A Superiority-Equivalence Approach to One-Sided Tests on Multiple Endpoints in Clinical Trials

Ajit C. Tamhane	and	Brent R. Logan
Department of Statistics		Division of Biostatistics
Northwestern University		Medical College of Wisconsin
2006 Sheridan Road		8701 Watertown Plank Rd.
Evanston, IL 60208, USA		Milwaukee, WI 53226, USA
E-mail: ajit@iems.northwestern.edu		E-mail: blogan@mcw.edu

SUMMARY. This paper gives a new formulation of the one-sided multivariate testing problem. This formulation (i) is practically relevant in the context of comparing a new treatment with a control on multiple endpoints, (ii) avoids the anomalies associated with the likelihood ratio test that uses the traditional null hypothesis formulation, and (iii) requires very little multiplicity adjustment. The hypotheses are formulated with the goal of showing that the treatment is equivalent (not inferior) on all endpoints and superior on at least one endpoint compared to the control, where thresholds for equivalence and superiority are specified for each endpoint. The union-intersection (Roy 1953) and intersection-union (Berger 1982) principles are employed to derive the basic test. It is shown that the critical constants required by this test can be sharpened by a careful analysis of its size. The test is illustrated by an example. Some extensions of the test are mentioned.

KEY WORDS: Multivariate one-sided test; union-intersection principle; intersection-union principle

1. Introduction

Many clinical trials are conducted to compare a treatment group with a control group

on multiple endpoints. Often, the treatment is expected to have a positive effect on all endpoints. If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ denotes the vector of mean differences between the treatment group and the control group on $m \geq 2$ endpoints then the hypotheses are usually formulated as

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \text{ vs. } H_1 : \boldsymbol{\theta} \in \mathcal{O}^+, \tag{1.1}$$

where $\mathbf{0}$ is the null vector and $\mathcal{O}^+ = \{\boldsymbol{\theta} | \theta_k \geq 0 \text{ for } 1 \leq k \leq m, \boldsymbol{\theta} \neq \mathbf{0}\}$ is the positive orthant. Kudô (1963) derived an exact likelihood ratio (LR) test for this testing problem assuming multivariate normality of the data and a common, known covariance matrix. Perlman (1969) extended this test to the unknown covariance matrix case.

Silvapulle (1997) showed, by means of an example, that the rejection region of the LR test is nonmonotone in that the outcomes that lie deeper in the complement of the alternative hypothesis region can produce more significant test statistics. Figure 1 shows the rejection region of an LR test for two endpoints with correlation 0.9. Note that the rejection region contains points in the negative quadrant. Perlman and Wu (2002) showed that these undesirable properties are the result of the use of a point null hypothesis that the mean vector is a null vector; if the complete complement of the positive orthant is used as the null hypothesis, then the rejection region of the LR test does not have the abovementioned undesirable properties. In particular, the rejection region is completely contained in the alternative hypothesis region.

Cohen and Sackrowitz (1998) proposed the so-called cone-ordered monotone (COM) tests to overcome the anomalies of the classical LR tests. The rejection region of a COM test is shown in Figure 2. As can be seen, this rejection region also includes points that have large negative coordinates. Therefore the COM test also will not be acceptable in practice.

In this paper we propose an alternative approach, which we believe will be appealing to clinical researchers. When comparing a treatment group with a control group, a researcher desires to show that the treatment is equivalent (not inferior) to the control on all end-

points and is superior on at least one endpoint, where the equivalency and superiority for each endpoint is specified in terms of nonnegative threshold values. We formulate this approach and analyze the test procedure obtained by applying the union-intersection (UI) and intersection-union (IU) principles of test construction due to Roy (1953) and Berger (1982), respectively.

This approach has at least three benefits: (i) It gives the researcher a more straightforward formulation of the decision rule for the one-sided multiple endpoint testing situation. (ii) It avoids the anomalies associated with the likelihood ratio test that uses the traditional null hypothesis formulation. (iii) It results in little to no multiplicity adjustment in many practical situations.

2. Preliminaries and Notation

Consider a treatment group (group 1) and a control group (group 2) with n_1 and n_2 patients. Suppose that on each patient $m \geq 2$ endpoints are measured. Denote the random data vectors from group i by $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijm}), \quad i = 1, 2; j = 1, 2, \ldots, n_i$. We assume that the $\boldsymbol{X}_{ij}$ are independent and identically distributed (i.i.d.) random vectors from an m-variate normal distribution with mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})$ and a common covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{k\ell}\}$ with $\sigma_{kk} = \sigma_k^2 = \mathrm{Var}(X_{ijk})$ and $\sigma_{k\ell} = \mathrm{Cov}(X_{ijk}, X_{ij\ell})$ for $k \neq \ell$. Denote the correlation matrix by $\boldsymbol{R}$ with off-diagonal entries $\rho_{k\ell} = \mathrm{Corr}(X_{ijk}, X_{ij\ell}) = \sigma_{k\ell}/\sigma_k\sigma_\ell$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ be the vector of mean differences between the treatment and the control group.

For the kth endpoint, let $\delta_k \geq 0$ be a specified threshold for superiority and let $\epsilon_k \geq 0$ be a specified threshold for equivalency (non-inferiority), i.e., the treatment is regarded as superior to the control on the kth endpoint if $\theta_k > \delta_k$ and equivalent (non-inferior) to the control if $\theta_k > -\epsilon_k$. The hypotheses for showing the superiority and equivalency of the

treatment on the kth endpoint are as follows:

$$H_{0k}^{(S)} : \theta_k \leq \delta_k \text{ vs. } H_{1k}^{(S)} : \theta_k > \delta_k \ \ (1 \leq k \leq m)$$

and

$$H_{0k}^{(E)} : \theta_k \leq -\epsilon_k \text{ vs. } H_{1k}^{(E)} : \theta_k > -\epsilon_k \ \ (1 \leq k \leq m).$$

Let

$$H_0^{(S)} = \bigcap_{k=1}^{m} H_{0k}^{(S)}, H_1^{(S)} = \bigcup_{k=1}^{m} H_{1k}^{(S)}, H_0^{(E)} = \bigcup_{k=1}^{m} H_{0k}^{(E)} \text{ and } H_1^{(E)} = \bigcap_{k=1}^{m} H_{1k}^{(E)}.$$

Suppose the treatment is regarded as more effective than the control if it is superior on at least one endpoint and equivalent on all others. Then the hypotheses to be tested are:

$$H_0 = H_0^{(S)} \bigcup H_0^{(E)} \text{ vs. } H_1 = H_1^{(S)} \bigcap H_1^{(E)}. \tag{2.1}$$

It is desired to test H_0 at a preassigned level α. For $m = 2$, the regions of the parameter space corresponding to H_0 and H_1 are shown in Figure 3. A similar formulation was also recently proposed by Bloch, Lai and Tubert-Bitter (2000), who used $\delta_k = 0$ and $\epsilon_k > 0$ in line with common practice.

Note that (2.1) is a combination of union-intersection (UI) and intersection-union (IU) testing problems. If $\delta_k = \epsilon_k = 0$ for all k then $H_{0k}^{(S)} = H_{0k}^{(E)} = H_{0k}$ (say) and $H_{1k}^{(S)} = H_{1k}^{(E)} = H_{1k}$ (say). Thus the above hypothesis testing problem reduces to an intersection-union (IU) testing problem: Test

$$H_0 = \bigcup_{k=1}^{m} H_{0k} \text{ vs. } H_1 = \bigcap_{k=1}^{m} H_{1k}.$$

3. Simultaneous Confidence Intervals Approach

Let $\overline{X}_{1\cdot k}$ and $\overline{X}_{2\cdot k}$ be the sample means for the kth endpoint for group 1 and group 2, respectively. Further let $S_1^2, S_2^2, \ldots, S_m^2$ be the pooled sample variances based on $\nu = n_1 + n_2 - 2$ degrees of freedom (d.f.). We follow the usual convention of upper case letters for

random variables (r.v.'s) and the corresponding lower case letters for the observed values of those random variables.

The pivotal r.v. for θ_k is

$$T_k = \frac{(\overline{X}_{1\cdot k} - \overline{X}_{2\cdot k}) - \theta_k}{S_k\sqrt{1/n_1 + 1/n_2}} \ \ (1 \le k \le m). \tag{3.1}$$

Each T_k is marginally t-distributed with ν d.f. The joint distribution of $(T_1, T_2, \ldots, T_m)$ is the multivariate generalization of a bivariate t-distribution considered by Siddiqui (1967).

Since the joint distribution of $(T_1, T_2, \ldots, T_m)$ depends on the unknown correlation matrix $\boldsymbol{R}$, the exact critical constant is not available to compute simultaneous $100(1-\alpha)\%$ confidence intervals on the θ_k. Using the Bonferroni method, conservative lower one-sided confidence intervals are given by

$$\theta_k \ge L_k = \overline{x}_{1\cdot k} - \overline{x}_{2\cdot k} - t_{\nu,\alpha/m} s_k \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \ \ (1 \le k \le m), \tag{3.2}$$

where $t_{\nu,\alpha/m}$ is the upper α/m critical point of Student's t-distribution with ν d.f. If all $\rho_{k\ell} \ge 0$, which is common for multiple endpoints, then, based on Šidák's (1968) inequality, one could use the critical point t_{ν,α^*}, where $\alpha^* = 1 - (1-\alpha)^{1/m}$, in the above formula. This gives a slightly sharper critical constant, but the difference is not large and we will not distinguish between the two constants. We reject H_0 if all $L_k > -\epsilon_k$ and at least one $L_k > \delta_k$. Defining the t-statistics for testing the superiority and equivalence of the treatment on the kth endpoint by

$$t_k^{(S)} = \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k} - \delta_k}{s_k\sqrt{1/n_1 + 1/n_2}} \text{ and } t_k^{(E)} = \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k} + \epsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}} \ \ (1 \le k \le m), \tag{3.3}$$

we see that the above test is equivalent to

$$\min_{1\le k\le m} t_k^{(E)} > t_{\nu,\alpha/m} \text{ and } \max_{1\le k\le m} t_k^{(S)} > t_{\nu,\alpha/m}. \tag{3.4}$$

In fact, since all inferences follow from a single set of simultaneous confidence bounds (3.2), one can classify all the endpoints with regard to the equivalence/superiority of the treatment

as follows: on the kth endpoint the treatment is not equivalent (inferior) if $L_k \leq -\epsilon_k$, is equivalent but not superior if $-\epsilon_k < L_k \leq \delta_k$, and is superior if $L_k > \delta_k$ for $1 \leq k \leq m$, while strongly controlling the type I familywise error rate (Hochberg and Tamhane 1987) for the family $\left\{\left(H_{0k}^{(S)}, H_{0k}^{(E)}\right),\ k = 1, \ldots, m)\right\}$.

In the next section we show that the test (3.4) can be sharpened by applying the UI and IU principles of test construction.

4. A Test Based on Union-Intersection (UI) and Intersection-Union (IU) Principles

4.1 UI-IU Test

An α-level test of (2.1) derived by applying the UI and IU principles is as follows: Test $H_0^{(S)} = \bigcap_{k=1}^m H_{0k}^{(S)}$ and $H_0^{(E)} = \bigcup_{k=1}^m H_{0k}^{(E)}$ separately at level α, and reject H_0 if both are rejected. The UI test (Roy 1953) of $H_0^{(S)}$ rejects at level α using the Bonferroni approximation if $\max_{1\leq k\leq m} t_k^{(S)} > t_{\nu,\alpha/m}$. The IU test (Berger 1982) of $H_0^{(E)}$ rejects at level α if $\min_{1\leq k\leq m} t_k^{(E)} > t_{\nu,\alpha}$. Notice the smaller critical constant for the equivalency test compared to that used by the simultaneous confidence interval test (3.4).

This UI-IU test is somewhat conservative because it requires that the type I error probability be separately controlled for $H_0^{(E)}$ and $H_0^{(S)}$, which assumes the least favorable configuration (LFC) that one of the two hypotheses is true and the other is infinitely false. It is possible to have $H_0^{(E)}$ true and $H_0^{(S)}$ infinitely false, e.g., we can have $\theta_k = -\epsilon_k$ and $\theta_\ell \to \infty$ for $\ell \neq k$. In fact, this is the LFC for the IU test, which requires that each equivalency hypothesis, $H_{0k}^{(E)}$, be tested separately at level α. However, we cannot have $H_0^{(S)}$ true and $H_0^{(E)}$ infinitely false because if $\theta_k \leq \delta_k$ for all k then it cannot be simultaneously true that $\theta_k \to \infty$ for some k. This suggests that although the critical constant $t_{\nu,\alpha}$ for the IU test

of $H_0^{(E)}$ cannot be reduced, it may be possible to reduce the critical constant $t_{\nu,\alpha/m}$ for the UI test of $H_0^{(S)}$. From now on, we will use a general notation, c and d (with $d \geq c$), for the critical constants in place of $t_{\nu,\alpha}$ and $t_{\nu,\alpha/m}$, respectively. In the next section we investigate how to find the smallest possible values of c and d.

4.2 Sharpened Critical Constants for the UI-IU Test

4.2.1 The Type I Error Probability of the UI-IU Test and Its Supremum

Note that

$$t_k^{(S)} = t_k^{(E)} - \frac{\delta_k + \epsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}} \ (1 \leq k \leq m).$$

Using this relationship, the above UI-IU test can be written as

$$\min_{1\leq k\leq m}\left\{t_k^{(S)} + \frac{\delta_k + \epsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}}\right\} > c \text{ and } \max_{1\leq k\leq m} t_k^{(S)} > d. \tag{4.1}$$

For $m = 2$ and known σ_k, this rejection region is shown in Figure 4, where δ_k^* and ϵ_k^* are defined in (4.2).

To analytically determine the smallest possible values of c and d, we first obtain an expression for the type I error probability of the UI-IU test (4.1) in Lemma 1. Next we find its LFC in Lemma 2.

Lemma 1: Define

$$Z_k = \frac{\overline{X}_{1\cdot k} - \overline{X}_{2\cdot k} - \theta_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}} \sim N(0,1) \text{ and } U_k = \frac{S_k}{\sigma_k} \sim \sqrt{\frac{\chi_\nu^2}{\nu}} \ (1 \leq k \leq m),$$

so that $\boldsymbol{Z} = (Z_1, \ldots, Z_m)$ has an m-variate standard normal distribution with correlation matrix $\boldsymbol{R}$ independently of $\boldsymbol{U} = (U_1, \ldots, U_m)$. Denote the p.d.f.'s of $\boldsymbol{Z}$ and $\boldsymbol{U}$ by $\phi_m(\boldsymbol{z}|\boldsymbol{R})$ and $h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R})$, respectively. For $1 \leq k \leq m$, let

$$\delta_k^* = \frac{\delta_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}}, \epsilon_k^* = \frac{\epsilon_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}}, \theta_k^* = \frac{\theta_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}} \tag{4.2}$$

$$a_k = \theta_k^* + \epsilon_k^* \text{ and } b_k = \theta_k^* - \delta_k^*,$$

and $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_m^*)$.

Then the type I error probability of the UI-IU test (4.1) can be written as

$$Q = \int_0^\infty \cdots \int_0^\infty \Psi(\boldsymbol{\theta}^*|\boldsymbol{u}) h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R}) d\boldsymbol{u}, \tag{4.3}$$

where

$$\Psi(\boldsymbol{\theta}^*|\boldsymbol{u}) = \int_{cu_1-a_1}^\infty \cdots \int_{cu_m-a_m}^\infty \phi_m(\boldsymbol{z}|\boldsymbol{R}) d\boldsymbol{z} - \int_{cu_1-a_1}^{du_1-b_1} \cdots \int_{cu_m-a_m}^{du_m-b_m} \phi_m(\boldsymbol{z}|\boldsymbol{R}) d\boldsymbol{z}. \tag{4.4}$$

Proof: We have

$$\begin{aligned} Q &= P\left\{T_k^{(S)} > c - (\delta_k^* + \epsilon_k^*)/U_k \ \ (1 \le k \le m)\right\} \\ &\quad -P\left\{c - (\delta_k^* + \epsilon_k^*)/U_k \le T_k^{(S)} \le d \ \ (1 \le k \le m)\right\} \\ &= P\left\{T_k > c - (\theta_k^* + \epsilon_k^*)/U_k \ \ (1 \le k \le m)\right\} \\ &\quad -P\left\{c - (\theta_k^* + \epsilon_k^*)/U_k \le T_k \le d - (\theta_k^* - \delta_k^*)/U_k \ \ (1 \le k \le m)\right\}, \end{aligned}$$

where the T_k's are defined in (3.1) and have the multivariate t-distribution referred to there. Substitute $T_k = Z_k/U_k$. Then by conditioning on $U_k = u_k$ $(1 \le k \le m)$, the above probability can be written as

$$\begin{aligned} Q &= \int_0^\infty \cdots \int_0^\infty \left[P\left\{Z_k > cu_k - (\theta_k^* + \epsilon_k^*) \ (1 \le k \le m)\right\}\right. \\ &\quad \left. - P\left\{cu_k - (\theta_k^* + \epsilon_k^*) \le Z_k \le du_k - (\theta_k^* - \delta_k^*) \ (1 \le k \le m)\right\}\right] h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R}) d\boldsymbol{u} \\ &= \int_0^\infty \cdots \int_0^\infty \Psi(\boldsymbol{\theta}^*|\boldsymbol{u}) h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R}) d\boldsymbol{u}, \end{aligned}$$

which is the expression (4.3). □

Lemma 2: The type I error probability of the UI-IU test is maximized at one or more of the following configurations:

$$\text{LFC}_0 = \{\theta_1 = \delta_1, \ldots, \theta_m = \delta_m\} \text{ or } \text{LFC}_k = \{\theta_k = -\epsilon_k, \theta_\ell \to \infty, \ \ell \ne k\} \ (1 \le k \le m). \tag{4.5}$$

Proof: To find the maximum of Q with respect to (w.r.t.) θ_k over $H_0 = H_0^{(S)} \bigcup \left\{ \bigcup_{k=1}^{m} H_{0k}^{(E)} \right\}$, we take the derivatives of $\Psi(\boldsymbol{\theta}^*|\boldsymbol{u})$ (cf. (4.4)) for fixed $\boldsymbol{u}$ w.r.t. $\theta_k^* \propto \theta_k$. In particular, for $k = 1$, using $\partial a_1/\partial\theta_1^* = \partial b_1/\partial\theta_1^* = 1$, we get

$$\begin{aligned}
\frac{\partial\Psi(\boldsymbol{\theta}^*|\boldsymbol{u})}{\partial\theta_1^*} &= \int_{cu_2-a_2}^{\infty}\cdots\int_{cu_m-a_m}^{\infty}\phi_m(cu_1-a_1,z_2,\ldots,z_m|\boldsymbol{R})dz_2\ldots dz_m \\
&\quad -\Bigg[-\int_{cu_2-a_2}^{du_2-b_2}\cdots\int_{cu_m-a_m}^{du_m-b_m}\phi_m(du_1-b_1,z_2,\ldots,z_m|\boldsymbol{R})dz_2\ldots dz_m \\
&\quad +\int_{cu_2-a_2}^{du_2-b_2}\cdots\int_{cu_m-a_m}^{du_m-b_m}\phi_m(cu_1-a_1,z_2,\ldots,z_m|\boldsymbol{R})dz_2\ldots dz_m\Bigg] \\
&= \Bigg[\int_{cu_2-a_2}^{\infty}\cdots\int_{cu_m-a_m}^{\infty}\phi_m(cu_1-a_1,z_2,\ldots,z_m|\boldsymbol{R})dz_2\ldots dz_m \\
&\quad -\int_{cu_2-a_2}^{du_2-b_2}\cdots\int_{cu_m-a_m}^{du_m-b_m}\phi_m(cu_1-a_1,z_2,\ldots,z_m|\boldsymbol{R})dz_2\ldots dz_m\Bigg] \\
&\quad +\int_{cu_2-a_2}^{du_2-b_2}\cdots\int_{cu_m-a_m}^{du_m-b_m}\phi_m(du_1-b_1,z_2,\ldots,z_m|\boldsymbol{R})dz_2\ldots dz_m \\
&> 0.
\end{aligned}$$

Thus $\Psi(\boldsymbol{\theta}^*|\boldsymbol{u})$ is increasing in each θ_k^*. Therefore $\Psi(\boldsymbol{\theta}^*|\boldsymbol{u})$ (and hence Q) is maximized over $H_0^{(S)}$ at LFC_0 and over $H_{0k}^{(E)}$ at LFC_k for $1 \le k \le m$. The global maximum is found by evaluating Q at each of these $m+1$ LFC's, and taking the maximum over all of them. □

Let

$$e_k = \delta_k^* + \epsilon_k^* = \frac{\delta_k + \epsilon_k}{\sigma_k}\sqrt{\frac{n_1 n_2}{n_1+n_2}} \quad (1 \le k \le m). \tag{4.6}$$

Then for LFC_0 we get

$$\sup_{\boldsymbol{\theta}\in H_0^{(S)}}\Psi(\boldsymbol{\theta}^*|\boldsymbol{u}) = \int_{cu_1-e_1}^{\infty}\cdots\int_{cu_m-e_m}^{\infty}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} - \int_{cu_1-e_1}^{du_1}\cdots\int_{cu_m-e_m}^{du_m}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z}.$$

By substituting this expression in (4.3), the corresponding maximum of the type I error probability equals

$$\begin{aligned}
Q_{\max,0} &= \int_0^{\infty}\cdots\int_0^{\infty}\left[\int_{cu_1-e_1}^{\infty}\cdots\int_{cu_m-e_m}^{\infty}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} - \int_{cu_1-e_1}^{du_1}\cdots\int_{cu_m-e_m}^{du_m}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z}\right]h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R})d\boldsymbol{u} \\
&= P\left\{\min_{1\le k\le m}\left(T_k + \frac{\delta_k+\epsilon_k}{S_k\sqrt{1/n_1+1/n_2}}\right) > c \text{ and } \max_{1\le k\le m} T_k > d\right\}.
\end{aligned} \tag{4.7}$$

For LFC$_k$ $(1 \le k \le m)$ we get $a_k = 0, b_k = -e_k$ and $a_\ell, b_\ell \to \infty$ for $\ell \ne k$. So

$$\begin{aligned}\sup_{\boldsymbol{\theta} \in H_{0k}^{(E)}} \Psi(\boldsymbol{\theta}^*|\boldsymbol{u}) &= \int_{cu_k}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} - \int_{cu_k}^{du_k - e_k}\int_{-\infty}^{-\infty}\cdots\int_{-\infty}^{-\infty}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} \\ &= 1 - \Phi(cu_k).\end{aligned}$$

It follows from (4.3) that for $1 \le k \le m$,

$$Q_{\max,k} = \int_0^{\infty}[1 - \Phi(cu_k)]h_\nu(u_k)du_k \tag{4.8}$$

$$= P\{T_\nu > c\}, \tag{4.9}$$

where $h_\nu(u_k)$ is the density function of $U_k \sim \sqrt{\chi^2_\nu/\nu}$ and T_ν is Student's t r.v. with ν d.f. Equating $Q_{\max,k}$ to α, we get $c = t_{\nu,\alpha}$.

Given c, one can solve for d by setting $Q_{\max,0} = \alpha$. However, the solution of this equation requires the knowledge of the covariance matrix $\boldsymbol{\Sigma}$. If the δ_k and ϵ_k are specified as multiples of the σ_k then only the knowledge of the correlation matrix $\boldsymbol{R}$ is required. We will first study the behavior of solution d in the asymptotic $(\nu \to \infty)$ case.

4.2.2 The Asymptotic $(\nu \to \infty)$ Case

For $\nu \to \infty$, $U_k \to 1$ for all k with probability 1, and hence $cu_k \to c$ and $du_k \to d$. In that case, (4.7) becomes

$$Q_{\max,0} = P\left\{\min_{1\le k\le m}(Z_k + e_k) > c \text{ and } \max_{1\le k\le m} Z_k > d\right\} \tag{4.10}$$

$$= \int_{c-e_1}^{\infty}\cdots\int_{c-e_m}^{\infty}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} - \int_{c-e_1}^{d}\cdots\int_{c-e_m}^{d}\phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z}, \tag{4.11}$$

and (4.9) becomes

$$Q_{\max,k} = 1 - \Phi(c) \quad (1 \le k \le m).$$

To make $Q_{\max,k} = \alpha$ for $1 \le k \le m$, we choose $c = z_\alpha$ (the upper α critical point of the standard normal distribution) analogous to the small sample choice of $c = t_{\nu,\alpha}$. Next, we address the question of how to choose $d \ge c$.

Lemma 3 shows that if the e_k are "large" then $d = z_{m,\boldsymbol{R},\alpha}$, which is its largest value (that required by the UI test of the hypothesis $H_0^{(S)}$). Lemma 4 shows that if the e_k are "small" and the endpoints are independent then $d = z_\alpha$, which is its smallest value (that required by the IU test of the hypothesis $H_0^{(E)}$).

Lemma 3: If $e_k = \delta_k^* + \epsilon_k^* \to \infty$ for all k then $d = z_{m,\boldsymbol{R},\alpha}$, the upper α critical point of $\max_{1\le k\le m} Z_k$, where $\boldsymbol{Z} = (Z_1, \ldots, Z_m)$ has an m-variate standard normal distribution with correlation matrix $\boldsymbol{R}$. If $\boldsymbol{R}$ is completely unknown then a conservative choice is $d = z_{\alpha/m}$.

Proof: For $e_k \to \infty \ \ (1 \le k \le m)$, we get

$$Q_{\max,0} = 1 - \int_{-\infty}^{d} \cdots \int_{-\infty}^{d} \phi_m(\boldsymbol{z}|\boldsymbol{R}) d\boldsymbol{z} = 1 - P\{\max(Z_1, \ldots, Z_m) \le d\}.$$

Setting the above equal to α gives $d = z_{m,\boldsymbol{R},\alpha}$. By the Bonferroni inequality, the conservative choice is $d = z_{\alpha/m} > z_{m,\boldsymbol{R},\alpha}$. □

Lemma 4: If the endpoints are mutually independent and if all $e_k \le c = z_\alpha$ then we can choose $d = c = z_\alpha$.

Proof: We only need to show that under the given conditions, choosing $d = c = z_\alpha$ makes $Q_{\max,0} \le \alpha$. To see this, put $d = c$ in the expression for $Q_{\max,0}$ and use the mutual independence to obtain

$$Q_{\max,0} = \prod_{k=1}^{m} [1 - \Phi(c - e_k)] - \prod_{k=1}^{m} [\Phi(c) - \Phi(c - e_k)].$$

This is an increasing function of each e_k since

$$\frac{\partial Q_{\max,0}}{\partial e_k} = \phi(c - e_k) \left\{ \prod_{\ell=1,\ell\ne k}^{m} [1 - \Phi(c - e_\ell)] - \prod_{\ell=1,\ell\ne k}^{m} [\Phi(c) - \Phi(c - e_\ell)] \right\} > 0,$$

which follows because $1 - \Phi(c - e_\ell) > \Phi(c) - \Phi(c - e_\ell)$. Therefore, by setting $e_k = c$ for all k, we obtain an upper bound on $Q_{\max,0}$:

$$Q_{\max,0} \le \left(\frac{1}{2}\right)^m - \left(\Phi(c) - \frac{1}{2}\right)^m.$$

To see that this upper bound is $\leq 1 - \Phi(c) = \alpha$ we now show that

$$f(c) = \Phi(c) - \left(\Phi(c) - \frac{1}{2}\right)^m \leq 1 - \left(\frac{1}{2}\right)^m.$$

This follows if we show that $f(c)$ is an increasing function of c. Note that

$$\begin{aligned} f'(c) = \phi(c) - \phi(c)m\left(\Phi(c) - \frac{1}{2}\right)^{m-1} &> 0 \\ \iff \frac{1}{m} &> \left(\Phi(c) - \frac{1}{2}\right)^{m-1}. \end{aligned}$$

The last inequality can be proved by induction as follows. Obviously, it is true for $m = 2$. Assuming that it is true for $m \geq 2$, we now show it to be true for $m + 1$:

$$\frac{1}{m+1} = \frac{1}{m} \cdot \frac{m}{m+1} > \left(\Phi(c) - \frac{1}{2}\right)^{m-1} \left(\Phi(c) - \frac{1}{2}\right),$$

which holds because

$$\frac{1}{m} > \left(\Phi(c) - \frac{1}{2}\right)^{m-1}$$

by the induction hypothesis and

$$\frac{m}{m+1} > \frac{1}{2} > \left(\Phi(c) - \frac{1}{2}\right).$$

This proves the lemma. □

Numerical Illustration of Lemma 4: Suppose that $\delta_k = 0$ and $\epsilon_k = \lambda\sigma_k$ for $1 \leq k \leq m$. Also assume that $n_1 = n_2 = n$. Then the condition $e_k \leq c$ becomes

$$n \leq \frac{2c^2}{\lambda^2}.$$

Let $\lambda = 0.1$ and $c = 1.645$ (for $\alpha = .05$). Then

$$n \leq \frac{2(1.645)^2}{(0.1)^2} = 541.2,$$

which is a fairly large upper bound. □

Lemma 4 has been shown only under the independence assumption. It does not hold in general under dependence. If all $e_k \to 0$ then the lemma holds since in that case

$$Q_{\max,0} = \int_c^\infty \cdots \int_c^\infty \phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} \leq 1 - \Phi(c) = \alpha.$$

However, if the e_k are large, e.g., if *all* $e_k = c$, then the lemma does not hold under dependence. In other words, we cannot choose $d = z_\alpha$. The proof of this result is omitted for brevity.

It is still possible to choose d smaller than $z_{\alpha/m}$ under dependence. Assuming known covariance matrix $\boldsymbol{\Sigma}$, this can be done by solving $Q_{\max,0} = \alpha$ for d via simulation using the probability representation in (4.7) as follows.

1. Generate i.i.d. m-variate normally distributed random vectors $\boldsymbol{X}_{ij}$ $(i = 1, 2; 1 \leq j \leq n_i)$ having a null mean vector and covariance matrix $\boldsymbol{\Sigma}$.

2. Calculate the sample means $\overline{X}_{i\cdot k}$ $(i = 1, 2)$ and sample variances S_k^2 $(1 \leq k \leq m)$.

3. Calculate the t-statistics

$$T_k = \frac{(\overline{X}_{1\cdot k} - \overline{X}_{2\cdot k})}{S_k\sqrt{1/n_1 + 1/n_2}} \quad (1 \leq k \leq m).$$

4. Calculate a random variable Y such that

$$Y = \begin{cases} c = t_{\nu,\alpha} & \text{if } \min_{1\leq k\leq m}\left\{T_k + \frac{\delta_k+\epsilon_k}{S_k\sqrt{1/n_1+1/n_2}}\right\} \leq c \\ \max\{c, \max_{1\leq k\leq m} T_k\} & \text{if } \min_{1\leq k\leq m}\left\{T_k + \frac{\delta_k+\epsilon_k}{S_k\sqrt{1/n_1+1/n_2}}\right\} > c. \end{cases}$$

5. Having generated N replications of Y, find the $(1-\alpha)$th quantile of Y which is the desired critical point d.

Table 1 gives the values of d evaluated assuming a common known correlation ρ among the endpoints, and $\delta_k = 0$ and $\epsilon_k = \lambda\sigma_k$. The following points are worth noting.

1. The d values for $n = \infty$ equal $z_{m,\boldsymbol{R},\alpha}$, where $\boldsymbol{R}$ is the correlation matrix with all off-diagonal elements equal to ρ. This follows as a result of Lemma 3 since all $e_k = \infty$ in this case.

2. The d values for $n = 25$ equal $c = t_{48,.05} = 1.6772 \approx 1.68$ as all e_k are small in this case, and although Lemma 4 does not strictly apply, analogous result appears to hold empirically.

3. As n increases, the d values generally decrease slightly (except when $\lambda = 0.2$ and $\rho = 0.75$) and then increase to their upper bound $= z_{m,\boldsymbol{R},\alpha}$. For most combinations with $\lambda = 0.1$ and $n \leq 200$, and even for many combinations with $\lambda = 0.2$, d is between 1.65 and 1.68, indicating that almost no multiplicity adjustment is required.

4. The initial decrease followed by an eventual increase in d as n increases deserves an explanation: The initial decrease is due to an increase in the degrees of freedom which reduces the value of $d = c = t_{\nu,\alpha}$; note that in this case the e_k are "small." For n sufficiently large, the effect on d of the increase in the degrees of freedom becomes negligible. As the e_k become large, we see from the equation (4.11) that the event of declaring all endpoints equivalent becomes more and more likely, which increases the type I error probability $Q_{\max,0}$. Hence d must be increased to maintain $Q_{\max,0} = \alpha$.

5. Example

We use the example from Tang, Geller and Pocock (1993) about the efficacy of an inhaled drug for asthma compared to placebo. Seventeen patients were randomized in a double-blind crossover trial. There were four standard respiratory function measures (endpoints): forced expiratory volume (FEV_1), forced vital capacity (FVC), peak expiratory flow rate (PEFR) and penetration index (PI). FEV_1 and FVC are expressed as percentages of the predicted

values for that patient's age, sex and height in the normal population. PEFR is expressed in per minute. PI measures the ability of a deep inhalation to reach small airways. There was no period or crossover effect, so the comparisons for individual endpoints could be performed using paired t-statistics. The summary statistics were as follows:

	FEV_1	FVC	PEFR	PI
Mean Difference	7.56	4.81	2.29	0.081
Std. Dev. of Difference	18.53	10.84	8.51	0.17
t-Statistic	1.682	1.830	1.110	1.965
p-Value	0.0560	0.0430	0.1417	0.0335

The estimated correlation matrix was

$$\begin{bmatrix} 1.000 & 0.095 & 0.219 & -0.162 \\ & 1.000 & 0.518 & -0.059 \\ & & 1.000 & 0.513 \\ & & & 1.000 \end{bmatrix}.$$

For these data the OLS and GLS statistics are highly significant indicating a global improvement. However, none of the individual endpoints can be identified as having significant improvement at $\alpha = 0.05$ using the Bonferroni procedure or one of its sharpened versions.

Suppose $\delta_k = 0$ and $\epsilon_k = \lambda\sigma_k$ with $\lambda = 0.20$ for $1 \leq k \leq 4$. Then

$$e_k = \frac{\delta_k + \epsilon_k}{\sigma_k\sqrt{1/n}} = 0.20\sqrt{17} = 0.825;$$

here $\sqrt{1/n_1 + 1/n_2}$ is changed to $\sqrt{1/n}$ since this is essentially a paired sample study with n patients. The t-statistics given in the above table are the superiority t-statistics, $t_k^{(S)}$ since the $\delta_k = 0$.

For $\alpha = 0.05$, we have $c = t_{16,.05} = 1.746$. Next we calculated d using the simulation method given in Section 4 by assuming the above correlation matrix, and obtained $d = c =$

1.746. Note that $e_k = 0.825 < c = 1.746$, so the condition in Lemma 4 for $d = c$ is satisfied, although that lemma is not strictly applicable here because (i) the lemma assumes that the endpoints are independent, and (ii) the lemma assumes $\nu = \infty$.

Taking $s_k \approx \sigma_k$ and applying the rule (4.1), we find that

$$\min_{1\le k\le 4}\left\{t_k^{(S)} + 0.825\right\} = \min\{2.507, 2.655, 1.935, 2.790\} > c = 1.746$$

and

$$\max_{1\le k\le 4}\left\{t_k^{(S)}\right\} = \max\{1.682, 1.830, 1.110, 1.965\} > d = 1.746.$$

Therefore H_0 is rejected and the inhaled drug is shown to be equivalent to the placebo on all endpoints and superior on at least one endpoint.

From the above calculation one can see that the smallest value of λ for which the equivalence holds for all endpoints is given by $t_{\text{PEFR}}^{(S)} + \lambda\sqrt{17} = 1.110 + \lambda\sqrt{17} \ge 1.746$ or $\lambda \ge 0.154$.

6. Generalizations and Extensions

The UI-IU procedure given above tests a single global null hypothesis in (2.1). However, it may be useful to be able to determine exactly which of the endpoints show a superior treatment effect. One way to do this is to define the family of individual endpoint hypotheses $H_{0k} = H_{0k}^{(S)} \bigcup H_0^{(E)}$. We can close this family by including in it all intersections of $H_{0k},\ k \in K$. Note that the overall intersection of all H_{0k} is just H_0, and that the intersection of the hypotheses in set K of endpoints can be written as

$$\bigcap_{k\in K} H_{0k} = \left(\bigcap_{k\in K} H_{0k}^{(S)}\right) \bigcup H_0^{(E)}.$$

If we use a union-intersection test for each superiority component of the hypotheses in this family and apply a closed test procedure, we obtain a step-down procedure as follows: Order the superiority t-statistics so that $t_{(1)}^{(S)} \le \cdots \le t_{(m)}^{(S)}$. Denote the corresponding hypotheses

by $H_{0(1)}, H_{0(2)}, \ldots, H_{0(m)}$. The adjusted p-value for testing $H_{0(m)}$ is

$$\tilde{p}_{(m)} = P\left\{\min_{1\le k\le m} T_k^{(E)} \ge \min_{1\le k\le m} t_k^{(E)}, \max_{1\le k\le m} T_k^{(S)} \ge t_{(m)}^{(S)}\right\},$$

and the adjusted p-value for testing $H_{0(\ell)}$ for $\ell < m$ is

$$\tilde{p}_{(\ell)} = \max\left[\tilde{p}_{(\ell+1)}, P\left\{\min_{\ell\le k\le m} T_k^{(E)} \ge \min_{\ell\le k\le m} t_k^{(E)}, \max_{1\le k\le \ell} T_k^{(S)} \ge t_{(\ell)}^{(S)}\right\}\right].$$

These adjusted p-values may be estimated through resampling. Note that if hypothesis H_{0k} is rejected, one can conclude that the treatment is superior to the control on endpoint k and equivalent on all others. While this step-down procedure allows decisions on individual endpoints for superiority, it does not allow such decisions for equivalence. The simultaneous confidence interval approach given in Section 3 does allow for decisions on both superiority and equivalence on individual endpoints; however, it is conservative. A more powerful step-down procedure, which classifies the treatment effect on each endpoint into superior, equivalent and inferior, is a problem for future research.

The multivariate normality assumption may not hold in practice. In that case a bootstrap version of the proposed test procedure can be easily applied.

Finally, the approach given here can be generalized to deal with the goal of showing that the treatment is equivalent on all endpoints and superior on at least r endpoints, where r is specified ($1 \le r < m$).

References

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, **24**, 295 -300.

Bloch, D. A., Lai, T. L. and Tubert-Bitter, P. (2000). One-sided tests in clinical trials with multiple endpoints. Technical Report No. 214, Department of Statistics, Stanford University, Stanford, CA.

Cohen, A. and Sackrowitz, H. B. (1998). Directional tests for one-sided alternatives in multivariate models. *Annals of Statistics*, **26**, 2321 -2338.

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

Lehmacher, W., Wassmer, G. and Reitmeir, P. (1991). Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics* **47**, 511–521.

Perlman, M. D. and Wu, L. (2002). A defense of the likelihood ratio test for one-sided and order-restricted alternatives. *Journal of Statistical Planning and Inference*, to appear.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics*, **24**, 220 -238.

Šidák, Z. (1968). On multivariate normal probabilities of rectangles: Their dependence on correlations. *Annals of Mathematical Statistics*, **39**, 1425 -1434.

Siddiqui, M. M. (1967). A bivariate t distribution. *Annals of Mathematical Statistics*, **38**, 162–166.

Silvapulle, M. J. (1997). A curious example involving the likelihood ratio test against one-sided hypotheses. *American Statistician*, **51**, 178 -180.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751 -754.

Tang, D-I, Geller, N. L. and Pocock, S. J. (1993). On the design and analysis of randomized clinical trials with multiple endpoints. *Biometrics*, **49**, 23–30.

Figure 1: Rejection Region of the LR Test for $m = 2$

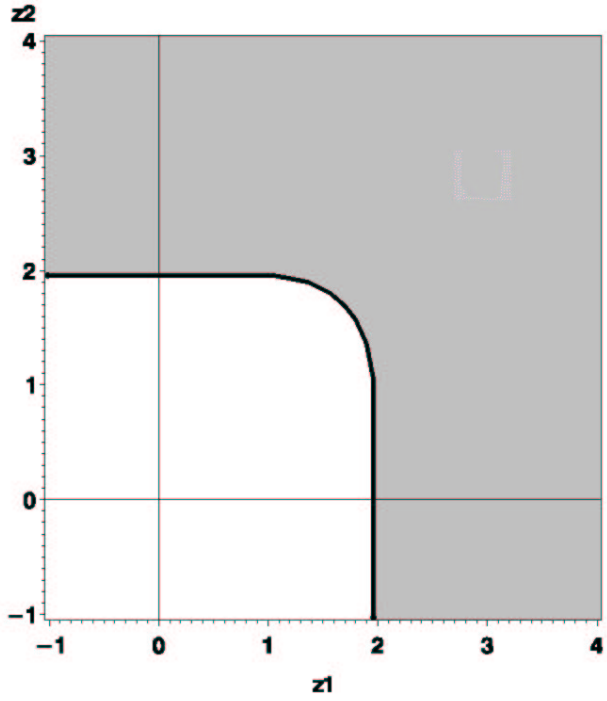

Figure 2: Rejection Region of the COM Test for $m = 2$

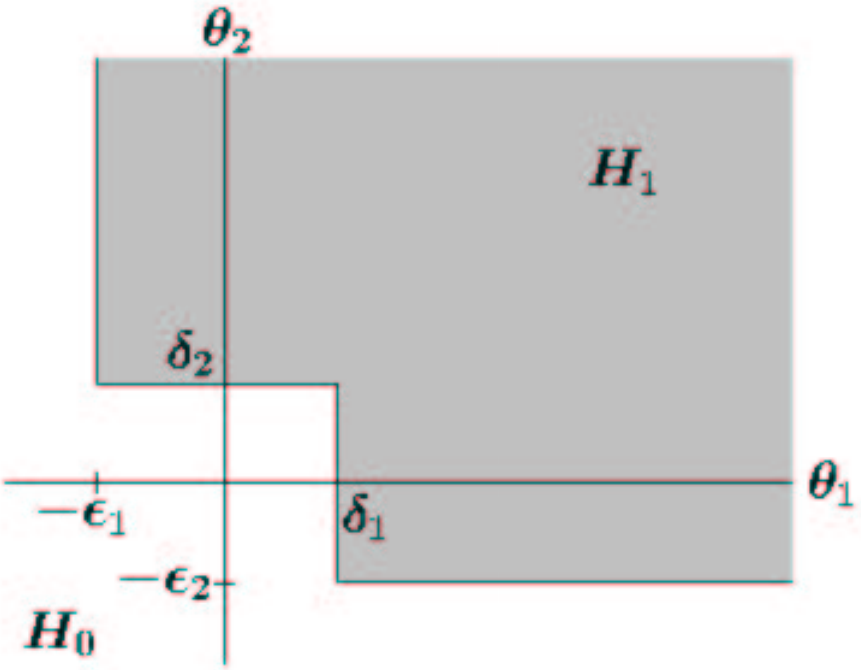

Figure 3: Hypotheses H_0 and H_1

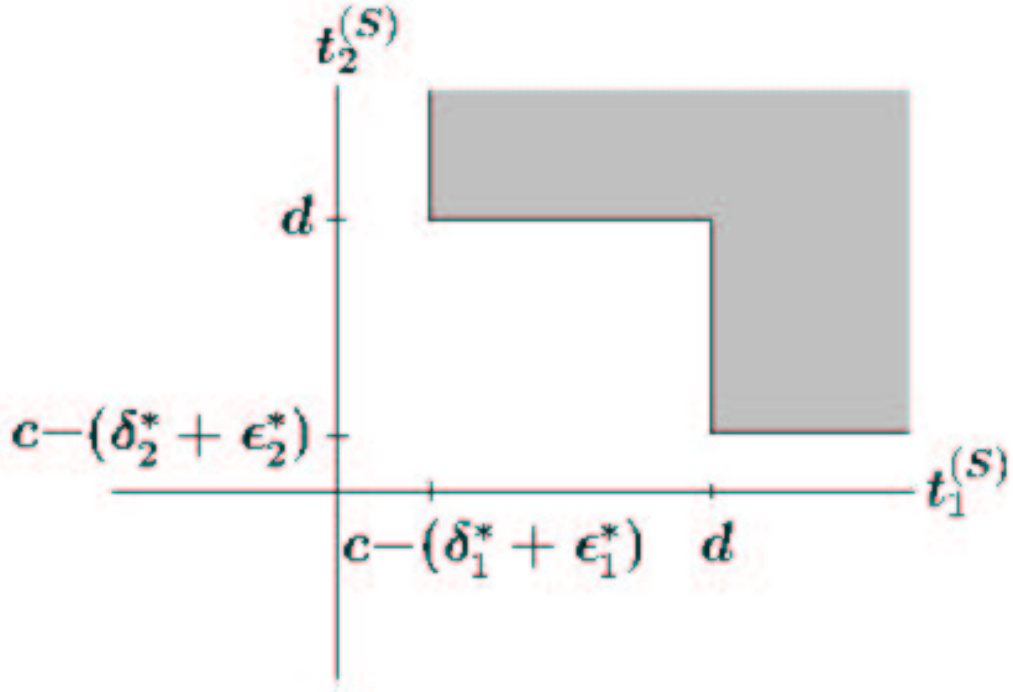

Figure 4: Rejection Region of the UI-IU Test for $m = 2$

Table 1: Simulated Values of the Critical Constant d for $\alpha = 0.05$.

			$n_1 = n_2 = n$				
m	λ	ρ	25	50	100	200	∞
2	0.1	0	1.68	1.66	1.65	1.65	1.96
		0.25	1.68	1.66	1.65	1.65	1.95
		0.5	1.68	1.66	1.65	1.70	1.92
		0.75	1.68	1.66	1.75	1.82	1.86
	0.2	0	1.68	1.66	1.65	1.76	1.96
		0.25	1.68	1.66	1.70	1.85	1.95
		0.5	1.68	1.71	1.83	1.90	1.92
		0.75	1.68	1.83	1.86	1.87	1.86
4	0.1	0	1.68	1.66	1.65	1.65	2.24
		0.25	1.68	1.66	1.65	1.65	2.21
		0.5	1.68	1.66	1.65	1.65	2.16
		0.75	1.68	1.66	1.67	1.96	2.06
	0.2	0	1.68	1.66	1.65	1.65	2.24
		0.25	1.68	1.66	1.65	1.99	2.21
		0.5	1.68	1.66	1.94	2.11	2.16
		0.75	1.68	1.97	2.06	2.06	2.06
8	0.1	0	1.68	1.66	1.65	1.65	2.49
		0.25	1.68	1.66	1.65	1.65	2.46
		0.5	1.68	1.66	1.65	1.65	2.38
		0.75	1.68	1.66	1.65	2.00	2.23
	0.2	0	1.68	1.66	1.65	1.65	2.49
		0.25	1.68	1.66	1.65	1.98	2.46
		0.5	1.68	1.66	1.86	2.30	2.38
		0.75	1.68	2.04	2.21	2.23	2.23