Published in final edited form as:

J Biopharm Stat. 2012 January ; 22(1): 93–108. doi:10.1080/10543406.2010.504905.

Simultaneous Confidence Bands for Comparisons to Placebo, with Application to Detecting the Minimum Effective Dose

Julia N. Soualakova¹, Allan R. Sampson², Gang Jia³, and Leon J. Gleser²

Julia N. Soualakova: jsoulakova2@unl.edu

¹Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall-North, Lincoln, NE 68583-0963, Phone: (402) 472-7231, Fax: (402) 472-5179

²Department of Statistics, University of Pittsburgh, 2717 Cathedral of Learning, Pittsburgh, PA 15260, Phone: (412) 624-8368, Fax: (412) 648-8814

³Biostatistics & Research Decision Sciences, Merck Research Laboratories, Merck & Co. Inc., Mailstop: UG1C-36, P.O. Box 1000, Upper Gwynedd, PA 19454-2505, Phone: (267) 305-5369, Fax: (267) 305-6395

Abstract

In this paper we consider a max-min approach to construct two-sided and one-sided simultaneous confidence bands in a dose-response study for the contrast in mean responses of each ascending dose versus placebo. The method utilizes the assumption of monotone non-decreasing dose-response curve. Also discussed is a step-down testing procedure which utilizes the lower bands to estimate the minimum effective dose (MED). The performance of our proposed step-down procedure is assessed by simulations and is shown to be superior to competitors when the MED is the lowest dose in a dose-response study.

Keywords

Closed Testing; Isotonic Response; Minimum Effective Dose

1. INTRODUCTION

A key component of modern drug development is to establish a compound's dose-response relationship. Based on this relationship; often the goal is to identify all effective doses of the drug, and, in addition, to identify the lowest dose among the effective ones. This lowest dose is known as the minimum effective dose (MED).

There are a number of known statistical procedures that are used to estimate the dose-response relationship and obtain confidence bands for it, but there is no one best procedure since the properties of each procedure depend on the shape of the true, but unknown, dose-response curve. Bretz, Hsu, Pinheiro and Liu (2008) provided a thorough overview of a number of statistical approaches commonly used in dose-response studies. In this paper, we present a new approach for constructing simultaneous confidence bands for each dose's mean effect relative to placebo assuming a monotone dose-response curve for the test compound. These bands can be used to more fully assess the behavior of the dose-response function; in particular, lower one-sided simultaneous confidence bands for efficacy relative to placebo are quite useful in biopharmaceutical development. In addition, we illustrate how these bands can be used to estimate the MED while controlling the overall error rate in the strong sense.

To handle the necessary dependent random variables, our methodology generalizes a method of Lee (1996) to construct two-sided simultaneous confidence bands and lower one-sided simultaneous confidence bands for the differences between the mean responses to the test drug dose levels and placebo (respectively, presented in Sections 3.1 and 3.2). These bands can be used to test certain hypotheses in a step-down manner, starting with the highest dose, so that the pattern of the rejected and accepted null hypotheses provides an estimate of the MED, if it exists (Section 3.3). In Section 4 we consider a simple example to illustrate the proposed approach for constructing the bands and their application to estimate the MED. Simulation results are provided in Section 5 and Section 6 concludes with a brief discussion.

2. BACKGROUND

2.1. Confidence Bands

It is well known that if a specific parametric model for a dose-response function is assumed then the shape of the response curve and confidence regions can be estimated using non-linear parametric regression. If no parametric family for the dose-response function is available, then a simple approach for constructing simultaneous confidence bands uses the studentized maximum modulus (SMM) distribution. However, in many situations a test drug's dose-response function can be assumed to be monotonic, so that the SMM technique can be improved by utilizing this assumption. Korn (1982) proposed a generalization of the SMM approach by incorporating the monotonicity assumption to obtain confidence bands for the test drug dose-response curve. However, these bands are not applicable to comparisons against placebo. Schoenfeld (1986) presented another method for constructing "isotonic confidence bounds" based on the likelihood ratio tests for a similar setting to Korn's. Following Korn (1982), we assume that responses of the test drug are monotone.

Lee (1996) considered a further modification of the SMM approach and proposed a max-min approach to construct simultaneous confidence bands for a monotonic curve, comparing his approach with Korn's (1982) and Schoenfeld's (1986) methods.

Subjects are assumed randomly allocated to one of $K + 1$ dose groups, where n_i , $i = 0, 1, \dots, K$, denotes the number of subjects allocated to the i -th group. In each group, subjects receive a particular dose x_i of the test drug or placebo, where $x_0 = 0$ denotes placebo; and $x_1 < x_2 < \dots < x_K$. Let μ_i denote the population mean dose-response for the i -th group, $i = 0, 1, \dots, K$, so that the monotonicity of the test drug dose-response can be written as

$$\mu_1 \leq \mu_2 \leq \dots \leq \mu_K. \quad (1)$$

We assume that $y_{ij} \sim (\mu_i, \sigma^2)$, where y_{ij} is the response of the j -th subject in the i -th group for $i = 0, 1, \dots, K$ and $j = 1, 2, \dots, n_i$, and the y_{ij} 's are mutually independent. Let $\bar{y}_i \sim N(\mu_i, \sigma^2/n_i)$ denote the sample mean at each dose level i , $i = 0, 1, \dots, K$. Let $\hat{\sigma}^2$ be an estimate of σ^2 ,

such that $\hat{\sigma}^2$ is independent of $\bar{y}_i$, $i = 0, 1, \dots, K$, and $\hat{\sigma}^2 \sim \sigma^2 \chi_{\xi}^2 / \xi$, with $\xi = \sum_{i=0}^K n_i - (K+1)$ degrees of freedom.

Under the above conditions Lee (1996) constructed max-min simultaneous confidence intervals for $\mu_1 \leq \mu_2 \leq \dots \leq \mu_K$, which are based on the generalized SMM, defined as

$$M = \max_{1 \leq i \leq j \leq K} (\sigma | \sum_{h=i}^j z_h |) / (\hat{\sigma} \sqrt{j-i+1}), \quad (2)$$

where $z_0, z_1, \dots, z_K$ are i.i.d. $N(0,1)$. In order to interpolate these confidence intervals to construct confidence bands, Lee (1996) defined certain left- and right-continuous step functions for the low and upper bands, respectively. These are further discussed in Section 3.1.

Jia (2004) introduced a method for obtaining simultaneous confidence intervals for the monotone mean differences

$$v_i \equiv \mu_i - \mu_0, i=1, 2, \dots, K. \quad (3)$$

To construct such intervals, Jia (2004) extended Korn's (1982) simultaneous confidence bands procedure to the case of correlated variables. In addition to estimating the behavior of a dose-response curve, the simultaneous confidence intervals can be used to identify the MED. Furthermore, Jia (2004) proposed a closed step-down testing procedure for identifying the lowest effective dose for an adaptive sampling design and illustrated its application to an adaptive design with a binary endpoint.

Extensions of Korn's (1982), Lee's (1996) and other methods to designs with two controls, placebo and an active drug, were addressed by Dilleen, Heimann and Hirsch (2003). The authors considered a problem of detecting a dose which provides the same response as the active control and proposed several approaches, including resampling methods, to estimate the monotone dose-response curve.

A more general problem of detecting a dose such that the mean response at this dose is significantly higher than the control mean response was discussed by Hsu and Berger (1999). These authors proposed a step-wise approach to detect the MED, which does not rely on monotonicity assumption.

2.2. Minimum Effective Dose

The MED is defined as $MED = \min_{1 \leq i \leq K} \{i: v_i > 0\}$, where the v_i is defined by (3). One popular approach to the problem of estimating the MED assuming (1) is to formulate it as a sequence of hypothesis testing problems as follows. Let $\mathfrak{J}$ be a family of the following hypotheses

$$\mathfrak{J} = \{H_i^0; i=1, 2, \dots, K\}, \text{ where } H_i^0: v_1 \leq v_2 \leq \dots \leq v_i \leq 0, \quad (4)$$

and the corresponding alternative hypotheses be $H_i^a: v_i > 0$. In this approach, if i^* is the smallest i such that H_i^0 is rejected, we estimate the MED as the $i^* - th$ dose.

A similar formulation of the problem was provided by Tamhane, Hochberg and Dunnett (1996), who posed the problem of detecting the MED as a sequence of hypothesis testing problems:

$$\begin{aligned} H_{oi}: \mu_0 = \mu_1 = \dots = \mu_i \text{ versus} \\ H_{ai}: \mu_0 = \mu_1 = \dots = \mu_{i-1} < \mu_i, i=1, 2, \dots, K. \end{aligned} \quad (5)$$

Tamhane *et al.* (1996) outlined a number of step-wise closed testing procedures based on contrasts among the sample means. None of these latter procedures, except for one based on Williams (1971, 1972) tests, requires the monotonicity assumption. Bauer (1997) pointed

out that Tamhane *et al.*'s (1996) procedures strongly control the overall error rate only if the order restriction (1) on the means is imposed.

The essence of each of the noted procedures discussed in Tamhane *et al.* (1996) is to propose a test statistic for each hypothesis H_{oi} in (5) and use an appropriate step-wise approach to sequentially test these hypotheses. One discussed procedure, Williams (1971, 1972) is a step-down procedure which starts with the largest dose and is based upon the order restricted (maximum likelihood) estimators of the μ_i 's. Another procedure is the step-down procedure based on pairwise contrasts (SDPC) introduced by Tamhane *et al.* (1996) which uses tests of pairwise dose comparisons with placebo and steps down from the dose with the largest test statistic.

Bretz, Hothorn and Hsu (2003) proposed a step-down testing approach (termed here as BHH) to estimate the MED when the hypotheses are stated in terms of ratios of the true means. This approach allows testing for any specific value of the ratio (including 1), and does not require monotonicity of mean response. Nonetheless, BHH is suitable only for studies, where the control mean is significantly higher than zero. In the following we briefly review BHH for the case of equal group sample sizes, $n_i = n$ for $i = 0, 1, \dots, K$.

Consider a set of hypotheses, where the null hypothesis $H_{0,i} : \mu_i / \mu_0 \leq 1$ is tested against the corresponding alternative $H_{a,i} : \mu_i / \mu_0 > 1$ for $i = 1, 2, \dots, K$. Then, starting with $k = K$ the individual $100(1-\alpha)\%$ lower confidence intervals for μ_i / μ_0 are obtained in a step-down manner with the lower bounds given by

$$L_i^{BHH} = \{\bar{y}_i \bar{y}_0 - (a \bar{y}_i^2 + a \bar{y}_0^2 - a^2)^{1/2} / (\bar{y}_0^2 - a), i=1, 2, \dots, K, \quad (6)$$

where $a = \hat{\sigma}^2 t_{1-\alpha, \xi}^2 / n$ and $t_{1-\alpha, \xi}$ denotes the α -th quantile of a t -distribution with ξ degrees of freedom. The null hypotheses are rejected in a step-down manner as long as the lower bounds exceed 1. The smallest index of rejected hypotheses corresponds to the estimated MED. Bretz, Hothorn and Hsu (2003) noted that due to the restrictions in Fieller's (1954) approach, e.g., Miller (1986), the method can be applied only in cases where $\bar{y}_0^2 > a$, which "can be regarded as a test on μ_0 being significantly larger than 0".

3. MAX-MIN SIMULTANEOUS CONFIDENCE BANDS FOR MEAN RESPONSE DIFFERENCES

3.1. Two-Sided Simultaneous Confidence Bands for Mean Response Differences

We extend Lee's (1996) method to construct two-sided confidence bands for the mean response contrast to placebo. First, let $w_h = (\bar{y}_h - \bar{y}_0 - v_h) / \sqrt{\sigma^2(1+n_h+1/n_0)}$, $h=1, 2, \dots, K$. Then $(w_1, \dots, w_K)' \sim MVN$ with mean $\mathbf{0}$ and $Cov(w_i, w_j) = 1$ for $i=j$, and $Cov(w_i, w_j) = \sqrt{n_i n_j / \{(n_0 + n_i)(n_0 + n_j)\}}$ for $i \neq j$.

Next, we construct the test statistics based on these random variables and the corresponding confidence intervals. Consider the test statistic

$$T = \max_{1 \leq i \leq j \leq K} (\sigma | \sum_{h=i}^j w_h |) / (\hat{\sigma} \sqrt{j-i+1}). \quad (7)$$

The distribution of this statistic differs from the distribution of Lee's (1996) generalized SMM in (2) because the w_h 's are not independent random variables. The density function of T is right skewed as we illustrate with $K = 10$ in Figure 1.

The critical values of the distribution of T , $c(K, \xi, \alpha)$, are computed using Lee's (1996) approach via simulations. This approach approximates the percentiles $c = c(K, \xi, \alpha)$ by solving the equation $P(T \leq |c|) = 1 - \alpha$ for c for specified values of K , ξ and α . First, under above assumptions random data are generated, and the test statistic T is calculated; the steps are replicated a large number of times to estimate the corresponding percentile. The critical values given in Table 1 are calculated for $K \leq 10$, $\xi = \infty$ and several values of α (100,000 replicates are used).

For simplicity of notation, let $c = c(K, \xi, \alpha)$,

$\lambda_h = 1 / \sqrt{1/n_h + 1/n_0}$, and $S_{pq} = \{ \sum_{h=p}^q \lambda_h (\bar{y}_h - \bar{y}_0) \} / \sum_{h=p}^q \lambda_h$. Then the simultaneous 100 (1- α) % two-sided confidence intervals for $v_l = \mu_l - \mu_0$, $l = 1, 2, \dots, K$, are given by

$$\max_{1 \leq p \leq q \leq l} (S_{pq} - c\hat{\sigma} \sqrt{q-p+1}) / \sum_{h=p}^q \lambda_h \leq v_l \leq \min_{l \leq p \leq q \leq K} (S_{pq} + c\hat{\sigma} \sqrt{q-p+1}) / \sum_{h=p}^q \lambda_h. \quad (8)$$

In order to construct simultaneous confidence bands, one can use left- and right-continuous step functions defined by the K confidence intervals given in (8) to obtain the low and upper bounds, respectively, for additional dose levels that are between the considered dose levels $l = 1, 2, \dots, K$. The Appendix details the derivation of 100 (1- α) % simultaneous confidence intervals (8) and the details for these bands.

3.2. One-Sided Simultaneous Confidence Bands for Mean Response Differences

To construct one-sided confidence bands, we consider the test statistic

$$Q = \max_{1 \leq i \leq j \leq k} \sigma \sum_{h=i}^j w_h / (\hat{\sigma} \sqrt{j-i+1}), \quad (9)$$

where k is any fixed index such that $1 \leq k \leq K$. The distribution of this test statistic depends on k . Note that for $k = 1$, Q has the t -distribution with $\xi = n_0 + n_1 - 2$ degrees of freedom. While being less skewed than T distribution, the distribution of Q is somewhat right skewed (see Figure 2).

The critical values of Q , denoted by $m(k, \xi, \alpha)$, are obtained, again, by simulation. Table 2 contains the critical values of Q for up to ten active dose levels for infinite degrees of freedom. Although, for relatively small α , $c(k, \xi, \alpha / 2) \approx m(k, \xi, \alpha)$, due to asymmetry of the distribution of Q the latter relationship does not hold in general (See Figure 2 for an illustration of the lack of symmetry).

Simultaneous 100 (1- α) % one-sided max-min confidence intervals for mean response differences, v_l , $l = 1, 2, \dots, k$, based on the tests statistic Q are given by

$$L_l = \max_{1 \leq p \leq q \leq l} (S_{pq} - m\hat{\sigma} \sqrt{q-p+1}) / \sum_{h=p}^q \lambda_h \leq v_l, l = 1, 2, \dots, k, \quad (10)$$

where $m = m(k, \xi, \alpha)$.

The proof that (10) are the $100(1-\alpha)\%$ confidence level intervals is similar to the one given in the Appendix. These lower bounds can readily be used to provide the simultaneous lower confidence statements for the mean response differences at all dose levels. For any given number of active drug levels, K , and confidence level, $100(1-\alpha)\%$, the positive values of the lower bounds indicate effective doses of the drug and the smallest such dose can be used as one possible MED estimate. In the next section we discuss a more effective approach to estimate the MED. While not derived here, one could similarly obtain upper simultaneous bands that may be useful when comparing dose levels to placebo when considering adverse effects that are suitably quantified.

3.3. Procedure for Identifying the Minimum Effective Dose

In this section we describe a step-down procedure to test the hypotheses in (4) based on constructing $100(1-\alpha)\%$ simultaneous one-sided max-min confidence intervals for mean differences, v_l 's, given by (10).

The hypothesis family $\mathfrak{I}$ given by (4) is closed, *i.e.*, any hypothesis which is obtained as an intersection of any number of the hypotheses belonging to the family belongs to $\mathfrak{I}$. Hence, similarly to Tamhane *et al.* (1996) we can apply Marcus, Peritz and Gabriel's (1976) closed testing principle, and test each hypothesis at level of significance α to strongly control the family-wise error rate (FWE) at level α .

To estimate the MED, start with $l = k = K$, and test H_k^0 against H_k^a by computing L_K using the corresponding critical value $m(K, \xi, \alpha)$. Positive L_K suggests that the K -th mean difference is positive so the corresponding null hypothesis H_K^0 is rejected. Next, the highest dose is omitted from consideration and the first $k = K - 1$ doses are considered and the lower bound for v_{K-1} is calculated based on $m(K-1, \xi, \alpha)$. If $L_{K-1} > 0$ then H_{K-1}^0 is rejected and the testing is continued with $k = K - 2$ and so on; unless $L_m \leq 0$ for some $m = 1, 2, \dots, K$, then H_m^0 is accepted and testing is completed. The MED is estimated as $MED^* = m + 1$. If $m = K$ then $MED^* = K + 1$, meaning that the procedure estimates all doses as non-effective. If all hypotheses are rejected, then all of the test drug doses are considered superior to placebo, *i.e.*, $MED^* = 1$.

The above procedure which we term the GLEE approach utilizes the same logic as the SDPC procedure of Tamhane *et al.* (1996) does. Moreover, it is interesting to note that if one were to apply Jia's (2004) modification of Korn (1982) to construct one-sided lower simultaneous confidence intervals for $v_1, v_2, \dots, v_K$ based on the pairwise test statistics t_i 's and then utilized the approach described by us in the preceding to estimate the MED, the resulting procedure can be shown to be equivalent to the SDPC.

A simpler single-step alternative approach, described in detail by Tamhane *et al.* (1996) and called SD2, is to construct the simultaneous lower bounds for L_l , $l = 1, 2, \dots, K$, with $k = K$ based on common critical value $m(K, \xi, \alpha)$. Then this procedure is applied to estimate the MED. This alternative testing approach results in a less powerful procedure than the method described above. Since the critical value $m(K, \xi, \alpha) > m(l, \xi, \alpha)$ for $1 \leq l < K$, our proposed procedure is more powerful than this alternative. Note that these two procedures perform similarly only in the case when the lower bound for the highest observed mean dose-response difference is non-positive, therefore, there are no effective doses detected by the procedures.

4. EXAMPLE

To illustrate the proposed two-sided and one-sided simultaneous confidence bands for a monotonic dose-response curve and the step-down approach to identify the MED, we consider hypothetical data analyzed by Tamhane *et al.* (1996). These hypothetical data correspond to a single drug trial with five test drug dose levels and a placebo, where the same number of subjects, n , is allocated to each of the six groups. The sample group means are given by $\bar{y}_0 = 0.0$, $\bar{y}_1 = 1.5$, $\bar{y}_2 = 2.1$, $\bar{y}_3 = 1.9$, $\bar{y}_4 = 2.3$ and $\bar{y}_5 = 2.1$ and variance is assumed to be known and to satisfy $2\sigma^2/n = 1$.

First, we construct the two-sided and one-sided simultaneous 95% confidence bands for the mean response contrasts versus placebo. To construct the two-sided and one-sided intervals we use critical values given in Tables 1 and 2, for a case of 5 active dose levels, *i.e.*, 3.47 and 2.96, respectively. Figure 3 illustrates our max-min simultaneous bands. The upper bounds of our two-sided bands are given by 3.53, 3.84, 4.10, 4.65 and 5.57 for dose levels 1 through 5, respectively, and the depicted bands correspond to a left-continuous step function. The lower bounds of our two-sided bands are given by -1.97, -0.65, -0.17, 0.22 and 0.43 for dose levels 1 through 5, respectively. And the respective lower one-sided bounds are -1.46, -0.29, 0.12, 0.47 and 0.66. Both the depicted two- and one-sided lower bounds correspond to the right-continuous step functions. Thus, if estimation of a mean response difference from below is of main concern, the one-sided simultaneous lower bounds provide a considerable improvement over the two-sided bounds.

Next we illustrate our approach to identify the MED. Based on the critical values $m(5, \infty, 0.05) = 2.96$, $m(4, \infty, 0.05) = 2.73$, $m(3, \infty, 0.05) = 2.45$, $m(2, \infty, 0.05) = 2.12$ and $m(1, \infty, 0.05) = 1.64$ we calculate the corresponding lower bounds as $L_5 = 0.66$, $L_4 = 0.59$, $L_3 = 0.42$, $L_2 = 0.30$ and $L_1 = -0.14$. Thus, we reject all null hypotheses except for H_1^0 and conclude that $MED^* = 2$. When Williams' and SDPC procedures are applied at $\alpha = 0.05$, they both result in $MED^* = 2$.

Note that if we apply the alternative approach SD2 then this procedure is equivalent to applying the proposed simultaneous one-sided confidence bounds noted in Section 3.2. Hence, since the resulting lower bounds for dose levels 5, 4, 3 and 2 are given by 0.66, 0.47, 0.12 and -0.29, we reject only H_5^0 , H_4^0 and H_3^0 , and thus, conclude that $MED^* = 3$.

In order to illustrate the BHH method, we shift the observed means by 1.5 units to satisfy $\bar{y}_0^2 > a \approx 1.34$. Therefore, the modified observed means are given by $\bar{y}_i = \bar{y}_i + 1.5$ for $i = 0, 1, \dots, 5$. Note that while the above methods are all invariant with respect to such a shift, the results of BHH highly depend on it. Then based on the confidence bounds

$$L_5^{BHH} = L_2^{BHH} = 1.19, \\ L_4^{BHH} = 1.28 \text{ and } L_3^{BHH} = 1.11 \text{ and } L_1^{BHH} = 0.94, \text{ we would conclude that } MED^* = 2.$$

5. SIMULATION STUDIES

Lee (1996) notes that his procedure is most efficient when a dose-response curve does not increase rapidly and it is less efficient than Korn's (1982) method when a curve is rapidly increasing. Particularly, Lee states that the lower bound provides an improvement when $\mu_1 - \mu_1$ is small. The case where Lee's bounds are most beneficial is when $\mu_1 = \mu_2 = \dots = \mu_K$. Thus, we expect that the proposed procedure for identifying the MED would be most advantageous when $v_1 = v_2 = \dots = v_K$. The latter condition implies that either there is no MED, $v_1 = v_2 = \dots = v_K \leq 0$, or the MED is given by the lowest dose level, $v_1 = v_2 = \dots = v_K >$

0, i.e., $MED = 1$. In both cases, the dose-response patterns correspond to a monotone step-function.

To compare the performance of our proposed GLEE approach to find the MED with Williams' (WILM), SDPC and BHH procedures, we perform a simulation study, similarly to Tamhane *et al.* (1996). Comparisons are made in terms of power, overall error rate, lack of power, and bias, defined below. In addition, as is requested by a reviewer, we considered several non-monotone mean response functions to examine robustness concerns. The critical values for WILM and SDPC procedures are given by Tamhane *et al.* (1996) and Williams (1971).

To mimic one of commonly used designs we consider a study with $K = 3$ test drug dose levels and placebo, Bretz, Hothorn and Hsu (2003). We use $\alpha = 0.05$, and fix $\mu_0 = 3$. For each case of given mean values described below, we generate the sample mean responses $\bar{y}_i \sim N(\mu_i, 0.5)$, $i = 0, 1, 2, 3$, and apply GLEE, WILM, BHH and SDPC procedures. Based on $N = 10,000$ replications for each case (hence, the standard error of estimation is $\sqrt{0.5 \cdot 0.5 / 10000} = 0.005$), we obtain the proportion when each dose is estimated as the MED.

We consider two types of monotone dose-response curves: linear and step. Let δ denote the largest value of the response function, this value corresponds to the 3rd dose level. We use $\delta = 6$ and $\delta = 8$. For each combination of the type of response and its maximum value, we vary the true MED over the three dose levels. For example, given linear response and $\delta = 6$, for the case of $MED = 2$, the means are given by $\mu_0 = \mu_1 = 3$, $\mu_2 = 4.5$, and $\mu_3 = 6$. Note that for $MED = 3$ the linear and step functions provide the same values for the means. In addition, we consider a case where there is no MED, i.e., all means are equal. We also considered several non-monotone configurations of the means, which are presented in Table 4.

Since we know a priori the true MED and what doses correspond to the effective and non-effective doses, we can simulate the values of family-wise error rate (FWE), Power, Lack of Power (LOP) and Bias in each case. We use formulas for calculation of these values given by Tamhane *et al.* (1996): $FWE = N_1 / N$, $Power = N_2 / N$, $LOP = N_3 / N$, and $Bias = E(MED^*) - MED$, where N_1 is the number of replications where non-effective doses estimated as the MED's; N_2 is the number of cases where MED's are correctly estimated as the MED's; N_3 is the number of cases where doses higher than MED's are estimated as the MED's (including cases where none of the doses is estimated as the MED); $E(MED^*)$ is the expected value of the estimated MED, based on the observed probabilities, if none of three doses is estimated as the MED then $MED^* = 4$ is used to calculate the expectation.

Note that for monotone dose-response curves FWE, power and LOP add up to 1; in addition, in the case when the true $MED = 1$, there are no non-effective doses, so there can be no FWE. In non-monotone settings there could be truly non-efficacious doses (as in a case with $\mu_0 = \mu_3 = 3$, $\mu_1 = \mu_2 = 5.5$, depicted in Table 4) thus, FWE is not necessarily zero. Moreover, proportions of cases where a non-efficacious dose higher than the true MED is estimated as the MED contribute simultaneously to LOP and FWE.

First, when there is no MED, the observed probability of correct decisions is about 95% for any method. Thus, we do not present detailed results for such a case as all procedures perform very well.

Table 3 summarizes the simulation results for other settings for monotone mean responses. As is noted above, values of FWE and power can be used to estimate LOP and therefore, LOP is not presented. For example, in a case of $\delta = 8$, $MED = 1$ and linear response the estimated lack of power is given by $1 - 0.000 - 0.507 = 0.493$ for BHH.

As the theory establishes, all three procedures strongly control the FWE at $\alpha = 0.05$. Indeed, in order to conclude that the FWE is inflated above 5% the estimated FWE must exceed $0.05 + 1.96 \sqrt{0.05 \cdot 0.95 / 10000} = 0.0543$, Tamhane *et al.* (1996).

As is expected, in the cases with the true $MED = 1$, our proposed procedure performs the best. The average power across such configurations equals 0.665 for GLEE, followed by 0.657 for the SDPC, by 0.654 for WILM and 0.632 for BHH. A similar pattern is observed in terms of the average bias and lack of power.

When the true MED is larger than 1, then BHH performs the best in terms of power and bias, followed by WILM. If we compare the results corresponding to the same shape of the mean dose-response function for different values of δ , then, obviously, all procedures perform better in the case of $\delta = 8$. Also, the average power of the procedures is higher for the step function. It can be explained by the flat shape of the step function with zero mean difference for the consecutive non-effective dose levels and a relatively large positive difference associated with the dose effectiveness.

Although, we considered additional types of monotone dose-response relationships, such as square-root function, the pattern of conclusions remained the same and thus, these cases are not discussed here.

Table 4 illustrates detailed results for non-monotone mean responses with $\mu_0 = 3$. The Table presents the proportions of cases where each dose is estimated as MED including instances where no dose is declared as MED for each considered simulation configuration. There are no indications that the overall error rate is inflated in non-monotone settings. A similar observation was made by Tamhane *et al.* (1996).

GLEE is still superior (in terms of power, bias and lack of power) to the other approaches when the MED is 1. In addition, BHH can be dominated by all three methods with a substantial difference in power performance. BHH does perform the best in the setting where the mean response is “close to be monotone”, i.e., when $\mu_1 = 3$, $\mu_2 = 2.8$ and $\mu_3 = 6.0$.

The performed simulation study is limited to the considered simulation configurations and equal sample size case, and the conclusions may be different for other parameter settings. Hence, we recommend detailed simulations be done for any particular study one is designing.

6. DISCUSSION

In this paper we extend an approach proposed by Lee (1996) in order to construct two-sided simultaneous confidence bands for the true mean response differences relative to placebo when the underlying test drug dose-response curve is monotone. The bands can be used to describe a general behavior of the mean response differences between test drug doses and placebo, and thus, to characterize the efficacy of a drug. We also present one-sided simultaneous confidence bands that can offer improvement over the two-sided bands when estimation of the mean response differences from below is of main interest.

In addition, we discuss an application of proposed max-min bands to estimate the MED. While the simultaneous lower confidence bands can be used directly to identify the MED, they offer a more conservative approach than the one proposed in this paper. Simulations show that our procedure performs better than its competitors when the true MED is relatively low. In other cases there is no unique best approach. As our simulations illustrate, there are cases where each of the four considered methods dominates but these cases highly

depend on the particular parameter values and therefore, no general recommendations can be made.

Note that while BHH performs the best in many settings, it utilizes an assumption of $\bar{y}_0^2 > a$ and thus, may not be applicable in some cases. Indeed, if we consider the original settings of the Example of Section 4, then the original placebo mean $\bar{y}_0 = 0.0$ does not satisfy $\bar{y}_0 > 1.163$. Therefore, this method is not applicable. Thus, care must be taken when designing a simulation study or using this method in practice to assure that the latter condition is satisfied.

Our methodological approach can also be extended for any newly developed suitable simultaneous banding procedures, e.g., Sampson, Singh and Whitaker (2009). In addition, future research can be targeted at examining performance of other suitable methods including the approach by Hsu and Berger (1999).

Acknowledgments

The authors want to thank the referees for their helpful comments. Research of Allan R. Sampson was supported in part by NIMH Grant MH08405301.

APPENDIX

We use the approach presented by Lee (1996) to construct two-sided simultaneous confidence intervals for $v_l = \mu_l - \mu_0$, $l = 1, 2, \dots, K$, given in (8) and then show that these are the 100 $(1 - \alpha)$ % confidence intervals. Then, we show that (8) can be used to construct the confidence bands as is discussed in Section 3.1.

First we consider and prove a simple lemma given by Lee (1996). Let a_h , $h = 1, 2, \dots, K$, be any non-negative constants. Then under monotonicity assumption (1)

$$\max_{1 \leq p \leq q \leq l} \sum_{h=p}^q a_h v_h / \sum_{h=p}^q a_h = v_l \text{ and } \min_{l \leq p \leq q \leq K} \sum_{h=p}^q a_h v_h / \sum_{h=p}^q a_h = v_l, \text{ where } l = 1, 2, \dots, K.$$

To prove this fact, note that $v_h \leq v_l$ for all $h \leq l \leq K$ and $\sum_{h=p}^q a_h v_h \leq \sum_{h=p}^q a_h v_l = v_l \sum_{h=p}^q a_h$ for all $1 \leq p \leq q \leq l$. Hence, $\sum_{h=p}^q a_h v_h / \sum_{h=p}^q a_h \leq v_l$ for all $1 \leq p \leq q \leq l$. The last inequality becomes an equality when $p = q = l$, which completes the proof of the first statement. The second statement can be proved similarly.

Next, consider an inequality $T \leq c$, where $c = c(K, \xi, \alpha)$. This inequality can be rewritten as

$$\max_{1 \leq i \leq j \leq K} \left| \sum_{h=i}^j \lambda_h (\bar{y}_h - \bar{y}_0 - v_h) \right| \leq c \hat{\sigma} \sqrt{j-i+1} \text{ and then as}$$

$$\sum_{h=i}^j \lambda_h (\bar{y}_h - \bar{y}_0) - c \hat{\sigma} \sqrt{j-i+1} \leq \sum_{h=i}^j v_h \lambda_h \leq \sum_{h=i}^j \lambda_h (\bar{y}_h - \bar{y}_0) + c \hat{\sigma} \sqrt{j-i+1} \text{ for all } 1 \leq i \leq j \leq K.$$

Next, we divide the above inequality by $\sum_{h=i}^j \lambda_h$ throughout. And then we take the maximum of the left hand side over all indices less or equal to l , and we take the minimum of the right hand side with respect to all indices greater or equal to l , and apply the lemma to obtain

$$\max_{1 \leq p \leq q \leq l} (S_{pq} - c\hat{\sigma} \sqrt{q-p+1} / \sum_{h=p}^q \lambda_h) \leq \max_{1 \leq p \leq q \leq l} \sum_{h=p}^q v_h \lambda_h / \sum_{h=p}^q \lambda_h = v_l \text{ and}$$

$$v_l = \min_{l \leq p \leq q \leq K} \sum_{h=p}^q v_h \lambda_h / \sum_{h=p}^q \lambda_h \leq \min_{l \leq p \leq q \leq K} (S_{pq} + c\hat{\sigma} \sqrt{q-p+1} / \sum_{h=p}^q \lambda_h).$$

Finally, we combine both parts and obtain the confidence intervals for v_l , $l = 1, 2, \dots, K$, given by (8).

Next, we show that the confidence level $C(v)$ of the max-min simultaneous confidence intervals (8) is at least $100(1 - \alpha)\%$. The proof is similar to the one given by Lee (1996). For convenience of illustration we let

$$A_l = \max_{1 \leq p \leq q \leq l} (S_{pq} - c\hat{\sigma} \sqrt{q-p+1} / \sum_{h=p}^q \lambda_h) \text{ and } B_l = \min_{l \leq p \leq q \leq K} (S_{pq} + c\hat{\sigma} \sqrt{q-p+1} / \sum_{h=p}^q \lambda_h).$$

Then $C(v) = P(A_l \leq v_l \leq B_l, l = 1, 2, \dots, K)$. From the lemma,

$$v_l = \max_{1 \leq p \leq q \leq l} \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h = \min_{l \leq p \leq q \leq K} \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h \text{ so}$$

$$C(v) = P(A_l \leq \max_{1 \leq p \leq q \leq l} \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h = \min_{l \leq p \leq q \leq K} \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h \leq B_l, l = 1, 2, \dots, K)$$

$$= P(A_l - \max_{1 \leq p \leq q \leq l} \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h \leq 0, B_l - \min_{l \leq p \leq q \leq K} \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h \geq 0, l = 1, 2, \dots, K).$$

$$A_{pq} = S_{pq} - c\hat{\sigma} \sqrt{q-p+1} / \sum_{h=p}^q \lambda_h - \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h \text{ and}$$

$$B_{pq} = S_{pq} + c\hat{\sigma} \sqrt{q-p+1} / \sum_{h=p}^q \lambda_h - \sum_{h=p}^q \lambda_h v_h / \sum_{h=p}^q \lambda_h \text{ then}$$

$$\text{Let } C(v) \geq P(\max_{1 \leq p \leq q \leq l} A_{pq} \leq 0, \min_{l \leq p \leq q \leq K} B_{pq} \geq 0, l = 1, 2, \dots, K).$$

Consider $\max_{1 \leq p \leq q \leq l} A_{pq} \leq 0$ which is identical to $\max_{1 \leq p \leq q \leq l} \sigma \sum_{h=p}^q w_h / \hat{\sigma} \sqrt{q-p+1} \leq c$. Similarly, an

inequality $\min_{l \leq p \leq q \leq K} B_{pq} \geq 0$ can be written as $\min_{l \leq p \leq q \leq K} \sigma \sum_{h=p}^q w_h / \hat{\sigma} \sqrt{q-p+1} \geq -c$. Hence,

$$C(v) \geq P(\max_{1 \leq p \leq q \leq l} \sigma \sum_{h=p}^q w_h / \hat{\sigma} \sqrt{q-p+1} \leq c, l = 1, 2, \dots, K) = P(T \leq c).$$

Since c satisfies $P(T \leq c) = 1 - \alpha$, $C(v) \geq 1 - \alpha$, which completes the proof.

If one uses left- and right-continuous step functions to obtain the low and upper bounds, respectively, at the doses other than the ones considered in the study, then the resulting simultaneous confidence bands will have the same level $C(v)$ as the confidence intervals (8). Indeed, consider any $\Delta \in (0, 1)$, i.e., $t < t + \Delta < t + 1$, $t = 1, 2, \dots, K - 1$. Let the corresponding

true mean response difference at the new dose level $t + \Delta$ be denoted by $v_{t+\Delta}$. Then the monotonicity assumption (1) implies that $v_t \leq v_{t+\Delta} \leq v_{t+1}$ for all $\Delta \in (0,1)$, $t = 1, 2, \dots, K - 1$.

Since $P\{A_l \leq v_l \leq B_l, l = 1, \dots, K\}$ is equal to $P\{A_l \leq v_l \leq B_l, v_t \leq v_{t+\Delta} \leq v_{t+1}\}$ for all $\Delta \in (0,1)$, $l = 1, \dots, K$, $t = 1, \dots, K - 1$, which does not exceed $P(A_l \leq v_{t+\Delta} \leq B_{t+1} \text{ for all } \Delta \in (0,1), t = 1, 2, \dots, K - 1)$, the resulting simultaneous confidence bands satisfy $P(A_l \leq v_{t+\Delta} \leq B_{t+1} \text{ for all } \Delta \in (0,1), t = 1, 2, \dots, K - 1) \leq 1 - \alpha$. This approach is illustrated in Section 4.

REFERENCES

1. Bauer P. A Note on Multiple Testing Procedures in Dose Finding. *Biometrics*. 1997; 53:1125–1128. [PubMed: 9333343]
2. Bretz F, Hothorn LA, Hsu JC. Identifying Effective and/or Safe Doses by Stepwise Confidence Intervals for Ratios. *Statistics in Medicine*. 2003; 22:847–858. [PubMed: 12627404]
3. Bretz F, Hsu JC, Pinheiro J, Liu Y. Dose Finding – A Challenge in Statistics. *Biometrical Journal*. 2008; 50(4):480–504. [PubMed: 18663758]
4. Dilleen M, Heimann G, Hirsch I. Non-parametric Estimators of a Monotone Dose-response Curve and Bootstrap Confidence Intervals. *Statistics in Medicine*. 2003; 22:869–882. [PubMed: 12627406]
5. Fieller EC. Some Problems in Interval Estimation. *Journal of the Royal Statistical Society, Series B*. 1954; 16:175–185.
6. Hsu JC, Berger RL. Stepwise Confidence Intervals without Multiplicity Adjustment for Dose-Response and Toxicity Studies. *Journal of American Statistical Association*. 1999; 94:468–482.
7. Jia, G. PhD Dissertation. University of Pittsburgh, Department of Statistics; 2004. Use of Simultaneous Inference Under Order Restriction, Step-Down Testing Procedure and Stage-Wise Sequential Optimal Design in Clinical Dose Study. Available at http://etd.library.pitt.edu/ETD/available/etd-10282004-142719/unrestricted/Jia_G_Dec2004.pdf
8. Korn EL. Confidence Bands for Isotonic Dose-Response Curves. *Applied Statistics*. 1982; 31:59–63.
9. Lee CC. On Estimation for Monotone Dose-Response Curves. *Journal of American Statistical Association*. 1996; 91:1110–1119.
10. Marcus R, Peritz E, Gabriel KR. On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance. *Biometrika*. 1976; 63:655–660.
11. Miller, R. *Beyond ANOVA: Basics of Applied Statistics*. New York: John Wiley and Sons, Inc; 1986.
12. Sampson AR, Singh H, Whitaker LR. Simultaneous Confidence Bands for Isotonic Functions. *Journal of Statistical Planning and Inference*. 2009; 139:828–842.
13. Schoenfeld DA. Confidence Bounds for Normal Means Under Order Restrictions, With Application to Dose-Response Curves, Toxicology Experiments, and Low-Dose Extrapolation. *Journal of the American Statistical Association*. 1986; 81(393):186–195.
14. Tamhane AC, Hochberg Y, Dunnett CW. Multiple Test Procedures for Dose Finding. *Biometrics*. 1996; 52:21–37. [PubMed: 8934584]
15. Williams DA. A Test for Differences between Treatment Means When Several Dose Levels are Compared with a Zero Dose Control. *Biometrics*. 1971; 27:103–117. [PubMed: 5547548]
16. Williams DA. The Comparison of Several Dose Levels with a Zero Dose Control. *Biometrics*. 1972; 28:519–531. [PubMed: 5037867]

Density of T for K=10

Figure 1.
Approximated density function of T statistic (7)

Density of Q for K=10

Figure 2.
Approximated density function of Q statistic (9)

Figure 3. Two-sided and one-sided simultaneous confidence bands for Example of Section 4. Note: The upper band is a left-continuous step function while the lower bands are the right-continuous step functions.

Table 1

Critical values of T statistic (7) in the case of equal sample sizes and infinite degrees of freedom obtained by simulation based on 100,000 replications

Number of Active Dose Levels (K)	Significance Level										
	0.90	0.80	0.70	0.60	0.50	0.30	0.20	0.10	0.05	0.025	0.01
1	0.13	0.25	0.38	0.52	0.67	1.03	1.28	1.64	1.96	2.24	2.58
2	0.40	0.58	0.75	0.91	1.07	1.45	1.72	2.12	2.47	2.77	3.19
3	0.60	0.80	0.98	1.13	1.31	1.72	2.01	2.45	2.85	3.24	3.67
4	0.75	0.96	1.13	1.29	1.48	1.91	2.24	2.73	3.18	3.58	4.09
5	0.87	1.08	1.26	1.43	1.62	2.08	2.42	2.97	3.47	3.94	4.49
6	0.96	1.17	1.35	1.54	1.73	2.23	2.59	3.20	3.73	4.26	4.86
7	1.04	1.25	1.44	1.62	1.82	2.35	2.76	3.39	3.99	4.56	5.18
8	1.10	1.32	1.51	1.69	1.91	2.47	2.89	3.59	4.21	4.80	5.49
9	1.16	1.38	1.58	1.77	1.99	2.58	3.04	3.78	4.45	5.05	5.77
10	1.22	1.43	1.63	1.83	2.06	2.69	3.17	3.94	4.65	5.31	6.08

Table 2

Critical values of Q statistic (9) in the case of equal sample sizes and infinite degrees of freedom obtained by simulation based on 100,000 replications

Number of Active Dose Levels (k)	Significance Level										
	0.50	0.45	0.40	0.35	0.30	0.25	0.20	0.10	0.05	0.025	0.01
1	0.00	0.13	0.26	0.37	0.52	0.67	0.83	1.28	1.64	1.96	2.33
2	0.40	0.53	0.65	0.78	0.93	1.07	1.26	1.72	2.12	2.45	2.89
3	0.62	0.74	0.86	1.01	1.15	1.32	1.49	2.01	2.45	2.89	3.34
4	0.77	0.90	1.02	1.16	1.31	1.48	1.68	2.24	2.73	3.18	3.72
5	0.88	1.00	1.14	1.28	1.44	1.62	1.83	2.42	2.96	3.46	4.06
6	0.97	1.10	1.23	1.39	1.55	1.73	1.96	2.59	3.19	3.73	4.38
7	1.03	1.17	1.30	1.46	1.64	1.82	2.06	2.76	3.39	4.02	4.71
8	1.10	1.22	1.37	1.53	1.70	1.91	2.16	2.89	3.59	4.23	4.99
9	1.15	1.29	1.44	1.59	1.77	1.99	2.24	3.04	3.79	4.43	5.23
10	1.20	1.34	1.47	1.66	1.83	2.05	2.33	3.17	3.94	4.67	5.50

Table 3

Simulation results for monotone dose-responses

Delta	True MED	Linear Response				Step Response			
		BHH	WILM	SDPC	GLEE	BHH	WILM	SDPC	GLEE
$\delta = 6$		0.052	0.051	0.049	0.050				
	3	0.860	0.846	0.776	0.668				
		0.025	0.036	0.098	0.206				
$\delta = 6$		0.041	0.043	0.044	0.048	0.050	0.051	0.051	0.051
	2	0.400	0.378	0.301	0.255	0.808	0.834	0.806	0.769
		0.599	0.627	0.767	0.854	0.172	0.113	0.159	0.191
$\delta = 6$		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.219	0.231	0.231	0.247	0.803	0.879	0.889	0.902
		1.259	1.244	1.353	1.343	0.434	0.226	0.208	0.158
$\delta = 6$		0.048	0.047	0.047	0.047				
	3	0.952	0.953	0.952	0.948				
		-0.060	-0.064	-0.075	-0.072				
$\delta = 8$		0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
	2	0.750	0.731	0.670	0.604	0.947	0.948	0.947	0.946
		0.147	0.167	0.228	0.295	-0.051	-0.052	-0.052	-0.051
$\delta = 8$		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.507	0.508	0.507	0.511	0.998	1.000	1.000	1.000
		0.541	0.545	0.565	0.568	0.004	0.000	0.000	0.000

NOTE: Family-wise error rate, power and bias are given by the upper, middle and lower entry, respectively.

Table 4

Simulation results for non-monotone dose-responses

True MED	Means μ_1, μ_2, μ_3	Procedure	Proportion of Cases where the Dose is Estimated as the MED				FWE	Power	LOP	Bias
			1	2	3	No MED				
3	3.0, 2.8, 6.0	BHH	0.009	0.025	0.877	0.089	0.034	0.877	0.089	0.046
		WILM	0.014	0.021	0.861	0.103	0.035	0.861	0.103	0.054
		SDPC	0.027	0.014	0.785	0.175	0.041	0.785	0.175	0.107
		GLEE	0.026	0.014	0.675	0.285	0.040	0.675	0.285	0.220
	3.0, 5.5, 3.0	BHH	0.011	0.036	0.000	0.953	0.011	0.036	0.953	1.895
		WILM	0.038	0.248	0.000	7130	0.038	0.248	0.713	1.388
		SDPC	0.046	0.624	0.000	0.329	0.047	0.624	0.330	0.613
		GLEE	0.047	0.494	0.003	0.456	0.050	0.494	0.459	0.867
2	3.0, 5.5, 5.0	BHH	0.044	0.526	0.069	0.360	0.044	0.526	0.430	0.746
		WILM	0.047	0.636	0.070	0.247	0.047	0.636	0.317	0.518
		SDPC	0.047	0.642	0.063	0.249	0.047	0.642	0.311	0.513
		GLEE	0.048	0.585	0.111	0.256	0.048	0.585	0.367	0.576
	3.0, 6.0, 4.5	BHH	0.040	0.391	0.010	0.559	0.040	0.039	0.569	1.088
		WILM	0.047	0.668	0.010	0.276	0.047	0.668	0.286	0.515
		SDPC	0.048	0.781	0.008	0.164	0.048	0.781	0.171	0.288
		GLEE	0.048	0.712	0.026	0.214	0.048	0.712	0.240	0.406
	5.7, 5.5, 5.3	BHH	0.606	0.045	0.093	0.256	0.000	0.606	0.394	0.999
		WILM	0.766	0.056	0.042	0.136	0.000	0.766	0.234	0.548
		SDPC	0.804	0.054	0.028	0.114	0.000	0.804	0.196	0.451
		GLEE	0.828	0.065	0.026	0.081	0.000	0.828	0.172	0.36
1	5.5, 5.5, 3.0	BHH	0.046	0.001	0.000	0.953	0.000	0.046	0.954	2.860
		WILM	0.453	0.010	0.000	0.537	0.000	0.453	0.547	1.622
		SDPC	0.743	0.074	0.000	0.183	0.000	0.743	0.257	0.625

True MED	Means μ_1, μ_2, μ_3	Procedure	Proportion of Cases where the Dose is Estimated as the MED					FWE	Power	LOP	Bias
			1	2	3	No MED					
		GLEE	0.757	0.063	0.000	0.181	0.000	0.757	0.243	0.605	