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

Ajit C. Tamhane

(Joint work with Brent R. Logan)

Department of IE/MS and Department of Statistics

Northwestern University

Evanston, IL 60208

1. Problem

- Compare a treatment (Treatment 1) with a control (Treatment 2) based on $m \geq 2$ endpoints.

- X_{ijk} = Obs. on the kth endpoint for the jth patient in the ith group ($i = 1, 2; 1 \leq j \leq n_i; 1 \leq k \leq m$).

$$\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijm}) \sim \mathrm{MVN}_m(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}),\ i = 1, 2; 1 \leq j \leq n_i.$$

- Further notation:

$$\boldsymbol{\theta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = (\theta_1, \ldots, \theta_m)$$

$$\boldsymbol{R} = \{\rho_{k\ell}\} = \text{Correlation matrix}$$

- The treatment is expected to have no negative effect on any endpoint and a positive effect on at least one endpoint.

- Traditional one-sided hypothesis testing formulation:

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

where $\mathbf{0}$ is the null vector and

$$\mathcal{O}^+ = \{\boldsymbol{\theta} | \theta_k \geq 0 \ \forall \ k, \boldsymbol{\theta} \neq \mathbf{0}\}$$

is the positive orthant.

- Likelihood ratio (LR) rejection region for this formulation has some undesirable properties, e.g., is nonmonotone, contains points with some or all negative coordinates. (Berger 1989, Silvapulle 1997)
- Perlman and Wu (2002) show that the LR test using the full complement of $\mathcal{O}^+$ as the null hypothesis does not have these drawbacks.
- Cone-ordered monotone (COM) rejection region (Cohen and Sackrowitz 1998) also contains points with some negative coordinates.

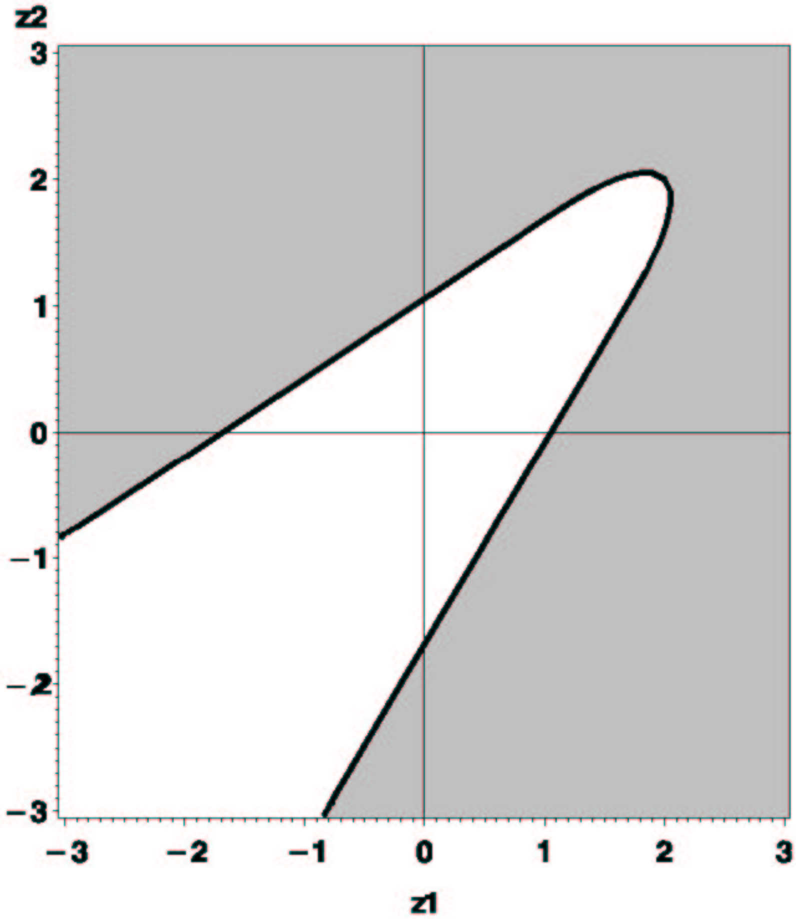

Fig. 1: Rejection Region of the LR Test for $m = 2$

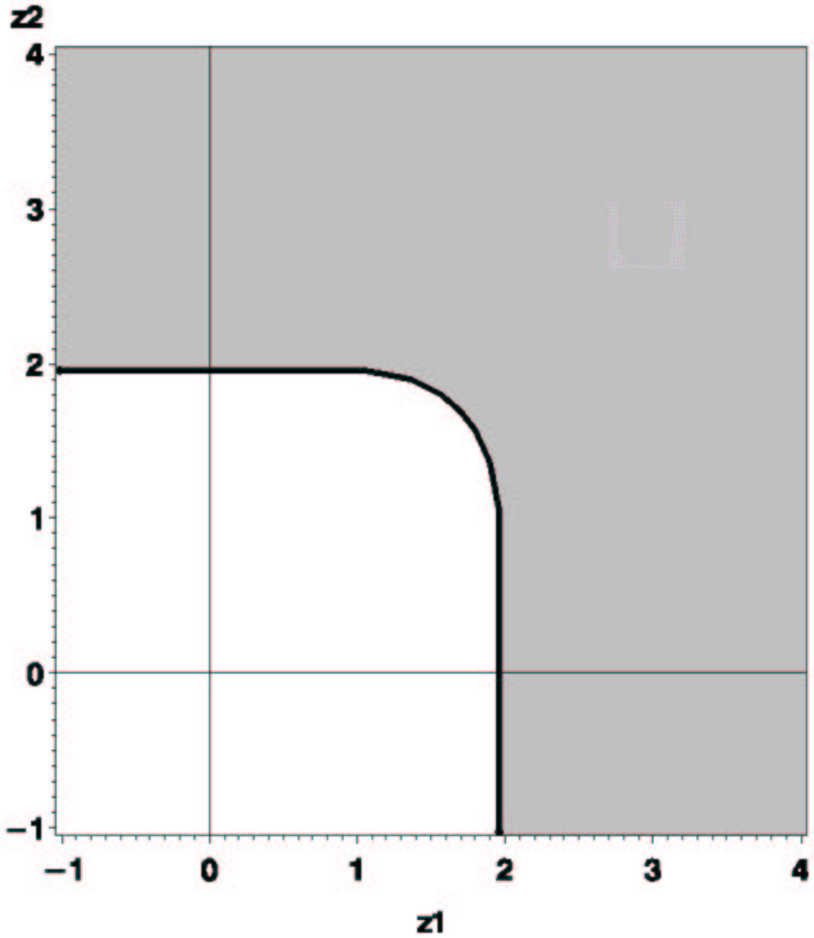

Fig. 2: Rejection Region of the COM Test for $m = 2$

2. Proposed Formulation

- The treatment is *superior* on the kth endpoint if $\theta_k > \delta_k$ and *equivalent* if $\theta_k > -\epsilon_k$, where $\delta_k, \epsilon_k \geq 0$ are specified constants.

- The treatment is deemed *effective* if it is equivalent on *all* endpoints and superior on *at least* one endpoint.

- Superiority Hypotheses:

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

  and

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

- Equivalence Hypotheses:

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

  and

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

- Hypothesis Testing Problem:

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

- Combination of union-intersection (UI) (Roy 1953) and intersection-union (IU) (Berger 1982) testing problems.

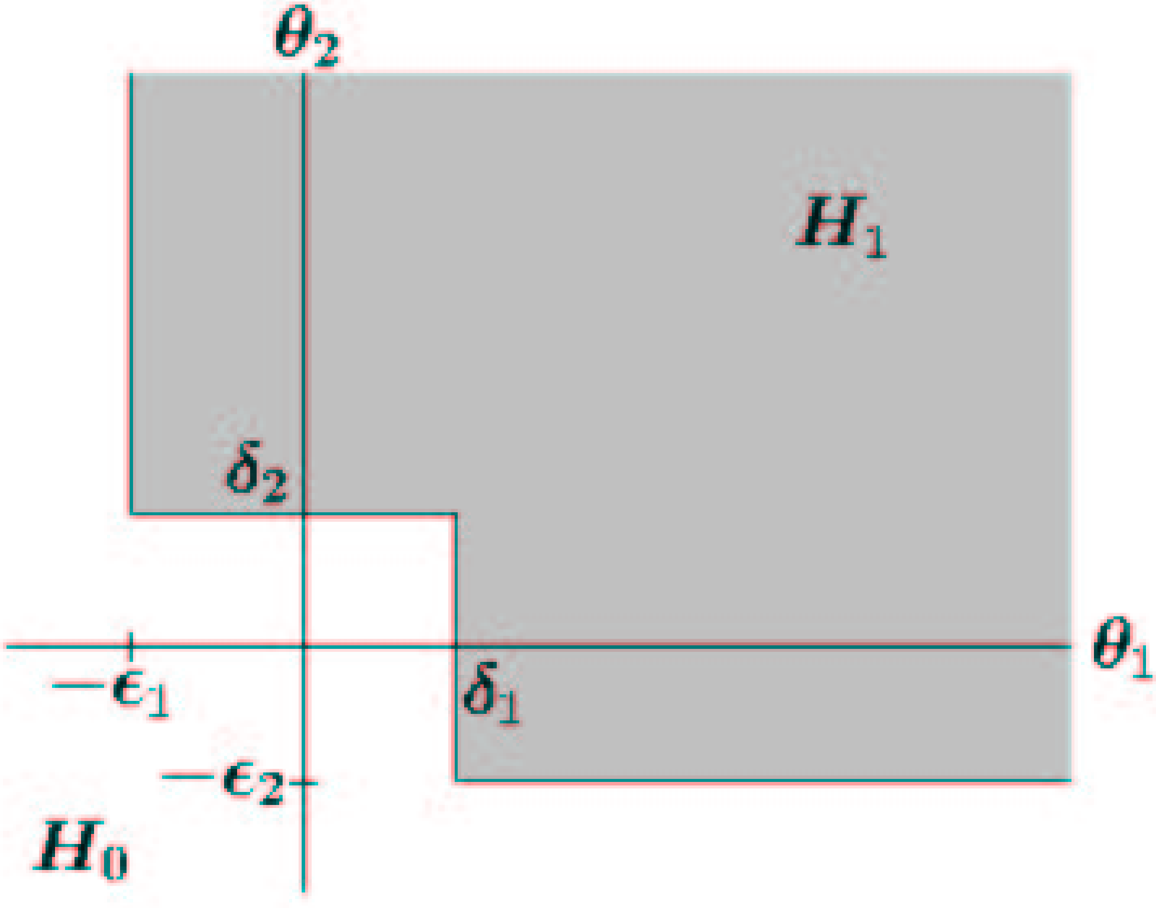

Fig. 3: Hypotheses H_0 and H_1 for $m = 2$

3. Simultaneous Confidence Intervals (SCI) Approach

- Denote by $\overline{X}_{1\cdot k}$ and $\overline{X}_{2\cdot k}$ the sample means for the kth endpoint for group 1 and group 2. Denote by $S_1^2, S_2^2, \ldots, S_m^2$ the pooled sample variances based on $\nu = n_1 + n_2 - 2$ degrees of freedom.

- The pivotal r.v. for θ_k is

$$T_k = \frac{(\overline{X}_{1\cdot k} - \overline{X}_{2\cdot k}) - \theta_k}{S_k\sqrt{1/n_1 + 1/n_2}} = \frac{Z_k}{U_k},$$

  where $\boldsymbol{Z} = (Z_1, \ldots, Z_k)$ is std. multivariate normal with correlation matrix $\boldsymbol{R}$. Denote the p.d.f. of $\boldsymbol{Z}$ by $\phi_m(\boldsymbol{z}|\boldsymbol{R})$. Next,

$$U_k = \frac{S_k}{\sigma_k} \sim \sqrt{\frac{\chi_\nu^2}{\nu}}.$$

  Denote the p.d.f. of $\boldsymbol{U} = (U_1, \ldots, U_m)$ by $h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R})$.

- Each $T_k \sim$ Student's t_ν. The joint distribution of $(T_1, T_2, \ldots, T_m)$ is a multivariate generalization of a bivariate t-distribution of Siddiqui (1967).

- Denote by $t_{\nu,\boldsymbol{R},\alpha} = (1-\alpha)$th quantile of $\max_{1\le k\le m} T_k$. The Bonferroni upper bound: $t_{\nu,\alpha/m} > t_{\nu,\boldsymbol{R},\alpha}$.

- $100(1-\alpha)\%$ SCI's on the θ_k:

$$\theta_k \geq L_k = \overline{x}_{1\cdot k} - \overline{x}_{2\cdot k} - t_{\nu,\alpha/m} s_k \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (1 \leq k \leq m).$$

- Treatment is equivalent on the kth endpoint if

$$L_k > -\epsilon_k \Longleftrightarrow t_k^{(E)} = \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k} + \epsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}} > t_{\nu,\alpha/m}.$$

- Treatment is superior on the kth endpoint if

$$L_k > \delta_k \Longleftrightarrow t_k^{(S)} = \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k} - \delta_k}{s_k\sqrt{1/n_1 + 1/n_2}} > t_{\nu,\alpha/m}.$$

- Reject H_0 if

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

- In addition, all endpoints can be classified with FWE $\leq \alpha$ into three groups: (i) not equivalent ($L_k \leq -\epsilon_k$), (ii) equivalent but not superior ($-\epsilon_k < L_k \leq \delta_k$), and (iii) superior ($L_k > \delta_k$).

4. A Combination Union-Intersection and Intersection-Union (UI-IU) Test

4.1 UI-IU Test

- Since $H_0 = H_0^{(S)} \cup H_0^{(E)}$, an α-level IU test rejects $H_0^{(S)}$ and $H_0^{(E)}$ each separately @ level α.

- Since $H_0^{(E)} = \cup_{k=1}^{m} H_{0k}^{(E)}$, an α-level IU test rejects @ level α if $\min_{1\le k\le m} t_k^{(E)} > t_{\nu,\alpha}$ (note smaller constant than that used by SCI's).

- Since $H_0^{(S)} = \cap_{k=1}^{m} H_{0k}^{(S)}$, an α-level UI test rejects @ level α if $\max_{1\le k\le m} t_k^{(S)} > t_{\nu,\alpha/m}$.

- The following argument shows that this test can be sharpened.

- Controlling α separately for $H_0^{(S)}$ and $H_0^{(E)}$ assumes that one hypothesis is true and the other is infinitely false, which is the Least Favorable Configuration (LFC).

- It is possible that $H_0^{(E)} = \cup_{k=1}^m (\theta_k \le -\epsilon_k)$ is true but $H_0^{(S)} = \cap_{k=1}^m (\theta_k \le \delta_k)$ is infinitely false. Therefore the IU test of $H_0^{(E)}$ can't be sharpened.

- It is not possible that $H_0^{(S)} = \cap_{k=1}^m (\theta_k \le \delta_k)$ is true but $H_0^{(E)} = \cup_{k=1}^m (\theta_k \le -\epsilon_k)$ is infinitely false. Therefore the UI test of $H_0^{(S)}$ can be sharpened.

- Denote the critical constant for the IU test of $H_0^{(E)}$ by $c = t_{\nu,\alpha}$ and the critical constant for the UI test of $H_0^{(S)}$ by $d \ge c$.

  Problem: Find the smallest possible d.

- Note
$$t_k^{(S)} = t_k^{(E)} - \frac{\delta_k + \epsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}}.$$
  Therefore the rejection region of the UI-IU test can be written as
$$\min_{1\le k\le m}\left\{t_k^{(S)} + \frac{\delta_k + \epsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}}\right\} > c \text{ and } \max_{1\le k\le m} t_k^{(S)} > d.$$

- Let
$$\delta_k^* = \frac{\delta_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}}, \epsilon_k^* = \frac{\epsilon_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}}, \theta_k^* = \frac{\theta_k}{\sigma_k\sqrt{1/n_1 + 1/n_2}}.$$

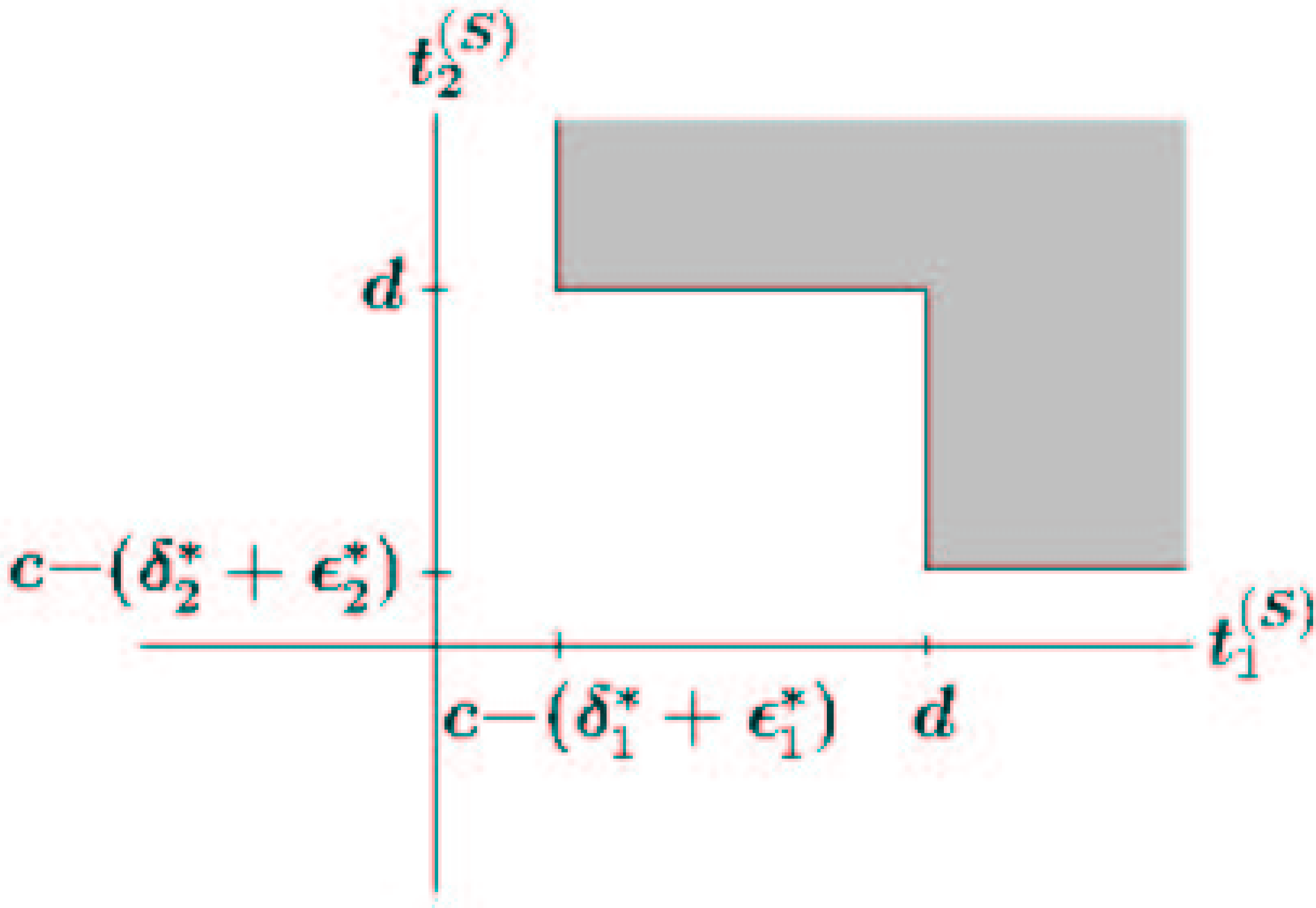

Fig. 4: Rejection Region of the UI-IU Test for $m = 2$

4.2 Sharpened Critical Constants for the UI-IU Test

For simplicity we consider the known σ_k ($\nu \to \infty$) case. For the finite ν case the probability expressions can be unconditioned w.r.t. the p.d.f. $h_{m,\nu}(\boldsymbol{u}|\boldsymbol{R})$.

Lemma 1: Let

$$a_k = \theta_k^* + \epsilon_k^*, \; b_k = \theta_k^* - \delta_k^*.$$

Then the type I error probability of the general UI-IU test equals

$$Q = \int_{c-a_1}^{\infty} \cdots \int_{c-a_m}^{\infty} \phi_m(\boldsymbol{z}|\boldsymbol{R}) d\boldsymbol{z} - \int_{c-a_1}^{d-b_1} \cdots \int_{c-a_m}^{d-b_m} \phi_m(\boldsymbol{z}|\boldsymbol{R}) d\boldsymbol{z}.$$

Lemma 2: The LFC of the UI-IU test is one or more of the following configurations:

$$\text{LFC}_0 = \{\theta_1 = \delta_1, \ldots, \theta_m = \delta_m\}$$

$$\text{LFC}_k = \{\theta_k = -\epsilon_k, \theta_\ell \to \infty, \; \ell \neq k\} \; (1 \leq k \leq m).$$

Denote

$$e_k = \delta_k^* + \epsilon_k^* = \frac{\delta_k + \epsilon_k}{\sigma_k}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

Then

$$Q_{\max,0} = \int_{c-e_1}^{\infty} \cdots \int_{c-e_m}^{\infty} \phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z} - \int_{c-e_1}^{d} \cdots \int_{c-e_m}^{d} \phi_m(\boldsymbol{z}|\boldsymbol{R})d\boldsymbol{z},$$

and

$$Q_{\max,k} = 1 - \Phi(c) \ \ (1 \leq k \leq m) \quad \Rightarrow c = z_\alpha.$$

Evaluation of d by solving $Q_{\max,0} = \alpha$ requires the knowledge of $\boldsymbol{R}$ and the σ_k (to calculate the e_k). For the known equicorrelated case with $\delta_k = 0$ and $\epsilon_k = \lambda\sigma_k$, we have calculated d via simulation for selected cases.

Note that the d-values do not involve much multiplicity adjustment except when ρ is large or when $n \to \infty$ ($e_k \to \infty$).

Simulated Values of d for $\alpha = 0.05$.

m	λ	ρ	n = 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	1.78	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

Lemma 3: If $e_k = \delta_k^* + \epsilon_k^* \to \infty$ for all k then $d = z_{m,\boldsymbol{R},\alpha}$ = the $(1-\alpha)$th quantile of $\max_{1\le k\le m} Z_k$. Use $d = z_{\alpha/m} \ge z_{m,\boldsymbol{R},\alpha}$.

Lemma 4: If all $\rho_{k\ell} = 0$ and all $e_k \le c = z_\alpha$ then $d = c = z_\alpha$.

Implications of Lemmas 3 and 4: If the e_k are large (e.g., if the n_k are large) then d is the largest possible $= d = z_{\alpha/m}$ ($t_{\nu,\alpha/m}$ for small samples). If the e_k are small then d is the smallest possible $= d = z_\alpha$ ($t_{\nu,\alpha}$ for small samples).

Numerical Illustration of Lemma 4: Suppose that $\delta_k = 0, \epsilon_k = \lambda\sigma_k$ and $n_1 = n_2 = n$. Then $e_k \le c$ is equivalent to

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

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

$$n \le \frac{2(1.645)^2}{(0.1)^2} = 541.2.$$

5. Example

- Randomized double-blind crossover asthma trial to compare an inhaled drug with placebo (Tang, Geller and Pocock 1993) with $n = 17$ patients.
- No period effect; hence analyzed as a paired sample study.
- Summary statistics for four endpoints:

	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 sample correlation matrix:

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

Suppose $\delta_k = 0$ and $\epsilon_k = \lambda\sigma_k$ with $\lambda = 0.20$. Then

$$\frac{\delta_k + \epsilon_k}{s_k\sqrt{1/n}} \approx 0.20\sqrt{17} = 0.825$$

(assuming $s_k \approx \sigma_k$). Finally, for $\alpha = 0.05$, $c = t_{16,.05} = 1.746$, and by solving $Q_{\max,0} = \alpha$ using $\boldsymbol{R}$ = sample correlation matrix, we obtained $d = c = 1.746$.

By applying the UI-IU test, we find that

$$\min_{1\le k\le 4}\left\{t_k^{(S)} + 0.825\right\} = \min\{2.506, 2.655, 1.935, 2.790\} > c = 1.746$$

and

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

Hence the drug is proven effective.

The smallest value of $\lambda = 0.155$ to conclude equivalence.

In this example both the Bonferroni and Westfall-Young resampling methods give nonsignificant results.

5. Generalizations and Extensions

1. The UI-IU procedure addresses a single global null hypotheseis:

$$H_0 = \left(\bigcap_{k=1}^{m} H_{0k}^{(S)} \right) \cup \left(\bigcup_{k=1}^{m} H_{0k}^{(E)} \right).$$

   Extend to pinpoint the endpoints that show a positive effect.

2. Bootstrap sampling to avoid the multivariate normality assumption and work in terms p-values.

3. Devise a procedure to show that the treatment is equivalent on all endpoints, and superior on at least r endpoints $(1 \leq r < m)$.