Recent Developments in Multiple Comparison Procedures
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 47 (2004) 76–88

On O'Brien's OLS and GLS tests for multiple endpoints

Brent R. Logan[1] **and Ajit C. Tamhane**[2]

Medical College of Wisconsin and Northwestern University

Abstract: In this article we obtain some new results and extensions of the OLS and GLS tests proposed by O'Brien (1984) for the one-sided multivariate testing problem. In particular, we empirically obtain an accurate small sample approximation to the critical point of the OLS test. Next we give a power comparison between the OLS test and a competing test proposed by Läuter(1996). Lastly, we extend the OLS and GLS tests to the heteroscedastic setup where the control and treatment populations have different covariance matrices.

1. Introduction

Most 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. O'Brien (1984) proposed two global tests, known as the ordinary least squares (OLS) and generalized least squares (GLS) tests, to demonstrate such an overall treatment effect. In this article we obtain some new results and extensions of these tests.

The following is an outline of the paper. Section 2 gives the notation, the problem formulation and the assumptions. Section 3 deals with the homoscedastic case. First it gives a review of the OLS and GLS tests, including an improved approximation to the small sample critical value of the OLS test. Next it gives a power comparison between the OLS test and a test proposed by Läuter. Section 4 derives extensions of the OLS and GLS tests to the heteroscedastic case. Section 5 gives some concluding remarks. The appendix gives derivations of asymptotic power expressions of the OLS and Läuter's tests required for the power comparison in Section 3.

2. Notation and preliminaries

Suppose that there are two independent treatment groups with n_1 and n_2 subjects on each of whom $m \geq 2$ endpoints are measured. Treatment 1 is the test treatment and treatment 2 is the control. Let x_{ijk} denote the measurement on the kth endpoint for the jth subject in the ith treatment group. For treatment group i, assume that $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijm})'$, $j = 1, 2, \ldots, n_i$, are independent and identically distributed (i.i.d.) random vectors from a multivariate normal (MVN) distribution with mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})'$ and covariance matrix $\boldsymbol{\Sigma}_i$ $(i = 1, 2)$. In the homoscedastic case, we assume $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ (say). The elements of $\boldsymbol{\Sigma}$ are

$$\sigma_{kk} = \mathrm{Var}(x_{ijk}) \text{ and } \sigma_{k\ell} = \mathrm{Cov}(x_{ijk}, x_{ij\ell}) \quad (1 \leq k < \ell \leq m).$$

[1]Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI 53226, USA. e-mail: `blogan@mcw.edu`

[2]Department of Statistics, Northwestern University, 2006 Sheridan Rd., Evanston, IL 60208, USA. e-mail: `ajit@iems.northwestern.edu`

Keywords and phrases: clinical trials, one-sided multivariate test, homoscedastic, heteroscedastic.

AMS 2000 subject classifications: primary 62J15; secondary 62P10.

The corresponding correlation matrix will be denoted by $\boldsymbol{R}$ with elements

$$\rho_{k\ell} = \text{Corr}(x_{ijk}, x_{ij\ell}) = \frac{\sigma_{k\ell}}{\sqrt{\sigma_{kk}\sigma_{\ell\ell}}} \quad (1 \leq k < \ell \leq m).$$

In the heteroscedastic case, the elements of $\boldsymbol{\Sigma}_i$ will be denoted by $\sigma_{i,k\ell}$ $(1 \leq k \leq \ell \leq m)$ and the corresponding correlation matrices will be denoted by $\boldsymbol{R}_i = \{\rho_{i,k\ell}\}$ $(i = 1, 2)$.

Let $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = (\delta_1, \delta_2, \ldots, \delta_m)'$ denote the vector of mean differences. To establish an overall treatment effect, a global null hypothesis of no difference is tested against a one-sided alternative

$$H_0 : \boldsymbol{\delta} = \mathbf{0} \text{ vs. } H_1 : \boldsymbol{\delta} \in \mathcal{O}^+, \tag{1}$$

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

$$\mathcal{O}^+ = \{\boldsymbol{\delta} | \boldsymbol{\delta} \geq \mathbf{0}, \boldsymbol{\delta} \neq \mathbf{0}\}$$

is the positive orthant.

Let $\overline{\boldsymbol{x}}_{i\cdot} = (\overline{x}_{i\cdot 1}, \overline{x}_{i\cdot 2}, \ldots, \overline{x}_{i\cdot m})'$ denote the vector of sample means of the n_i subjects from the ith group and let $\widehat{\boldsymbol{\Sigma}}_i$ denote the sample covariance matrix from the ith group with $\nu_i = n_i - 1$ degrees of freedom (d.f.) $(i = 1, 2)$. In the homoscedastic case, we use the pooled estimate of $\boldsymbol{\Sigma}$ given by $\widehat{\boldsymbol{\Sigma}} = \{(n_1 - 1)\widehat{\boldsymbol{\Sigma}}_1 + (n_2 - 1)\widehat{\boldsymbol{\Sigma}}_2\}/(n_1 + n_2 - 2)$ with $n_1 + n_2 - 2$ d.f. Denote the elements of $\widehat{\boldsymbol{\Sigma}}$ by $\widehat{\sigma}_{k\ell}$ $(1 \leq k \leq \ell \leq m)$.

3. Homoscedastic case

3.1. OLS and GLS Tests

O'Brien (1984) considered a simplified version of the hypothesis testing problem (1) obtained by restricting the mean difference vector $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ to a ray $\lambda(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{mm}})'$ where $\lambda \geq 0$. In other words, if $\delta_k/\sqrt{\sigma_{kk}} = \lambda_k$ denotes the standardized treatment effect for the kth endpoint then O'Brien assumed that $\lambda_k = \lambda \geq 0$ for all k. In that case the hypothesis testing problem (1) simplifies to

$$H_0 : \lambda = 0 \quad \text{vs.} \quad H_1 : \lambda > 0. \tag{2}$$

O'Brien solved this problem by using a univariate regression framework that models the standardized responses as

$$y_{ijk} = \frac{x_{ijk}}{\sqrt{\sigma_{kk}}} = \frac{\mu_k}{\sqrt{\sigma_{kk}}} + \frac{\lambda}{2} I_{ijk} + \epsilon_{ijk} \quad (i = 1, 2;\ 1 \leq j \leq n_i;\ 1 \leq k \leq m), \tag{3}$$

where $\mu_k = (\mu_{1k} + \mu_{2k})/2, I_{ijk} = +1$ if $i = 1$ and -1 if $i = 2$, and $\epsilon_{ijk} \sim N(0, 1)$ r.v.'s with correlations

$$\text{Corr}(\epsilon_{ijk}, \epsilon_{i'j'\ell}) = \rho_{k\ell} \quad \text{if} \quad i = i' \quad \text{and} \quad j = j', \text{Corr}(\epsilon_{ijk}, \epsilon_{i'j'\ell}) = 0 \quad \text{otherwise.}$$

Note that the vectors $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijm})'$ are independent, each with correlation matrix $\boldsymbol{R} = \{\rho_{k\ell}\}$.

Assuming that $\boldsymbol{R}$ is known, O'Brien showed that the OLS estimate of λ and its standard deviation (SD) equal

$$\widehat{\lambda}_{\text{OLS}} = \frac{\boldsymbol{j}'(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{m} = \overline{y}_{1\cdot\cdot} - \overline{y}_{2\cdot\cdot} \quad \text{and} \quad \text{SD}(\widehat{\lambda}_{\text{OLS}}) = \frac{1}{m}\sqrt{\left(\frac{n_1 + n_2}{n_1 n_2}\right)(\boldsymbol{j}'\boldsymbol{R}\boldsymbol{j})},$$

where $\boldsymbol{j}$ is a vector of all 1's of an appropriate dimension. Therefore the OLS statistic with $\boldsymbol{R}$ replaced by the sample correlation matrix $\widehat{\boldsymbol{R}}$ equals

$$t_{\text{OLS}} = \frac{\widehat{\lambda}}{\widehat{\text{SD}}(\widehat{\lambda})} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left[\frac{\boldsymbol{j}'(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{\sqrt{\boldsymbol{j}'\widehat{\boldsymbol{R}}\boldsymbol{j}}} \right] = \frac{\boldsymbol{j}'\boldsymbol{t}}{\sqrt{\boldsymbol{j}'\widehat{\boldsymbol{R}}\boldsymbol{j}}}, \tag{4}$$

where $\boldsymbol{t}$ is a vector of the t-statistics,

$$t_k = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k}}{\sqrt{\widehat{\sigma}_{kk}}} \right) \quad (1 \le k \le m), \tag{5}$$

for comparing the treatment and control groups on the individual endpoints. Each t_k is marginally t-distributed under H_{0k} with $n_1 + n_2 - 2$ d.f.

Since the errors ϵ_{ijk} in the regression model (3) are correlated, one may prefer the generalized least squares (GLS) estimate of λ (which is also its MLE) to the OLS estimate. Assuming that $\boldsymbol{R}$ is known, O'Brien showed that

$$\widehat{\lambda}_{\text{GLS}} = \frac{\boldsymbol{j}'\boldsymbol{R}^{-1}(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{\boldsymbol{j}'\boldsymbol{R}^{-1}\boldsymbol{j}} \quad \text{and} \quad \text{SD}(\widehat{\lambda}_{\text{GLS}}) = \sqrt{\left(\frac{n_1 + n_2}{n_1 n_2}\right)\left(\frac{1}{\boldsymbol{j}'\boldsymbol{R}^{-1}\boldsymbol{j}}\right)}.$$

The test statistic using this GLS estimate with the estimated correlation matrix $\widehat{\boldsymbol{R}}$ substituted in place of $\boldsymbol{R}$ equals

$$t_{\text{GLS}} = \frac{\widehat{\lambda}}{\widehat{\text{SD}}(\widehat{\lambda})} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\boldsymbol{j}'\widehat{\boldsymbol{R}}^{-1}(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{\sqrt{\boldsymbol{j}'\widehat{\boldsymbol{R}}^{-1}\boldsymbol{j}}} \right) = \frac{\boldsymbol{j}'\widehat{\boldsymbol{R}}^{-1}\boldsymbol{t}}{\sqrt{\boldsymbol{j}'\widehat{\boldsymbol{R}}^{-1}\boldsymbol{j}}}. \tag{6}$$

Both the OLS and GLS statistics are standardized weighted sums of the individual t-statistics for the m endpoints. The OLS statistic uses equal weights, while the GLS statistic uses unequal weights determined by the sample correlation matrix $\widehat{\boldsymbol{R}}$. If some endpoint is highly correlated with the others then the GLS statistic gives a correspondingly lower weight to its t-statistic. The convergence of t_{GLS} to the standard normal distribution is slower than that of t_{OLS} because of the use of the estimated correlation matrix $\widehat{\boldsymbol{R}}$ both in the calculation of $\widehat{\lambda}_{\text{GLS}}$ and $\widehat{\text{SD}}(\widehat{\lambda}_{\text{GLS}})$. Also, the simulation study by Reitmeir and Wassmer (1996) has shown that the powers of the OLS and GLS tests are comparable when used to test subset hypotheses in closed testing procedures. Finally, the linear combination of the t_k-statistics used in the GLS test can have some negative weights, which can lead to anomalous results; this problem does not occur with the OLS test. For all these reasons, the OLS test is preferred.

The exact small sample null distribution of t_{OLS} is intractable. O'Brien (1984) proposed to approximate it by the t-distribution with $n_1 + n_2 - 2m$ d.f. For large sample sizes, the standard normal (z) distribution may be used as an approximation; however, this approximation is liberal for small sample sizes. The t-approximation is exact for $m = 1$ and conservative for $m > 1$ if the d.f. is small. For example, if $n_1 = n_2 = 10$ and $m = 8$, which gives $\nu = 4$, the type I error rate is around 0.025 when nominal $\alpha = 0.05$. Therefore we investigated a better approximation to the d.f. of the t-distribution obtained by empirically matching the second moment with the actual distribution of t_{OLS} (generated via simulation assuming independence of the endpoints). The resulting approximation is given by

$$\nu = 0.5(n_1 + n_2 - 2)(1 + 1/m^2).$$

Table 1: Simulated type I error probability of the OLS test using the proposed approximation for the degrees of freedom of the t-distribution (nominal $\alpha = 0.05$)

		Results for $\rho = 0.0$					Results for $\rho = 0.5$				
		m					m				
n_1	n_2	2	4	6	8	10	2	4	6	8	10
5	5	0.051	0.049	0.053	0.051	0.050	0.043	0.043	0.041	0.040	0.042
10	10	0.051	0.050	0.049	0.047	0.051	0.049	0.046	0.047	0.047	0.047
15	15	0.050	0.052	0.050	0.051	0.050	0.044	0.050	0.044	0.047	0.046
20	20	0.050	0.051	0.046	0.052	0.053	0.049	0.047	0.050	0.049	0.048
25	25	0.051	0.048	0.047	0.050	0.048	0.054	0.050	0.051	0.047	0.046
5	10	0.050	0.050	0.052	0.052	0.048	0.045	0.045	0.042	0.042	0.043
5	15	0.053	0.045	0.049	0.050	0.051	0.045	0.047	0.044	0.048	0.044
5	20	0.053	0.050	0.049	0.048	0.048	0.050	0.048	0.052	0.045	0.044
10	15	0.048	0.052	0.051	0.049	0.052	0.050	0.045	0.048	0.047	0.048
10	20	0.050	0.052	0.051	0.050	0.050	0.044	0.048	0.049	0.049	0.049

All estimates are based on 10,000 replications.

This approximation is exact for $m = 1$. For large m, we get $\nu \approx 0.5(n_1 + n_2 - 2)$. Simulation results in Table 1 using 10,000 simulated datasets indicate that this approximation controls the type I error probability very accurately for uncorrelated endpoints, within $\pm 2\text{SE} = \pm 0.004$ of the nominal 0.05 level for all configurations. For correlated endpoints with equal $\rho = 0.5$, the approximation was found to be slightly conservative for some configurations, with type I error rates ranging between 0.04 and 0.05 for the settings studied. Simulations for other type I error probabilities ($\alpha = 0.01, 0.10$) are omitted for brevity, but they also indicate accurate control of the type I error rate.

3.2. Comparison of the OLS test with Läuter's SS test

Läuter (1996) proposed a class of test statistics for the hypotheses (1) having the property that they are exactly t-distributed with $n_1 + n_2 - 2$ d.f. under H_0. Recall that $\bar{\boldsymbol{x}}_{i\cdot} = (\bar{x}_{i\cdot 1}, \bar{x}_{i\cdot 2}, \ldots, \bar{x}_{i\cdot m})'$ denotes the vector of sample means for the ith group ($i = 1, 2$) and let

$$\bar{\boldsymbol{x}}_{\cdot\cdot} = \frac{n_1 \bar{\boldsymbol{x}}_{1\cdot} + n_2 \bar{\boldsymbol{x}}_{2\cdot}}{n_1 + n_2} = \left(\bar{x}_{\cdot\cdot 1}, \bar{x}_{\cdot\cdot 2}, \ldots, \bar{x}_{\cdot\cdot m}\right)'$$

denote the vector of overall sample means. Define the total cross-products matrix by

$$\boldsymbol{V} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot}\right)\left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot}\right)' = (n_1 + n_2 - 2)\widehat{\boldsymbol{\Sigma}} + \sum_{i=1}^{2} n_i \left(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot}\right)\left(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot}\right)'.$$

Let $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{V})$ be any m-dimensional vector of weights depending solely on $\boldsymbol{V}$ such that $\boldsymbol{w} \neq \boldsymbol{0}$ with probability 1. Using the results from the theory of spherical distributions (Fang and Zhang 1990), Läuter (1996) showed that

$$t_{\boldsymbol{w}} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\boldsymbol{w}'\boldsymbol{t}}{\sqrt{\boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{w}}} \right)$$

is exactly t-distributed with $n_1 + n_2 - 2$ d.f. under H_0. Various choices for $\boldsymbol{w}$ were discussed by Läuter, Kropf and Glimm (1998). We will focus on the standardized sum (SS) statistic (denoted by t_{SS}) for which $\boldsymbol{w} = (1/\sqrt{v_{11}}, 1/\sqrt{v_{22}}, \ldots, 1/\sqrt{v_{mm}})'$, where

$$v_{kk} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(x_{ijk} - \bar{x}_{\cdot\cdot k}\right)^2$$

is the kth diagonal element of $\boldsymbol{V}$.

The SS statistic can be expressed as a t-statistic for comparing the treatment and control groups based on the sum of the standardized observations for each patient

$$y_{ij} = \sum_{k=1}^{m} \frac{x_{ijk}}{\sqrt{v_{kk}}} \quad (i = 1, 2; 1 \le j \le n_i).$$

Thus

$$t_{\mathrm{SS}} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{\hat{\sigma}_y} \right)$$

where

$$\bar{y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad (i = 1, 2) \text{ and } \hat{\sigma}_y = \sqrt{\frac{\sum_{i=1}^{2} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2}{n_1 + n_2 - 2}}.$$

The OLS statistic is the sum of the t_k-statistics (5), which are obtained by standardizing the individual endpoints by their pooled *within* group sample standard deviations. On the other hand, the SS statistic is obtained by standardizing the data on each endpoint by its pooled *total* group sample standard deviation and then computing an overall t-statistic. Because the total pooled standard deviation overestimates the true standard deviation since it includes the between treatment group difference, the power of the SS test would be expected to be lower. We show this in a special case by comparing the powers of the two tests when $n_1 = n_2 = n$ (say) and $n \to \infty$.

The limiting null and non-null distributions of t_{OLS} and t_{SS} are normal, and their asymptotic powers for α-level tests can be expressed as follows (for derivations, see the Appendix). Let

$$a_k = \frac{1}{\sqrt{2\sigma_{kk}}} \quad \text{and} \quad b_k = \frac{1}{\sqrt{(2 + \lambda_k^2/2)\sigma_{kk}}} \quad (1 \le k \le m),$$

where $\lambda_k = \delta_k / \sqrt{\sigma_{kk}}$ as defined before. Then

$$\text{Power}_{\mathrm{OLS}} = \Phi\left(-z_\alpha + \frac{\boldsymbol{a}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} \sqrt{\frac{n}{2}} \right)$$

and

$$\text{Power}_{\mathrm{SS}} = \Phi\left(-z_\alpha + \frac{\boldsymbol{b}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}} \sqrt{\frac{n}{2}} \right),$$

where $\boldsymbol{a} = (a_1, a_2, \ldots, a_m)'$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)'$ and z_α is the $(1-\alpha)$th quantile of the standard normal distribution,.

Therefore

$$\text{Power}_{\text{OLS}} \geq \text{Power}_{\text{SS}} \iff \frac{\boldsymbol{a}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} \geq \frac{\boldsymbol{b}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}}. \tag{7}$$

It is easy to show that

$$\begin{aligned}
\frac{\boldsymbol{a}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} &= \frac{\sum_{k=1}^{m}\lambda_k}{\sqrt{\sum_{k=1}^{m}\sum_{\ell=1}^{m}\rho_{k\ell}}} \quad \text{and} \\
\frac{\boldsymbol{b}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}} &= \frac{\sum_{k=1}^{m}\lambda_k/\sqrt{1+\lambda_k^2/4}}{\sqrt{\sum_{k=1}^{m}\sum_{\ell=1}^{m}\rho_{k\ell}/\sqrt{(1+\lambda_k^2/4)(1+\lambda_\ell^2/4)}}},
\end{aligned}$$

where $\rho_{k\ell} = 1$ if $k = \ell$. Comparison of the powers of the two tests reduces to comparing the two expressions above.

Consider the case $\lambda_1 > 0$ and $\lambda_k = 0$ for $k > 1$. Then we have

$$\frac{\boldsymbol{a}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} = \frac{\lambda_1}{\sqrt{\sum_{k=1}^{m}\sum_{\ell=1}^{m}\rho_{k\ell}}}$$

and

$$\frac{\boldsymbol{b}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}} = \frac{\lambda_1/\sqrt{1+\lambda_1^2/4}}{\sqrt{\sum_{k=2}^{m}\sum_{\ell=2}^{m}\rho_{k\ell} + 2\sum_{k=2}^{m}\left(\rho_{1k}/\sqrt{1+\lambda_1^2/4}\right) + 1/(1+\lambda_1^2/4)}}.$$

Simple algebra shows that the second inequality in (7) is strict in this case. Thus, if only one endpoint has a positive treatment effect then the OLS test is asymptotically more powerful to detect this effect than the SS test. In fact,

$$\begin{aligned}
&\lim_{\lambda_1\to\infty}\frac{\boldsymbol{b}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}} \\
&= \lim_{\lambda_1\to\infty}\frac{\lambda_1/\sqrt{1+\lambda_1^2/4}}{\sqrt{\sum_{k=2}^{m}\sum_{\ell=2}^{m}\rho_{k\ell} + 2\sum_{k=2}^{m}\left(\rho_{1k}/\sqrt{1+\lambda_1^2/4}\right) + 1/(1+\lambda_1^2/4)}} \\
&= \frac{2}{\sqrt{\sum_{k=2}^{m}\sum_{\ell=2}^{m}\rho_{k\ell}}} < \infty.
\end{aligned}$$

Therefore the asymptotic power of the SS test is strictly less than 1 when $\lambda_1 \to \infty$. This undesirable property of the SS test has been noted by Frick (1996).

Next consider the case $\lambda_k = \lambda > 0$ for all k, which is the assumption underlying the OLS test. Here we have

$$\frac{\boldsymbol{a}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} = \frac{\boldsymbol{b}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}} = \frac{m\lambda}{\sqrt{\sum_{k=1}^{m}\sum_{\ell=1}^{m}\rho_{k\ell}}},$$

and therefore $\text{Power}_{\text{OLS}} = \text{Power}_{\text{SS}}$ asymptotically. Note that this configuration is typically of most interest, since both tests are designed to have high power when $\lambda_k = \lambda$ for all k, and are not necessarily designed to perform well when the treatment effects are highly variable.

Table 2: Simulated powers of the OLS and SS tests (no. of replications = 10,000, $\alpha = 0.05$)

			$\rho = 0.0$		$\rho = 0.5$	
m	n	$\boldsymbol{\delta}'$	OLS	SS	OLS	SS
4	10	(3,0,0,0)	0.920	0.614	0.628	0.323
		(1.5,1.5,0,0)	0.930	0.894	0.633	0.564
		(1.0,1.0,1.0,0)	0.931	0.930	0.635	0.634
		(0.7,0.7,0.7,0.7)	0.905	0.908	0.581	0.602
		(1.0,1.0,0.5,0.5)	0.935	0.933	0.637	0.646
	50	(1.2,0,0,0)	0.904	0.845	0.594	0.516
		(0.6,0.6,0,0)	0.909	0.903	0.594	0.582
		(0.4,0.4,0.4,0)	0.912	0.912	0.584	0.585
		(0.3,0.3,0.3,0.3)	0.903	0.904	0.588	0.591
		(0.4,0.4,0.2,0.2)	0.911	0.911	0.595	0.596
8	10	(2,0,0,0)	0.903	0.747	0.393	0.259
		(1,1,0,0)	0.908	0.892	0.393	0.372
		(0.7,0.7,0.7,0)	0.931	0.930	0.426	0.437
		(0.5,0.5,0.5,0.5)	0.912	0.913	0.399	0.416
		(0.6,0.6,0.3,0.3)	0.858	0.858	0.349	0.363
	50	(0.9,0,0,0)	0.932	0.907	0.441	0.401
		(0.4,0.4,0,0)	0.882	0.878	0.374	0.372
		(0.3,0.3,0.3,0)	0.936	0.936	0.431	0.431
		(0.2,0.2,0.2,0.2)	0.876	0.876	0.382	0.379
		(0.3,0.3,0.15,0.15)	0.934	0.933	0.437	0.439

The $\boldsymbol{\delta}$-vector for $m = 8$ equals the two $\boldsymbol{\delta}$ vectors for $m = 4$ put together, i.e., $\boldsymbol{\delta}'_8 = (\boldsymbol{\delta}'_4, \boldsymbol{\delta}'_4)$.

Table 2 gives simulation results for the powers of the OLS and SS tests conducted at $\alpha = 0.05$ for some selected cases with $m = 4$ and 8 endpoints. The n i.i.d. data vectors for the treatment group $\boldsymbol{x}_{1j}$, $j = 1, \ldots, n$, are each generated from an MVN$(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}_{ii} = 1$ and $\boldsymbol{\Sigma}_{ij} = \rho$ for $i \neq j$. Similarly, the n i.i.d. vectors for the control group, $\boldsymbol{x}_{2j}$, $j = 1, \ldots, n$, are generated from an MVN$(\mathbf{0}, \boldsymbol{\Sigma})$ distribution. Correlation values of $\rho = 0$ and 0.5, and sample sizes of $n = 10$ and 50 were investigated. Four dimensional vectors denoting $\boldsymbol{\delta}$ are given in the table. The $\boldsymbol{\delta}$ vector for $m = 8$ equals the two $\boldsymbol{\delta}$ vectors for $m = 4$ put together, i.e., $\boldsymbol{\delta}_8 = (\boldsymbol{\delta}'_4, \boldsymbol{\delta}'_4)'$. A total of $N = 10,000$ replications were generated for each simulation run.

We see that the difference in the powers of the OLS and SS tests is not very large for most configurations. When there is a large treatment effect on one or two endpoints but no treatment effect on the other endpoints, the SS test suffers from the anomalies discussed before, resulting in a substantial loss of power relative to the OLS test; however, the OLS test is also not suited well for this configuration. When all the endpoints have an effect, the correlation is moderate, and the sample size is small, the SS test tends to perform slightly better than the OLS test (1–2% higher power). This is probably due to the slightly conservative nature of the t-approximation to the OLS statistic for correlated endpoints. In all other situations, the two procedures perform similarly, and both are adequately suited for comparing two groups when the treatment effect is expected to be similar across endpoints.

4. Heteroscedastic case

Pocock, Geller and Tsiatis (1987) proposed an ad-hoc extension of O'Brien's GLS test to the heteroscedastic case as follows. Assume that $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are known. Then the statistic for comparing the treatment with the control on the kth endpoint is

$$z_k = \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k}}{\sqrt{\sigma_{1,kk}/n_1 + \sigma_{2,kk}/n_2}} \quad (1 \le k \le m). \tag{8}$$

Let $\boldsymbol{z} = (z_1, z_2, \ldots, z_m)'$ and $\bar{\boldsymbol{R}} = (n_1\boldsymbol{R}_1 + n_2\boldsymbol{R}_2)/(n_1 + n_2)$. In analogy with (6), Pocock et al. proposed the statistic

$$z_{\text{GLS}} = \frac{\boldsymbol{j}'\bar{\boldsymbol{R}}^{-1}\boldsymbol{z}}{\sqrt{\boldsymbol{j}'\bar{\boldsymbol{R}}^{-1}\boldsymbol{j}}}.$$

Unfortunately, this statistic does not have the standard normal distribution under H_0 as claimed by Pocock et al. because the covariance (correlation) matrix of $\boldsymbol{z}$ is not $\bar{\boldsymbol{R}}$, but $\mathbf{\Gamma} = \{\gamma_{k\ell}\}$ with elements

$$\gamma_{k\ell} = \frac{\sigma_{1,k\ell}/n_1 + \sigma_{2,k\ell}/n_2}{\sqrt{(\sigma_{1,kk}/n_1 + \sigma_{2,kk}/n_2)(\sigma_{1,\ell\ell}/n_1 + \sigma_{2,\ell\ell}/n_2)}} \quad (1 \le k < \ell \le m).$$

In the following we correctly derive the OLS and GLS tests in the heteroscedastic case.

4.1. OLS test

We use the following definition for the standardized treatment effect

$$\lambda_k = \frac{\delta_k}{\sqrt{\sigma_{1,kk} + \sigma_{2,kk}}} \quad (1 \le k \le m).$$

As in O'Brien (1984), assume that $\lambda_k = \lambda \ge 0$ for all k. To test the hypotheses (2), standardize the observations as

$$y_{ijk} = \frac{x_{ijk}}{\sqrt{\sigma_{1,kk} + \sigma_{2,kk}}} \quad (i = 1, 2;\ 1 \le j \le n_i; 1 \le k \le m).$$

Then $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijm})'$ are independently distributed as $\text{MVN}(\boldsymbol{\xi}_i, \mathbf{\Gamma}_i)$ where $\boldsymbol{\xi}_i$ has the elements

$$\xi_{ik} = \frac{\mu_{ik}}{\sqrt{\sigma_{1,kk} + \sigma_{2,kk}}} \quad (1 \le k \le m)$$

and $\mathbf{\Gamma}_i$ has the elements

$$\gamma_{i,k\ell} = \frac{\sigma_{i,k\ell}}{\sqrt{(\sigma_{1,kk} + \sigma_{2,kk})(\sigma_{1,\ell\ell} + \sigma_{2,\ell\ell})}} \quad (i = 1, 2;\ 1 \le k \le \ell \le m).$$

Note that $\xi_{1k} - \xi_{2k} = \lambda$ for all k. Also note that $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are not correlation matrices, and $\mathbf{\Gamma} = \mathbf{\Gamma}_1 + \mathbf{\Gamma}_2$ if $n_1 = n_2$.

The hypotheses (2) can be tested by using a univariate regression framework analogous to (3)

$$y_{ijk} = \xi_k + \frac{\lambda}{2} I_{ijk} + \epsilon_{ijk} \quad (i = 1, 2; 1 \le j \le n_i;\ 1 \le k \le m), \tag{9}$$

where $\xi_k = (\xi_{1k} + \xi_{2k})/2, I_{ijk} = +1$ if $i = 1$ and -1 if $i = 2$, and $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \ldots, \epsilon_{ijm})'$ are independently distributed as $N(\mathbf{0}, \boldsymbol{\Gamma}_i)$.

Let $\beta = \lambda/2$ and let $\boldsymbol{\theta} = (\beta, \xi_1, \ldots, \xi_m)'$ be the vector of unknown parameters. Then the above model can be written as

$$\boldsymbol{y} = \boldsymbol{D\theta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y} = (\boldsymbol{y}'_{11}, \ldots, \boldsymbol{y}'_{1n_1}, \boldsymbol{y}'_{21}, \ldots, \boldsymbol{y}'_{2n_2})'$, $\boldsymbol{D} = (\underbrace{\boldsymbol{D}'_1, \ldots, \boldsymbol{D}'_1}_{n_1}, \underbrace{\boldsymbol{D}'_2, \ldots, \boldsymbol{D}'_2}_{n_2})'$, $\boldsymbol{\theta} = (\beta, \xi_1, \ldots, \xi_m)'$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}'_{11}, \ldots, \boldsymbol{\epsilon}'_{1n_1}, \boldsymbol{\epsilon}'_{21}, \ldots, \boldsymbol{\epsilon}'_{2n_2})'$. In the above, $\boldsymbol{D}_1 = (\boldsymbol{j}, \boldsymbol{I})$ and $\boldsymbol{D}_2 = (-\boldsymbol{j}, \boldsymbol{I})$ where $\boldsymbol{j}$ is an m-dimensional vector of 1's and $\boldsymbol{I}$ is the identity matrix of dimension m.

The OLS estimator of β is the first component of $\widehat{\boldsymbol{\theta}} = (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{y}$. Now,

$$\boldsymbol{D}'\boldsymbol{D} = \begin{bmatrix} (n_1 + n_2)m & (n_1 - n_2)\boldsymbol{j}' \\ (n_1 - n_2)\boldsymbol{j} & (n_1 + n_2)\boldsymbol{I} \end{bmatrix}.$$

The first row of $(\boldsymbol{D}'\boldsymbol{D})^{-1}$ required to compute $\widehat{\beta}$ equals

$$\left(\frac{n_1 + n_2}{4n_1 n_2 m}, \frac{-(n_1 - n_2)\boldsymbol{j}'}{4n_1 n_2 m} \right).$$

Also,

$$\boldsymbol{D}'\boldsymbol{y} = \begin{bmatrix} \boldsymbol{j}'(n_1\overline{\boldsymbol{y}}_{1\cdot} - n_2\overline{\boldsymbol{y}}_{2\cdot}) \\ n_1\overline{\boldsymbol{y}}_{1\cdot} + n_2\overline{\boldsymbol{y}}_{2\cdot} \end{bmatrix},$$

where $\overline{\boldsymbol{y}}_{1\cdot}$ and $\overline{\boldsymbol{y}}_{2\cdot}$ are the vectors of sample means of the standardized data. Hence,

$$\widehat{\beta} = \left[(\boldsymbol{D}'\boldsymbol{D})^{-1} \boldsymbol{D}'\boldsymbol{y} \right]_1 = \frac{\boldsymbol{j}'(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{2m}.$$

So the OLS estimate of λ and its standard deviation equal

$$\widehat{\lambda} = 2\widehat{\beta} = \frac{\boldsymbol{j}'(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{m} \quad \text{and} \quad \mathrm{SD}(\widehat{\lambda}) = \frac{\{\boldsymbol{j}'(\boldsymbol{\Gamma}_1/n_1 + \boldsymbol{\Gamma}_2/n_2)\boldsymbol{j}\}^{1/2}}{m}.$$

Then the OLS test statistic, using the estimated covariance matrices, is

$$t_{\mathrm{OLS}} = \frac{\widehat{\lambda}}{\widehat{\mathrm{SD}}(\widehat{\lambda})} = \frac{\boldsymbol{j}'(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{\left\{ \boldsymbol{j}'(\widehat{\boldsymbol{\Gamma}}_1/n_1 + \widehat{\boldsymbol{\Gamma}}_2/n_2)\boldsymbol{j} \right\}^{1/2}},$$

where $\widehat{\boldsymbol{\Gamma}}_i = \{\widehat{\gamma}_{i,kl}\}$ and

$$\widehat{\gamma}_{i,k\ell} = \frac{\widehat{\sigma}_{i,k\ell}}{\sqrt{(\widehat{\sigma}_{1,kk} + \widehat{\sigma}_{2,kk})(\widehat{\sigma}_{1,\ell\ell} + \widehat{\sigma}_{2,\ell\ell})}}.$$

This statistic is asymptotically standard normal under H_0.

Let

$$t_k = \frac{(\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k})}{\sqrt{\widehat{\sigma}_{1,kk}/n_1 + \widehat{\sigma}_{2,kk}/n_2}} \quad (1 \le k \le m)$$

be the t-statistics for comparing the treatment and control groups on the individual endpoints. They are marginally approximately t-distributed under H_{0k} with d.f. estimated by the Welch-Satterthwaite formula

$$\nu_k = \frac{(\widehat{\sigma}_{1,kk}/n_1 + \widehat{\sigma}_{2,kk}/n_2)^2}{\widehat{\sigma}^2_{1,kk}/n_1^2(n_1-1) + \widehat{\sigma}^2_{2,kk}/n_2^2(n_2-1)} \quad (1 \le k \le m).$$

For $n_1 = n_2 = n$, analogous to (4), the t_{OLS} test statistic simplifies to

$$t_{\text{OLS}} = \frac{\widehat{\lambda}}{\widehat{\text{SD}}(\widehat{\lambda})} = \frac{\boldsymbol{j}'\boldsymbol{t}}{(\boldsymbol{j}'\widehat{\boldsymbol{\Gamma}}\boldsymbol{j})^{1/2}},$$

where $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}}_1 + \widehat{\boldsymbol{\Gamma}}_2$ is the sample estimate of the correlation matrix $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2$ between the numerators of the t_k statistics.

4.2. GLS test

Next we obtain the generalized least squares (GLS) estimate of λ. The GLS estimate of $\boldsymbol{\theta}$ is given by $(\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{y}$, where $\boldsymbol{V}$ is the covariance matrix of the ϵ's, which has a block diagonal structure given by

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Gamma}_2 \end{bmatrix}.$$

Then

$$\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D} = \begin{bmatrix} n_1\boldsymbol{j}'\boldsymbol{\Gamma}_1^{-1}\boldsymbol{j} + n_2\boldsymbol{j}'\boldsymbol{\Gamma}_2^{-1}\boldsymbol{j} & n_1\boldsymbol{j}'\boldsymbol{\Gamma}_1^{-1} - n_2\boldsymbol{j}'\boldsymbol{\Gamma}_2^{-1} \\ n_1\boldsymbol{\Gamma}_1^{-1}\boldsymbol{j} - n_2\boldsymbol{\Gamma}_2^{-1}\boldsymbol{j} & n_1\boldsymbol{\Gamma}_1^{-1} + n_2\boldsymbol{\Gamma}_2^{-1} \end{bmatrix}.$$

The first row of $(\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}$ required to compute $\widehat{\beta}$ equals

$$\left(\frac{1}{d}, \frac{-\boldsymbol{j}'\boldsymbol{C}}{d}\right),$$

where

$$\begin{aligned} \boldsymbol{C} &= \left(n_1\boldsymbol{\Gamma}_1^{-1} - n_2\boldsymbol{\Gamma}_2^{-1}\right)\left(n_1\boldsymbol{\Gamma}_1^{-1} + n_2\boldsymbol{\Gamma}_2^{-1}\right)^{-1} \quad \text{and} \\ d &= \boldsymbol{j}'\left[(\boldsymbol{I}-\boldsymbol{C})\boldsymbol{\Gamma}_1^{-1}/n_1 + (\boldsymbol{I}+\boldsymbol{C})\boldsymbol{\Gamma}_2^{-1}/n_2\right]\boldsymbol{j}. \end{aligned}$$

Then

$$\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{y} = \begin{bmatrix} \boldsymbol{j}'\left(n_1\boldsymbol{\Gamma}_1^{-1}\overline{\boldsymbol{y}}_{1\cdot} - n_2\boldsymbol{\Gamma}_2^{-1}\overline{\boldsymbol{y}}_{2\cdot}\right) \\ n_1\boldsymbol{\Gamma}_1^{-1}\overline{\boldsymbol{y}}_{1\cdot} + n_2\boldsymbol{\Gamma}_2^{-1}\overline{\boldsymbol{y}}_{2\cdot} \end{bmatrix}$$

and

$$\widehat{\beta} = \left[(\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{y}\right]_1 = \frac{2\boldsymbol{j}'(\boldsymbol{\Gamma}_1/n_1 + \boldsymbol{\Gamma}_2/n_2)^{-1}(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{d}.$$

So the GLS estimate of λ and its standard deviation equal

$$\widehat{\lambda} = \frac{4\boldsymbol{j}'(\boldsymbol{\Gamma}_1/n_1 + \boldsymbol{\Gamma}_2/n_2)^{-1}(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{d} \quad \text{and}$$

$$\mathrm{SD}(\widehat{\lambda}) = \frac{4\{\boldsymbol{j}'(\boldsymbol{\Gamma}_1/n_1 + \boldsymbol{\Gamma}_2/n_2)^{-1}\boldsymbol{j}\}^{1/2}}{d}.$$

Hence the GLS test statistic, using the estimated covariance matrices, is

$$t_{\mathrm{GLS}} = \frac{\widehat{\lambda}}{\widehat{\mathrm{SD}}(\widehat{\lambda})} = \frac{\boldsymbol{j}'(\widehat{\boldsymbol{\Gamma}}_1/n_1 + \widehat{\boldsymbol{\Gamma}}_2/n_2)^{-1}(\overline{\boldsymbol{y}}_{1\cdot} - \overline{\boldsymbol{y}}_{2\cdot})}{\{\boldsymbol{j}'(\widehat{\boldsymbol{\Gamma}}_1/n_1 + \widehat{\boldsymbol{\Gamma}}_2/n_2)^{-1}\boldsymbol{j}\}^{1/2}}.$$

This statistic is also asymptotically standard normal under H_0. However, because it uses estimates of the covariance matrices in the weights, it has a slower convergence to the standard normal.

Our simulations show that use of the standard normal critical points in performing the t_{OLS} or t_{GLS} tests give too high type I error rates for small sample sizes ($n_1, n_2 < 50$). Unfortunately, better small sample approximations are not available at this time.

In the case of equal sample sizes, analogous to (6), this reduces to

$$t_{\mathrm{GLS}} = \frac{\boldsymbol{j}'\widehat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{t}}{(\boldsymbol{j}'\widehat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{j})^{1/2}}$$

with $\boldsymbol{t}$ and $\widehat{\boldsymbol{\Gamma}}$ defined as above. We see that, as in the homoscedastic case, under equal sample sizes, both methods are based on weighted sums of the t-statistics for testing each endpoint individually. The OLS statistic uses equal weights, while the GLS statistic uses unequal weights determined by the two covariance matrices.

5. Concluding remarks

In this paper we presented some refinements and extensions of the OLS and GLS tests. These tests are thus made more widely applicable. In future research it would be useful to find a good small sample approximation to the critical points of t_{OLS} and t_{GLS} in the heteroscedastic case.

Appendix

Derivation of the power expressions for Läuter's SS test and O'Brien's OLS test

Let

$$\overline{y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{k=1}^{m}\frac{x_{ijk}}{\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ijk} - \overline{x}_{\cdot\cdot k})^2}} = \sum_{k=1}^{m}\frac{\overline{x}_{i\cdot k}}{\sqrt{\mathrm{SST}_{kk}}},$$

where $\mathrm{SST}_{kk} = v_{kk}$ is the corrected total sum of squares for the kth endpoint. Then Läuter's SS test statistic equals

$$t_{\mathrm{SS}} = \frac{\overline{y}_{1\cdot} - \overline{y}_{2\cdot}}{\widehat{\mathrm{SD}}(\overline{y}_{1\cdot} - \overline{y}_{2\cdot})}.$$

Thus the SS test statistic is a standardized version of

$$\overline{y}_{1\cdot} - \overline{y}_{2\cdot} = \sum_{k=1}^{m}\frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k}}{\sqrt{\mathrm{SST}_{kk}}}.$$

In contrast, the OLS test statistic is a standardized version of

$$\overline{z}_{1\cdot} - \overline{z}_{2\cdot} = \sum_{k=1}^{m} \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k}}{\sqrt{\sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ijk} - \overline{x}_{i\cdot k})^2}} = \sum_{k=1}^{m} \frac{\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k}}{\sqrt{\mathrm{SSE}_{kk}}},$$

where SSE_{kk} is the pooled error sum of squares for the kth endpoint. Note that the OLS statistic uses the within group sum of squares to scale each endpoint, while the SS statistic uses the total sum of squares.

We next examine the asymptotic distribution of each test statistic. Assuming $n_1 = n_2 = n$ for simplification, note that

$$\boldsymbol{u}_n = \sqrt{n}\big(\overline{\boldsymbol{x}}_{1\cdot} - \overline{\boldsymbol{x}}_{2\cdot}\big) \sim \mathrm{MVN}\Big(\sqrt{n}\boldsymbol{\delta}, 2\boldsymbol{\Sigma}\Big).$$

Now consider Läuter's test. First, for large n,

$$\begin{aligned} E(\mathrm{SST}_{kk}) &= E\Big(\sum_{i,j}(x_{ijk} - \overline{x}_{i\cdot k})^2\Big) + E\Big(\sum_{i,j}(\overline{x}_{i\cdot k} - \overline{x}_{\cdot\cdot k})^2\Big) \\ &= 2(n-1)\sigma_{kk} + \sigma_{kk} + \frac{n\delta_k^2}{2} \\ &\approx n\sigma_{kk}\big(2 + \lambda_k^2/2\big), \end{aligned}$$

where $\lambda_k = \delta_k / \sqrt{\sigma_{kk}}$. We know that

$$c_{k,n} = \sqrt{\frac{n}{\mathrm{SST}_{kk}}} \xrightarrow{p} c_k = \frac{1}{\sqrt{(2+\lambda_k^2/2)\sigma_{kk}}} \quad \text{for } k = 1, \ldots, m.$$

Let $\boldsymbol{c}_n = (c_{1,n}, \ldots, c_{m,n})$ and $\boldsymbol{c} = (c_1, \ldots, c_m)$. Then by Slutsky's Theorem,

$$n \sum_{k=1}^{m} \frac{(\overline{x}_{1\cdot k} - \overline{x}_{2\cdot k})}{\sqrt{v_{kk}}} = \boldsymbol{c}_n' \boldsymbol{u}_n \xrightarrow{\mathcal{L}} N\Big(\sqrt{n}\boldsymbol{c}'\boldsymbol{\delta}, 2\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}\Big)$$

and therefore,

$$\overline{y}_{1\cdot} - \overline{y}_{2\cdot} = \frac{\boldsymbol{c}_n'\boldsymbol{u}_n}{n} \xrightarrow{\mathcal{L}} N\left(\frac{\boldsymbol{c}'\boldsymbol{\delta}}{\sqrt{n}}, \frac{2\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}}{n^2}\right).$$

Thus, under H_1,

$$t_{\mathrm{SS}} \xrightarrow{\mathcal{L}} N\left(\frac{\boldsymbol{c}'\boldsymbol{\delta}\sqrt{n}}{\sqrt{2\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}}}, 1\right).$$

Next consider O'Brien's test. Since

$$E(\mathrm{SSE}_k) \approx 2n\sigma_{kk},$$

for large n, we know that

$$d_{k,n} = \sqrt{\frac{n}{\mathrm{SSE}_k}} \xrightarrow{p} d_k = \frac{1}{\sqrt{2\sigma_{kk}}} \quad \text{for } k = 1, \ldots, m.$$

Let $\boldsymbol{d}_n = (d_{1,n}, \ldots, d_{m,n})$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$. Then by Slutsky's Theorem,

$$\boldsymbol{d}_n'\boldsymbol{u}_n \xrightarrow{\mathcal{L}} N\Big(\sqrt{n}\boldsymbol{d}'\boldsymbol{\delta}, 2\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}\Big),$$

and therefore

$$\bar{z}_{1\cdot} - \bar{z}_{2\cdot} = \frac{\boldsymbol{d}_n' \boldsymbol{u}_n}{n} \xrightarrow{\mathcal{L}} N\left(\frac{\boldsymbol{d}'\boldsymbol{\delta}}{\sqrt{n}}, \frac{2\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}}{n^2}\right).$$

Thus, under H_1,

$$t_{\text{OLS}} \xrightarrow{\mathcal{L}} N\left(\frac{\boldsymbol{d}'\boldsymbol{\delta}\sqrt{n}}{\sqrt{2\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}}}, 1\right).$$

The asymptotic power of the Läuter test is

$$\begin{aligned}
\text{Power}_{\text{SS}} &= P(t_{\text{SS}} > z_\alpha | \boldsymbol{\delta}) \\
&= 1 - \Phi\left(z_\alpha - \frac{\boldsymbol{c}'\boldsymbol{\delta}\sqrt{n}}{\sqrt{2\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}}}\right) \\
&= \Phi\left(-z_\alpha + \frac{\boldsymbol{c}'\boldsymbol{\delta}\sqrt{n}}{\sqrt{2\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}}}\right).
\end{aligned}$$

Similarly, the asymptotic power of the O'Brien test is

$$\begin{aligned}
\text{Power}_{\text{OLS}} &= P(t_{\text{OLS}} > z_\alpha | \boldsymbol{\delta}) \\
&= 1 - \Phi\left(z_\alpha - \frac{\boldsymbol{d}'\boldsymbol{\delta}\sqrt{n}}{\sqrt{2\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}}}\right) \\
&= \Phi\left(-z_\alpha + \frac{\boldsymbol{d}'\boldsymbol{\delta}\sqrt{n}}{\sqrt{2\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}}}\right).
\end{aligned}$$

References

[1] Fang, K.-T. and Zhang, Y.-T. (1990). *Generalized Multivariate Analysis.* Berlin, Heidelberg, Springer. MR1079542

[2] Frick, H. (1996). On the power behaviour of Läuter's exact multivariate one-sided tests. *Biom. J.* **38**, 405–414. MR1405666

[3] Läuter, J. (1996). Exact t and F tests for analyzing studies with multiple endpoints. *Biometrics* **52**, 964–970. MR1411742

[4] Läuter, J., Kropf, S. and Glimm, E. (1998). Exact stable multivariate tests for applications in clinical research. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 46–55.

[5] O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, 1079–1087. MR786180

[6] Pocock, S. J., Geller, N. L. and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* **43**, 487–498. MR909756

[7] Reitmeir, P. and Wassmer, G. (1996). One-Sided multiple endpoint testing in two-sample comparisons. *Comm. Statist. Simulation Comput.* **25**, 99–117.