Single and multiple comparison procedures for partial covariance matrices of two treatment groups in clinical trials

Yoshiomi Nakazuru and Takashi Seo

Abstract. The main focus of therapeutic confirmatory clinical trials, generally, is to demonstrate the efficacy of a new or experimental treatment in terms of mean(s) of the primary endpoint(s). In some cases, however, variance(s) of the endpoints also may be the focus of interest. For example, given two treatments with equal efficacy in terms of the means, a treatment with smaller variability, and thus more predictable efficacy, may be preferable. In this paper we consider single and multiple comparison procedures for partial covariance matrices of endpoints in clinical trials that can be used to demonstrate the superiority of a new treatment to an active comparator in terms of the variability. First, we review and discuss the existing methods for comparing two covariance matrices based on the union-intersection test procedure. Second, we propose a single comparison procedure and a multiple comparison procedure for partial covariance matrices that limits the number of comparisons and compare its performance with the tests discussed in the first section, using Monte Carlo simulations. The simulation results suggest that power of the proposed procedure is generally higher than those of the previous methods, while keeping the type I error rates nearly within the nominal level.

AMS 2010 *Mathematics Subject Classification.* 62H15, 62H10.

Key words and phrases. Clinical trial, covariance matrix, multivariate normal distribution.

§1.   Introduction

We consider the following situation in a clinical development of an experimental treatment. First, we assume that efficacy of the new or experimental treatment has been established in a placebo-controlled trial. In addition, we assume that non-inferiority of the new treatment to an active comparator has

been established. Given these two pieces of evidence, we seek for differentiation of the new treatment to the active comparator. Such differentiation is important because it can provide additional benefit or mitigate unnecessary risk for the patients. Possible differentiation to an active comparator is, for example, better safety profile or greater convenience to use. Another possible point of differentiation might be smaller variability.

For example, Rothwell et al. (2010) points out "Visit-to-visit variability in systolic blood pressure (SBP) and maximum SBP are strong predictors of stroke, independent of mean SBP. Increased residual variability in SBP in patients with treated hypertension is associated with a high risk of vascular events.". In Danne et al. (2008), the within-subject variability of detemir and insulin glargine in subjects with type 1 diabetes mellitus was evaluated in a crossover clinical trial. In the statistical analyses of this study, "the null hypothesis was that the within-subject variability was the same for the two insulin preparations (i.e., the ratio between the variances was equal to 1)" (Danne et al. (2008)). Chow and Liu (2000) gives an example in which variability of the PK parameter is compared between test and reference formulations. Their reason is that "the safety of the test formulation may be of concern, and the exchangeability between two formulations is questionable, even when the two formulations are equivalent in average bioavailability" (Chow and Liu (2000)).

In this paper, we focus on differentiation based on variability of effect. That is, a treatment with smaller variability, and thus more predictable efficacy, is preferable given two treatments with equal efficacy in terms of the means. For the problem at hand, standard tests of difference of means or of location parameters (e.g. a t-test) clearly do not provide statistical power to differentiate the two treatments. What we require is a test or a confidence interval of a parameter that characterizes variability.

In the framework of a univariate k-sample problem, Hartley (1950) proposed a procedure for the equality of variances in a one-way layout which is based on the union-intersection principle. This test can be used to make all pairwise comparisons among the variances. Also, O'Brien test, Brown-Forsythe test, Levene test, and Bartlett test can be used for test of equality of variances. According to Proust (2012), "The concept behind the first three tests of equal variances is to perform an analysis of variance on a new response variable constructed to measure the spread in each group. The fourth test is the Bartlett's test, which is similar to the likelihood-ratio test under normal distributions" (Proust (2012)).

For the multivariate version of the k-sample problem, a likelihood ratio test and a modified likelihood ratio test are available as tests for the equality of variance-covariance matrices (see, e.g., Nagao (1973), Muirhead (1982)). These tests are affected by larger elements of the covariance matrix because all

variances are assumed to have the same weight. Another point is that variance is not standardized as it is for the test of means. For example, if a covariance matrix consist of endpoints with different units that are not interconvertible, the former property may become an issue.

Specifically, for a multivariate two-sample problem, Siotani et al. (1985) proposed a test of $\{H_0 : (\boldsymbol{a}^t\boldsymbol{\Sigma}_1\boldsymbol{a})/(\boldsymbol{a}^t\boldsymbol{\Sigma}_2\boldsymbol{a}) = 1,$ for all non-zero vector $\boldsymbol{a}$ versus $H_1 : (\boldsymbol{a}^t\boldsymbol{\Sigma}_1\boldsymbol{a})/(\boldsymbol{a}^t\boldsymbol{\Sigma}_2\boldsymbol{a}) \neq 1,$ for at least one non-zero vector $\boldsymbol{a}\}$ based on the union-intersection principle, where $\boldsymbol{\Sigma}_i$ is a covariance matrix. This test can assign separate weight to each element of covariance matrix with suitable choice of $\boldsymbol{a}$. They show that multiple comparisons can be conducted with any non-zero vector $\boldsymbol{a}$ while maintaining the type I error rate within the nominal level. The test, however, has important practical shortcomings. One is the computational inconvenience and another is the potentially low statistical power due to an allowance for unlimited number of non-zero vector $\boldsymbol{a}$, as in the case of the Scheffe method for comparison of means in ANOVA.

In this paper, we propose a weighted test for covariance matrices for a single comparison, and for multiple comparisons, a procedure for partial covariance matrices that limits the number of comparisons. Both procedures are computationally simple to carryout. The performance of the proposed multiple comparison procedure is compared with that of the existing methods using Monte Carlo simulations. Discussions and concluding remarks are presented at the end.

§2.   Procedures for comparing two partial covariance matrices

In the first section, we briefly review previous methods for comparing partial covariance matrices that do not limit the number of comparisons. In the subsequent two sections, we present the proposed weighted test for covariance matrices for a single comparison and a multiple comparison procedure for partial covariance matrices that limits the number of comparisons.

We consider a clinical trial with two treatment groups: the control (group 1) and the test (group 2). Let n_1 and n_2 denote the number of subjects in the respective treatment groups. We assume that the benefit (i.e. differentiation) of the test over the control is demonstrated when the variance of the test group is smaller than that of the control group. Let p denote the number of endpoints or timepoints that define the dimension of the problem. Let $\boldsymbol{x}_j^{(i)}$ denote a sample vector $(p \times 1)$ of the jth subject of the ith treatment group $(i = 1, 2; j = 1, \cdots, n_i)$ which is assumed to be independent and identically distributed as normal random vector with a mean vector $\boldsymbol{\mu}_i$ and a positive definite covariance matrix $\boldsymbol{\Sigma}_i$. Let $\bar{\boldsymbol{x}}^{(i)}$ and $\boldsymbol{S}_i$ denote the sample mean vector and the sample covariance matrix of the ith treatment. We note that $(n_i-1)\boldsymbol{S}_i$

has as a Wishart distribution $W_p(\mathbf{\Sigma}_i, n_i - 1)$. Let $\boldsymbol{a}$, $\boldsymbol{a}_j$, $\boldsymbol{b}$, and $\boldsymbol{b}_j$ be arbitrary vectors $(p \times 1)$ which will be used to extract partial covariance matrix of interest from $\mathbf{\Sigma}_i$. We assume the above setting throughout the rest of this paper.

2.1. Previous procedure for comparison of two covariance matrices

Siotani et al. (1985) proposed a test for multivariate two sample problems. Their test is based on the union-intersection principle. Multiple comparisons can be conducted with any non-zero vector $\boldsymbol{a}$ while keeping the type I error rates within the nominal level. The hypotheses, test statistics and the associated critical values, and confidence interval are given as follows. The proof of the critical values is referred from Siotani et al. (1985), Section 8.3.2.

(i) Hypotheses:

$$H_0 : \frac{\boldsymbol{a}^t \mathbf{\Sigma}_1 \boldsymbol{a}}{\boldsymbol{a}^t \mathbf{\Sigma}_2 \boldsymbol{a}} = 1, \text{for all non-zero vector } \boldsymbol{a}$$

$$H_1 : \frac{\boldsymbol{a}^t \mathbf{\Sigma}_1 \boldsymbol{a}}{\boldsymbol{a}^t \mathbf{\Sigma}_2 \boldsymbol{a}} \neq 1, \text{for at least one non-zero vector } \boldsymbol{a}$$

(ii) Test:
  Test statistics:

$$T_1 = \sup_{\boldsymbol{a} \in \mathbf{R}^p - \{\mathbf{0}\}} \left\{ \frac{\boldsymbol{a}^t \boldsymbol{S}_1 \boldsymbol{a}}{\boldsymbol{a}^t \boldsymbol{S}_2 \boldsymbol{a}} \right\}, \ T_2 = \inf_{\boldsymbol{a} \in \mathbf{R}^p - \{\mathbf{0}\}} \left\{ \frac{\boldsymbol{a}^t \boldsymbol{S}_1 \boldsymbol{a}}{\boldsymbol{a}^t \boldsymbol{S}_2 \boldsymbol{a}} \right\}$$

  Critical values:

  If $T_1 > c_{\max}^{(1-\alpha/2)}$ or $T_2 < c_{\min}^{(\alpha/2)}$, then reject H_0,

  where $c_{\max}^{(1-\alpha/2)}$ and $c_{\min}^{(\alpha/2)}$ are the $100(1-\alpha/2)$th and the $100(\alpha/2)$th percentiles, respectively, of the distribution of maximum and minimum eigenvalue of $\boldsymbol{S}_2^{-1/2^t} \boldsymbol{S}_1 \boldsymbol{S}_2^{-1/2}$ under H_0.

(iii) $100(1-\alpha)\%$ confidence interval for $\boldsymbol{a}^t \mathbf{\Sigma}_1 \boldsymbol{a}/\boldsymbol{a}^t \mathbf{\Sigma}_2 \boldsymbol{a}$:

$$\left[\frac{\boldsymbol{a}^t \boldsymbol{S}_1 \boldsymbol{a}}{\boldsymbol{a}^t \boldsymbol{S}_2 \boldsymbol{a}} \cdot \frac{1}{c_{\max}^{(1-\alpha/2)}}, \ \frac{\boldsymbol{a}^t \boldsymbol{S}_1 \boldsymbol{a}}{\boldsymbol{a}^t \boldsymbol{S}_2 \boldsymbol{a}} \cdot \frac{1}{c_{\min}^{(\alpha/2)}} \right] , \text{ for the given non-zero vector } \boldsymbol{a},$$

  where $c_{\max}^{(1-\alpha/2)}$ and $c_{\min}^{(\alpha/2)}$ are as in (ii) above.

Multiple dimensions can arise from, for example, multiple endpoints or multiple timepoints in longitudinal data. It is noted that this test can be used for a single comparison when specific weights are given to the endpoints or timepoints of interest by choosing appropriate $\boldsymbol{a}$. Giving differing level of priority or weights to endpoints or timepoints is not unusual in practice.

Despite its flexibility to allow unlimited comparisons and its allowance for a single weighted comparison, the test also has at least two shortcomings. First, it is not easy to use in practice, since one needs to calculate the critical values of the test from the distribution of maximum and minimum eigenvalues of the $\boldsymbol{S}_2^{-1/2^t}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}$. Second, the test may have low power as the number of comparisons is unlimited by its construction. The unlimited number of comparisons allowed by the procedure is unnecessary in most practical applications arising in clinical trials, where the number of comparisons is limited to only a few.

In the following sections, first, we propose a single weighted comparison procedure whose critical values are easier to compute. Second, we expand the idea to a multiple comparison procedure for partial covariance matrices that limits the number of comparisons.

2.2. A weighted procedure for comparison of two covariance matrices (single comparison)

The following is a weighted test for covariance matrices which involves a single comparison. We derive the test statistics and its distribution in this section. Let $\boldsymbol{a}$ be a single non-zero vector that is chosen *a priori*. The hypotheses, test, and confidence interval are given below. Critical values can be obtain easily from the standard F-distribution.

(i) Hypotheses:

$$H_0 : \frac{\boldsymbol{a}^t\boldsymbol{\Sigma}_1\boldsymbol{a}}{\boldsymbol{a}^t\boldsymbol{\Sigma}_2\boldsymbol{a}} = 1, \text{for all non-zero vector } \boldsymbol{a}$$

$$H_1 : \frac{\boldsymbol{a}^t\boldsymbol{\Sigma}_1\boldsymbol{a}}{\boldsymbol{a}^t\boldsymbol{\Sigma}_2\boldsymbol{a}} \neq 1, \text{for at least one non-zero vector } \boldsymbol{a}$$

(ii) Test:
Test statistics:
$$T = \frac{\boldsymbol{a}^t\boldsymbol{S}_1\boldsymbol{a}}{\boldsymbol{a}^t\boldsymbol{S}_2\boldsymbol{a}}$$
$T \sim F_{n_1-1,n_2-1}$, under H_0

If T is greater than the critical value from the standard F-distribution, then reject H_0.

Detail:
From the definition of Wishart distribution,

$$(2.1) \qquad (n_i - 1)\boldsymbol{S}_i = \sum_{j=1}^{n_i-1} \boldsymbol{y}_j \boldsymbol{y}_j^t,$$

where $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{n_i-1}$ are independent and identically distributed as $\boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$. Since $\boldsymbol{a}^t$ is a constant vector, we have

$$\boldsymbol{a}^t \boldsymbol{y}_1, \cdots, \boldsymbol{a}^t \boldsymbol{y}_{n_i-1} \overset{\text{i.i.d.}}{\sim} N(0, \boldsymbol{a}^t \boldsymbol{\Sigma}_i \boldsymbol{a}),$$

and

$$(\boldsymbol{a}^t \boldsymbol{\Sigma}_i \boldsymbol{a})^{-1/2} \boldsymbol{a}^t \boldsymbol{y}_1, \cdots, (\boldsymbol{a}^t \boldsymbol{\Sigma}_i \boldsymbol{a})^{-1/2} \boldsymbol{a}^t \boldsymbol{y}_{n_i-1} \overset{\text{i.i.d.}}{\sim} N(0, 1)$$

where "i.i.d." is a shorthand for "independent and identically distributed". Moreover, from a characterization of the Chi-square distribution, we obtain the following.

$$(2.2) \qquad (\boldsymbol{a}^t \boldsymbol{\Sigma}_i \boldsymbol{a})^{-1} \sum_{j=1}^{n_i-1} (\boldsymbol{a}^t \boldsymbol{y}_j)^2 \sim \chi^2_{n_i-1}$$

On the other hand, $\boldsymbol{a}^t \boldsymbol{S}_i \boldsymbol{a}$ can be transformed using equation (2.1) as follows.

$$(\boldsymbol{a}^t \boldsymbol{S}_i \boldsymbol{a}) = \frac{1}{n_i - 1} \left[\boldsymbol{a}^t \left\{ (n_i - 1)\boldsymbol{S}_i \right\} \boldsymbol{a} \right] = \frac{1}{n_i - 1} \left[\boldsymbol{a}^t (\sum_{j=1}^{n_i-1} \boldsymbol{y}_j \boldsymbol{y}_j^t) \boldsymbol{a} \right]$$

$$= \frac{1}{n_i - 1} \sum_{j=1}^{n_i-1} (\boldsymbol{a}^t \boldsymbol{y}_j)^2$$

Moreover, we can transform the above equation as follows and confirm that left-hand expression has a Chi-square distribution based on (2.2).

$$\frac{(n_i - 1)(\boldsymbol{a}^t \boldsymbol{S}_i \boldsymbol{a})}{\boldsymbol{a}^t \boldsymbol{\Sigma}_i \boldsymbol{a}} = (\boldsymbol{a}^t \boldsymbol{\Sigma}_i \boldsymbol{a})^{-1} \sum_{j=1}^{n_i-1} (\boldsymbol{a}^t \boldsymbol{y}_j)^2 \sim \chi^2_{n_i-1}$$

Since the ratio of two independent chi-squares, each divided by its degrees of freedom, is known to have an F-distribution, we have

$$\frac{\dfrac{(n_1 - 1)(\boldsymbol{a}^t \boldsymbol{S}_1 \boldsymbol{a})}{\boldsymbol{a}^t \boldsymbol{\Sigma}_1 \boldsymbol{a}} / (n_1 - 1)}{\dfrac{(n_2 - 1)(\boldsymbol{a}^t \boldsymbol{S}_2 \boldsymbol{a})}{\boldsymbol{a}^t \boldsymbol{\Sigma}_2 \boldsymbol{a}} / (n_2 - 1)} \sim F_{n_1-1, n_2-1}$$

Finally, under the null hypothesis H_0, i.e. $\boldsymbol{a}^t\boldsymbol{\Sigma}_1\boldsymbol{a} = \boldsymbol{a}^t\boldsymbol{\Sigma}_2\boldsymbol{a}$, we have

$$\frac{\boldsymbol{a}^t\boldsymbol{S}_1\boldsymbol{a}}{\boldsymbol{a}^t\boldsymbol{S}_2\boldsymbol{a}} \sim F_{n_1-1,n_2-1}, \text{ under } H_0$$

(iii) $100(1 - \alpha)\%$ confidence interval for $\boldsymbol{a}^t\boldsymbol{\Sigma}_1\boldsymbol{a}/\boldsymbol{a}^t\boldsymbol{\Sigma}_2\boldsymbol{a}$:

Based on the distribution of the test statistics, we see that following condition holds with probability $(1 - \alpha) \times 100\%$, where $F_{df_1,df_2,\alpha}$ indicates $100\alpha^{th}$ percentile of the F-distribution.

$$\left[\frac{\boldsymbol{a}^t\boldsymbol{S}_1\boldsymbol{a}}{\boldsymbol{a}^t\boldsymbol{S}_2\boldsymbol{a}}F_{n_2-1,n_1-1,\alpha/2}, \ \frac{\boldsymbol{a}^t\boldsymbol{S}_1\boldsymbol{a}}{\boldsymbol{a}^t\boldsymbol{S}_2\boldsymbol{a}}F_{n_2-1,n_1-1,1-\alpha/2}\right]$$

As an illustrative example, consider a clinical trial with 5 visits, where the measurement of interest is measured twice on each visit. We assume that the number of the subjects who have second measurement is half that of the first measurement. The following matrix is an example of covariance matrix arising from such a scenario. We focus only on one visit (the last visit), and the first measurement on last visit is given more weight than the second measurement according to the number of subjects with measurements. If we set $\boldsymbol{a}^t = (00000000\sqrt{2}1)$, then we can extract the partial covariance matrix of interest as shown by the underlined elements below with the indicated weights.

$$\begin{pmatrix} s_{11} & s_{12} & \cdots & \cdots & s_{1,10} \\ s_{21} & s_{22} & & & \\ \vdots & & \ddots & & \\ \vdots & & & \underline{s_{99}} & \underline{s_{9,10}} \\ s_{10,1} & \cdots & \cdots & \underline{s_{10,9}} & \underline{s_{10,10}} \end{pmatrix}$$

We note that this method can be applied also to a situation where there are two primary endpoints one of which is given more weight than the other. Also, if we set $\boldsymbol{a}^t = (1111111111)$, then we can compare the entire covariance matrix.

2.3. A weighted procedure for comparison of two covariance matrices (multiple comparisons)

The last example in Section 2.2 considered only one visit, namely, the last visit. When we are interested in multiple visits, however, we need to consider multiplicity. In this case, we have multiple $\boldsymbol{a}_\ell$'s ($\ell = 1, \cdots, k$) of interest. The hypotheses, test statistics and the associated critical values, and confidence interval are given below.

(i) Hypotheses:

$$H_0 : \frac{\boldsymbol{a}_\ell^t \boldsymbol{\Sigma}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{\Sigma}_2 \boldsymbol{a}_\ell} = 1, \text{for non-zero vectors } \boldsymbol{a}_\ell, \ell = 1, \cdots, k$$

$$H_1 : \frac{\boldsymbol{a}_\ell^t \boldsymbol{\Sigma}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{\Sigma}_2 \boldsymbol{a}_\ell} \neq 1, \text{for at least one non-zero vector } \boldsymbol{a}_\ell, \ell = 1, \cdots, k$$

(ii) Test:

Test statistics:

$$T_1 = \max_{\ell=1,\cdots,k} \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell}, \ \ T_2 = \min_{\ell=1,\cdots,k} \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell}$$

Critical values:

If $T_1 > c_{\max}^{(1-\alpha/2)}$ or $T_2 < c_{\min}^{(\alpha/2)}$, then reject H_0,

where $c_{\max}^{(1-\alpha/2)}$ and $c_{\min}^{(\alpha/2)}$ are the $100(1 - \alpha/2)$th and the $100(\alpha/2)$th percentiles, respectively, of the distribution of T_1 and T_2 under H_0.

Various methods to control multiplicity have been developed over the years such as Bonferroni, Sidak, Holm, and Hochberg. And, methods to control multiplicity based on the former methods are still a topic of research such as that in Sarkar et al. (2012). In this paper, we calculated approximate $c_{\max}$s and $c_{\min}$s based on Sidak and Bonferroni methods. Let us denote the approximate $c_{\max}$s and $c_{\min}$s based on Sidak inequality and Bonferroni inequality as: $c_{\max,S}^{(1-\alpha/2)}, c_{\min,S}^{(\alpha/2)}$ and $c_{\max,B}^{(1-\alpha/2)}, c_{\min,B}^{(\alpha/2)}$, respectively.

Based on the Sidak inequality:

$$\Pr\left\{ \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell} > c_{\max,S}^{(1-\alpha/2)} \right\} = \Pr\left\{ \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell} < c_{\min,S}^{(\alpha/2)} \right\} = 1 - (1 - \frac{\alpha}{2})^{1/k}, \ell = 1, \cdots, k$$

Based on the Bonferroni inequality:

$$\Pr\left\{ \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell} > c_{\max,B}^{(1-\alpha/2)} \right\} = \Pr\left\{ \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell} < c_{\min,B}^{(\alpha/2)} \right\} = \frac{1}{k} \cdot \frac{\alpha}{2}, \ell = 1, \cdots, k$$

(iii) $100(1 - \alpha)\%$ confidence interval for $\boldsymbol{a}_\ell^t \boldsymbol{\Sigma}_1 \boldsymbol{a}_\ell / \boldsymbol{a}_\ell^t \boldsymbol{\Sigma}_2 \boldsymbol{a}_\ell$:

We can construct an approximate confidence interval by using the critical values based on Sidak or Bonferroni inequality.

$$\left[\frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell} \cdot \frac{1}{c_{\max}^{(1-\alpha/2)}}, \ \frac{\boldsymbol{a}_\ell^t \boldsymbol{S}_1 \boldsymbol{a}_\ell}{\boldsymbol{a}_\ell^t \boldsymbol{S}_2 \boldsymbol{a}_\ell} \cdot \frac{1}{c_{\min}^{(\alpha/2)}} \right], \ \ell = 1, \cdots, k,$$

where $c_{\max}^{(1-\alpha/2)}$ and $c_{\min}^{(\alpha/2)}$ are as in (ii) above.

The Bonferroni method is generally applicable. For the Sidak method, however, the family-wise error rate is strongly controlled only when the statistics are independent or follows multivariate normal distribution (see, e.g., Dmitrienko et al. (2005)). Thus the critical values and confidence interval given above do not hold exactly. However, the Sidak method does have the following property.

Consider two statistics, $(\boldsymbol{a}_1^t \boldsymbol{S}_1 \boldsymbol{a}_1)/(\boldsymbol{a}_1^t \boldsymbol{S}_2 \boldsymbol{a}_1)$ and $(\boldsymbol{a}_2^t \boldsymbol{S}_1 \boldsymbol{a}_2)/(\boldsymbol{a}_2^t \boldsymbol{S}_2 \boldsymbol{a}_2)$. We focus on the numerators, and calculate the covariance of $\boldsymbol{a}_1^t \boldsymbol{S}_1 \boldsymbol{a}_1$ and $\boldsymbol{a}_2^t \boldsymbol{S}_1 \boldsymbol{a}_2$ by using Gupta and Nagar (2000), Theorem 3.3.15.

$$
\begin{aligned}
&\mathrm{Cov}(\boldsymbol{a}_1^t \boldsymbol{S}_1 \boldsymbol{a}_1, \boldsymbol{a}_2^t \boldsymbol{S}_1 \boldsymbol{a}_2) \\
&= \mathrm{E}[\mathrm{tr}(\boldsymbol{a}_1 \boldsymbol{a}_1^t \boldsymbol{S}_1)\mathrm{tr}(\boldsymbol{a}_2 \boldsymbol{a}_2^t \boldsymbol{S}_1)] - \mathrm{E}[\mathrm{tr}(\boldsymbol{a}_1 \boldsymbol{a}_1^t \boldsymbol{S}_1)]\mathrm{E}[\mathrm{tr}(\boldsymbol{a}_2 \boldsymbol{a}_2^t \boldsymbol{S}_1)] \\
&= \frac{1}{n_1 - 1}\left[\mathrm{tr}(\boldsymbol{a}_1 \boldsymbol{a}_1^t \boldsymbol{\Sigma}_1 \boldsymbol{a}_2 \boldsymbol{a}_2^t \boldsymbol{\Sigma}_1) + \mathrm{tr}\{(\boldsymbol{a}_1 \boldsymbol{a}_1^t)^t \boldsymbol{\Sigma}_1 \boldsymbol{a}_2 \boldsymbol{a}_2^t \boldsymbol{\Sigma}_1\}\right] \\
&= \frac{2}{n_1 - 1}(\boldsymbol{a}_1^t \boldsymbol{\Sigma}_1 \boldsymbol{a}_2)^2
\end{aligned}
$$

This shows that, although not independent, the numerators are asymptotically independent. Similarly, we obtain $\mathrm{Cov}(\boldsymbol{a}_1^t \boldsymbol{S}_2 \boldsymbol{a}_1, \boldsymbol{a}_2^t \boldsymbol{S}_2 \boldsymbol{a}_2) = 2(n_2 - 1)^{-1}(\boldsymbol{a}_1^t \boldsymbol{\Sigma}_2 \boldsymbol{a}_2)^2$ for the denominators, thus the denominators are also asymptotically independent. Furthermore, because the numerators and the denominators are independent, $(\boldsymbol{a}_1^t \boldsymbol{S}_1 \boldsymbol{a}_1)/(\boldsymbol{a}_1^t \boldsymbol{S}_2 \boldsymbol{a}_1)$ and $(\boldsymbol{a}_2^t \boldsymbol{S}_1 \boldsymbol{a}_2)/(\boldsymbol{a}_2^t \boldsymbol{S}_2 \boldsymbol{a}_2)$ are asymptotically independent.

§3.　Simulations

We conducted simulations to compare the performance of the proposed test and the Siotani test, and to assess the impact of non-independence. Monte Carlo simulation consisting of one million random samplings was carried out for each setting.

We considered the following set-up. Number of visits was 5 and the number of measurements at each visit was 2. Sample sizes considered were $n_1 = n_2 = 25$, 50, or 100. Let $\sigma_{11}^{(i)}, \cdots, \sigma_{10,10}^{(i)}$ be the diagonal elements of $\boldsymbol{\Sigma}_i$, ρ_1 be the between-visit correlation, and ρ_2 be the within-visit correlation. Since it is likely that the within-visit correlation is equal to or higher than the between-visit correlation, we set (ρ_1, ρ_2) to be $(0,0)$, $(0.4, 0.4)$, $(0.8, 0.8)$, or $(0.4, 0.7)$.

We focused on the four timepoints (visits), and we gave the first measurement on each visit more weight than the second measurement. In particular, $\boldsymbol{a}_j$'s were set to the following four vectors, $\boldsymbol{a}_1^t = (0\ 0\ \sqrt{2}\ 1\ 0\ 0\ 0\ 0\ 0\ 0)$, $\boldsymbol{a}_2^t = (0\,0\,0\,0\,\sqrt{2}\,1\,0\,0\,0\,0)$, $\boldsymbol{a}_3^t = (0\,0\,0\,0\,0\,0\,\sqrt{2}\,1\,0\,0)$, and $\boldsymbol{a}_4^t = (0\,0\,0\,0\,0\,0\,0\,0\,\sqrt{2}\,1)$.

We assumed $(\boldsymbol{x}_1^{(i)}, \cdots, \boldsymbol{x}_{n_i}^{(i)})$ to be independently distributed as $N_{10}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, where

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_{11}^{(i)} & \rho_2\sqrt{\sigma_{11}^{(i)}\sigma_{22}^{(i)}} & \rho_1\sqrt{\sigma_{11}^{(i)}\sigma_{33}^{(i)}} & \cdots & \cdots \\ \rho_2\sqrt{\sigma_{11}^{(i)}\sigma_{22}^{(i)}} & \sigma_{22}^{(i)} & \rho_2\sqrt{\sigma_{22}^{(i)}\sigma_{33}^{(i)}} & \cdots & \cdots \\ \rho_1\sqrt{\sigma_{11}^{(i)}\sigma_{33}^{(i)}} & \rho_2\sqrt{\sigma_{22}^{(i)}\sigma_{33}^{(i)}} & \sigma_{33}^{(i)} & \rho_2\sqrt{\sigma_{33}^{(i)}\sigma_{44}^{(i)}} & \\ \vdots & \vdots & \rho_2\sqrt{\sigma_{33}^{(i)}\sigma_{44}^{(i)}} & \sigma_{44}^{(i)} & \\ \vdots & \vdots & & & \ddots \end{pmatrix}$$

Settings of $\sigma_{11}^{(i)}, \cdots, \sigma_{10,10}^{(i)}$ are shown in each table.

Table 1 displays the 97.5% critical values of each test statistics under the null hypothesis for the four scenarios. The true critical values and the critical values for the Siotani test in Section 2.1 were based on simulations. Specifically, the true critical values were calculated from the 97.5th percentiles of the statistics, $\max_{j=1,\cdots,4} \boldsymbol{a}_j^t \boldsymbol{S}_1 \boldsymbol{a}_j / \boldsymbol{a}_j^t \boldsymbol{S}_2 \boldsymbol{a}_j$, based on simulations. Similarly, the critical values for Siotani method were calculated from the 97.5th percentiles of the distribution of maximum eigenvalue of $\boldsymbol{S}_2^{-1/2^t} \boldsymbol{S}_1 \boldsymbol{S}_2^{-1/2}$. The critical values for the Sidak and Bonferroni method were from analytic calculations. The main message from Table 1 is that the critical values for the Siotani test are larger than those of the other tests.

Table 2 displays the actual type I error rates corresponding to the critical values in Table 1. The type I error rates for the Siotani test are much lower than the others. The type I error rates for the Sidak and Bonferroni methods are conservative when there are high positive correlations.

Table 3 shows the powers of the tests for sample size of 100 under the four settings of $\boldsymbol{\Sigma}$ and (ρ_1, ρ_2). The results show properties similar to that of the type I error rates. The powers for the Siotani test are much lower than the others. The powers of the Sidak and Bonferroni methods are low when there are high positive correlations.

Table 1: 97.5% critical values for $p = 4$

(ρ_1, ρ_2)	Subjects	True	Sidak	Bonferroni	Siotani
$\rho_1 = \rho_2 = 0$	$n_1 = n_2 = 25$	2.862	2.862	2.866	13.800
	$n_1 = n_2 = 50$	2.061	2.063	2.065	4.944
	$n_1 = n_2 = 100$	1.657	1.657	1.659	2.895
$\rho_1 = \rho_2 = 0.4$	$n_1 = n_2 = 25$	2.837	2.862	2.866	13.806
	$n_1 = n_2 = 50$	2.048	2.063	2.065	4.941
	$n_1 = n_2 = 100$	1.650	1.657	1.659	2.897
$\rho_1 = \rho_2 = 0.8$	$n_1 = n_2 = 25$	2.684	2.862	2.866	13.830
	$n_1 = n_2 = 50$	1.971	2.063	2.065	4.946
	$n_1 = n_2 = 100$	1.604	1.657	1.659	2.895
$\rho_1 = 0.4, \rho_2 = 0.7$	$n_1 = n_2 = 25$	2.850	2.862	2.866	13.803
	$n_1 = n_2 = 50$	2.057	2.063	2.065	4.949
	$n_1 = n_2 = 100$	1.653	1.657	1.659	2.897

$\sigma_{11}^{(i)} = \cdots = \sigma_{10,10}^{(i)} = 1$ are set for $i = 1, 2$.

Table 2: Type I error rates with 97.5% critical values from simulations for $p = 4$

(ρ_1, ρ_2)	Subjects	True	Sidak	Bonferroni	Siotani
$\rho_1 = \rho_2 = 0$	$n_1 = n_2 = 25$	2.50%	2.50%	2.48%	<0.01%
	$n_1 = n_2 = 50$	2.50%	2.49%	2.47%	<0.01%
	$n_1 = n_2 = 100$	2.50%	2.49%	2.46%	<0.01%
$\rho_1 = \rho_2 = 0.4$	$n_1 = n_2 = 25$	2.50%	2.36%	2.34%	<0.01%
	$n_1 = n_2 = 50$	2.50%	2.35%	2.33%	<0.01%
	$n_1 = n_2 = 100$	2.50%	2.35%	2.33%	<0.01%
$\rho_1 = \rho_2 = 0.8$	$n_1 = n_2 = 25$	2.50%	1.72%	1.71%	<0.01%
	$n_1 = n_2 = 50$	2.50%	1.69%	1.68%	<0.01%
	$n_1 = n_2 = 100$	2.50%	1.66%	1.65%	<0.01%
$\rho_1 = 0.4, \rho_2 = 0.7$	$n_1 = n_2 = 25$	2.50%	2.43%	2.41%	<0.01%
	$n_1 = n_2 = 50$	2.50%	2.44%	2.42%	<0.01%
	$n_1 = n_2 = 100$	2.50%	2.41%	2.39%	<0.01%

$\sigma_{11}^{(i)} = \cdots = \sigma_{10,10}^{(i)} = 1$ are set for $i = 1, 2$.

Table 3: Powers with 97.5% critical values from simulations for $p = 4$ and $n_1 = n_2 = 100$

Diagonal elements for Σ_1	(ρ_1, ρ_2)	True	Sidak	Bonfe rroni	Siotani
$\sigma_{11} = \cdots = \sigma_{88} = 1$,	$\rho_1 = \rho_2 = 0$	82.8	82.7	82.6	3.4
$\sigma_{99} = \sigma_{10,10} = 2$	$\rho_1 = \rho_2 = 0.4$	83.0	82.5	82.4	3.3
	$\rho_1 = \rho_2 = 0.8$	86.3	82.4	82.3	3.3
	$\rho_1 = 0.4, \rho_2 = 0.7$	82.9	82.6	82.5	3.3
$\sigma_{11} = \cdots = \sigma_{88} = 1$,	$\rho_1 = \rho_2 = 0$	32.4	32.3	32.2	0.1
$\sigma_{99} = \sigma_{10,10} = 1.5$	$\rho_1 = \rho_2 = 0.4$	32.5	31.7	31.5	0.1
	$\rho_1 = \rho_2 = 0.8$	37.0	31.0	30.9	0.1
	$\rho_1 = 0.4, \rho_2 = 0.7$	32.3	31.8	31.7	0.1
$\sigma_{11} = \cdots = \sigma_{44} = 1$,	$\rho_1 = \rho_2 = 0$	67.4	67.4	67.2	0.2
$\sigma_{55} = \cdots = \sigma_{10,10}$	$\rho_1 = \rho_2 = 0.4$	60.2	59.1	59.0	0.2
$= 1.5$	$\rho_1 = \rho_2 = 0.8$	52.5	45.6	45.5	0.2
	$\rho_1 = 0.4, \rho_2 = 0.7$	62.4	61.7	61.6	0.2
$\sigma_{11} = 1, \sigma_{22} = 1$,	$\rho_1 = \rho_2 = 0$	77.7	77.6	77.5	0.7
$\sigma_{33} = 1.2, \sigma_{44} = 1$,	$\rho_1 = \rho_2 = 0.4$	70.4	69.5	69.4	0.6
$\sigma_{55} = 1.4, \sigma_{66} = 1.2$,	$\rho_1 = \rho_2 = 0.8$	66.0	59.7	59.6	0.5
$\sigma_{77} = 1.6, \sigma_{88} = 1.4$,	$\rho_1 = 0.4, \rho_2 = 0.7$	71.5	70.9	70.8	0.6
$\sigma_{99} = 1.8, \sigma_{10,10} = 1.6$					

$\sigma_{11}^{(2)} = \cdots = \sigma_{10,10}^{(2)} = 1$ are set for Σ_2.

Each power is shown as a percentage.

§4.   An example of application

We applied our test (single comparison) to a four-period crossover bioavailability study described in Example 4.3 of Patterson and Jones. (1985). The data are from a trial with four periods that used the sequence groups RTTR (Group 1) and TRRT (Group 2) with 8 subjects and 9 subjects, respectively (R: Reference, T: Test). One subject in Group 2 has a missing observation in period 4. In this application, we conducted a complete-case analysis to assess the within-subject variability. Values of AUC and Cmax were log-transformed prior to analysis, as it is a standard practice for these pharmacokinetic parameters. The data for each sequence are shown in Figure 1.
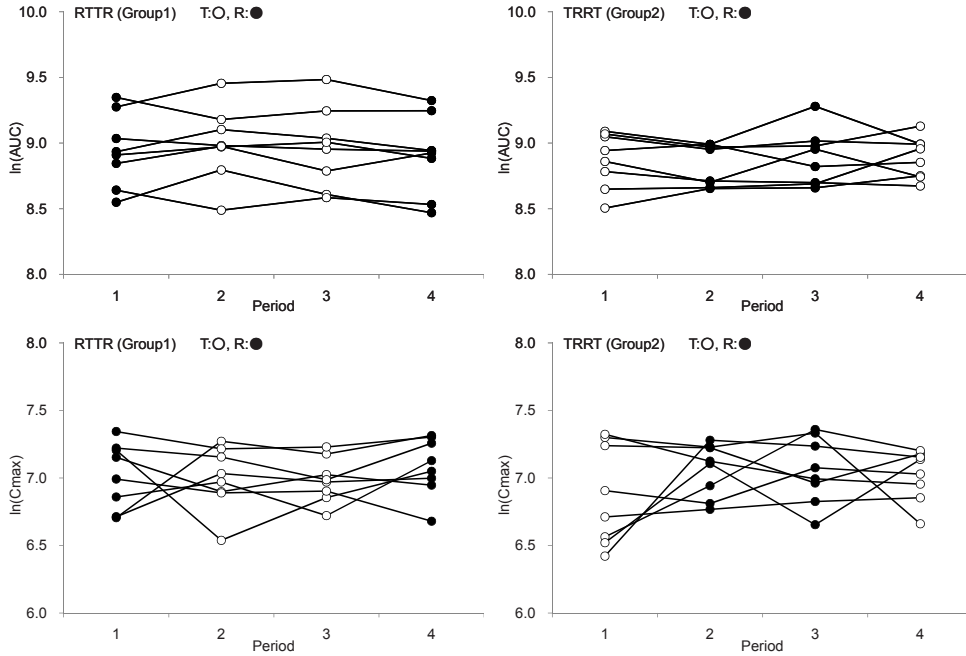

Figure 1: Subject profile plots of the log-transformed AUC and the log-transformed Cmax

Although the test as shown in Section 2.2 applies only to independent samples, the test can be adapted to data arising from a crossover study such as the one described above. The proof for the extension follows the arguments of Chow and Liu (2000), Theorem 9.4.1. Let the vector $\boldsymbol{Y}_{jqi} = \left(Y_{jqi}^{AUC}, Y_{jqi}^{Cmax} \right)^t$ denote log-transformed data for the period q of the jth subject in the sequence group i $[j = 1, \cdots, n_i; q = 1, \cdots, 4; i = 1(= \text{Group 1}), 2(= \text{Group 2})]$. Define $\boldsymbol{D}_{1i} = ((\boldsymbol{Y}_{11i} - \boldsymbol{Y}_{14i})| \cdots |(\boldsymbol{Y}_{n_i1i} - \boldsymbol{Y}_{n_i4i}))^t$ and $\boldsymbol{D}_{2i} = ((\boldsymbol{Y}_{12i} -$

$\boldsymbol{Y}_{13i})| \cdots |(\boldsymbol{Y}_{n_i2i} - \boldsymbol{Y}_{n_i3i}))^t$; that is, $\boldsymbol{D}_{1i}$ and $\boldsymbol{D}_{2i}$ are $n_i \times 2$ matrices of period 1-4 differences and period 2-3 differences, respectively. Under the model assumption described in Appendix A, the vector $((\boldsymbol{Y}_{j1i} - \boldsymbol{Y}_{j4i})^t, (\boldsymbol{Y}_{j2i} - \boldsymbol{Y}_{j3i})^t)^t$ follows a multivariate normal distribution with the following mean and the covariance matrix.

Mean:
$$i = 1 : (\mu_{AUC,R}, \mu_{Cmax,R}, \mu_{AUC,T}, \mu_{Cmax,T})^t,$$
$$i = 2 : (\mu_{AUC,T}, \mu_{Cmax,T}, \mu_{AUC,R}, \mu_{Cmax,R})^t$$

Covariance matrix:
$$i = 1 : \begin{pmatrix} 2\boldsymbol{\Sigma}_R & \boldsymbol{0} \\ \boldsymbol{0} & 2\boldsymbol{\Sigma}_T \end{pmatrix}, i = 2 : \begin{pmatrix} 2\boldsymbol{\Sigma}_T & \boldsymbol{0} \\ \boldsymbol{0} & 2\boldsymbol{\Sigma}_R \end{pmatrix},$$

where $\boldsymbol{\Sigma}_z = \begin{pmatrix} \sigma^2_{AUC,z} & \tau_z \\ \tau_z & \sigma^2_{Cmax,z} \end{pmatrix}, z = T\,or\,R,$

$\sigma^2_{AUC,z}$ and $\sigma^2_{Cmax,z}$ are within-subject error variances,

and τ_z is the within-subject error covariance described in Appendix A.

Let

$$\boldsymbol{S}_T = \frac{1}{2(n_1 + n_2 - 2)}[(\boldsymbol{D}_{21} - \boldsymbol{J}_{n_1}\bar{\boldsymbol{D}}_{21})^t(\boldsymbol{D}_{21} - \boldsymbol{J}_{n_1}\bar{\boldsymbol{D}}_{21})$$
$$+ (\boldsymbol{D}_{12} - \boldsymbol{J}_{n_2}\bar{\boldsymbol{D}}_{12})^t(\boldsymbol{D}_{12} - \boldsymbol{J}_{n_2}\bar{\boldsymbol{D}}_{12})],$$

and

$$\boldsymbol{S}_R = \frac{1}{2(n_1 + n_2 - 2)}[(\boldsymbol{D}_{11} - \boldsymbol{J}_{n_1}\bar{\boldsymbol{D}}_{11})^t(\boldsymbol{D}_{11} - \boldsymbol{J}_{n_1}\bar{\boldsymbol{D}}_{11})$$
$$+ (\boldsymbol{D}_{22} - \boldsymbol{J}_{n_2}\bar{\boldsymbol{D}}_{22})^t(\boldsymbol{D}_{22} - \boldsymbol{J}_{n_2}\bar{\boldsymbol{D}}_{22})],$$

where $\bar{\boldsymbol{D}}_{1i}$ and $\bar{\boldsymbol{D}}_{2i}$ are sample mean vectors of $\boldsymbol{D}_{1i}$ and $\boldsymbol{D}_{2i}$, respectively; and $\boldsymbol{J}_{n_i}$ is the $n_i \times 1$ matrix of 1's.

$\boldsymbol{S}_T$ and $\boldsymbol{S}_R$ are the estimators of $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\Sigma}_R$, respectively. Then, owing to the normality of the vector $((\boldsymbol{Y}_{j1i} - \boldsymbol{Y}_{j4i})^t, (\boldsymbol{Y}_{j2i} - \boldsymbol{Y}_{j3i})^t)^t$, $(n_1 + n_2 - 2)\boldsymbol{S}_T$ and $(n_1 + n_2 - 2)\boldsymbol{S}_R$ have $W_p(\boldsymbol{\Sigma}_T, n_1 + n_2 - 2)$ and $W_p(\boldsymbol{\Sigma}_R, n_1 + n_2 - 2)$ distributions, respectively. The proof of the independence between $\boldsymbol{S}_T$ and $\boldsymbol{S}_R$ are shown in Appendix A. As $(n_1 + n_2 - 2)\boldsymbol{S}_T$ and $(n_1 + n_2 - 2)\boldsymbol{S}_R$ are independent and have Wishart distribution, the proposed method can be applied to this example.

Estimated $\boldsymbol{S}_T$ and $\boldsymbol{S}_R$ in this example were $\begin{pmatrix} 0.0117 & -0.0015 \\ -0.0015 & 0.0697 \end{pmatrix}$ and $\begin{pmatrix} 0.0069 & 0.0066 \\ 0.0066 & 0.0433 \end{pmatrix}$, respectively. We set $\boldsymbol{a} = (1,1)^t$ to compare the above covariance matrices. The value of statistics $(= \boldsymbol{a}^t\boldsymbol{S}_T\boldsymbol{a}/\boldsymbol{a}^t\boldsymbol{S}_R\boldsymbol{a})$ and its P-value were 0.8087 and 0.6517, respectively. The result of our test was not statistically significant at 5% level.

Note that we can set $\boldsymbol{a}$ as a weight defined from *a priori* scientific considerations. For example, we can select $\boldsymbol{a}$ based on conversion factor between two different units, e.g., mg/dL and ng/dL. Also, we may be able to set $\boldsymbol{a}$ based on historical data.

§5.   Discussion and conclusion

Demonstration of efficacy and safety with appropriate risk-benefit balance may be sufficient for regulatory approval of a drug; however, this may not be enough for all customers. In particular, differentiating features of a test drug with respect to comparators can provide valuable information to physicians and patients. One such point of differentiation is smaller variability of effect. A drug or a formulation with smaller variability, which implies larger number of patients would be more likely to experience clinical effect near the central value, is expected to be easier for a physician to use.

In this paper, firstly, we introduced a single comparison procedure for variance-covariance matrices in a form of hypothesis test and confidence interval. This procedure can be applied in a situation where there is a single endpoint measured at multiple timepoints or a multiple endpoints measured at a single timepoint. The procedure is characterized by weighting of the variance-covariance matrix elements in the comparison. This procedure can be thought of as an extension of the F-test, and in fact, the test statistic has an exact F-distribution under the null hypothesis.

Secondly, we introduced a multiple comparison procedure for partial covariance matrices that limits the number of comparisons. The simulation results suggest that powers of the proposed procedure are higher than those of the Siotani method, while keeping the type I error rates within the nominal level. As can be seen from Table 1, the critical values for the Siotani test are larger than those of the Bonferroni method. In particular, when $n_1 = n_2 = 100$, $\rho_1 = \rho_2 = 0$, and the number of comparison $p = 4$, the critical value for the Siotani test is 2.895 compared to 1.659 for the Bonferroni method. In fact, even when $p = 1000$, the critical value for Bonferroni is 2.290 which is still smaller than the critical value for the Siotani test when $p = 4$. Thus when the number of comparisons is small, as often is the case in clinical trials, utilization of a common method for multiplicity adjustment (e.g. Bonferroni or Sidak method) is recommended compared to the Siotani method. Although there was no apparent impact of "non-independence" between statistics under our simulation settings, it may have a noticeable impact in other situations (e.g. when there are many comparisons.). Thus, it is recommended that impact of non-independence be assessed under settings close to the actual problem before choosing the method of multiplicity adjustment.

Before actual implementation, it is important to investigate the robustness of the test and assess the impact to cases of non-normality of the data. In the special case of univariate data, our test reduces to the F test for equal variances, which is known to be non-robust. This suggests that our test may also lack robustness. Regarding non-normal data, the impact of non-normality may be mitigated, if a transformation such as the log-transformation is applied.

The proposed method is recommended for multiple comparisons of variance-covariance matrices when the number of comparisons is small, as often is the case in clinical trials. Future considerations include investigation of other application areas for the method. Some possible areas might be problems in drug resistance, assessment of stability of drug effect overtime, and assessment of medical devices. Another consideration for the future is extention of the method to a within-subject variance-covariance matrix estimated from a longitudinal model fitted to data arising from a parallel group study.

Acknowledgements

References

[1] Chow, S. C. and Liu, J. P. (2000). *Design and Analysis of Bioavailability and Bioequivalence Studies, Second Edition.* Marcel Dekker, Inc., New York.

[2] Danne, T., Datz, N., Endahl, L., Haahr, H., Nestoris, C., Westergaard, L., Fjording, M. S., and Kordonouri, O. (2008). Insulin detemir is characterized by a more reproducible pharmacokinetic profile than insulin glargine in children and adolescents with type 1 diabete: results from a randomized, double-blind, controlled trial. *Pediatric Diabetes*, 9:554-560.

[3] Dmitrienko, A., Molenberghs, G., Chuang-Stein, C., and Offen, W. (2005). *Analysis of Clinical Trials Using SAS: A Practical Guide.* SAS Institute Inc., Cary, NC.

[4] Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions.* Chapman and Hall/CRC, Florida.

[5] Hartley, H. O. (1950). The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika*, 37:308-312

[6] Khuri, A. (2009). *Linear Model Methodology*. Chapman and Hall/CRC, London.

[7] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc., New York.

[8] Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics*, 1:700-709.

[9] Patterson, S. D. and Jones, B. (2005). *Bioequivalence and Statistics in Clinical Pharmacology*. Chapman and Hall/CRC, Florida.

[10] Proust, M. (2012). *Basic Analysis and Graphing Second Edition: Unequal Variances in JMP 10.0.2 Online Documentation*. SAS Institute Inc., Cary, NC. Available from: URL: http://support.sas.com/documentation/onlinedoc/jmp/index.html (final access 5SEP2014)

[11] Rothwell, P. M., Howard, S. C., Dolan, E., O'Brien, E., Dobson, J. E., Dahlof, B., Sever, P. S., and Poulter, N. R. (2010). Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. *The Lancet*, 375:895-905.

[12] Sarkar, S. K., Guo, W., and Finner H. (2012). On adaptive procedures controlling the familywise error rate. *Journal of Statistical Planning and Inference*, 142:65-78.

[13] Siotani M., Hayakawa T., and Fujikoshi Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*. American Science Press Inc., Ohio.

§A.   Proof of the independence to two covariance matrices in a 2×4 crossover study

Settings follows that of the section 4. Assume that log-transformed AUC values follows the model.

$Y_{jqi}^{AUC} = \mu_{AUC,qi} + S_{AUC,ji} + \varepsilon_{AUC,jqi},$

where $j = 1, \cdots, n_i$; $q = 1, \cdots, 4$; and $i = 1 (= \text{Group 1}), 2 (= \text{Group 2})$ $\mu_{AUC,qi}$ denotes a mean that depends on the period q and the sequence group i. $S_{AUC,ji}$ denotes the between-subject variability which follows $N(0, \sigma_{AUC,s}^2)$. $\varepsilon_{AUC,jqi}$ denotes the within-subject error term which follows,

$N(0, \sigma_{AUC,T}^2)$ for $i = 1$ and $q = 2$ or 3; or $i = 2$ and $q = 1$ or 4

$N(0, \sigma_{AUC,R}^2)$ for $i = 1$ and $q = 1$ or 4; or $i = 2$ and $q = 2$ or 3.

$\varepsilon_{AUC,jqi}$ for $q = 1, \cdots, 4$ are assumed to be mutually independent. Also, any combination of $\varepsilon_{AUC,jqi}$ and $S_{AUC,ji}$ are assumed to be independent.

Similar assumption are made for, $Y_{jqi}^{Cmax} = \mu_{Cmax,qi} + S_{Cmax,ji} + \varepsilon_{Cmax,jqi}$ for Cmax. Furthermore, we set τ_{1i} and τ_{2i} to be the within-subject error covariance, and ξ_i to be the between-subject covariance.

Note that $(\sigma_{AUC,T}^2, \sigma_{Cmax,T}^2)$ and $(\sigma_{AUC,R}^2, \sigma_{Cmax,R}^2)$ are diagonal element of $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_R$ of Section 4, respectively.

Let $\boldsymbol{Y}_{ji} = (Y_{j1i}^{AUC}, Y_{j2i}^{AUC}, Y_{j3i}^{AUC}, Y_{j4i}^{AUC}, Y_{j1i}^{Cmax}, Y_{j2i}^{Cmax}, Y_{j3i}^{Cmax}, Y_{j4i}^{Cmax})^t$ and $\boldsymbol{Y}_i = (\boldsymbol{Y}_{1i}^t, \boldsymbol{Y}_{2i}^t, \cdots, \boldsymbol{Y}_{n_i i}^t)^t$. The covariance matrix of $\boldsymbol{Y}_i$ is given by

$$\boldsymbol{V}_i \equiv \text{Cov}(\boldsymbol{Y}_i) = \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma}_i,$$

where $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{AUC,i} & \boldsymbol{Cov}_i \\ \boldsymbol{Cov}_i & \boldsymbol{\Sigma}_{Cmax,i} \end{pmatrix}$,

$\boldsymbol{\Sigma}_{endp,1} =$
$$\begin{pmatrix} \sigma_{endp,R}^2 + \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 \\ \sigma_{endp,s}^2 & \sigma_{endp,T}^2 + \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 \\ \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,T}^2 + \sigma_{endp,s}^2 & \sigma_{endp,s}^2 \\ \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,R}^2 + \sigma_{endp,s}^2 \end{pmatrix},$$

$\boldsymbol{\Sigma}_{endp,2} =$
$$\begin{pmatrix} \sigma_{endp,T}^2 + \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 \\ \sigma_{endp,s}^2 & \sigma_{endp,R}^2 + \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 \\ \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,R}^2 + \sigma_{endp,s}^2 & \sigma_{endp,s}^2 \\ \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,s}^2 & \sigma_{endp,T}^2 + \sigma_{endp,s}^2 \end{pmatrix},$$

$\boldsymbol{Cov}_i = \begin{pmatrix} \tau_{1i} + \xi_i & \xi_i & \xi_i & \xi_i \\ \xi_i & \tau_{2i} + \xi_i & \xi_i & \xi_i \\ \xi_i & \xi_i & \tau_{2i} + \xi_i & \xi_i \\ \xi_i & \xi_i & \xi_i & \tau_{1i} + \xi_i \end{pmatrix},$

and $\tau_{21} = \tau_{12}(\equiv \tau_T)$ and $\tau_{11} = \tau_{22}(\equiv \tau_R)$; $endp = AUC$ or $Cmax$; $i = 1, 2$.

Now we consider the sum of squares $\boldsymbol{SSD}_{wi} = (\boldsymbol{D}_{wi} - \boldsymbol{J}_{n_i}\bar{\boldsymbol{D}}_{wi})^t(\boldsymbol{D}_{wi} - \boldsymbol{J}_{n_i}\bar{\boldsymbol{D}}_{wi})$, $i = 1, 2$; $w = 1, 2$, which are the component of $\boldsymbol{S}_T$ and $\boldsymbol{S}_R$ in the Section 4. Each element of the $\boldsymbol{SSD}_{wi}$ can be expressed by using $s(i, \boldsymbol{C}_1, \boldsymbol{C}_2) = \boldsymbol{Y}_i^t[\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_1][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_2^t]\boldsymbol{Y}_i$, where $\boldsymbol{I}_{n_i}$ is the $n_i \times n_i$ identity matrix, $\boldsymbol{J}_{n_i}^*$ is the $n_i \times n_i$ matrix of 1's. $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are selected from following the four comparison vectors, $\boldsymbol{C}_{AU,1} = (1, 0, 0, -1, 0, 0, 0, 0)^t$, $\boldsymbol{C}_{AU,2} = (0, 1, -1, 0, 0, 0, 0, 0)^t$, $\boldsymbol{C}_{Cm,1} = (0, 0, 0, 0, 1, 0, 0, -1)^t$, or $\boldsymbol{C}_{Cm,2} = (0, 0, 0, 0, 0, 1, -1, 0)^t$.
Therefore,

$$2(n_1 + n_2 - 2)\boldsymbol{S}_T = \boldsymbol{SSD}_{21} + \boldsymbol{SSD}_{12} =$$
$$\begin{pmatrix} s(1, \boldsymbol{C}_{AU,2}, \boldsymbol{C}_{AU,2}) & s(1, \boldsymbol{C}_{AU,2}, \boldsymbol{C}_{Cm,2}) \\ s(1, \boldsymbol{C}_{AU,2}, \boldsymbol{C}_{Cm,2}) & s(1, \boldsymbol{C}_{Cm,2}, \boldsymbol{C}_{Cm,2}) \end{pmatrix}$$
$$+ \begin{pmatrix} s(2, \boldsymbol{C}_{AU,1}, \boldsymbol{C}_{AU,1}) & s(2, \boldsymbol{C}_{AU,1}, \boldsymbol{C}_{Cm,1}) \\ s(2, \boldsymbol{C}_{AU,1}, \boldsymbol{C}_{Cm,1}) & s(2, \boldsymbol{C}_{Cm,1}, \boldsymbol{C}_{Cm,1}) \end{pmatrix}$$
$$2(n_1 + n_2 - 2)\boldsymbol{S}_R = \boldsymbol{SSD}_{11} + \boldsymbol{SSD}_{22} =$$
$$\begin{pmatrix} s(1, \boldsymbol{C}_{AU,1}, \boldsymbol{C}_{AU,1}) & s(1, \boldsymbol{C}_{AU,1}, \boldsymbol{C}_{Cm,1}) \\ s(1, \boldsymbol{C}_{AU,1}, \boldsymbol{C}_{Cm,1}) & s(1, \boldsymbol{C}_{Cm,1}, \boldsymbol{C}_{Cm,1}) \end{pmatrix}$$
$$+ \begin{pmatrix} s(2, \boldsymbol{C}_{AUC,2}, \boldsymbol{C}_{AU,2}) & s(2, \boldsymbol{C}_{AU,2}, \boldsymbol{C}_{Cm,2}) \\ s(2, \boldsymbol{C}_{AUC,2}, \boldsymbol{C}_{Cm,2}) & s(2, \boldsymbol{C}_{Cm,2}, \boldsymbol{C}_{Cm,2}) \end{pmatrix}$$

$\boldsymbol{SSD}_{w1}$ and $\boldsymbol{SSD}_{w2}$ are independent since subjects in two groups are distinct. To show the indenpendence between $\boldsymbol{S}_T$ and $\boldsymbol{S}_R$, we show that each element of $\boldsymbol{SSD}_{1i}$ and each element of $\boldsymbol{SSD}_{2i}$ are mutually independent as follows.

Generally, covariance of the quadratic forms $\boldsymbol{Y}^t\boldsymbol{A}\boldsymbol{Y}$ and $\boldsymbol{Y}^t\boldsymbol{B}\boldsymbol{Y}$ has the following property. $\text{Cov}(\boldsymbol{Y}^t\boldsymbol{A}\boldsymbol{Y}, \boldsymbol{Y}^t\boldsymbol{B}\boldsymbol{Y}) = 2tr(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}^t\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\mu}$, when $\boldsymbol{Y}$ is distributed as $\boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (see for example Khuri (2009)). It is known that $\boldsymbol{Y}^t\boldsymbol{A}\boldsymbol{Y}$ and $\boldsymbol{Y}^t\boldsymbol{B}\boldsymbol{Y}$ are independent if and only if $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = \boldsymbol{0}$ (see for example Muirhead (1982)).

Now we set $\boldsymbol{A} = [\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_1][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_2^t]$ and $\boldsymbol{B} = [\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{1*}][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{2*}^t]$. Then, we have

$\boldsymbol{A}\boldsymbol{V}_i\boldsymbol{B}$
$$= [\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_1][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_2^t]\text{Cov}(\boldsymbol{Y}_i)$$
$$[\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{1*}][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{2*}^t]$$
$$= [\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_1][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_2^t](\boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma}_i)$$
$$[\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{1*}][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{2*}^t]$$
$$= [\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_1][\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*](\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_2^t\boldsymbol{\Sigma}_i\boldsymbol{C}_{1*})[\boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}^*][\boldsymbol{I}_{n_i} \otimes \boldsymbol{C}_{2*}^t],$$
where $\boldsymbol{C}_{1*}$ and $\boldsymbol{C}_{2*}$ are selected from $\boldsymbol{C}_{AU,1}, \boldsymbol{C}_{AU,2}, \boldsymbol{C}_{Cm,1}$, or $\boldsymbol{C}_{Cm,2}$.

The last line above follows from the following property of kronecker product: $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D})(\boldsymbol{E} \otimes \boldsymbol{F}) = \boldsymbol{ACE} \otimes \boldsymbol{BDF}$. When we consider each element of $\boldsymbol{SSD}_{1i}$ and each element of $\boldsymbol{SSD}_{2i}$, $\boldsymbol{C}_2^t \boldsymbol{\Sigma}_i \boldsymbol{C}_{1*} = \boldsymbol{0}$. Therefore, $(n_1+n_2-2)\boldsymbol{S}_T$ and $(n_1 + n_2 - 2)\boldsymbol{S}_R$ are independent.

Yoshiomi Nakazuru
Pfizer Japan Inc.
3-22-7, Yoyogi, Shibuya-ku, Tokyo 151-8589, Japan
E-mail: yoshiomi.nakazuru@pfizer.com

Takashi Seo
Tokyo University of Science
Department of Mathematical Information Science, Faculty of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
E-mail: seo@rs.tus.ac.jp