Journal of Statistical Planning and Inference 140 (2010) 180 – 188

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

# Comparing multiple treatments to both positive and negative controls

Nairanjana Dasgupta[a,*], Eleanne Solorzano[b], Tiejun Tong[c]

[a] Department of Statistics, Washington State University, Pullman, WA 99164, USA
[b] Decision Sciences Department, University of New Hampshire, Durham, NH 03824, USA
[c] Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, USA

ARTICLE INFO

*Article history:*
Received 8 October 2008
Received in revised form
30 June 2009
Accepted 2 July 2009
Available online 9 July 2009

*Keywords:*
Least favorable configuration
Log concavity
Multiple comparison
Negative control
Positive control

ABSTRACT

In the past, most comparison to control problems have dealt with comparing $k$ test treatments to either positive or negative controls. Dasgupta et al. [2006. Using numerical methods to find the least favorable configuration when comparing $k$ test treatments to both positive and negative controls. Journal of Statistical Computation and Simulation 76, 251–265] enumerate situations where it is imperative to compare several test treatments to both a negative as well as a positive control simultaneously. Specifically, the aim is to see if the test treatments are worse than the negative control, or if they are better than the positive control when the two controls are sufficiently apart. To find critical regions for this problem, one needs to find the least favorable configuration (LFC) under the composite null. In their paper, Dasgupta et al. [2006. Using numerical methods to find the least favorable configuration when comparing $k$ test treatments to both positive and negative controls. Journal of Statistical Computation and Simulation 76, 251–265] came up with a numerical technique to find the LFC. In this paper we verify their result analytically. Via Monte Carlo simulation we compare the proposed method to the logical single step alternatives: Dunnett's [1955. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50, 1096–1121] or the Bonferroni correction. The proposed method is superior in terms of both the Type I error and the marginal power.



## 1. Introduction

Since Dunnett's (1955) breakthrough paper, multiple comparisons to control (MCC) or many-to-one comparisons have been a widely researched topic. A control is one of two types: "positive control" and "negative control". A "positive control" is generally the standard treatment. It could be the "ideal" treatment in some cases. A "negative control" generally implies no application of treatment. Under some circumstances, this could be the least favorable treatment. Dunnett (1955) asserts when a control is present in the study, it is of interest to compare several treatments to the control. He provided one- and two-sided confidence intervals for this purpose. There has been a spate of work on the topic of comparing several treatments with one control. More recently, still another aspect of MCC is being studied, the problem comparing multiple controls to multiple treatments. Some papers on this topic include Shaffer (1997), Hoover (1991), and Solorzano and Spurrier (1999). The multiple controls studied by the papers mentioned had multiple controls that were all in one group and did not differentiate between a positive control and a negative control.

Literature differentiating both types of controls is not as prolific. D'Agostino and Hareen (1991) discuss a problem involving both positive and negative controls for over-the-counter drugs. They discuss multiplicity and recommend control of family-wise

* Corresponding author.
*E-mail address:* dasgupta@wsu.edu (N. Dasgupta).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.07.001

**Table 1**
Data from the allergy testing example.

| Treatments | Saline (negative control) | T1 | T2 | T3 | T4 | d farinae (positive control) |
|---|---|---|---|---|---|---|
| Means | 0.522 | 2.522 | 2.565 | 1.783 | 10.304 | 20.609 |
| Stan. dev. | 1.755 | 3.175 | 2.936 | 3.444 | 3.655 | 10.701 |
| Sample size | 23 | 23 | 23 | 23 | 23 | 23 |

error as opposed to per-comparison-wise error. Dunnett and Tamhane (1992) discuss a problem in medical settings where one compares both positive and negative controls to a single treatment. Bauer et al. (1998) discuss a situation involving both controls for a multi-dose experiment. Hothrorn et al. (2000) discuss a situation involving both controls in a dose–response mutagenicity study. Dasgupta et al. (2006) discuss a problem where the motivating examples required testing a composite hypothesis involving multiple treatments. It was of interest to see if at least one treatment is either better than the positive control or worse than the negative control (different directions). To find critical points, they established the least favorable configuration (LFC) for the composite hypothesis numerically. In this paper we follow up on the problem and provide an analytical proof for the LFC.

In Section 2, we discuss in brief the data examples motivating this problem. We briefly describe an example related to allergy testing. In Section 3, we set up notation and definitions used in this manuscript. In Section 4 we present our main result illustrated with a numerical example. Section 5 details the results from a simulation study comparing our method to both Dunnett's method and the Bonferroni method. The conclusion, discussion and future research are given in Section 6.

## 2. Data examples

Applications involving both positive and negative controls abound in clinical trials. Generally, a pharmaceutical company faces two issues: passing the drug through the regulatory agency and marketing the drug. For a drug to pass the regulatory agency, it must be as viable (at least as good) as the existing standard. For marketing purposes, it is desirable to show that the drug is better than the "most effective" one in the market. The problem of interest is to allow both decisions to be made simultaneously while controlling the overall Type I error rate.

Dasgupta et al. (2006) provide two examples to motivate their research. We, very briefly, discuss one other example to give the reader a flavor of the real life application of this problem.

*The allergy testing example*: The presence of both positive and negative controls is common in allergy skin tests. Here various allergens are administered through skin punctures along with a positive control (generally histamine or a common allergen) and a negative control (saline/glycerin). The idea is: most people will react to the positive control and the skin will produce a wheal or a hive, and few should show reaction to the negative control. Other controls, generally common allergens, are included along with the suspected allergen(s) and the positive and negative controls. The response is the size of the wheal from the punctures. The suspected allergen(s) is compared to the size of the controls for classification as negative, positive or intermediate results. The data set used here was collected by a private Allergy Clinic in collaboration with the Department of Health Research Education Center at Washington State University. It was a part of a larger data collection effort and came as a consulting problem to the first author. The researchers were interested in identifying possible allergens that could be used as controls for future reference. Here 23 subjects were exposed to six skin pricks, negative control (saline), positive control *d farinae* (dust mites), and four test treatments. The four test treatments were all common allergens and it was of interest in this design study to compare the mean wheal size of the four test treatments with both the negative and positive controls. If the mean wheal size was smaller than the negative control it could be included in future tests as a potential negative control. If the mean wheal size was larger than the positive control it could be used as a potential positive control. Since we only use the data set for illustration purposes, we ignore the complications like the repeated measures/block design nature of the data and consider a one-way structure with four test treatments and two controls. The means and standard deviations are given in Table 1.

## 3. Notation and hypothesis

We use the same notation as Dasgupta et al. (2006). We briefly define the terms required for the analytical result. We assume a one-way layout with *i.i.d.* normal errors. Let $Y_{ij}$ be the response for the $j$ th observation of the $i$ th treatment. The model is given as

$$Y_{ij} = \theta_i + \varepsilon_{ij}, \quad i = 0, \ldots, k+1, \ j = 1, \ldots, n_i \tag{1}$$

with $n_0 = n_{k+1} = m \geq 2$ and $n_i = n$ for $i = 1, \ldots, k$, where

$\theta_0$: represents the mean effect for the negative control,
$\theta_i$: represents the mean effect for the $i$ th treatment, $i = 1, \ldots, k$,
$\theta_{k+1}$: represents the mean effects of the positive control,
$\varepsilon_{ij}$: the *i.i.d.* random errors.

Our assumptions for this model are

(1) $\varepsilon_{ij} \sim$ *i.i.d.* $N(0, \sigma^2)$;
(2) $\theta_{k+1} - \theta_0 > 3\sigma$.

We are assuming that the sample sizes associated with the two controls are equal but possibly different from the equal sample sized $k$ treatments. Having equal sample sizes for the controls and equal sample sizes for the treatments is routinely done when the study encompasses multiple controls and multiple treatments for ease of exact critical point calculations. Both Solorzano and Spurrier (1999) and Hoover (1991) made the same assumption. We also note that in assigning equal sample sizes to the controls we are tacitly assuming that both the controls are of equal importance. Gupta et al. (2002) discuss the presence of multiple controls of varying importance in the context of requiring different precisions for the comparison of treatments with different controls. Dasgupta and SahaRay (2007) discuss optimally allocating observations when there are two controls of unequal importance in the study. Here we are not considering these situations.

Assumption (1) follows from the usual set of ANOVA assumptions. Assumption (2) follows if we take into account the fact that this problem only makes sense when there is considerable distance between the two controls. If the means of the two controls are close to each other, it is not logical to compare the treatments for viability to the negative control and applicability to the positive control. In that case, when the distance between the two controls is close to zero, the ideal critical value is Dunnett's two-sided critical point.

Let $\hat{\theta}_i, i=0,\ldots,k+1$, represent the corresponding sample means for the effects and $s$ denote the overall pooled sample standard deviation. Let $v = k(n-1) + 2(m-1)$ be the corresponding degrees of freedom. The hypothesis of interest is given by

$$\begin{aligned} &H_0 : \theta_0 \le \theta_i \le \theta_{k+1}, \quad i = 1,\ldots,k, \\ &H_a : \theta_i < \theta_0 \quad \text{or} \quad \theta_i > \theta_{k+1} \quad \text{for at least one } i = 1,\ldots,k. \end{aligned} \tag{2}$$

We define our usual $t$ statistics as follows:

$$T_i = \frac{\hat{\theta}_i - \hat{\theta}_0}{s\sqrt{\frac{1}{m}+\frac{1}{n}}}, \quad U_i = \frac{\hat{\theta}_i - \hat{\theta}_{k+1}}{s\sqrt{\frac{1}{m}+\frac{1}{n}}} \quad \text{for } i = 1,\ldots,k. \tag{3}$$

In addition, we let $\eta = s/\sigma$ and

$$Z_i = \frac{\hat{\theta}_i - \theta_i}{\sigma/\sqrt{n_i}} \quad \text{for } i = 0,\ldots,k+1. \tag{4}$$

The hypothesis (2) is composite and to control overall Type I error strongly we need to calculate our critical points under the least favorable configuration. Let us consider all the configurations that are possible under the null:

1. All $k$ test treatment means are between the positive and negative control means.
2. Some of the test treatment means are equal to the negative control mean, some of the test treatment means are between the negative and positive controls means, and the rest of the test treatment means are equal to the positive control mean.
3. Some of the test treatment means are equal to the negative control mean and the other test treatment means are equal to the positive control mean.

## 4. Least favorable configurations and critical points

Our rejection criterion is: Reject $H_0$ if

$$T_i < -t \quad \text{or} \quad U_i > t \quad \text{for at least one } i,\ i = 1,\ldots,k. \tag{5}$$

As a result we define our "acceptance probability"

$$1 - P(\text{Type I error}) = P(T_i \ge -t \text{ and } U_i \le t \text{ for all } i,\ i = 1,\ldots,k). \tag{6}$$

With (3) the right hand side of (6) reduces to

$$P\left(\hat{\theta}_0 - ts\sqrt{\frac{1}{m}+\frac{1}{n}} \le \hat{\theta}_i \le \hat{\theta}_{k+1} + ts\sqrt{\frac{1}{m}+\frac{1}{n}},\ i = 1,\ldots,k\right). \tag{7}$$

As in Dasgupta et al. (2006), for ease of notation, we define

$$B_0 = t\eta\sqrt{1+n/m},$$

$$B_1 = \sqrt{n/m},$$

$$C = \sqrt{n}(\theta_{k+1}-\theta_0)/\sigma,$$

$$\delta_i = (\theta_i-\theta_0)/(\theta_{k+1}-\theta_0), \quad i=1,\ldots,k.$$

Using (3) and (4) and the notation defined above (7) reduces to

$$P(-B_0+B_1Z_0-\delta_iC \le Z_i \le B_0+B_1Z_{k+1}+(1-\delta_i)C,\ i=1,\ldots,k). \tag{8}$$

Finally, let $p(\delta_i,z_0,z_{k+1},\eta)=\Phi(B_0+B_1z_{k+1}+(1-\delta_i)C)-\Phi(-B_0+B_1z_0-\delta_iC)$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Then (8) reduces to

$$G(\boldsymbol{\delta}) = \int_0^\infty \int\int_\Omega \prod_{i=1}^{k} p(\delta_i,z_0,z_{k+1},\eta)\cdot\phi(z_0)\phi(z_{k+1})f(\eta)\,dz_0\,dz_{k+1}\,d\eta, \tag{9}$$

where $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_k)$, $\phi(\cdot)$ is the p.d.f. of the standard normal distribution, $f(\cdot)$ is the p.d.f. of the chi distribution with $v$ degrees of freedom, and $\Omega=\{(z_0,z_{k+1},\eta): z_0-z_{k+1}\le(2B_0+C)/B_1\}$.

To find the LFC we need to minimize $G(\boldsymbol{\delta})$ in terms of $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_k)$. At first we need to clarify that we are discounting the possibility of having "$T_i<0$ and $U_i>0$ for at least one $i$". This is a consequence of Assumption (2).

$$\begin{aligned}
P(T_i<0 \text{ and } U_i>0 \text{ for at least one } i,\ i=1,\ldots,k) &= 1-P(\hat\theta_i>\hat\theta_0 \text{ or } \hat\theta_i<\hat\theta_{k+1} \text{ for all } i)\\
&= 1-[P(\hat\theta_0-\hat\theta_{k+1}\le 0)+P(\hat\theta_i>\hat\theta_0 \text{ or } \hat\theta_i<\hat\theta_{k+1}\\
&\quad \text{for all } i|\hat\theta_0-\hat\theta_{k+1}>0)P(\hat\theta_0-\hat\theta_{k+1}>0)]\\
&= P(\hat\theta_0-\hat\theta_{k+1}>0)[1-P(\hat\theta_i>\hat\theta_0 \text{ or } \hat\theta_i<\hat\theta_{k+1} \text{ for all } i|\hat\theta_0-\hat\theta_{k+1}>0)]\\
&\le P(\hat\theta_0-\hat\theta_{k+1}>0)=1-\Phi\left(\frac{\theta_{k+1}-\theta_0}{\sigma\sqrt{2/m}}\right).
\end{aligned} \tag{10}$$

For any given $k$, under Assumption (2) when $\theta_{k+1}-\theta_0\ge 3\sigma$,

$$P(T_i<0 \text{ and } U_i>0 \text{ for at least one } i,\ i=1,\ldots,k|\theta_{k+1}-\theta_0\ge 3\sigma)\le 1-\Phi\left(3\sqrt{\frac{m}{2}}\right)\le 0.001.$$

The above derivation implies that under Assumption (2), the likelihood that a treatment will result in a negative $T_i$ and a positive $U_i$ for at least one treatment will be very small and so is negligible. To verify our analytical result, we first set up some definitions and present two lemmas.

*Log-concavity*: A non-negative real-valued function $f$ on $\mathscr{R}^k$ is called log-concave, if, for each $\mathbf{x}_1,\mathbf{x}_2\in\mathscr{R}^k$ and every $\varepsilon\in(0,1)$, we have $f(\varepsilon\mathbf{x}_1+(1-\varepsilon)\mathbf{x}_2)\ge f^{\varepsilon}(\mathbf{x}_1)f^{1-\varepsilon}(\mathbf{x}_2)$.

*Convexity*: Suppose that $f$ is twice-differentiable function in the convex domain $\mathscr{D}$. Then $f$ is a convex function if and only if the Hessian matrix

$$H=\left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i\,\partial x_j}\right)_{i,j=1,\ldots,n}$$

is positive semi-definite almost everywhere.

**Lemma 1.** *Define $h(x)=\Phi(x)+\phi(x)/x$. We have*

(a) *$h(x)$ is a decreasing function of $x$ in $(-\infty,0)$ and $(0,\infty)$;*
(b) *for any $x<0$, $h(x)<0$.*

**Proof.** For any $x\neq 0$,

$$h'(x)=\phi(x)-\frac{\phi(x)}{x^2}-\phi(x)=-\frac{\phi(x)}{x^2}<0.$$

This implies that $h(x)$ is a decreasing function of $x$ in $(-\infty,0)$ and $(0,\infty)$. Note that $h(-\infty)=0$, therefore, $h(x)<0$ for any $x<0$. □

**Lemma 2.** *$p(\mathbf{x})=\Phi(x_1-x_3+x_4)-\Phi(x_2-x_3-x_4)$ is a log-concave function of $\mathbf{x}\in\mathscr{D}$, where $\mathscr{D}$ is the convex domain $\mathscr{D}=\{(x_1,x_2,x_3,x_4)\in\mathscr{R}^4 : x_1+x_4\geq x_2-x_4\}$.*

**Proof.** To show $p(\mathbf{x})$ is log-concave, it suffices to show that $g(\mathbf{x})=-\log p(\mathbf{x})$ is a convex function of $\mathbf{x}\in\mathscr{D}$, where "log" denotes the natural logarithm. The Hessian matrix is given as

$$H(g)=\left(\frac{\partial^2 g(\mathbf{x})}{\partial x_i\,\partial x_j}\right)_{i,j=1,\ldots,4},$$

where $\partial^2 g(\mathbf{x})/\partial x_i\,\partial x_j=\partial^2 g(\mathbf{x})/\partial x_j\,\partial x_i$ for any $i\neq j$. For ease of notation, we denote $\phi_1=\phi(x_1-x_3+x_4)$, $\phi_2=\phi(x_2-x_3-x_4)$, $\Phi_1=\Phi(x_1-x_3+x_4)$ and $\Phi_2=\Phi(x_2-x_3-x_4)$.

Let $H_i$ be the $i$ th principal minor of $H(g)$ formed by the first $i$ rows and columns of $H(g)$. Noting that $H(g)$ is a symmetric matrix, to show $H(g)$ is positive semi-definite, it suffices to show that the determinants, $\det(H_i)$, $i=1,\ldots,4$, are all nonnegative. Simple calculations lead to

$$\begin{aligned}
\det(H_1)&=\frac{(x_1-x_3+x_4)\phi_1(\Phi_1-\Phi_2)+\phi_1^2}{(\Phi_1-\Phi_2)^2},\\
\det(H_2)&=\frac{\phi_1\phi_2[(x_1-x_3+x_4)\phi_2-(x_2-x_3-x_4)\phi_1-(x_1-x_3+x_4)(x_2-x_3-x_4)(\Phi_1-\Phi_2)]}{(\Phi_1-\Phi_2)^3},\\
\det(H_3)&=0,\\
\det(H_4)&=0.
\end{aligned}$$

To show $\det(H_1)\geq 0$, it suffices to show that

$$(x_1-x_3+x_4)(\Phi(x_1-x_3+x_4)-\Phi(x_2-x_3-x_4))+\phi(x_1-x_3+x_4)\geq 0. \tag{11}$$

Noting that $\Phi(x_1-x_3+x_4)-\Phi(x_2-x_3-x_4)\geq 0$ for any $\mathbf{x}\in\mathscr{D}$, (11) holds when $x_1-x_3+x_4\geq 0$. While for $x_1-x_3+x_4<0$, by Lemma 1 we have

$$\begin{aligned}
(x_1-x_3+x_4)(\Phi(x_1-x_3+x_4)-\Phi(x_2-x_3-x_4))+\phi(x_1-x_3+x_4)&\geq(x_1-x_3+x_4)\Phi(x_1-x_3+x_4)+\phi(x_1-x_3+x_4)\\
&=(x_1-x_3+x_4)\left(\Phi(x_1-x_3+x_4)+\frac{\phi(x_1-x_3+x_4)}{(x_1-x_3+x_4)}\right)\\
&\geq 0.
\end{aligned}$$

To show $\det(H_2)\geq 0$, it suffices to show that

$$(x_1-x_3+x_4)\phi_2-(x_2-x_3-x_4)\phi_1-(x_1-x_3+x_4)(x_2-x_3-x_4)(\Phi_1-\Phi_2)\geq 0. \tag{12}$$

Note that $\mathscr{D}=\{x_2-x_3-x_4\leq 0\leq x_1-x_3+x_4\}\cup\{(x_1-x_3+x_4)(x_2-x_3-x_4)>0\}$. It is clear that (12) holds when $x_2-x_3-x_4\leq 0\leq x_1-x_3+x_4$. Otherwise, to show (12) is equivalent to showing

$$\begin{aligned}
&\frac{\phi(x_2-x_3-x_4)}{(x_2-x_3-x_4)}-\frac{\phi(x_1-x_3+x_4)}{(x_1-x_3+x_4)}-(\Phi(x_1-x_3+x_4)-\Phi(x_2-x_3-x_4))\\
&\quad=\left(\Phi(x_2-x_3-x_4)+\frac{\phi(x_2-x_3-x_4)}{(x_2-x_3-x_4)}\right)-\left(\Phi(x_1-x_3+x_4)+\frac{\phi(x_1-x_3+x_4)}{(x_1-x_3+x_4)}\right)\geq 0,
\end{aligned}$$

which is guaranteed by noting that $h(x)=\Phi(x)+\phi(x)/x$ is a strictly decreasing function of $x$ in $(-\infty,0)$ and $(0,\infty)$. □

**Theorem 1.** *The least favorable configuration of the null hypothesis is given under Configuration* 3.

**Proof.** We first show that $G(\boldsymbol{\delta})$ is log-concave on $\boldsymbol{\delta}\in[0,1]^k$. By Theorem 2.16 of Dharmadhikari and Joag-dev (1988) it suffices to show that the density

$$g(\boldsymbol{\delta},z_0,z_{k+1},\eta)=c\prod_{i=1}^{k}p(\delta_i,z_0,z_{k+1},\eta)\cdot\phi(z_0)\phi(z_{k+1})f(\eta)$$

is a log-concave density of $(\boldsymbol{\delta},z_0,z_{k+1},\eta)\in[0,1]^k\times\Omega$, where $c$ is a constant to validate the density function $g(\boldsymbol{\delta},z_0,z_{k+1},\eta)$. By Lemma 2, it is easy to see that $p(\delta_i,z_0,z_{k+1},\eta)$ is log-concave function of $(\delta_i,z_0,z_{k+1},\eta)\in[0,1]\times\Omega$, where $X_1=B_1z_{k+1}+C$, $X_2=B_1z_0$, $X_3=\delta_iC$ and $X_4=B_0$. Therefore, noting that both the normal distribution and the chi distribution are log-concave (see Table 1 of Bagnoli and Bergstrom, 2005), by Theorem 2.7 of Dharmadhikari and Joag-dev (1988) we have $g(\boldsymbol{\delta},z_0,z_{k+1},\eta)$ is log-concave.

Now since $G(\boldsymbol{\delta})$ is log-concave on $\boldsymbol{\delta} \in [0,1]^k$, for any $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ in $[0,1]^k$ and for all $\varepsilon \in (0,1)$, we have

$$G(\varepsilon\boldsymbol{\delta}_1 + (1-\varepsilon)\boldsymbol{\delta}_2) \geq (G(\boldsymbol{\delta}_1))^{\varepsilon}(G(\boldsymbol{\delta}_2))^{1-\varepsilon} \geq (\min(G(\boldsymbol{\delta}_1), G(\boldsymbol{\delta}_2)))^{\varepsilon}(\min(G(\boldsymbol{\delta}_1), G(\boldsymbol{\delta}_2)))^{1-\varepsilon} = \min(G(\boldsymbol{\delta}_1), G(\boldsymbol{\delta}_2)).$$

This indicates that $G(\boldsymbol{\delta})$ is minimized under Configuration 3. Or equivalently, the Type I error, $1 - G(\boldsymbol{\delta})$, is maximized under Configuration 3.

Theorem 1 implies that some (say $r$) of the test treatment means are equal to the negative control and the other $k - r$ test treatment means are equal to the positive control. To find the critical points it is thus necessary to find the optimal value of $r$, where $r = 0, \ldots, k$. □

**Theorem 2.** *The Type I error of the null hypothesis is minimized under $r = k/2$ when $k$ is even and $(k+1)/2$ when $k$ is odd.*

**Proof.** Under Configuration 3 we define our "acceptance probability" as a function of $r$, $G(r)$. We further define

$$V_1 = B_0 + B_1 z_{k+1} + C, \quad V_2 = -B_0 + B_1 z_0, \quad V_3 = B_0 + B_1 z_{k+1} = V_1 - C, \quad V_4 = -B_0 + B_1 z_0 - C = V_2 - C \tag{13}$$

and

$$p_1 = [\Phi(V_1) - \Phi(V_2)], \quad p_2 = [\Phi(V_3) - \Phi(V_4)]. \tag{14}$$

Now,

$$G(r) = \int_0^{\infty} \int \int_{\Omega} [p_1]^r [p_2]^{k-r} \cdot \phi(z_0)\phi(z_{k+1}) f(\eta)\, dz_0\, dz_{k+1}\, d\eta. \tag{15}$$

By symmetry it is easy to verify that $G(r) = G(k-r)$, hence, $G(r) = \frac{1}{2}[G(r) + G(k-r)]$. The integrand of $G(r)$ can be written as

$$[\Phi(V_1) - \Phi(V_2)]^{k/2} \cdot [\Phi(V_3) - \Phi(V_4)]^{k/2} \cdot \left\{ \left[ \frac{\Phi(V_1) - \Phi(V_2)}{\Phi(V_3) - \Phi(V_4)} \right]^{(k/2)-r} + \left[ \frac{\Phi(V_3) - \Phi(V_4)}{\Phi(V_1) - \Phi(V_2)} \right]^{(k/2)-r} \right\} \tag{16}$$

which is of the form:

$$U(y, w) = \kappa \cdot (y^w + y^{-w}), \tag{17}$$

where

$$y = \frac{[\Phi(V_1) - \Phi(V_2)]}{[\Phi(V_3) - \Phi(V_4)]}, \quad w = \frac{k}{2} - r \quad \text{and} \quad \kappa = [[\Phi(V_1) - \Phi(V_2)] \cdot [\Phi(V_3) - \Phi(V_4)]]^{k/2}.$$

We note that $\kappa$ is a constant with respect to $w$. First we need to show, $\kappa \geq 0$ and $y \geq 0$. From (13) we can write,

$$\kappa = [[\Phi(V_1) - \Phi(V_2)] \cdot [\Phi(V_1 - C) - \Phi(V_2 - C)]]^{k/2}. \tag{18}$$

It is easy to show $\kappa \geq 0$, specifically note that $[\Phi(V_1) - \Phi(V_2)]$ and $[\Phi(V_1 - C) - \Phi(V_2 - C)]$ are both positive or both negative. Therefore the product is always non-negative. To show $y \geq 0$, we note that the numerator and denominator of $y$ are the two terms in $\kappa$, and the result follows. Now

$$\frac{\partial U}{\partial w} = \kappa \ln(y)(y^w - y^{-w}). \tag{19}$$

Now, if $w > 0$, then $\partial U/\partial w \geq 0$ and if $w < 0$, then $\partial U/\partial w \leq 0, y \geq 0, w \in \mathscr{R}$. So for any $y \geq 0$, $U(y, w)$ is non-increasing as $|w|$ decreases. Hence, $G(r)$ is non-increasing in as $|r - k/2|$ decreases. Therefore the least favorable configuration is the middle value of the treatments: $r = k/2$ when $k$ is even or $r = (k+1)/2$ when $k$ is odd.

We establish analytically the LFC found numerically in Dasgupta et al. (2006) which implies that our critical points will be the same as theirs. To find the critical points we need to equate Eq. (15) to $1 - \alpha$ and then solve for $t_\alpha$, where $r = k/2$ for even $k$ and $r = (k+1)/2$ for odd $k$. Recall by Assumption (2), $C = (\theta_{k+1} - \theta_0)/\sigma > 3$. For conservative estimates $C = 3$ should be used. A table of critical points appears in Dasgupta et al. (2006). Other critical values and the program finding critical values are available from the authors upon request. □

*Example revisited*: In the allergy example (see Table 1), we were interested in comparing the four test treatments to both the positive control (d farinae) and the negative control (saline). Here the pooled standard deviation is 5.188. The computed critical point for $k=4, m=n=23$ is $t_\alpha = 2.230$. The corresponding test statistics are $T_1 = 1.307, T_2 = 1.335, T_3 = 0.824, T_4 = 6.39, U_1 = -11.82$, $U_2 = -11.79, U_3 = -12.30$ and $U_4 = -6.73$. Thus, $T_i > -t$ and $U_i < t$ for $i = 1, 2, 3, 4$ and we do not reject the null hypothesis.

## 5. Simulation results

We now conduct a simulation study to show the method proposed by Dasgupta et al. (2006) which we analytically examine, is an improvement over its single-step competitors (Dunnett's method and the Bonferroni correction) in terms of the Type I error and the power. For comparing the Type I error, we assume negative control mean, $\theta_0 = 1$, positive control mean $\theta_{k+1} = 5$ and $\sigma = 1$. Treatments $i = 1, \ldots, (k+1)/2$ (or $k/2$) were generated with a mean of 1 and the remaining treatments were generated with a mean of 5. We calculated our $T_i$ and $U_i$ and found the number of cases with at least one rejection. The proportion of times there was at least one rejection is given as the first lines (corresponding to $\gamma = 0$) in Table 2.

We use Spurrier's (1992) definition of marginal power and define the marginal power for treatment $i$ as the probability of declaring it smaller (larger) than the negative (positive) control. Without loss of generality, let us assume we are interested in the marginal power of Treatment 1. Like the simulation for the Type I error, other than Treatment 1, all treatments were generated with a mean of $\theta_0$ or $\theta_{k+1}$. Treatment 1 was generated with a mean that equaled $\theta_0 - \gamma$ (for negative marginal power) or $\theta_{k+1} + \gamma$ (for positive marginal power), where the effect size $\gamma$ is ranging from 0.4 to 2. The simulation was repeated 10,000 times. The nominal error rate was kept at 0.05. It is to be noted that the standard error of simulation was 0.002 for the computation of the Type I error. For sample size consideration we noted that minimizing the sum of variances of all elementary contrast yields $m/n = \sqrt{k/2}$. In Table 2 we consider two different numbers of treatments ($k = 3, 6$) and three different combinations of sample sizes with $m/n = \sqrt{k/2}$, with $n = 5$, 10, and 17. Results for equal sample sizes ($m = n$) and other values of $k$ do not show any appreciable departure from these findings and are not included for space considerations. These are available from the authors. It must be mentioned that Dasgupta et al. (2006) used simulation to verify the LFC in their numerical technique but did not perform a simulation for comparison with competitors.

The simulation indicates that the proposed method controls the Type I error at levels close to the nominal level for finite sample sizes. Dunnett's method, tailor made for comparing $k$ treatments to a single control, is somewhat liberal when being used for two controls. The Bonferroni correction works reasonably well when the number of test treatments is small ($k = 3$). However, as the number of test treatments increases, the Bonferroni correction becomes conservative.

**Table 2**
Type I error and marginal power comparison.

| $k$ | $n$ | $m$ | $\gamma$ | Negative ours | Negative Dunnett | Negative Bonferroni | Positive ours | Positive Dunnett | Positive Bonferroni |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 5 | 6 | 0.0 | 0.0500 | 0.0046 | 0.0558 | | | |
| | | | 0.4 | 0.0693 | 0.0642 | 0.0772 | 0.0737 | 0.0695 | 0.0804 |
| | | | 0.8 | 0.1969 | 0.1880 | 0.2124 | 0.2041 | 0.1936 | 0.2185 |
| | | | 1.2 | 0.4164 | 0.4041 | 0.4378 | 0.4206 | 0.3992 | 0.4319 |
| | | | 1.6 | 0.6558 | 0.6432 | 0.6741 | 0.6466 | 0.6426 | 0.6727 |
| | | | 2.0 | 0.8506 | 0.8406 | 0.8662 | 0.8489 | 0.8416 | 0.8626 |
| | 10 | 12 | 0.0 | 0.0509 | 0.0438 | 0.0540 | | | |
| | | | 0.4 | 0.1207 | 0.1120 | 0.1311 | 0.1224 | 0.1129 | 0.1320 |
| | | | 0.8 | 0.3920 | 0.3757 | 0.4109 | 0.3883 | 0.3704 | 0.4056 |
| | | | 1.2 | 0.7384 | 0.7207 | 0.7531 | 0.7453 | 0.7294 | 0.7588 |
| | | | 1.6 | 0.9358 | 0.9295 | 0.9422 | 0.9419 | 0.9357 | 0.9486 |
| | | | 2.0 | 0.9928 | 0.9923 | 0.9943 | 0.9935 | 0.9931 | 0.9941 |
| | 17 | 21 | 0.0 | 0.0508 | 0.0460 | 0.0558 | | | |
| | | | 0.4 | 0.1886 | 0.1774 | 0.1996 | 0.1874 | 0.1764 | 0.1995 |
| | | | 0.8 | 0.6209 | 0.6041 | 0.6394 | 0.6249 | 0.6098 | 0.6412 |
| | | | 1.2 | 0.9351 | 0.9301 | 0.9416 | 0.9336 | 0.9282 | 0.9388 |
| | | | 1.6 | 0.9967 | 0.9965 | 0.9971 | 0.9964 | 0.9961 | 0.9967 |
| | | | 2.0 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 6 | 5 | 9 | 0.0 | 0.0486 | 0.0453 | 0.0586 | | | |
| | | | 0.4 | 0.0494 | 0.0449 | 0.0568 | 0.0507 | 0.0470 | 0.0596 |
| | | | 0.8 | 0.1675 | 0.1585 | 0.1878 | 0.1698 | 0.1609 | 0.1906 |
| | | | 1.2 | 0.3909 | 0.3753 | 0.4193 | 0.3902 | 0.3789 | 0.4262 |
| | | | 1.6 | 0.6587 | 0.6454 | 0.6901 | 0.6588 | 0.6430 | 0.6873 |
| | | | 2.0 | 0.8674 | 0.8591 | 0.8845 | 0.8626 | 0.8540 | 0.8822 |
| | 10 | 17 | 0.0 | 0.0501 | 0.0458 | 0.0590 | | | |
| | | | 0.4 | 0.0822 | 0.0768 | 0.0984 | 0.0863 | 0.0805 | 0.1004 |
| | | | 0.8 | 0.3457 | 0.3343 | 0.3760 | 0.3435 | 0.3303 | 0.3747 |
| | | | 1.2 | 0.7315 | 0.7483 | 0.7570 | 0.7233 | 0.7108 | 0.7470 |
| | | | 1.6 | 0.9453 | 0.9410 | 0.9540 | 0.9450 | 0.9402 | 0.9542 |
| | | | 2.0 | 0.9942 | 0.9938 | 0.9956 | 0.9946 | 0.9875 | 0.9958 |
| | 17 | 29 | 0.0 | 0.0511 | 0.0473 | 0.0607 | | | |
| | | | 0.4 | 0.1404 | 0.1326 | 0.1565 | 0.1499 | 0.1428 | 0.1678 |
| | | | 0.8 | 0.5872 | 0.5744 | 0.6121 | 0.5983 | 0.5855 | 0.6271 |
| | | | 1.2 | 0.9976 | 0.9973 | 0.9982 | 0.9366 | 0.9317 | 0.9448 |
| | | | 1.6 | 1.0000 | 1.0000 | 1.0000 | 0.9980 | 0.9976 | 0.9983 |
| | | | 2.0 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

In terms of power, the same trend is evident. Using Dunnett's method would result in higher power, but as noted in the previous paragraph, the Type I error is not controlled at the nominal level for this procedure and so this alternative should be discounted. The Bonferroni method, which is conservative, has lower power. Due to the symmetric nature of the procedure, the negative marginal power and the positive marginal power are about the same for the same $\gamma$, $n$ and $k$. The power calculation also shows that our procedure is quite sensitive to departures from the null. For example, for a departure of $0.4\sigma$, for comparing three treatments to the positive and negative controls (Table 2), for a sample size of $n=5$, $m=6$, we have a marginal power of 7 percent. For a sample size of $n=17$, $m=21$ this power increases to 19 percent. The power depends upon the number of treatments $k$, the sample size $n$, and the degree of departure $\gamma$. The marginal power increases with $n$ and $\gamma$ and decreases with $k$.

It must be noted that the data in Table 2 was generated under Assumption (2) that $\theta_{k+1}-\theta_0>3\sigma$. We looked at several situations, varying the distance between the two controls. If Assumption (2) is satisfied our proposed method is superior. However, when the distance between the controls is less than $3\sigma$, our method can be conservative. In that case, the set "$T_i<0$ and $U_i>0$ for at least one $i$" is not negligible anymore. Specifically, if we ignore this set, we reject less often than we should. When the distance between the two controls is 0, the ideal critical value is the Dunnett's two-sided method. Exact power calculations can be done for our method, and therefore, the simulation here is only for the purpose of comparison of the three methods.

## 6. Discussion and continuing work

In this manuscript, we analytically prove a result that was established numerically in the literature. In terms of the analytical result it is interesting to note that $P(\hat{\theta}_i>\hat{\theta}_0 \text{ or } \hat{\theta}_i<\hat{\theta}_{k+1} \text{ for all } i|\hat{\theta}_0-\hat{\theta}_{k+1}>0)$ is a decreasing function of $k$, which reduces to zero when $k$ is sufficiently large. While for small $k$, this conditional probability can be large. Another point we would like to make to clarify our procedure is that our test statistics $T_i$ and $U_i$ are not linearly independent. For example, for $k=2$,

$$U_1-U_2+T_2=\frac{\hat{\theta}_1-\hat{\theta}_{k+1}}{s\sqrt{\frac{1}{m}+\frac{1}{n}}}-\frac{\hat{\theta}_2-\hat{\theta}_{k+1}}{s\sqrt{\frac{1}{m}+\frac{1}{n}}}+\frac{\hat{\theta}_2-\hat{\theta}_0}{s\sqrt{\frac{1}{m}+\frac{1}{n}}}=T_1.$$

This implies that the covariance matrix of $\{T_i,U_i,i=1,\ldots,k\}$ is singular. This makes the standard logconcavity results of the non-central multivariate $t$ distribution that require non-singularity inapplicable, see for example Lemma 3.2 of Giani and Strassburger (2000).

We show via Monte Carlo simulation that the proposed method has an advantage over both Dunnett's method and the Bonferroni correction. Two of the configurations ($r=0,k$) lead us to Dunnett's critical points. However, since these are not the least favorable configurations, the overall Type I error is not maintained. The Bonferroni method, which would be a logical alternative, is quite conservative as the number of test treatments becomes large. This research is based on equal sample sizes for the test treatments and equal sample sizes for both positive and negative controls. The next step would be to investigate the problem when the sample sizes are different. This would make finding the least favorable configuration a more challenging problem. Investigating optimal design issues with both a positive and a negative control is another interesting open question. There is a fairly large body of literature on optimal design when there is a control group and a treatment group (see for example, Hedayat et al., 1988; Jacroux, 2001, 2002). The question of two controls (positive and negative or equal or varying importance) was studied by Dasgupta and SahaRay (2007) in terms of A and MV-optimality. There are various open problems in optimal design stemming from this problem. Due to the composite nature of the null, there is some ambiguity in inverting the test to form one-sided confidence intervals. It may be possible to derive confidence sets for the problem at hand in a stepwise fashion, as in Hayter and Hsu (1994). Another immediate problem stemming from our example is the problem of two controls in block design/repeated measures. Hence, we believe that this manuscript opens up possibilities for various directions of new research.

## Acknowledgments

The authors thank the editor, the associate editor and one reviewer for their constructive comments and suggestions that have led to a substantial improvement in the article.

## References

Bagnoli, M., Bergstrom, T., 2005. Log-concave probability and its applications. Economic Theory 26, 445–469.
Bauer, P., Rohmel, J., Maurer, W., Hothorn, L., 1998. Testing strategies in multi-dose experiments including active control. Statistics in Medicine 17, 2133–2146.
D'Agostino, R.B., Hareen, T.C., 1991. Multiple comparisons in over-the-counter drug clinical trials with both positive and placebo controls. Statistics in Medicine 10, 1–6.
Dasgupta, N., Solorzano, E., Lazar, N.A., 2006. Using numerical methods to find the least favorable configuration when comparing k test treatments to both positive and negative controls. Journal of Statistical Computation and Simulations 76, 251–265.
Dasgupta, N., SahaRay, R., 2007. Optimal allocation for comparing $k$ test treatments to positive and negative control with unequal weighting under A-optimality and MV-optimality. Metrika 65, 83–92.
Dharmadhikari, S., Joag-dev, K., 1988. Unimodality, Convexity and Applications. Academic Press, Boston.
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., Tamhane, A.C., 1992. Comparisons between a new drug and active and placebo controls in an efficacy clinical trial. Statistics in Medicine 11, 1057–1063.

Giani, G., Strassburger, K., 2000. Multiple comparison procedures for optimally discriminating between good, equivalent, and bad treatments with respect to a control. Journal of Statistical Planning and Inference 83, 413–440.
Gupta, V.K., Ramana, D.V.V., Prasad, R., 2002. Weighted A-optimal block designs for comparing treatments with controls with unequal precision. Journal of Statistical Planning and Inference 106, 159–175.
Hedayat, A.S., Jacroux, M., Majumdar, D., 1988. Optimal designs for comparing test treatments with controls. Statistical Science 3, 462–476.
Hayter, A.J, Hsu, J.C., 1994. On the relationship between step-wise decision procedures and confidence sets. Journal of the American Statistical Association 89, 778–785.
Hoover, D.R., 1991. Simultaneous comparisons of multiple treatments to two controls. Biometrical Journal 33, 913–921.
Hothrorn, L.A., Hayashi, M., Seidel, D., 2000. Dose–response relationships in mutagenicity assays including an appropriate positive control group: a multiple testing approach. Environmental and Ecological Statistics 7, 27–42.
Jacroux, M., 2001. A and MV-efficient block designs for comparing a set of controls to a set of test treatments. Sankhyä 64, 141–161.
Jacroux, M., 2002. Determination and construction of A-optional designs for comparing two sets of treatments. Sankhyä 63, 351–361.
Shaffer, J.P., 1977. Multiple comparisons emphasizing selected contrasts: an extension and generalization of Dunnett's procedure. Biometrics 33, 293–303.
Solorzano, E., Spurrier, J.D., 1999. One-sided simultaneous comparisons with more than one control. Journal of Statistical Computation and Simulation 63, 37–46.
Spurrier, J.D., 1992. Optimal designs for comparing variances of several treatments with that of a standard treatment. Technometrics 34, 332–339.