Journal of Statistical Planning and Inference 142 (2012) 2993–2998

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

A note on confidence bounds after fixed-sequence multiple tests

Yi-Hsuan Tu [a,*], Bin Cheng [b], Ying Kuen Cheung [b]

[a] Department of Statistics, National Cheng Kung University, 70101 Tainan, Taiwan
[b] Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032, USA

ARTICLE INFO

Article history:
Received 17 March 2011
Received in revised form
9 March 2012
Accepted 12 May 2012
Available online 23 May 2012

Keywords:
Dose–response
Familywise error rate
Minimum effective dose
Monotonicity
Multiple comparisons

ABSTRACT

We are concerned with the problem of estimating the treatment effects at the effective doses in a dose-finding study. Under monotone dose–response, the effective doses can be identified through the estimation of the minimum effective dose, for which there is an extensive set of statistical tools. In particular, when a fixed-sequence multiple testing procedure is used to estimate the minimum effective dose, Hsu and Berger (1999) show that the confidence lower bounds for the treatment effects can be constructed without the need to adjust for multiplicity. Their method, called the dose–response method, is simple to use, but does not account for the magnitude of the observed treatment effects. As a result, the dose–response method will estimate the treatment effects at effective doses with confidence bounds invariably identical to the hypothesized value. In this paper, we propose an error-splitting method as a variant of the dose–response method to construct confidence bounds at the identified effective doses after a fixed-sequence multiple testing procedure. Our proposed method has the virtue of simplicity as in the dose–response method, preserves the nominal coverage probability, and provides sharper bounds than the dose–response method in most cases.

1. Introduction

In early stage of drug development, a primary scientific objective is to identify doses that exhibit adequate drug activity, indicated by a shift of mean response from the control group by a margin greater than a practically significant value δ. When the dose–response is monotone increasing, the effective doses can be identified through the estimation of the minimum effective dose. At the same time, for the purposes of planning future experiments, it is also important to precisely assess the effect size at the identified effective doses. There is a long history and large literature on the estimation of minimum effective dose, including single-step procedures such as Bonferroni's adjustment and Dunnett's (1955) procedure for many-to-one comparisons, step-down methods due to Naik (1975) and Marcus et al. (1976), and a variety of stepwise procedures described in Tamhane et al. (1996).

On the other hand, relatively little attention has been given to the joint estimation of effective doses and their effect sizes. In fact, it has long been thought that stepwise procedures do not naturally yield confidence sets through inversion (Lehmann, 1986) until Bofinger (1987) and Stefansson et al. (1988) who derive confidence bounds following a step-down test by partitioning principle. Subsequently, Hsu and Berger (1999) propose a dose–response method to find stepwise confidence bounds without multiplicity adjustment. A difficulty associated with these methods is that the confidence

* Corresponding author.
E-mail addresses: yhtu@stat.ncku.edu.tw (Y.-H. Tu), bc2159@columbia.edu (B. Cheng), yc632@columbia.edu (Y.K. Cheung).

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.05.002

Table 1
Dose response data from Hsu and Berger (1999), and 95% simultaneous lower confidence bounds on the treatment effects γ_i by the Hsu–Berger dose–response method (HB), the error-splitting method, the partitioning principle by Stefansson et al. (1988, MPGN), and Dunnett's (1955) method. The sample size is $n=6$ per group.

Dose i	Mean	SD	HB[a]	Lower bounds of γ_i for $\alpha_t =$					MPGN	Dunnett
				0.045	0.040	0.035	0.030	0.025		
0	25.5	2.6								
1	23.9	4.0							−11.5	−12.7
2	27.7	3.3							−15.9	−8.9
3	33.4	2.3	0.4	0.2	−0.1				−2.0	−3.2
4	40.5	10.5	7.0	7.0	7.0	6.7	6.4	6.0	5.1	3.9
5	57.9	9.9	7.0	7.0	7.0	22.5	23.0	23.4	7.0	21.3
6	74.4	14.6	7.0	7.0	7.0	22.5	23.0	23.4	7.0	37.8
7	73.4	7.6	7.0	7.0	7.0	22.5	23.0	23.4	7.0	36.8
8	73.5	4.5	7.0	7.0	7.0	22.5	23.0	23.4	7.0	36.9
9	76.2	7.9	7.0	7.0	7.0	22.5	23.0	23.4	7.0	39.6

[a] HB is identical to error-splitting method with $\alpha_t = 0.050$ in this example.

bounds for the treatment effects at the identified effective doses always equal the hypothesized value irrespective of the data. To illustrate, Table 1 extracts the dose–response data and the confidence bounds given in Hsu and Berger (1999) who consider $\delta = 7$. The minimum effective dose is estimated to be dose 4 by the dose–response method of Hsu and Berger (1999), and dose 5 by the confidence bounds of Stefansson et al. (1988). The treatment effects at all identified effective doses are estimated with a confidence lower bound of $\delta = 7$ by the dose–response method, despite the fact that the observed effect sizes at the higher doses (dose 5 and above) are apparently much larger than 7. In practice, if we use this lower bound (i.e. 7) as an assumed effect size in the planning of a future study, we will unduly require a much larger sample size than needed. The fundamental problem is that these procedures use up the error in the testing procedure to establish confidence direction (i.e. whether there is an effect of the dose) with no margin of error left for estimation of the effect size. In this paper, we propose a two-step procedure that "splits the error rate" in two parts respectively for testing and estimation, and apply this procedure to modify Hsu and Berger's dose–response method.

2. Methods

Consider the balanced one-way layout $Y_{ij} = \mu_i + \epsilon_{ij}$ for $i = 0,\ldots,k$ and $j = 1,\ldots,n$, where Y_{ij} denotes the response of subject j in dose i, and ϵ_{ij} is a normal random noise with mean 0 and variance σ_e^2. Let $\gamma_i = \mu_i - \mu_0$ be the treatment effect of dose i relative to the control, which is commonly estimated by the pairwise statistic $T_i = \overline{Y}_i - \overline{Y}_0$, where $\overline{Y}_i = \sum_{j=1}^{n} Y_{ij}/n$, so that $\mathrm{var}(T_i) = \sigma^2 = 2\sigma_e^2/n$ and $\mathrm{corr}(T_i,T_j) = \rho = 1/2$. In addition, under monotonicity, i.e., $\gamma_1 \le \cdots \le \gamma_k$, the minimum effective dose $v \equiv \min\{i : \gamma_i > \delta\}$ is equal to $\max\{i : \gamma_i \le \delta\}+1$; we adopt the convention that $\min\{\emptyset\} = k+1$ and $\max\{\emptyset\} = 0$. Then the $(1-\alpha)$-upper confidence bound for v is

$$\hat{v}_\alpha = \max\{i : T_i \le \delta + \sigma z_\alpha\} + 1, \tag{1}$$

where z_α denotes the upper α critical point of standard normal distribution. In general, the variance σ^2 can be consistently estimated; here, it is assumed known for brevity. If $\hat{v}_\alpha \ge 2$, the dose–response method asserts that $\gamma_i > \delta$ for all $i \ge \hat{v}_\alpha$ and $\gamma_{\hat{v}_\alpha - 1} > T_{\hat{v}_\alpha - 1} - z_\alpha\sigma$. When $\hat{v}_\alpha = 1$, the stepwise confidence bound is $\gamma_i > \min_{i=1,\ldots,k}\{T_i - z_\alpha\sigma\}$ for all i. The confidence bounds thus obtained will achieve a $100(1-\alpha)\%$ coverage probability by Theorem 1 in Hsu and Berger (1999) when σ^2 is assumed known.

To improve the precision of the confidence bounds at the effective doses while maintaining the same coverage probability, we propose to estimate v at a slightly more conservative significance level, i.e., $\hat{v}_{\alpha_t}$ where $\alpha_t \le \alpha$. With this estimate of v, the confidence bounds for the effect sizes can be constructed as follows:

1. if $\hat{v}_{\alpha_t} = k+1$, then $\gamma_k > T_k - \sigma z_\alpha$;
2. if $\hat{v}_{\alpha_t} \in \{2,\ldots,k\}$, then $\gamma_{\hat{v}_{\alpha_t}-1} > T_{\hat{v}_{\alpha_t}-1} - \sigma z_t, \gamma_i > \max\{T_{\hat{v}_{\alpha_t}} - \sigma z_e, \delta\}$ for all $i \ge \hat{v}_{\alpha_t}$;
3. if $\hat{v}_{\alpha_t} = 1$, then $\gamma_i > \max\{T_1 - \sigma z_e, \delta\}$ for all $i \ge 1$,

where z_t and z_e are the abbreviations of z_{α_t} and z_{α_e}, respectively, for brevity. Let $\mathfrak{C}(T)$ denote the confidence bound for $\gamma = (\gamma_1,\ldots,\gamma_k)'$ by this two-step error-splitting procedure.

Proposition 1. *For any $\alpha_t \in [0,\alpha]$, choose α_e such that*

$$\mathrm{pr}\{Z_1 < \min(z_t, z_e), Z_2 < z_e\} = 1-\alpha, \tag{2}$$

where $(Z_1,Z_2)'$ *is distributed as bivariate normal with standard normal marginal and a correlation coefficient of* 0.5 *between* Z_1 *and* Z_2. *Then*, $\mathrm{pr}\{\gamma \in \mathfrak{C}(T)\} \geq 1-\alpha$.

The condition (2) in Proposition 1 provides some guidance on how the error probabilities, α_t and α_e, should be chosen. First, consider the case $\alpha_t = \alpha$, under which condition (2) imposes $\alpha_e = 0$. In this case, the bound $\mathfrak{C}(T)$ is identical to the confidence bound due to the Hsu–Berger dose–response method when $\hat{v}_\alpha > 1$; however, it is easy to verify that the former is more conservative than the latter when $\hat{v}_\alpha = 1$. Therefore, setting $\alpha_t = \alpha$ is not an admissible choice.

Second, consider the case $\alpha_t \leq \alpha_e < \alpha$ so that $z_e \leq z_t$. Then condition (2) does not depend on α_t. In other words, we can increase the test level α_t without affecting the overall coverage probability, as long as $\alpha_t \leq \alpha_e$. A practical implication is that we should set α_t to be at least as large as α_e.

From now on , we will focus on $\alpha_e \leq \alpha_t < \alpha$ so that $\hat{v}_\alpha \leq \hat{v}_{\alpha_t}$, under which it is possible for the error-splitting approach to yield a more conservative estimate of v than the Hsu–Berger dose–response method. The motivation, on the other hand, is to improve the confidence lower bounds at the estimated effective doses. When $\alpha_t = \alpha_e$, the lower bound $\mathfrak{C}(T)$

$$\gamma_i > \max(T_{\hat{v}_{\alpha_t}} - \sigma z_e, \delta) = \max(T_{\hat{v}_{\alpha_t}} - \sigma z_t, \delta) > \delta \tag{3}$$

for the estimated effective doses, i.e., $i \geq \hat{v}_{\alpha_t}$, if $2 \leq \hat{v}_\alpha \leq \hat{v}_{\alpha_t} \leq k$. The last inequality in (3) is a result of the definition of $\hat{v}_{\alpha_t}$ in (1). If the true γ_i at the minimum effective dose is far greater than δ, the gain over the dose–response method (which gives a lower bound of δ) is potentially substantial. To preserve the gain for the general case $\alpha_e \leq \alpha_t$, the comparison (3) suggest choosing a value of α_e that is not much smaller than α_t; see also the simulation results in Section 3. It is important to note that our proposed $\mathfrak{C}(T)$ is not uniformly better than the Hsu–Berger dose–response method in terms of the lower bound. In particular, when $\hat{v}_\alpha = \hat{v}_{\alpha_t} = 1$, the lower bound $\mathfrak{C}(T)$ gives $\gamma_i > \max(T_1 - \sigma z_e, \delta)$ for all i, whereas the dose–response method gives $\gamma_i > \min_{i=1,\ldots,k}\{T_i - \sigma z_\alpha\}$. Depending on the shape of the dose–response curve, the dose–response method can likely be superior to $\mathfrak{C}(T)$ on the event $\{\hat{v}_\alpha = \hat{v}_{\alpha_t} = 1\}$ since the Hsu–Berger method does not require the monotonicity assumption.

Our discussion has been assuming a known σ_e^2. In situations when the variance σ_e^2 is unknown, we could apply the error-splitting approach in an analogous manner by replacing σ_e^2 with the pooled sample variance, and using the critical values with respect to a t distribution with $(n-1)(k+1)$ degrees of freedom. It is easy to see that this procedure will achieve the nominal coverage probability asymptotically. For small-to-moderate sample sizes, intuitively, this approach will be conservative because t distribution has a heavier tail than the normal distribution. In the simulation study that follows, we will implement this procedure and examine the performance of the error splitting procedure in finite sample settings.

3. Numerical studies

We first consider the dose response data in Hsu and Berger (1999), who compared nine doses to a placebo with six subjects per dose level. The goal was to identify doses having a mean that is 7 mg/kg greater than that of the placebo, i.e., $\delta = 7$. Table 1 shows the 95% confidence lower bounds given by the error-splitting method with $\alpha_t = 0.045, 0.040, 0.035, 0.030, 0.025$, with respective $\alpha_e = 0.008, 0.014, 0.020, 0.026, 0.025$ so that overall $\alpha = 0.05$ in accord with condition (2). The results due to the Hsu–Berger dose–response method, the partitioning principle (Stefansson et al., 1988), and Dunnett's (1955) method are also given. When α_t is close to the nominal α, the error-splitting method behaves similarly to the dose–response method, and selects dose 4 as the minimum effective dose. However, it does not materialize the advantage of our proposed method as α_e is extremely small. When $\alpha_t \leq 0.035$, the error-splitting method estimates v with dose level 5 but provides a much more encouraging effect size at this dose (and above) than the dose–response method. The partitioning principle also estimates v with dose 5 but fails to use the observed data to estimate the effect size. In this example, Dunnett's method appears to be superior to the error-splitting approach. As we will see in the following simulation, Dunnett's method tends to yield sharper lower bounds on the higher doses but also tends to over-estimate v more often than the error-splitting approach.

Based on the dose–response data in Table 1, we next simulated data from the normal distributions with a common standard deviation $\sigma_e = 7.8$ at $k = 9$ doses with six subjects at each dose; we considered three sets of mean dose–response so that the true $v = 4$. The first scenario has a linear dose–response with $\gamma = (-7, 0, 7, 14, 21, 28, 35, 42, 49)'$. The second scenario is obtained from an $E_{\max}$ model (Ting, 2006) with $\mu_i = 25 + 51 i^{6.5}/(i^{6.5} + 4^{6.5})$, $i = 0, \ldots, 9$, yielding $\gamma = (0, 1, 7, 26, 41, 48, 50, 50, 51)'$. The third scenario has a plateau dose–response pattern with $\gamma = (3, 3, 3, 15, 46, 46, 46, 46, 46)'$. Table 2 gives the coverage probability (cov), the probability of selecting the true minimum effective dose (pcs), and the median of the lower bound at the truly effective doses based on 5000 simulation runs under each scenario. The coverage probability is estimated by the proportion when the lower bounds cover the true means of the nine doses simultaneously. Because the lower bounds may take values on negative infinity, we use medians (instead of means) of the lower bounds in the comparison.

As expected, the precision of $\mathfrak{C}(T)$ improves for all γ_i's with $i > \hat{v}_{\alpha_t}$ as α_t decreases and α_e increases; on the other hand, we are slightly surprised to see non-trivial improvement even with small α_e's, namely when $\alpha_t = 0.040$ and 0.045. Also as expected, the pcs tends to increase with α_t. However, we find that its impact on the error-splitting approach is mild when compared to Dunnett's method. In general, the performance of the error-splitting approach is somewhere between Hsu–Berger dose–response method and Dunnett's method. Based on this simulation, setting $\alpha_t = 0.035$ appears to strike a good balance between lower bound precision and pcs.

Table 2
Simulation results of Hsu–Berger dose–response method, Dunnett's method, and the error-splitting method for $\alpha_t = 0.045, 0.040, , 035, , 030, , 025$. The error-splitting method with $\alpha_t = 0.050$ yields identical results to the dose–response method.

α_t	cov	pcs	lb4	lb5	lb6	lb7	lb8	lb9
(a) Linear dose–response, $\gamma = (-7, 0, 7, 14, 21, 28, 35, 42, 49)'$								
HB	0.957	0.392	6.2	7.0	7.0	7.0	7.0	7.0
0.045	0.954	0.375	6.0	7.1	7.8	7.8	7.8	7.8
0.040	0.953	0.359	5.7	8.3	9.1	9.1	9.1	9.1
0.035	0.951	0.334	5.4	9.2	10.1	10.1	10.1	10.1
0.030	0.951	0.313	5.0	9.9	10.8	10.9	10.9	10.9
0.025	0.955	0.290	4.6	10.0	11.1	11.2	11.2	11.2
Dunnett	0.942	0.188	3.0	10.0	17.2	24.1	31.1	38.0
(b) $E_{\max}$ dose–response, $\gamma = (0, 1, 7, 26, 41, 48, 50, 50, 51)'$								
HB	0.950	0.943	7.0	7.0	7.0	7.0	7.0	7.0
0.045	0.949	0.945	14.2	14.3	14.3	14.3	14.3	14.3
0.040	0.949	0.950	15.3	15.4	15.4	15.4	15.4	15.4
0.035	0.950	0.952	16.0	16.2	16.2	16.2	16.2	16.2
0.030	0.950	0.955	16.6	16.7	16.7	16.7	16.7	16.7
0.025	0.954	0.957	16.6	16.8	16.8	16.8	16.8	16.8
Dunnett	0.942	0.950	15.0	30.0	37.2	39.1	39.1	40.0
(c) Plateau dose–response, $\gamma = (3, 3, 3, 15, 46, 46, 46, 46, 46)'$								
HB	0.954	0.522	7.0	7.0	7.0	7.0	7.0	7.0
0.045	0.952	0.499	7.0	15.5	15.5	15.5	15.5	15.5
0.040	0.951	0.479	6.8	25.9	25.9	25.9	25.9	25.9
0.035	0.950	0.457	6.6	28.5	28.5	28.5	28.5	28.5
0.030	0.950	0.426	6.2	30.3	30.3	30.3	30.3	30.3
0.025	0.954	0.402	5.9	31.3	31.3	31.3	31.3	31.3
Dunnett	0.942	0.262	4.0	35.0	35.2	35.1	35.1	35.0

cov, coverage probability; pcs, probability of selecting v; lb$k$, the median of lower bound of treatment effect at dose k, for $k = 4, \ldots, 9$.

Table 3
Coverage probability of the error-splitting method under non-normal data for given α_t.

α_t	Linear		$E_{\max}$		Plateau	
	Scaled t_3	DE	Scaled t_3	DE	Scaled t_3	DE
0.050	0.944	0.958	0.944	0.952	0.946	0.956
0.045	0.944	0.958	0.944	0.953	0.946	0.957
0.040	0.944	0.958	0.944	0.954	0.946	0.958
0.035	0.944	0.957	0.944	0.955	0.947	0.955
0.030	0.946	0.955	0.946	0.956	0.947	0.954
0.025	0.952	0.959	0.952	0.959	0.951	0.958

While the error-splitting procedure is asymptotically valid as long as the random error ϵ_{ij} has zero mean and finite variance, we examined its robustness in finite sample sizes in terms of coverage probability when the error distribution is non-normal. In particular, we performed additional simulations under the same sets of means with the errors generated from scaled t_3 and double exponential (DE) distributions with standard deviation $\sigma_e = 7.8$. As indicated in Table 3, the error-splitting approach achieves nominal coverage even when the sample size is as small as $n = 6$ per group.

We also examined the robustness of the error-splitting method under unequal sample sizes in the one-way layout. While Proposition 1 does not apply to these scenarios, the simulation results in Table 4 shows that mild imbalance in the one-way layout does not affect the coverage probability. It will indeed be interesting to extend our results to cover unbalanced one-way layout. Intuitively, we may extend Proposition 1 using the generalized Slepian's inequality for elliptically contoured distributions (Tong, 1980), although the partitioning arguments used in the proof will require nontrivial manipulations, and warrant further investigation.

4. Proof of Proposition 1

Lemma. *Let Y_i, $i = 0, \ldots, n$, $n \geq 3$, be independent real valued random variables and let Y_i, $i = 1, \ldots, n$, be identically distributed. Then*

$$\mathrm{pr}\{X_1 > a_1, X_2 < a_2, X^{(12)} \in \mathfrak{B}\} \geq \mathrm{pr}\{a_1 < X_1 < a_2, X^{(12)} \in \mathfrak{B}\},$$

where $X = (X_1, \ldots, X_n)' = (Y_1 - Y_0, \ldots, Y_n - Y_0)'$, $X^{(12)} = (X_3, \ldots, X_n)'$ and $\mathfrak{B}$ is a Borel set in the $n-2$-dimensional Euclidean space.

Table 4
Simulation results of the error-splitting method under various dose–response curves with $v=4$. Sample size for control and each level are 6, 5, 4, 3, 4, 4, 7, 8, 8, and 6.

α_t	cov	pcs	lb4	lb5	lb6	lb7	lb8	lb9
(a) Linear dose–response, $\gamma=(-7,0,7,14,21,28,35,42,49)'$								
HB	0.968	0.340	5.1	7.0	7.0	7.0	7.0	7.0
0.045	0.964	0.321	4.7	7.0	7.7	7.7	7.7	7.7
0.040	0.962	0.300	4.4	7.6	9.2	9.2	9.2	9.2
0.035	0.960	0.279	4.0	8.5	10.3	10.3	10.3	10.3
0.030	0.959	0.254	3.4	9.2	11.2	11.2	11.2	11.2
0.025	0.962	0.234	2.9	9.3	11.5	11.5	11.5	11.5
Dunnett	0.942	0.156	1.8	8.8	17.3	24.8	31.7	38.1
(b) $E_{\max}$ dose–response, $\gamma=(0,1.7,26,41,48,50,50,51)'$								
HB	0.954	0.935	7.0	7.0	7.0	7.0	7.0	7.0
0.045	0.952	0.936	11.9	12.2	12.2	12.2	12.2	12.2
0.040	0.952	0.937	13.3	13.6	13.6	13.6	13.6	13.6
0.035	0.952	0.937	14.1	14.5	14.5	14.5	14.5	14.5
0.030	0.955	0.936	14.8	15.2	15.2	15.2	15.2	15.2
0.025	0.959	0.932	14.7	15.2	15.2	15.2	15.2	15.2
Dunnett	0.942	0.897	13.8	28.8	37.3	39.8	39.7	40.1
(c) Plateau dose–response, $\gamma=(3,3,3,15,46,46,46,46)'$								
HB	0.963	0.461	6.6	7.0	7.0	7.0	7.0	7.0
0.045	0.961	0.443	6.3	25.6	25.6	25.6	25.6	25.6
0.040	0.960	0.420	6.0	27.7	27.7	27.7	27.7	27.7
0.035	0.959	0.397	5.7	29.3	29.3	29.3	29.3	29.3
0.030	0.958	0.371	5.3	30.6	30.6	30.6	30.6	30.6
0.025	0.961	0.339	4.9	31.3	31.3	31.3	31.3	31.3
Dunnett	0.942	0.212	2.8	33.8	35.3	35.8	35.7	35.1

cov, coverage probability; pcs, probability of selecting v; lb$k$, the median of lower bound of treatment effect at dose k, for $k=4,\ldots,9$.

Proof. Define the cdf's of Y_0 and Y_i, $i=1,\ldots,n$, by F_0 and F, respectively, then it holds

$$\begin{aligned}\mathrm{pr}\{X_1>a_1,X_2<a_2,X^{(12)}\in\mathfrak{B}\} &= \int_{-\infty}^{\infty}\mathrm{pr}\{Y_1>a_1+t,Y_2<a_2+t,(Y_3-t,\ldots,Y_n-t)\in\mathfrak{B}\}\,dF_0(t)\\ &= \int_{-\infty}^{\infty}[F(a_2+t)-F(a_1+t)F(a_2+t)]\mathrm{pr}\{(Y_3-t,\ldots,Y_n-t)\in\mathfrak{B}\}\,dF_0(t)\\ &\geq \int_{-\infty}^{\infty}[F(a_2+t)-F(a_1+t)]\mathrm{pr}\{(Y_3-t,\ldots,Y_n-t)\in\mathfrak{B}\}\,dF_0(t)\\ &= \mathrm{pr}\{a_1<X_1<a_2,X^{(12)}\in\mathfrak{B}\}.\end{aligned}$$

To prove Proposition 1, we define $R_i=\{\text{Reject } H_{0j} \text{ for } i\leq j\leq k;\ \text{accept } H_{0,i-1}\}$, for $i=1,\ldots,k+1$, be the event of detecting dose i as the minimum effective dose in the testing procedure. In particular, $R_1=\{\text{Reject } H_{01},\ldots,H_{0k}\}=\{T_i>\delta+\sigma z_t \text{ for all } i\in\{1,\ldots,k\}\}$, and $R_{k+1}=\{\text{Accept } H_{0k}\}=\{T_k\leq\delta+\sigma z_t\}$. Let $\mathfrak{C}(T)$ denote the confidence bound for $\gamma=(\gamma_1,\ldots,\gamma_k)'$ generated by the error-splitting method based on $T=(T_1,\ldots,T_k)'$ and $\mathfrak{C}_i(T)$ denote the conference bound for γ when the event R_i occurs such that $\mathrm{pr}\{\gamma\in\mathfrak{C}(T)\}=\sum_{i=1}^{k+1}\mathrm{pr}\{\gamma\in\mathfrak{C}_i(T),R_i\}$. Partition the parameter space, $\Theta=\{\gamma:-\infty<\gamma_1\leq\cdots\leq\gamma_k<\infty\}$, into $k+1$ disjoint parameter space, Θ_i, where $\Theta_{k+1}=\{\gamma:\gamma_k\leq\delta+\sigma(z_t-z_\alpha)\}$ and $\Theta_i=\{\gamma:\gamma_k>\delta+\sigma(z_t-z_\alpha),\gamma_i>\delta,\gamma_{i-1}\leq\delta\}$, for $i=1,\ldots,k$. Define $a_{j,i}=P\{\gamma\in\mathfrak{C}_j(T),R_j|\gamma\in\Theta_i\}$, $1\leq i,j\leq k+1$, where $P\{\cdot|\gamma\in\Theta_i\}$ denotes probability computed when $\gamma\in\Theta_i$. Then $P\{\gamma\in\mathfrak{C}(T)|\gamma\in\Theta_i\}=\sum_{j=1}^{k+1}a_{j,i}$. It is immediately seen that $a_{j,i}=0$ for $1\leq j<i\leq k$. □

Proof of Proposition 1. Without loss of generality, we set $\delta=0$, $\sigma=1$. Let $Z_i=(T_i-\gamma_i)$, $i=1,\ldots,k$. When $\gamma\in\Theta_{k+1}$, $P\{\gamma\in\mathfrak{C}(T)|\gamma\in\Theta_{k+1}\}\geq\mathrm{pr}\{Z_k\leq z_\alpha,Z_k\leq z_t-\gamma_k\}=\mathrm{pr}\{Z_k\leq z_\alpha\}=1-\alpha$.

Define $A_j=\{Z_j>z_t-\gamma_j\}$, $B_j=\{z_t-\gamma_j<Z_j<z_e\}$, $B'_j=\{z_t-\gamma_j<Z_j<\min(z_t,z_e)\}$, $C_j=\{Z_j<z_t\}$, $C'_j=\{Z_j<z_e\}$, and $C''_j=\{Z_j<\min(z_t,z_e)\}$, $j=1,\ldots,k$. For $\gamma\in\Theta_k$, $P\{\gamma\in\mathfrak{C}(T)|\gamma\in\Theta_k\}=\mathrm{pr}\{B_k\cap C_{k-1}\}+\mathrm{pr}\{A_k^c\}\geq\mathrm{pr}\{C'_k\cap C''_{k-1}\}=1-\alpha$, where the last equality is from (2).

For $1\leq i\leq k-1$ and $i\leq j\leq k+1$, $a_{j,i}$ can be expressed as

$$a_{j,i}=\begin{cases}\mathrm{pr}\{A_k\cap\cdots\cap A_{i+1}\cap B_i\cap C_{i-1}\}, & j=i,\\ \mathrm{pr}\{A_k\cap\cdots\cap A_{j+1}\cap B_j\cap A_{j-1}^c\}, & i+1\leq j\leq k,\\ \mathrm{pr}\{A_k^c\}, & j=k+1,\end{cases}$$

where A_j^c denotes the complement of A_j. A set of the form $A_k \cap \cdots \cap A_s$ is understood as Ω, the whole sample space, whenever $s > k$. We prove by induction that for any l such that $i \le l \le k-1$

$$\sum_{j=i}^{l} a_{j,i} \ge \begin{cases} \text{pr}\{A_k \cap \cdots \cap A_{l+2} \cap B'_{l+1} \cap B_l\}, & \text{if } l-i \text{ is even,} \\ \text{pr}\{A_k \cap \cdots \cap A_{l+2} \cap B_{l+1} \cap B'_l\}, & \text{if } l-i \text{ is odd.} \end{cases} \tag{4}$$

First, $a_{i,i} = \text{pr}\{A_k \cap \cdots \cap A_{i+1} \cap B_i \cap C_{i-1}\} \ge \text{pr}\{A_k \cap \cdots \cap A_{i+2} \cap B'_{i+1} \cap B_i\}$ by the Lemma, hence (4) holds when $l = i$, $i > k-1$. Note that, when $l = i = k-1$, $a_{k-1,k-1} = \text{pr}\{B'_k \cap B_{k-1}\}$. Suppose (4) holds for l, we will prove that it also holds for $l+1 \le k-1$. If $l-i$ is an even number, then

$$\begin{aligned} \sum_{j=i}^{l} a_{j,i} + a_{l+1,i} &\ge \text{pr}\{A_k \cap \cdots \cap A_{l+2} \cap B'_{l+1} \cap B_l\} + \text{pr}\{A_k \cap \cdots \cap A_{l+2} \cap B_{l+1} \cap A_l^c\} \\ &\ge \text{pr}\{A_k \cap \cdots \cap A_{l+2} \cap B'_{l+1} \cap (B_l \cup A_l^c)\} \\ &= \text{pr}\{A_k \cap \cdots \cap A_{L+2} \cap B'_{l+1} \cap C'_l\} \\ &\ge \text{pr}\{A_k \cap \cdots \cap A_{l+3} \cap (A_{l+2} \cap C'_{l+2}) \cap B'_{l+1}\} \\ &= \text{pr}\{A_k \cap \cdots \cap A_{l+3} \cap B_{l+2} \cap B'_{l+1}\}, \end{aligned}$$

where the third inequality is by the Lemma. The case when $l-i$ is an odd number can be proved using the above technique and the fact that $A_j^c \cup B'_j \supseteq C''_j$ for $j = 1, \ldots, k$. Therefore, (4) holds for all $l \le k-1$ by the mathematical induction principle.

For $\gamma \in \Theta_i$, $1 \le i \le k-1$, set $l = k-1$ in (4) and assume that $k-1-i$ is even

$$\begin{aligned} P\{\gamma \in \mathfrak{C}(T) | \gamma \in \Theta_i\} &= \sum_{j=i}^{k-1} a_{j,i} + a_{k,i} + a_{k+1,i} \\ &\ge \text{pr}\{B'_k \cap B_{k-1}\} + \text{pr}\{B_k \cap A_{k-1}^c\} + \text{pr}\{A_k^c\} \\ &\ge \text{pr}\{B'_k \cap C'_{k-1}\} + \text{pr}\{A_k^c\} \\ &\ge \text{pr}\{C''_k \cap C'_{k-1}\} \\ &= 1-\alpha, \end{aligned}$$

where the first inequality follows from (4) and the last equality is by (2). The result can be proved similarly when $k-1-i$ is odd. □

Acknowledgments

We thank the Editor, Associated Editor, and two referees for helpful comments. Dr. Cheung's work was partially supported by NIH Grant R01 NS055809. Dr. Tu's work was partially supported by National Science Council under Grant NSC 96-2118-M-006-008.

References

Bofinger, F., 1987. Stepdown procedures for comparison with a control. Australian Journal of Statistics 29, 348–364.
Dunnett, C.W., 1955. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50, 1096–1121.
Hsu, J.C., 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.
Lehmann, E.L., 1986. Testing Statistical Hypothesis, second ed. Springer, New York.
Marcus, R., Peritz, E., Gabriel, K.R., 1976. On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63, 655–660.
Naik, U.D., 1975. Some selection rules for comparing p processes with a standard. Communications in Statistics: Theory and Methods 4, 519–535.
Stefansson, G., Kim, W., Hsu, J.C., 1988. On confidence sets in multiple comparisons. In: Gupta, S.S., Berger, J.O. (Eds.), Statistical Decision Theory and Related Topics IV, vol. 2. , Springer-Verlag, New York, pp. 89–104.
Tamhane, A.C., Hochberg, Y., Dunnett, C.W., 1996. Multiple test procedures for dose finding. Biometrics 52, 21–37.
Ting, N., 2006. Dose Finding in Drug Development. Springer-Verlag, New York.
Tong, Y.L., 1980. Probability Inequalities in Multivariate Distributions. Academic Press, New York.