}_i / \sum_{i=1}^{k} n_i$ and μ_i^* are the isotonic (maximum likelihood) estimates of μ_i under the monotonicity assumption, which can be computed by the pool-adjacent-violators algorithm (PAVA) (see Robertson, Wright, and Dykstra (1988) or Silvapulle and Sen (2005)). The null distribution of S_{01} under H_{0k} is given by

$$P(S_{01} > s) = \sum_{j=2}^{k} P(j, k; \mathbf{w}) P\left\{ F_{j-1,N-j} > \frac{s(N-j)}{\nu(j-1)} \right\} \quad (2)$$

for any $s > 0$, where $N = \sum_{i=1}^{k} n_i, \mathbf{w} = (n_1, \ldots, n_k)$, and $P(j, k; \mathbf{w})$ is the level probability under H_{0k} that $\boldsymbol{\mu}^*$ takes j distinct values. For the equal weights case, $P(j, k; \mathbf{w})$ and the critical value $s_{k,\nu,\alpha}$, the solution when one equates (2) to α, can be found in Tables A.10 and A.6 ($w = 1$) of Robertson et al. (1988), respectively.

2.2 *The Multiple Contrast Test Statistic* T_k

When $S_{01} > s_{k,\nu,\alpha}$, one rejects H_{0k} and concludes that there is indeed a DR effect; in other words, at least one mean response $\mu_i (2 \leq i \leq k)$ is significantly larger than μ_1. However, there is no corresponding lower confidence bounds for $\mu_k - \mu_1$ when $k > 2$. We introduce the following MC test statistic:

$$T_k = \max_{\mathbf{c} \in \mathbf{C_k}} \sum_{i=1}^{k} n_i c_i \bar{Y}_i \Bigg/ S\left(\sum_{i=1}^{k} n_i c_i^2 \right)^{1/2}, \quad (3)$$

where $\mathbf{C}_k = \{\mathbf{c} = (c_1, \ldots, c_k) : \sum_{i=1}^k n_i c_i = 0, c_1 \le \ldots \le c_k\}$. Following a similar argument as in Hogg (1965), we have

$$T_k^2 = \sum_{i=1}^k n_i(\mu_i^* - \hat{\mu})^2/S^2. \tag{4}$$

The right-hand side of (4) was given by Wright (1988) but it was derived for the purpose of testing H_{0k} versus H_{1k}. Let $t_{k,\nu,\alpha}$ be the critical value of T_k, then

$$P_{\boldsymbol{\mu}}\left\{\sum_{i=1}^k n_i c_i \mu_i \ge \sum_{i=1}^k n_i c_i \bar{Y}_i - t_{k,\nu,\alpha} S\left(\sum_{i=1}^k n_i c_i^2\right)^{1/2}, \text{ for all } \mathbf{c} \in \mathbf{C_k}\right\} = 1-\alpha. \tag{5}$$

When the variance σ^2 is known, Marcus and Peritz (1976) derived (5). The left-hand side of (5) can be rewritten as

$$\begin{aligned} &P_{\boldsymbol{\mu}}\left\{\max_{\mathbf{c}\in\mathbf{C_k}} \sum_{i=1}^k n_i c_i(\bar{Y}_i - \mu_i)/S\left(\sum_{i=1}^k n_i c_i^2\right)^{1/2} \le t_{k,\nu,\alpha}\right\} \\ &= P_{\mathbf{0}}\left\{\max_{c\in\mathbf{C_k}} \sum_{i=1}^k n_i c_i \bar{Y}_i/S\left(\sum_{i=1}^k n_i c_i^2\right)^{1/2} \le t_{k,\nu,\alpha}\right\} \\ &= P_{\mathbf{0}}\left\{T_k^2 \le t_{k,\nu,\alpha}^2\right\}. \end{aligned}$$

The statistic T_k^2 is asymptotically equivalent to S_{01} (see Wright, 1988). The null distribution of T_k under H_{0k} is given by

$$P(T_k \ge t) = \sum_{j=2}^k P(j,k;\mathbf{w}) P\left\{F_{j-1,\nu} \ge \frac{t^2}{j-1}\right\} \tag{6}$$

for any $t > 0$. The critical value $t_{k,\nu,\alpha}$ is the solution when one equates (6) to α. For equal sample size case, one can obtain $t_{k,\nu,\alpha}$ by applying Table A.10 in Robertson et al. (1988) to (6); if further the variance σ^2 is known, then $t_{k,\nu,\alpha}$ is the square root of the corresponding critical value in Table A.4 of Robertson et al. (1988). For unequal sample sizes case, one can use the method in Miwa, Hayter, and Liu (2000) to obtain the critical values of T_k. The `R` code to obtain the p-value of statistic T_k is available from the second author upon request.

2.3 *The Lower Confidence Bound for* $\mu_k - \mu_1$

According to (5), the $100(1-\alpha)\%$ one-sided simultaneous confidence bounds for any contrast $\sum_{i=1}^k n_i c_i \mu_i$ with $c_1 \le \ldots \le c_k$, is given by

$$l\left(\sum_{i=1}^k n_i c_i \mu_i\right) = \sum_{i=1}^k n_i c_i \bar{Y}_i - t_{k,\nu,\alpha} S\left(\sum_{i=1}^k n_i c_i^2\right)^{1/2}.$$

However, we only focus on the lower confidence bound $\mu_k - \mu_1$ because its size may be useful to the experimenter in assessing the actual treatment effect difference between the largest dose and the control. Let $\mathcal{K}_k = \{\mathbf{c} : \mathbf{c} \in \mathbf{C}_k, \Sigma_{i=1}^k n_i c_i \mu_i \le \mu_k - \mu_1, \text{ for all } \boldsymbol{\mu} \in \Omega_k\}$. The largest lower confidence bound for $\mu_k - \mu_1$ is given by

$$L(\mu_k - \mu_1) = \max_{\mathbf{c}\in\mathcal{K}_k} l\left(\sum_{i=1}^k n_i c_i \mu_i\right). \tag{7}$$

The Web Appendix provides a proof that the optimization solution to (7) is equivalent to the solution to the following optimization problem:

$$\max\left\{\sum_{i=1}^k n_i c_i \mu_i^* - t_{k,\nu,\alpha} S\left(\sum_{i=1}^k n_i c_i^2\right)^{1/2}\right\}$$

subject to $\mathbf{c} \in \mathbf{C}_k$ and $\sum_{i=j}^k n_i c_i \le 1,\ j = 2, \ldots, k$.

The following theorem establishes a duality between the test statistic T_k in (3) and the above maximized lower bound in (7). Its proof can be found in the Web Appendix.

THEOREM 2.1. *When $\boldsymbol{\mu} \in \Omega_k$, we have that $T_k > t_{k,\nu,\alpha}$ if and only if $L(\mu_k - \mu_1) > 0$.*

Marcus and Peritz (1976) calculated the optimal solution by trying out all the partitions of $\{1, 2, \ldots, k\}$. For large values of k, their method becomes tedious, and a more efficient method is necessary. The following theorem establishes a necessary and sufficient condition for the optimization solution and its proof is provided in the Web Appendix.

THEOREM 2.2. *Suppose that $T_k > t_{k,\nu,\alpha}$. The vector $\mathbf{c}^o \in \mathcal{K}_k$ is an optimal solution to (7) if and only if there exist positive integers p and q, $1 \le p < q \le k$, such that $\mu_p^* < \hat{\mu} < \mu_q^*$, $S_{1p}^2 + S_{qk}^2 < S^2 t_{k,\nu,\alpha}^2$ and $c^o{}_1 \le \cdots \le c_p^o < c_{p+1}^o = \cdots = c_{q-1}^o = 0 < c_q^o \le \cdots \le c_k^o$, where $c_i^o = -N_{1p}^{-1} + b^{-1}(\mu_i^* - \bar{Y}_{1p}), i = 1, \ldots, p$, and $c_i^o = N_{qk}^{-1} + b^{-1}(\mu_i^* - \bar{Y}_{qk}), i = q, \ldots, k$, with*

$$\begin{aligned} &\max\{N_{1p}(\mu_p^* - \bar{Y}_{1p}), N_{qk}(\bar{Y}_{qk} - \mu_q^*)\} < b \\ &\quad \le \min\{N_{1(p+1)}(\mu_{p+1}^* - \bar{Y}_{1(p+1)}), N_{(q-1)k}(\bar{Y}_{(q-1)k} - \mu_{q-1}^*)\}, \end{aligned} \tag{8}$$

where

$$b^2 = \left(t_{k,\nu,\alpha}^2 S^2 - S_{1p}^2 - S_{qk}^2\right) \Big/ \left(N_{1p}^{-1} + N_{qk}^{-1}\right), \tag{9}$$

and $N_{ab} = \sum_{i=a}^b n_i$, $\bar{Y}_{ab} = \sum_{i=a}^b n_i \mu_i^*/N_{ab}$, $S_{ab}^2 = \sum_{i=a}^b n_i \times (\mu_i^* - \bar{Y}_{ab})^2$. *When $q = p+1$, the upper bound for b in (8) is replaced by $(\bar{Y}_{qk} - \bar{Y}_{1p})/(N_{1p}^{-1} + N_{qk}^{-1})$.*

The optimized solution $\mathbf{c}^o$ can be obtained iteratively in a few steps by the following algorithm, which is a vast improvement over Marcus and Peritz's (1976) method no matter whether the variance σ^2 is known or not.

1. Set $i = 0$, $p_0 = \max\{1 \le j < k : \mu_j^* < \hat{\mu}\}$ and $q_0 = \min\{2 \le j \le k : \mu_j^* > \hat{\mu}\}$.
2. Let $\beta_{i+1} = \max\{N_{1p_i}(\mu_{p_i}^* - \bar{Y}_{1p_i}), N_{q_i k}(\bar{Y}_{q_i k} - \mu_{q_i}^*)\}$, $t_{k,\nu,\alpha_{i+1}} = \{S_{1p_i}^2 + S_{q_i k}^2 + (N_{1p_i}^{-1} + N_{q_i k}^{-1})\beta_{i+1}^2\}^{1/2}/S$. If $t_{k,\nu,\alpha_{i+1}} < t_{k,\nu,\alpha}$, the optimization solution is $\mathbf{c}^o$ with $p = p_i$ and $q = q_i$. Otherwise, go to (ii).
3. If $N_{1p_i}(\mu_{p_i}^* - \bar{Y}_{1p_i}) > N_{q_i k}(\bar{Y}_{q_i k} - \mu_{q_i}^*)$, then set $p_{i+1} = \max\{j : 1 \le j < p_i, \mu_j^* < \mu_{p_i}^*\}$ and $q_{i+1} = q_i$. Otherwise, set $p_{i+1} = p_i$ and $q_{i+1} = \min\{j : q_i < j \le k, \mu_j^* > \mu_{q_i}^*\}$. Set $i = i+1$, go to step (i).

3. The Stepwise Confidence Interval Procedure

With a sequence of test statistics T_k, T_{k-1}, ... , T_2, we propose the following procedure to find the MED by

making use of the monotonicity assumption. This procedure follows the partitioning principle in Hsu and Berger (1999). It should be pointed out that when computing $L(\mu_i - \mu_1) = \max_{\mathbf{c}\in\mathbf{C_i}} \{ \sum_{j=1}^{i} n_j c_j \mu_j^* - t_{i,\alpha,\nu} S \sqrt{\sum_{j=1}^{i} n_j c_j^2} \}$ subject to $\sum_{j=l}^{i} n_j c_j \le 1, l = 2, \ldots, i,$ where $\mathbf{C}_i = \{\mathbf{c} = (c_1, \ldots, c_i) : \sum_{j=1}^{i} n_j c_j = 0, c_1 \le \ldots \le c_i\}$, it is necessary to recalculate μ_1^*, $\ldots$, μ_i^* using only the data from the dose levels, 1, $\ldots$, i (i = 2, $\ldots$, k). However, the same estimate S^2 of σ^2 may be used.

1. If $L(\mu_k - \mu_1) \ge \delta$, then claim $\mu_k > \mu_1 + \delta$ and go to Step 2; else claim that there is no nonzero-dose level which is significantly better than the control and $\mu_k - \mu_1 > L(\mu_k - \mu_1)$ and stop.

2. If $L(\mu_{k-1} - \mu_1) \ge \delta$, then claim $\mu_{k-1} > \mu_1 + \delta$ and go to Step 3; else claim $\widehat{MED} = k$ and $\mu_{k-1} - \mu_1 > L(\mu_{k-1} - \mu_1)$, then stop.

$\vdots$

k-1. If $L(\mu_2 - \mu_1) \ge \delta$, then claim $\mu_2 > \mu_1 + \delta$ and go to Step k; else claim $\widehat{MED} = 3$ and $\mu_2 - \mu_1 > L(\mu_2 - \mu_1)$ (when $\bar{Y}_2 > \bar{Y}_1$, this lower bound is the same as $\bar{Y}_2 - \bar{Y}_1 - t_{\alpha,\nu} S \sqrt{1/n_1 + 1/n_2}$, where $t_{\alpha,\nu}$ is the t critical value), then stop.

k. Claim every dose level is significantly better than the control, i.e., $\widehat{MED} = 2$ and stop.

To help understand how this stepwise method works, let step M ($1 \le M \le k$) be the step at which the stepwise method stops. If $M > 1$, then the stepwise method declares doses $k - M + 2, \ldots, k$ to be efficacious. If $M < k$, then the stepwise method fails to declare doses 2, $\ldots$, $k - M + 1$ to be efficacious, and gives a lower confidence bound (which is less than δ) for $\mu_{k-M+1} - \mu_1$. If $M = k$, then the stepwise method gives a lower bound on how efficacious every dose is (this lower bound is greater than δ).

4. Examples

The DR method, the Dunnett (1955) method (in the step-down fashion), and our MC method are the methods that can be used to construct the lower confidence bound for the efficacy of every dose. To illustrate their differences, we consider two numerical examples in this section.

Example 1. Consider the data set in Table 1 which is taken from Table 1 of Wöhr et al. (2005). As mentioned in Section 1, it is believed that the DR is monotone. The critical values of our MC test statistic computed by our `R` program are $t_{5,28,0.05} = 2.370$, $t_{4,28,0.05} = 2.221$, and $t_{3,28,0.05} = 2.034$. The critical values of Dunnett's statistic computed by Dunnett (1989) are $d_{5,28,0.05} = 2.268$, $d_{4,28,0.05} = 2.154$, and $d_{3,28,0.05} = 1.994$. The one-sided two sample t critical value $t_{28,0.05} = 1.701$. Table 2 shows that the step-down 95% lower confidence bounds on $\mu_i - \mu_1$, i = 5, 4, 3, given by the three methods. For illustration, if $\delta = 10.0$, then $\widehat{MED} = 3$ by these three methods; if $\delta = 11.0$, then $\widehat{MED} = 3$ by the DR method and the MC method but not by the Dunnett method; if $\delta = 13.0$, then $\widehat{MED} = 3$ only by the MC method.

Table 1
Immobility/min during the context phase (1–3 min) of the retest day

Dosage	Sample size	Mean response	SEM response
0.0 mA	7	8.89	3.96
0.2 mA	7	5.36	1.87
0.5 mA	7	32.01	6.29
0.8 mA	7	42.75	4.93
1.1 mA	5	48.06	3.55

Table 2
Step-down 95% lower confidence bounds for $\mu_i - \mu_1$ in example 1

$\mu_i - \mu_1$	DUNNETT	DR	MC
$\mu_5 - \mu_1$	23.82	27.66	27.67
$\mu_4 - \mu_1$	20.55	23.35	23.74
$\mu_3 - \mu_1$	10.80	12.61	14.00

Example 2. An example in Williams (1971) compares six dose levels with a zero-dose control under the monotonicity assumption. The sample means are $\bar{Y}_1 = 10.4, \bar{Y}_2 = 9.9, \bar{Y}_3 = 10.0, \bar{Y}_4 = 10.6, \bar{Y}_5 = 11.4, \bar{Y}_6 = 11.9, \bar{Y}_7 = 11.7$ with equal sample size 8, where $S^2 = 1.16$ and degrees of freedom 42. Williams (1971) concluded that dose 5 is the MED at 5% level for $\delta = 0$. The critical values of our MC test statistic computed by (6) in Section 2.2 are $t_{7,42,0.05} = 2.486$, $t_{6,42,0.05} = 2.410$, and $t_{5,42,0.05} = 2.315$. The critical values of Dunnett's statistic computed by Dunnett (1989) are $d_{7,42,0.05} = 2.368$, $d_{6,42,0.05} = 2.306$, and $d_{5,42,0.05} = 2.227$. The one-sided two sample t critical value $t_{42,0.05} = 1.682$. The step-down 95% lower confidence bounds on $\mu_i - \mu_1$, i = 7, 6, 5 are 0.22, 0.26, and -0.20 by the Dunnett method, 0.39, 0.59, and 0.09 by the DR method, and 0.84, 0.78, and 0.28 by the MC method, respectively. For illustration, if $\delta = 0.2$, then both the DR method and the Dunnett method infer that $\widehat{MED} = 6$, but the MC procedure infers that $\widehat{MED} = 5$; for any $0 \le \delta \le 0.05$, both the DR method and the MC procedure infer that $\widehat{MED} = 5$, but the Dunnett procedure infers that $\widehat{MED} = 6$. Also note that the MC procedure gives the sharpest lower bound for the example and these lower bounds for $\mu_i - \mu_1$ are decreasing for the MC procedure but not for the DR method and the Dunnett method.

5. Simulation Results

A simulation study was conducted to compare the behavior of the newly defined MC method with methods based on linear contrasts (denoted by LC), Helmert contrasts (HC), reverse Helmert contrasts (RH) (see Tamhane et al., 1996), Williams' (1971) procedure (W), and the DR method. We fixed $\alpha = 0.05, \sigma/\sqrt{n} = 1, \nu = \infty, k = 6, \mu_1 = 0$. Both monotone and nonmonotone dose responses were considered. The monotone DR functions were of two types, linear and step. For a given type of monotone response, $\mu_k = 5$. For the nonmonotone case, the configurations were both mild and severe violations

Table 3
Estimated FWER and probability of identifying true MED under monotone configurations

Configuration	δ	MED	HC	RH	LC	DR	W	MC
(0,1,2,3,4,5)	2.5	4	0.0002	0.0093	0.0290	0.0484	0.0523	0.0293
		FWER	0.0000	0.0056	0.0060	0.0073	0.0093	0.0075
	2.0	4	0.0014	0.0296	0.0720	0.0910	0.0980	0.0693
		FWER	0.0000	0.0162	0.0235	0.0243	0.0268	0.0191
	1.5	3	0.0004	0.0315	0.0529	0.0458	0.0506	0.0395
		FWER	0.0001	0.0139	0.0053	0.0065	0.0079	0.0080
	1.0	3	0.0043	0.0646	0.1053	0.0909	0.0988	0.0820
		FWER	0.0001	0.0381	0.0182	0.0208	0.0268	0.0272
(0,0,0,0,0,5)	2.5	6	0.7351	0.0009	0.0880	0.5433	0.4983	0.5540
		FWER	0.0001	0.0000	0.0000	0.0001	0.0001	0.0001
	2.0	6	0.8632	0.0044	0.2059	0.6794	0.6375	0.7204
		FWER	0.0002	0.0002	0.0002	0.0009	0.0006	0.0001
	1.5	6	0.9351	0.0145	0.3849	0.7894	0.7571	0.8507
		FWER	0.0011	0.0010	0.0011	0.0030	0.0023	0.0008
	1.0	6	0.9699	0.0407	0.5967	0.8701	0.8475	0.9271
		FWER	0.0060	0.0050	0.0049	0.0093	0.0068	0.0042
(0,0,0,0,5,5)	2.5	5	0.2839	0.0021	0.1871	0.3803	0.4200	0.5429
		FWER	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
	2	5	0.4891	0.0089	0.3437	0.5302	0.5758	0.7210
		FWER	0.0001	0.0002	0.0001	0.0006	0.0002	0.0000
	1.5	5	0.6836	0.0263	0.5319	0.6747	0.7139	0.8504
		FWER	0.0011	0.0013	0.0010	0.0029	0.0024	0.0010
	1.0	5	0.8353	0.0705	0.7108	0.7917	0.8224	0.9259
		FWER	0.0054	0.0045	0.0045	0.0079	0.0069	0.0057
(0,0,0,5,5,5)	2.5	4	0.0243	0.0073	0.2809	0.2911	0.3832	0.5254
		FWER	0.0000	0.0000	0.0005	0.0005	0.0003	0.0002
	2.0	4	0.0893	0.0239	0.4506	0.4402	0.5436	0.7020
		FWER	0.0001	0.0004	0.0016	0.0016	0.0012	0.0007
	1.5	4	0.2309	0.0605	0.6254	0.5977	0.6916	0.8373
		FWER	0.0005	0.0025	0.0039	0.0039	0.0035	0.0026
	1.0	4	0.4350	0.1323	0.7744	0.7303	0.8058	0.9169
		FWER	0.0033	0.0063	0.0099	0.0099	0.0090	0.0074
(0,0,5,5,5,5)	2.5	3	0.0000	0.0471	0.3704	0.2366	0.3693	0.4933
		FWER	0.0000	0.0004	0.0002	0.0004	0.0004	0.0004
	2.0	3	0.0016	0.1045	0.5621	0.3830	0.5351	0.6694
		FWER	0.0000	0.0008	0.0008	0.0008	0.0008	0.0008
	1.5	3	0.0111	0.1920	0.7302	0.5453	0.6882	0.8162
		FWER	0.0000	0.0028	0.0027	0.0028	0.0028	0.0028
	1.0	3	0.0508	0.3263	0.8477	0.6883	0.8097	0.9040
		FWER	0.0005	0.0076	0.0075	0.0076	0.0076	0.0076
(0,5,5,5,5,5)	2.5	2	0.0000	0.4806	0.0340	0.2018	0.3743	0.3646
	2.0	2	0.0000	0.6346	0.1053	0.3396	0.5378	0.5314
	1.5	2	0.0000	0.7655	0.2437	0.5046	0.6956	0.6915
	1.0	2	0.0001	0.8672	0.4472	0.6596	0.8207	0.8187

of the monotone assumption. For illustration purposes, we considered values of the clinically significant difference, $\delta = 1.0$, 1.5, 2.0, *and* 2.5, respectively. For each configuration, 10,000 replications using R2.2.1 were made. Table 3 gives estimates of the familywise error rate (FWER) and the probability of identifying the true MED for each procedure in the monotone case. Here we adopted the FWER definition of Tamhane et al. (1996) and Dunnett and Tamhane (1998) as the proportion of replications corresponding to identifying noneffective doses as the MED in the monotone case (the same as Hsu and Berger, 1999). As DR researchers are also interested in obtaining a therapeutic window of effective doses, the estimated probabilities of correctly identifying at least one effective dose are given in Table 4 for the monotone case. Tables 5 and 6 are the corresponding results for the nonmonotone case.

First let us look at the monotone case. Configurations with true MED = 2 involve no FWER and the corresponding entry is omitted in Table 3. It is seen from Table 3 that all the procedures control the FWER quite well under the monotonicity assumption. The HC and RH contrasts perform very poorly in most cases and they are not recommended in a general MED identification. The MC procedure is best in 16 of 24 cases and worst in only 3 of 24 cases among the remaining four methods. Furthermore, the maximum gain in probability over the DR, Williams, and LC procedures are 0.2864 ($\delta =$ 2.0 at (0, 0, 5, 5, 5, 5)), 0.1584 ($\delta = 2.0$) at (0, 0, 0, 5, 5, 5)), and 0.5145 ($\delta =$ 2.0 at (0, 0, 0, 0, 0, 5)), respectively. The

Table 4
Estimated probability of identifying at least one effective dose under monotone configurations

Case	δ	MED	HC	RH	LC	DR	W	MC
(0,1,2,3,4,5)	2.5	4	0.1177	0.1084	0.4284	0.5361	0.5176	0.4102
	2.0	4	0.2312	0.2098	0.6248	0.6560	0.6457	0.5874
	1.5	3	0.3878	0.3710	0.8179	0.7859	0.7868	0.7651
	1.0	3	0.5691	0.5288	0.9075	0.8586	0.8620	0.8692
(0,0,0,0,0,5)	2.5	6	0.7351	0.0009	0.0880	0.5433	0.4983	0.5540
	2.0	6	0.8632	0.0044	0.2059	0.6794	0.6375	0.7204
	1.5	6	0.9351	0.0145	0.3849	0.7894	0.7571	0.8507
	1.0	6	0.9699	0.0407	0.5967	0.8701	0.8475	0.9271
(0,0,0,0,5,5)	2.5	5	0.3879	0.0155	0.6703	0.5433	0.5845	0.7935
	2.0	5	0.5961	0.0455	0.8355	0.6797	0.7271	0.9162
	1.5	5	0.7341	0.1127	0.9317	0.7895	0.8379	0.9734
	1.0	5	0.8580	0.2215	0.9776	0.8715	0.9146	0.9888
(0,0,0,5,5,5)	2.5	4	0.1177	0.1140	0.8512	0.5429	0.6195	0.8439
	2.0	4	0.2311	0.2256	0.9395	0.6789	0.7618	0.9416
	1.5	4	0.3874	0.3824	0.9802	0.7885	0.8648	0.9844
	1.0	4	0.5659	0.5606	0.9871	0.8695	0.9302	0.9902
(0,0,5,5,5,5)	2.5	3	0.0167	0.3845	0.6701	0.5430	0.6416	0.7917
	2.0	3	0.0510	0.5661	0.8348	0.6795	0.7804	0.9117
	1.5	3	0.1177	0.7291	0.9300	0.7896	0.8795	0.9677
	1.0	3	0.2307	0.8517	0.9746	0.8718	0.9408	0.9867
(0,5,5,5,5,5)	2.5	2	0.0011	0.7319	0.0880	0.5434	0.6539	0.5585
	2.0	2	0.0049	0.8593	0.2061	0.6803	0.7904	0.7169
	1.5	2	0.0167	0.9411	0.3860	0.7924	0.8912	0.8471
	1.0	2	0.0510	0.9802	0.6016	0.8794	0.9542	0.9320

maximum loss in probability over these procedures (all at (0, 1, 2, 3, 4, 5)) is 0.0217 ($\delta = 2.0$), 0.0287 ($\delta = 2.0$), and 0.0233 ($\delta = 1.0$), respectively. For the step response, the MC procedure is better than the DR method and the LC method with the gain at least 0.0107 and 0.0563, respectively. Although the MC procedure does not demonstrate a uniform gain in probability of detecting the true MED, the improvements are substantial over the losses in probability.

From Table 4, the MC procedure is best in 15 of 24 cases and worst in 3 of 24 cases. The maximum gain in probability over the DR, Williams, and LC procedures is 0.301 ($\delta = 2.5$ at (0, 0, 0, 5, 5, 5)), 0.2244 ($\delta = 2.5$ at (0, 0, 0, 5, 5, 5)), and 0.5474 ($\delta = 2.5$ at (0, 5, 5, 5, 5, 5)), respectively, while the maximum loss in probability to the procedures (all at (0, 1, 2, 3, 4, 5)) is 0.1259 ($\delta = 2.5$), 0.1074 ($\delta = 2.5$), and 0.0528 ($\delta = 1.5$), respectively. The MC procedure is better than the DR method with the gain at least 0.0107 for the step response.

It is of interest to see the performance of the MC procedure when the monotone assumption is violated unexpectedly. To study this, we turn to Tables 5 and 6. For the nonmonotone case, the error rate (denoted as ERROR in Table 5) is the same as Hsu and Berger (1999) but is different from Tamhane et al. (1996) and Dunnett and Tamhane (1998). To illustrate, for $\boldsymbol{\mu} = (0, 1, 2, 3, 7, 1)$ and $\delta = 1.5$, a method commits an error if it infers any dose i ($2 \leq i \leq 6$) to be the MED but Tamhane et al. (1996) and Dunnett and Tamhane (1998) define the FWER as if a method infers doses 2 and 6 to be effective (Dunnett and Tamhane (1998) state that for the selected nonmonotone configurations FWER is controlled by those procedures in their papers but Hsu and Berger [1999] point out that the DR method is the only method that always controls the error rate for each of the response shapes studied). If a nonmonotone configuration has the MED, then ERROR and FWER are the same and is denoted as FWER in Table 5. From Table 5, we see that for the selected nonmonotone configurations which have the MED for a given δ FWER is controlled by the methods. However, if the nonmontone configuration is mildly violated, the MC procedure may or may not control the error rate (not a surprise by Theorem 1.1). For example, for configuration (0, 1, 2, 3, 4, 2) with $\delta = 2.5$ and $\delta = 2.0$, the error rate by the MC procedure is 0.0445 and 0.1074, respectively. From Table 5, the MC procedure is best in 10 of 17 cases and worst in 0 of 17 cases regarding the probability of identifying the true MED. If the monotonic assumption is mildly violated, the MC procedure may still be used with caution. If the monotonic assumption is severely violated such as (0, 1, 2, 3, 7, 1), the MC procedure and the Williams method have very excessive error rates, which confirms the results in Hsu and Berger (1999). The MC procedure should not be used at all in this case. In contrast, the DR method always controls the error rate. From Table 6, the MC procedure is best in 11 of 17 cases and worst in 0 of 17 cases. Once a configuration has the MED, the MC procedure is generally preferable to the DR method and the Williams method even if the monotone assumption is mildly violated.

The DR method is the only method that always controls the error rate for each of the response shapes studied. Under the monotone assumption, every method controls the error rate. Because the MC procedure is constructed to maximize the difference between the jth treatment mean and the control mean over all linear combinations, it performs very well for a wide range of types of response curves and is recommended in identifying the MED if the assumption of monotonicity is

Table 5
Estimated FWER/ERROR and probability of identifying true MED under nonmonotone configurations

Configuration	δ	MED	HC	RH	LC	DR	W	MC
(0,1,2,3,4,3)	2.5	4	0.0000	0.0090	0.0140	0.0176	0.0353	0.0203
		FWER	0.0000	0.0056	0.0029	0.0042	0.0083	0.0071
	2.0	4	0.0000	0.0292	0.0445	0.0395	0.0739	0.0551
		FWER	0.0000	0.0162	0.0143	0.0157	0.0244	0.0173
	1.5	3	0.0001	0.0313	0.0393	0.0318	0.0478	0.0382
		FWER	0.0000	0.0139	0.0045	0.0057	0.0078	0.0078
	1	3	0.0005	0.0646	0.0916	0.0666	0.0937	0.0787
		FWER	0.0000	0.0380	0.0158	0.0185	0.0265	0.0268
(0,0,0,0,5,4)	2.5	5	0.0879	0.0021	0.1697	0.2153	0.3183	0.4641
		FWER	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
	2.0	5	0.1993	0.0089	0.3245	0.3435	0.4755	0.6657
		FWER	0.0001	0.0002	0.0001	0.0006	0.0002	0.0000
	1.5	5	0.3619	0.0263	0.5189	0.4890	0.6278	0.8249
		FWER	0.0005	0.0013	0.0010	0.0029	0.0024	0.0010
	1.0	5	0.5528	0.0705	0.7034	0.6303	0.7564	0.9161
		FWER	0.0030	0.0045	0.0045	0.0079	0.0069	0.0057
(0,0,4,4,3,3)	2.5	3	0.0000	0.0155	0.0262	0.0169	0.0704	0.0978
		FWER	0.0000	0.0004	0.0000	0.0001	0.0004	0.0004
	2.0	3	0.0000	0.0440	0.0888	0.0465	0.1544	0.2267
		FWER	0.0000	0.0008	0.0001	0.0004	0.0006	0.0006
	1.5	3	0.0000	0.0995	0.2071	0.1000	0.2769	0.4089
		FWER	0.0000	0.0028	0.0010	0.0016	0.0026	0.0026
	1.0	3	0.0002	0.1839	0.3924	0.1970	0.4375	0.6042
		FWER	0.0000	0.0075	0.0050	0.0059	0.0074	0.0074
(0,1,2,3,4,2)	2.5	ERROR	0.0001	0.0380	0.0252	0.0203	0.0624	0.0445
	2.0	ERROR	0.0003	0.0967	0.0781	0.0464	0.1328	0.1074
	1.5	3	0.0000	0.0312	0.0245	0.0169	0.0442	0.0352
		FWER	0.0000	0.0137	0.0031	0.0042	0.0076	0.0077
	1.0	3	0.0000	0.0642	0.0659	0.0380	0.0857	0.0724
		FWER	0.0000	0.0380	0.0125	0.0142	0.0260	0.0260
(0,1,2,3,3,1.5)	2.5	ERROR	0.0000	0.0199	0.0047	0.0086	0.0209	0.0119
	2.0	ERROR	0.0002	0.0570	0.0192	0.0203	0.0544	0.0394
	1.5	ERROR	0.0005	0.1328	0.0618	0.0464	0.1192	0.1026
	1.0	3	0.0000	0.0616	0.0355	0.0211	0.0661	0.0588
		FWER	0.0000	0.0372	0.0067	0.0101	0.0240	0.0233
(0,0,0,0,5,2)	2.5	ERROR	0.0011	0.0035	0.0880	0.0203	0.1093	0.1755
	2.0	ERROR	0.0049	0.0123	0.2061	0.0464	0.2050	0.3421
	1.5	5	0.0158	0.0237	0.3303	0.0909	0.3390	0.5605
		FWER	0.0001	0.0013	0.0010	0.0014	0.0023	0.0010
	1.0	5	0.0499	0.0657	0.5389	0.1601	0.4944	0.7541
		FWER	0.0003	0.0045	0.0043	0.0046	0.0069	0.0056
(0,1,2,3,7,1)	2.5	ERROR	0.0000	0.0809	0.0708	0.0034	0.2117	0.1771
	2.0	ERROR	0.0000	0.1756	0.1722	0.0086	0.3530	0.3272
	1.5	ERROR	0.0000	0.3170	0.3401	0.0203	0.5171	0.5295
	1.0	ERROR	0.0001	0.4961	0.5563	0.0464	0.6776	0.7295

satisfied. If the monotone assumption is not satisfied, one should use the DR method if the error rate control is of primary concern.

6. Discussion

In DR studies, it is desirable for a method to not declare a lower dose to be efficacious if it does not declare a higher dose to be efficacious (Hsu and Berger, 1999). This can be achieved by examining $\mu_i - \mu_1 > \delta$ in a step-down fashion from $i = k$, continuing only while the answer is in the affirmative. The proposed method, like the DR method in Hsu and Berger (1999), maintains the error rate at its nominal level. Furthermore, our method accounts for monotone DR shapes and incorporates a nonzero clinically significant difference ($\delta > 0$). Because the square of the MC test T_k is asymptotically equivalent to the LRT and T_k can be considered the maximum of an infinite number of contrast statistics (Robertson et al., 1988), its power performance is superior to many other procedures for a wide range of monotone DR profiles. Utilizing the Kuhn–Tucker equivalence theorem in Theorem 2.2 is the key to the optimization problem and the proposed algorithm, which improves the method of Marcus and Peritz (1976) significantly. As Phillips and Haudiquet (2003) pointed out that the magnitude of the treatment effect at each dose is

Table 6
Estimated probability of identifying at least one effective dose under nonmonotone configurations

Configuration	δ	MED	HC	RH	LC	DR	W	MC
(0,1,2,3,4,3)	2.5	4	0.0011	0.0514	0.0851	0.0899	0.1285	0.0881
	2.0	4	0.0049	0.1166	0.1918	0.1517	0.2199	0.1845
	1.5	3	0.0167	0.2389	0.3815	0.2674	0.3863	0.3587
	1.0	3	0.0510	0.3840	0.5858	0.3845	0.5334	0.5456
(0,1,2,3,4,2)	1.5	3	0.0011	0.1873	0.1851	0.0899	0.2368	0.2164
	1.0	3	0.0049	0.3131	0.3497	0.1532	0.3696	0.3810
(0,1,2,3,3,1.5)	1.0	3	0.0020	0.2156	0.1516	0.0840	0.2016	0.1983
(0,0,4,4,3,3)	2.5	3	0.0006	0.0805	0.0708	0.0940	0.1290	0.1636
	2.0	3	0.0029	0.1748	0.1721	0.1670	0.2393	0.3332
	1.5	3	0.0109	0.3142	0.3391	0.2715	0.3878	0.5534
	1.0	3	0.0342	0.4886	0.5513	0.3971	0.5608	0.7524
(0,0,0,0,5,4)	2.5	5	0.1177	0.0099	0.4344	0.2730	0.3724	0.5798
	2.0	5	0.2311	0.0301	0.6482	0.4024	0.5333	0.7752
	1.5	5	0.3874	0.0796	0.8222	0.5405	0.6813	0.9058
	1.0	5	0.5662	0.1711	0.9212	0.6724	0.8032	0.9648
(0,0,0,0,5,2)	1.5	5	0.0166	0.0367	0.3850	0.0927	0.3403	0.5679
	1.0	5	0.0507	0.0922	0.5973	0.1628	0.4868	0.7632

important to regulators. The clinically significant difference δ should be considered. Finally, `R` code is developed and its application should not pose any problems. The procedure described in the article should be helpful for regulatory decision making and is recommended.

7. Supplementary Materials

Proofs of the theorems outlined in Section 2 are available under the Paper Information link of the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

We thank the associate editor, a referee, and the coeditor for very helpful suggestions that considerably improved the presentation of this article. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and Acadia University Research Fund for JP. Peng thanks Dr Tetsuhisa Miwa for sharing his computing level probabilities `C` code. KAD received support from NSERC and a graduate fellowship from Memorial University of Newfoundland. The coauthors dedicate this article to JP and KAD's mentor Dr C. I. C. Lee, who passed away in a car accident on August 30, 2006. JP and KAD thank him for the support he gave us.

References

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

Bretz, F., Pinheiro, J. C., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61,** 738–748.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50,** 1096–1121.

Dunnett, C. W. (1989). Multivariate normal probability integrals with product correlation structure, algorithm AS251. *Applied Statistics* **38,** 564–579.

Dunnett, C. and Tamhane, A. C. (1998). Some new multiple-test procedures for dose finding. *Journal of Biopharmaceutical Statistics* **8,** 353–366.

Hellmich, M. and Lehmacher, W. (2005). Closure procedures for monotone bi-factorial dose-response designs. *Biometrics* **61,** 269–276.

Hogg, R. V. (1965). On models and hypotheses with restricted alternatives. *Journal of the American Statistical Association* **60,** 1153–1162.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods.* London: Chapman & Hall/CRC.

Hsu, J. C. and Berger, R. L. (1999). Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *Journal of the American Statistical Association* **94,** 468–482.

International Conference on Harmonization Guidances E4, E9, and E10 (2005) at www.ich.org.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *2nd Berkeley Symposium, Mathematics, Statistics, and Probability,* 481–492. Berkeley: University of California Press.

Lee, C. I. C., Peng, J., and Liu, L. (2006). Statistical inference for the difference between the best treatment mean and a control mean. *Journal of the American Statistical Association* **101,** 1050–1058.

Marcus, R. and Peritz, E. (1976). Some simultaneous confidence bounds in normal models with restricted alternatives. *Journal of the Royal Statistical Society, Series B* **38,** 157–165.

Miwa, T., Hayter, A., and Liu, W. (2000). Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models. *Computational Statistics and Data Analysis* **34,** 17–32.

Phillips, A. and Haudiquet, V. (2003). ICH E 9 guideline 'Statistical Principles for Clinical Trials': A case study. *Statistics in Medicine* **22,** 1–11.

Robertson, T., Wright, F., and Dykstra, R. (1988). *Order Restricted Statistical Inference.* New York: Wiley.

Ruberg, S. J. (1989). Contrasts for identifying the minimum effective dose. *Journal of the American Statistical Association* **84,** 816–822.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference*. Hoboken, New Jersey: Wiley.

Tamhane, A. C., Hochberg, Y., and Dunnett, C. W. (1996). Multiple test procedures for dose finding. *Biometrics* **52,** 21–37.

Tukey, J. W., Ciminera, J. L., and Heyse, J. W. (1985). Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics* **41,** 295–301.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics* **27,** 103–117.

Williams, D. A. (1972). The comparison of several dose levels with a zero dose control. *Biometrics* **28,** 519–531.

Wöhr, M., Borta, A., and Schwarting, R. (2005). Overt behavior and ultrasonic vocalization in a fear conditioning paradigm: A dose-response study in the rat. *Neurobiology of Learning and Memory* **84,** 228–240.

Wright, F. T. (1988). The one-way analysis of variance with ordered alternatives: A modification of Bartholomew's $\bar{E}^2$ test. *The Canadian Journal of Statistics* **1,** 75–85.

Received November 2006. *Revised October* 2007.
Accepted October 2007.