Mathematical Statistics
Stockholm University

Multivariate bioequivalence with respect to both means and variability for general equivalence restrictions

Niclas Sjögren

Research Report 2002:1

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.matematik.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2002:1**,
http://www.matematik.su.se/matstat

Multivariate bioequivalence with respect to both means and variability for general equivalence restrictions

Niclas Sjögren*

January 2002

Abstract

In a 2×2 crossover trial with multivariate response, it is shown how to test for bioequivalence when the prespecified bioequivalence restrictions can concern differences in fixed treatment effects as well as in within subject variability. The restrictions are expressed as intervals, within which expectations and variances for a set of linear combinations of the examined pharmacokinetic parameters should be to claim bioequivalence. As in the univariate case, the intersection-union test and the confidence set approach are the same test if the level of the confidence set is chosen appropriately. The probability rejecting the null hypotheses erroneously are controlled in a conservative way. In the special case, when the restrictions only involve individual pharmacokinetic parameters, it is shown how to construct a test at desired size. The statistical tests are semiparametric in that no distributional assumption is made about the between subject variability, whereas a normality assumption is made about the within subject variability.

KEY WORDS: Bioequivalence; bioavailability; intersection-union; semiparametric; simultaneous inference.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: Niclas.Sjogren@matematik.su.se.

1 Introduction

In recent years, generic drug products have become very popular. For the approval of a new generic drug product, the United States Food and Drug Administration (FDA) usually does not require a new drug application, if the generic drug company can provide the evidence of bioequivalence between the generic drug product and the innovator. Thus, bioequivalence studies are important since a new drug application submission is very time consuming and costly to obtain.

The definition of bioavailability has evolved over time with different meanings by different individuals and organizations. However, the most common used definition is that adopted by the FDA in 1983: the bioavailability of a drug product is defined as "*the rate and extent to which the active drug ingredient or therapeutic moiety of the drug product becomes available at the site of drug action.*" For the approval of generic drugs, a bioequivalence assessment, as a surrogate for the clinical evaluation of the generic drug products, is based on the following fundamental bioequivalence assumption: "*when two drug products are equivalent in the rate and extent to which the active drug ingredient or therapeutic moiety is absorbed and becomes available at the site of drug action, it is assumed that they will be therapeutically equivalent.*" The concepts of bioavailability and bioequivalence became a public issue in the late 1960's when concern was raised that a generic product might not be as bioavailable as that manufactured by the innovator.

In bioavailability studies, the blood or plasma concentration-time curve is often used to study the absorption and elimination of the drug. The curve can be characterized by taking blood samples immediately prior to and at various time points after drug administration. The profile of the curve is then usually studied by means of several pharmacokinetic parameters, such as the area under the blood or plasma concentration-time curve (AUC), maximum concentration (C_{max}) and time to reach maximum concentration (T_{max}), etc. The measurements of these pharmacokinetic parameters can be estimated either directly from the observed curve or by fitting the observed concentrations to a one or multicompartment pharmacokinetic model, Okpako (1991). Among the pharmacokinetic parameters, AUC is considered the primary measure for the extent of absorption, which provides information regarding the total amount of the drug absorbed in the body.

Statistical tests for the assessment of bioequivalence use the estimated pharmacokinetic parameters mentioned above without taking any account of the measured concentration-time points. It is obvious that we loose information when we replace raw data with some estimated parameters. Even though we think that we model the estimated pharmacokinetic parameters in the first place, our multivariate model and results are also applicable on the raw data.

Although the FDA still recommends testing for average bioequivalence in most bioequivalence studies, the present FDA guideline (FDA (2001)) also describes "population" and "individual" bioequivalence that besides averages of the measures also focus on the variances of the measures. In the last years, a lot of work have been done on how to test for "population" and "individual" bioequivalence, see for example the special issue on bioavailability/bioequivalence in Journal of Biopharmaceutical Statistics, Chow (1997). The FDA recommends testing for "population" respective "individual" bioequivalence by using single criterions that contain both means and variances of the measurement. That is done for individual pharmacokinetic parameters one at a time. We find it interesting to make a simultaneous test that includes restrictions on both means and variability for all interesting measurements. Wang *et al.* (1999) did that for rectangular bioequivalence restrictions on the means. We first generalize this for other restrictions on the means. Then we show how to construct a size α test for testing equality of within subject variability and also how to combine this to a test concerning both means and variances for all examined characteristics simultaneously.

The outline of this report is as follows. We describe the multivariate statistical model, assumptions and notations in section 2; In section 3 we show how to test for bioequivalence with respect to fixed treatment effects for general equivalence regions. We also show that the confidence interval approach and the intersection-union approach are equivalent if the confidence level in the confidence set is chosen appropriately; In Section 4 we test for equality in variability. In the univariate case, we show how to construct an exact size α test without distributional assumption made about the between subject variability. Further on, we show that intersection-union tests can be used to construct a size α test for equality in variability for all measured characteristics simultaneously. Finally in section 5, we combine the results in section 3 and section 4 ending up in a test for bioequivalence that includes restrictions on both direct treatments effects and within subject variability.

2 Statistical model and assumptions

In a two-sequence, two-period crossover design with multivariate responses, we assume that p characteristics are measured after each of the two treatments A and B. The characteristics can for example be pharmacokinetic parameters such as AUC; C_{max}; and T_{max}, etc, but can also be repeated measurements of the concentration of the active drug at fixed time points, $t_1, t_2, \ldots, t_p$ after drug administration. However, each subject gives rise to two multivariate responses, one in each period. Let $\mathbf{Y}_{ij1}$ and $\mathbf{Y}_{ij2}$ denote the $p \times 1$ response vectors of the ith subject in the jth sequence group in period 1 and period 2 respectively, $i = 1, 2, \ldots, n_j$ and $j = 1, 2$. Sequence group 1 corresponds to treatment sequence AB whereas sequence group 2 corresponds to treatment sequence BA. The $p \times 2$

matrix, $(\mathbf{Y}_{ij1}, \mathbf{Y}_{ij2})$ can be described by the following statistical model

$$(\mathbf{Y}_{ij1}, \mathbf{Y}_{ij2}) = (\boldsymbol{\mu}_{j1}, \boldsymbol{\mu}_{j2}) + (\boldsymbol{\xi}_{ij} + \boldsymbol{\varepsilon}_{ij1}, \boldsymbol{\xi}_{ij} + \boldsymbol{\varepsilon}_{ij2}), \tag{1}$$

where $\boldsymbol{\mu}_{j1}$ and $\boldsymbol{\mu}_{j2}$ are nonrandom vectors reflecting fixed effects such as direct treatment effects and period effects; $\boldsymbol{\xi}_{ij}$ is the random effect vector of the ith subject in the jth sequence; and $\boldsymbol{\varepsilon}_{ij1}$ and $\boldsymbol{\varepsilon}_{ij2}$ are the vectors of within-subject random errors in observing $\mathbf{Y}_{ij1}$ respective $\mathbf{Y}_{ij2}$. $n_j \geq 2$ denotes the number of subjects randomized into sequence group j, $j = 1, 2$. Moreover, let

$$n = n_1 + n_2$$

denote the total number of subjects in the study.

We assume that the direct treatment effects and the period effects are fixed and additive and that there are no carry-over effects. These assumptions imply that the fixed effects $\boldsymbol{\mu}_{j1}$ and $\boldsymbol{\mu}_{j2}$ in model (1) satisfy

$$\boldsymbol{\mu}_{j1} - \boldsymbol{\mu}_{j2} = \begin{cases} \boldsymbol{\pi} + \boldsymbol{\tau}, & \text{if } j = 1 \\ \boldsymbol{\pi} - \boldsymbol{\tau}, & \text{if } j = 2 \end{cases}, \tag{2}$$

where $\boldsymbol{\pi}$ is the constant vector equaling the (period 1 - period 2) difference of the fixed period effects and $\boldsymbol{\tau}$ is the constant vector equaling the direct (treatment A - treatment B) fixed effects, $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_p)'$

The only distributional assumption made in this report concerns the n within random vectors $\boldsymbol{\varepsilon}_{i1k}$ and $\boldsymbol{\varepsilon}_{i2k}$ in model (1). They are assumed to be mutually independent and independent of the n between random vectors $\boldsymbol{\xi}_{ij}$. Moreover, they are assumed to be normally distributed with expectation zero and covariance matrices depending on the treatment in the corresponding period. Thus, with $=_d$ denoting equality in distribution, it is assumed that for $i = 1, 2, \ldots, n_j$,

$$(\boldsymbol{\varepsilon}_{ij1}, \boldsymbol{\varepsilon}_{ij2}) =_d \begin{cases} (\boldsymbol{\varepsilon}_A, \boldsymbol{\varepsilon}_B), & \text{if } j = 1 \\ (\boldsymbol{\varepsilon}_B, \boldsymbol{\varepsilon}_A), & \text{if } j = 2 \end{cases} \tag{3}$$

where $\boldsymbol{\varepsilon}_A$ and $\boldsymbol{\varepsilon}_B$ are independent multivariate normal distributed stochastic variables with zero means and covariance matrices $\boldsymbol{\Lambda}_A$ respective $\boldsymbol{\Lambda}_B$, we write $\boldsymbol{\varepsilon}_A \sim N_p(\mathbf{0}, \boldsymbol{\Lambda}_A)$ respective $\boldsymbol{\varepsilon}_B \sim N_p(\mathbf{0}, \boldsymbol{\Lambda}_B)$. The indexes $_A$ and $_B$ indicates treatments. Note that no assumption is made about the joint distribution of the n between subject random vectors, $\boldsymbol{\xi}_{ij}$ in model (1). In particular, they need neither be independent nor identically distributed.

Let $\mathbf{Y}^+_{ij}$ and $\mathbf{Y}^-_{ij}$, $i = 1, 2, \ldots, n_j$ denote the column vectors containing within-subject sums and (A-B)-differences, i.e.

$$\mathbf{Y}^+_{ij} = \mathbf{Y}_{ij1} + \mathbf{Y}_{ij2} \tag{4}$$

$$\mathbf{Y}_{ij}^{-} = \begin{cases} \mathbf{Y}_{ij1} - \mathbf{Y}_{ij2}, & \text{if } j = 1 \\ \mathbf{Y}_{ij2} - \mathbf{Y}_{ij1}, & \text{if } j = 2 \end{cases} \tag{5}$$

Statistical inference about the fixed (treatment A- treatment B) effects, $\boldsymbol{\tau}$, is made by analyzing the within subject (A-B)-differences, $\mathbf{Y}_{ij}^{-}$, whereas both $\mathbf{Y}_{ij}^{+}$ and $\mathbf{Y}_{ij}^{-}$ are used to make statistical inference about the within subject variability.

3 Testing for bioequivalence with respect to differences in fixed treatment effects

We begin this section with a description on how to make statistical inference about differences in fixed treatment effects. This is easily done using similar transformations as that of Jones and Kenward (1989) for the univariate case. It ends up in a test based on the Hotelling´s T^2 statistic instead of Student´s t-distribution. Then we show how to test for bioequivalence for general bioequivalence restrictions on differences in fixed treatment effects. Further on, we describe the intersection-union test, likelihood ratio test and the confidence interval approach for testing bioequivalence under this circumstances and we claim that these three approaches in fact ends up in the same test.

3.1 Statistics to make inferences about difference in fixed treatment effects

Statistical inference about the difference in fixed treatment effects, $\boldsymbol{\tau}$, is made through the within subject (A-B)-differences, $\mathbf{Y}_{ij}^{-}$, defined in (5). The statistical model for $\mathbf{Y}_{ij}^{-}$ can be written as

$$\mathbf{Y}_{ij}^{-} = \boldsymbol{\mu}_j^{-} + \boldsymbol{\varepsilon}_{ij}^{A} - \boldsymbol{\varepsilon}_{ij}^{B}, \tag{6}$$

where the vectors $\boldsymbol{\mu}_1^{-}$ and $\boldsymbol{\mu}_2^{-}$ are fixed effects and $\boldsymbol{\varepsilon}_{ij}^{A}$ and $\boldsymbol{\varepsilon}_{ij}^{B}$ are independent multivariate normal distributed with zero means and covariance matrices $\boldsymbol{\Lambda}_A$ and $\boldsymbol{\Lambda}_B$ respectively. This mean that $\mathbf{Y}_{ij}^{-}$ is normally distributed with mean $\boldsymbol{\mu}_j^{-}$ and covariance matrix $\boldsymbol{\Sigma}_{--} = \boldsymbol{\Lambda}_A + \boldsymbol{\Lambda}_B$, i.e.

$$\mathbf{Y}_{ij}^{-} \sim N_p(\boldsymbol{\mu}_j^{-}, \boldsymbol{\Sigma}_{--}),$$

where the fixed effects $\boldsymbol{\mu}_1^{-}$ and $\boldsymbol{\mu}_2^{-}$ equals $\boldsymbol{\tau} + \boldsymbol{\pi}$ respective $\boldsymbol{\tau} - \boldsymbol{\pi}$ because of (5) and (2). Exact multivariate statistical inference about $\boldsymbol{\tau}$ can be made using the statistic $\mathbf{D}$ defined by

$$\mathbf{D} = (\bar{\mathbf{Y}}_1^{-} + \bar{\mathbf{Y}}_2^{-})/2,$$

where $\bar{\mathbf{Y}}_1^- = (1/n_1)\sum_{i=1}^{n_1}\mathbf{Y}_{i1}^-$ and $\bar{\mathbf{Y}}_2^- = (1/n_2)\sum_{i=1}^{n_2}\mathbf{Y}_{i2}^-$ are the mean differences of (treatment A- treatment B) effects in sequence group 1 respective sequence group 2. Now we have that

$$\mathbf{D} \sim N_p(\boldsymbol{\tau}, \boldsymbol{\Sigma}_\mathbf{D}), \tag{7}$$

where the covariance matrix $\boldsymbol{\Sigma}_\mathbf{D}$ is

$$\boldsymbol{\Sigma}_\mathbf{D} = (1/n_1 + 1/n_2)\boldsymbol{\Sigma}_{--}/4,$$

and is estimated by

$$\widehat{\boldsymbol{\Sigma}}_\mathbf{D} = (1/n_1 + 1/n_2)\widehat{\boldsymbol{\Sigma}}_{--}/4,$$

where

$$\widehat{\boldsymbol{\Sigma}}_{--} = 1/(n-2)\sum_{j=1}^{2}\sum_{i=1}^{n_j}(\mathbf{Y}_{ij}^- - \bar{\mathbf{Y}}_j^-)(\mathbf{Y}_{ij}^- - \bar{\mathbf{Y}}_j^-)'.$$

Note that $(n-2)\widehat{\boldsymbol{\Sigma}}_{--}$ is Wishart distributed with $n-2$ degrees of freedom, i.e.

$$(n-2)\widehat{\boldsymbol{\Sigma}}_{--} \sim W_p(n-2, \boldsymbol{\Sigma}_{--}).$$

3.2 A general criteria for bioequivalence

Wang *et al.* (1999) showed how to test for bioequivalence simultaneously, when bioequivalence is claimed if all elements in the vector of the fixed direct treatment effects, $\boldsymbol{\tau}$, lies inside a cube. Assume that Δ is the cut-off number specifying bioequivalence. Then, testing for bioequivalence is to test

$$H_0 : \max_{1\le k\le p}|\tau_k| \ge \Delta \qquad \text{versus} \qquad H_A : \max_{1\le k\le p}|\tau_k| < \Delta. \tag{8}$$

For the one-dimensional case ($p = 1$), Schuirmann (1981) proposed what has become a standard test of (8). It is called the "two one-sided tests" (TOST) and is equivalent to the symmetric confidence interval approach proposed by Westlake (1981). To test (8) when $p > 1$, Wang *et al.* (1999) proposed a test based on the intersection-union method. This test has rejection region

$$R^I = \bigcap_{k=1}^{p}\left\{|\mathbf{D}_k| < \Delta - t_{n-2}(\alpha)\sqrt{(\widehat{\boldsymbol{\Sigma}}_\mathbf{D})_{kk}}\right\},$$

which defines a size-α test for (8). Wang *et al.* (1999) also conclude that this intersection-union test is equivalent to the likelihood ratio test as well as the confidence set approach for testing (8).

We show how to generalize these results for a more general bioequivalence criteria on the direct fixed (treatment A- treatment B) effects, $\boldsymbol{\tau}$. A criteria defined by putting restrictions on a number of linear combinations of $\boldsymbol{\tau}$. That is a test of

$$H_0 : \max_{1\le k\le q} |\mathbf{a}_k'\boldsymbol{\tau}| \ge \Delta \qquad \text{versus} \qquad H_A : \max_{1\le k\le q} |\mathbf{a}_k'\boldsymbol{\tau}| < \Delta, \tag{9}$$

where the $\mathbf{a}_k'$'s are row vectors of unit length, i.e. $\mathbf{a}_k \cdot \mathbf{a}_k = 1$, $k = 1, 2, \ldots, q$.

If $\mathbf{a}_k$ is the kth coordinate vector, $k = 1, 2, \ldots, p$, then the region of equivalence is simply the original one described by Wang *et al.* (1999), i.e. we test (8). We can specify other equivalence criteria by using other linear combinations. The following are examples in the 2-dimensional case, i.e. when two characteristics are measured.

1. $\mathbf{a}_1 = (1, 0)'$ and $\mathbf{a}_2 = (0, 1)'$ gives the quadratic region of equivalence, see Figure 1.
2. $\mathbf{a}_1 = (1, 0)'$, $\mathbf{a}_2 = (0, 1)'$, $\mathbf{a}_3 = (1/\sqrt{2}, 1/\sqrt{2})'$ and $\mathbf{a}_4 = (-1/\sqrt{2}, 1/\sqrt{2})'$ gives the region in Figure 2.
3. Including all vectors of unit length (if possible), gives a circle with radius Δ, see Figure 3.

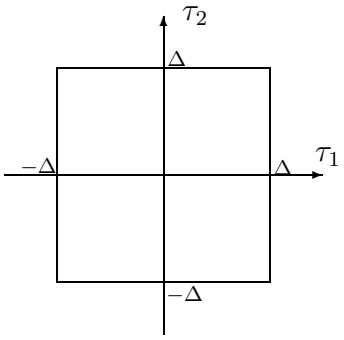

Figure 1:

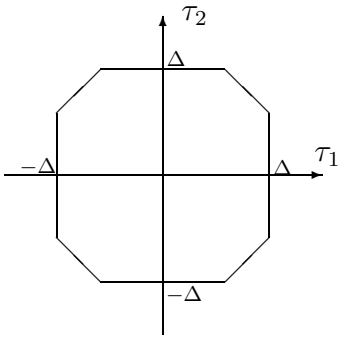

Figure 2:

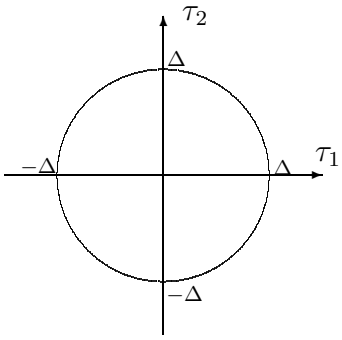

Figure 3:

If we let $\mathbf{A}$ denote the $q \times p$ matrix with row vectors $\mathbf{a}_k'$, $k = 1, 2, \ldots, q$, then we can use the transformed variables, $\mathbf{AY}_{ij1}$ and $\mathbf{AY}_{ij2}$ instead of the original ones to test (9). The $\mathbf{D}$ statistic for the transformed variables equals $\mathbf{AD}$ and is an observation from the q-variate normal distribution with mean $\mathbf{A}\boldsymbol{\tau}$ and covariance matrix $\mathbf{A}\boldsymbol{\Sigma}_D\mathbf{A}'$, i.e.

$$\mathbf{AD} \sim N_q\Big(\mathbf{A}\boldsymbol{\tau}, \mathbf{A}\boldsymbol{\Sigma}_\mathbf{D}\mathbf{A}'\Big).$$

Note that the covariance matrix, $\mathbf{A}\boldsymbol{\Sigma}_\mathbf{D}\mathbf{A}'$, is singular if $q > p$.

We can use the intersection-union test to test the hypothesis (9) under model (1). Let us consider the q sets of hypotheses,

$$H_{0_k} : |\mathbf{a}_k^{'}\boldsymbol{\tau}| \geq \Delta \quad \text{versus} \quad H_{A_k} : |\mathbf{a}_k^{'}\boldsymbol{\tau}| < \Delta, \text{ for } k = 1, 2, \ldots, q. \tag{10}$$

Note that the test proposed by Wang *et al.* (1999) is a test of (9) in the special case when the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_q$ equals the coordinate vectors and $q = p$. The null hypothesis and alternative hypothesis in (9) can be seen as the union respective intersection of the q sets of hypothesis in (10), i.e. $H_0 = \bigcup_{k=1}^{q} H_{0_k}$ and $H_0 = \bigcap_{k=1}^{q} H_{A_k}$ respectively. For each set of the hypothesis in (10), the size-α TOST has rejection region

$$R_k = \left\{ |\mathbf{a}_k'\mathbf{D}| < \Delta - t_{n-2}(\alpha)(\mathbf{a}_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2} \right\}. \tag{11}$$

Using the intersection-union method, it follows that the rejection region

$$R^I_{\boldsymbol{\tau}} = \bigcap_{k=1}^{q} R_k = \left\{ |\mathbf{a}_k'\mathbf{D}| < \Delta - t_{n-2}(\alpha)(\mathbf{a}_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2}, \text{ for all k} \right\} \tag{12}$$

defines a level-α test for (9).

Theorem 3.1:
The intersection-union test, $R^I_{\boldsymbol{\tau}}$, for testing (9) under model (1) has size α.

Proof: As noted before, each R_k defines a TOST of size α. The theorem follows from theorem 2 in Berger and Hsu (1996), which is applicable here because there is a sequence of parameter points, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_{2,\ldots}$, in the parameter space of the alternative, $H_{A_{k'}}$ such that

$$\lim_{l \to \infty} P_{\boldsymbol{\theta}_l}(\mathbf{Y} \in R_{k'}) = \alpha \tag{13}$$

and, for every $k = 1, 2, \ldots, q$, $k \neq k'$,

$$\lim_{l \to \infty} P_{\boldsymbol{\theta}_l}(\mathbf{Y} \in R_k) = 1. \tag{14}$$

Let $\boldsymbol{\theta}_l = (\boldsymbol{\tau}_l = \Delta\mathbf{a}_{k'}, \boldsymbol{\Sigma}^l_{\mathbf{D}} = \delta_l\boldsymbol{\Sigma}'_{\mathbf{D}})$ be a sequence of parameter points, where $\boldsymbol{\Sigma}'_{\mathbf{D}}$ is a fixed proper covariance matrix and δ_l are small positive numbers that goes to zero as l goes to infinity, i.e. $\delta_l \to 0$ as $l \to \infty$. Using (11) and that $\mathbf{a}_k'\mathbf{D}$ is normally distributed with mean $\Delta\mathbf{a}_k'\mathbf{a}_{k'}$ and variance $\delta\mathbf{a}_k'\boldsymbol{\Sigma}'_{\mathbf{D}}\mathbf{a}_k$ for all points in the parameter sequence $\boldsymbol{\theta}_l$, we have that

$$\begin{aligned} P_{\boldsymbol{\theta}_l}(\mathbf{Y} \in R_k) \; &= P_{\boldsymbol{\theta}_l}\left(T_{n-2} < -t_{n-2}(\alpha) + (\Delta - \Delta\mathbf{a}_k'\mathbf{a}_{k'})/(\mathbf{a}_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2} \right) - \\ &\quad - P_{\boldsymbol{\theta}_l}\left(T_{n-2} < t_{n-2}(\alpha) - (\Delta + \Delta\mathbf{a}_k'\mathbf{a}_{k'})/(\mathbf{a}_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2} \right), \end{aligned} \tag{15}$$

where

$$T_{n-2} = (\mathbf{a}_k'\mathbf{D} - \Delta\mathbf{a}_k'\mathbf{a}_{k'})/(\mathbf{a}_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2} \sim t_{n-2}.$$

For $k = k'$ we have that $\mathbf{a}'_{k'}\mathbf{a}_{k'} = 1$ so the first term in (15) equals α. The second term goes to zero because $\mathbf{a}'_k\widehat{\mathbf{\Sigma}}_{\mathbf{D}}\mathbf{a}_k \to 0$ with probability 1 as $l \to \infty$. Thus, (13) holds. On the other hand if $k \neq k'$, we have that $\mathbf{a}'_k\mathbf{a}_{k'} < 1$ and it follows that the first term in (15) goes to one and the second term goes to zero as $l \to \infty$. Thus, (14) holds for $k \neq k'$.

□

The intersection-union test (12) is basically a likelihood ratio test. The likelihood ratio, $\lambda(\mathbf{D})$ equals (see for example Krzanowski (1988)).

$$\lambda(D) = \frac{\sup_{H_0} L(\boldsymbol{\tau}, \mathbf{\Sigma}_{\mathbf{D}}; \mathbf{D})}{\sup_{H_0 \cup H_A} L(\boldsymbol{\tau}, \mathbf{\Sigma}_{\mathbf{D}}; \mathbf{D})} = \left(\frac{1}{1 + \frac{1}{n-2}\inf_{\boldsymbol{\tau} \in H_0}(\mathbf{D} - \boldsymbol{\tau})'\widehat{\mathbf{\Sigma}}_{\mathbf{D}}^{-1}(\mathbf{D} - \boldsymbol{\tau})} \right)^{n/2},$$

where $L(.)$ denotes the likelihood function for model (7). The likelihood ratio test rejects H_0 if and only if $\lambda(D) < K$ or equivalent if and only if

$$\inf_{\boldsymbol{\tau} \in H_0}(\mathbf{D} - \boldsymbol{\tau})'\mathbf{\Sigma}_{\mathbf{D}}^{-1}(\mathbf{D} - \boldsymbol{\tau}) > (K^{-2/n} - 1)(n-2), \tag{16}$$

where K determines the level of the test. The following theorem implies that the intersection-union test and the likelihood ratio test are equivalent for proper choice of K.

Theorem 3.2:
For $0 < K < 1$, $\lambda(D) < K$ if and only if

$$|\mathbf{a}'_k\mathbf{D}| < \Delta - t_{n-2}(\alpha)(\mathbf{a}'_k\widehat{\mathbf{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2}, \text{ for } k = 1, 2, \ldots, q \tag{17}$$

where $K = \left(t^2_{n-2}(\alpha)/(n-2) + 1\right)^{-n/2}$.

Proof: Wang *et al.* (1999) proved this in the case when the row vectors in the $p \times p$ matrix $\mathbf{A}$ equal the coordinate vectors. If $\mathbf{A}$ is of full rang, but not includes all coordinate vectors, then the results of Wang *et al.* (1999) holds if we change coordinate system, i.e if we transform the original p characteristics into p new ones. The result of Wang *et al.* (1999) holds of course for the p new characteristics. Thus, it is evident that the result of Wang *et al.* (1999) holds for all $p \times p$ matrices of full rang. It remains to prove that the theorem holds even for other matrices.

Let $\mathbf{A}_m$ denote the $p \times p$ matrix containing the p different row vectors, $\mathbf{a}'_{m_k}$, $k = 1, 2, \ldots, p$, from $\mathbf{A}$. It is possible to construct $\binom{q}{p}$ different such $p \times p$ matrices of full rank. We have

$$\mathbf{A}_m\mathbf{D} \sim N_p(\mathbf{A}_m\boldsymbol{\tau}, \mathbf{A}_m\mathbf{\Sigma}_{\mathbf{D}}\mathbf{A}'_m), \text{ for } m = 1, 2, \ldots, \binom{q}{p},$$

where $\mathbf{A}_m\mathbf{\Sigma}_{\mathbf{D}}\mathbf{A}'_m$ is a covariance matrix. From Wang *et al.* (1999) we have that the intersection-union test and the likelihood ratio test for testing

$$H_{0_m} : \max_{1 \leq i \leq p} |\mathbf{a}'_{m_i}\boldsymbol{\tau}| \geq \Delta \qquad \text{versus} \qquad H_{A_m} : \max_{1 \leq i \leq p} |\mathbf{a}'_{m_i}\boldsymbol{\tau}| < \Delta \tag{18}$$

are equivalent, i.e.

$$\lambda^{(m)}(\mathbf{D}) < K \iff |\mathbf{a}'_{m_i} D| < \Delta - t_{n-2}(\alpha)(\mathbf{a}'_{m_i} \widehat{\mathbf{\Sigma}}_{\mathbf{D}} \mathbf{a}_{m_i})^{1/2}, \text{ for } i = 1, 2, \ldots, p, \tag{19}$$

where $\lambda^{(m)}(\mathbf{D})$ denotes the likelihood ratio test corresponding to matrix $\mathbf{A}_m$ and

$$K = \left(t^2_{n-2}(\alpha)/(n-2) + 1\right)^{-n/2}.$$

Further, note that the restrictions in the hypothesis (18) can be expressed as restrictions on $\boldsymbol{\tau}$ instead of those on $\mathbf{a}'_m \boldsymbol{\tau}$. This mean that the hypothesis (18) can be written as

$$H_{0_m} : \boldsymbol{\tau} \notin \bigcap_{i=1}^{p} r(\Delta, \mathbf{a}_{m_i}) \qquad \text{versus} \qquad H_{A_m} : \boldsymbol{\tau} \in \bigcap_{i=1}^{p} r(\Delta, \mathbf{a}_{m_i}), \tag{20}$$

where $r(\Delta, \mathbf{a}_{m_i})$ is the region in R^p where $|\mathbf{a}'_{m_i} \boldsymbol{\tau}| < \Delta$ for $i = 1, 2, \ldots, p$. Thus, we have that

$$\lambda^{m}(\mathbf{D}) < K \iff \inf_{\boldsymbol{\tau} \in H_{0_m}} (\mathbf{D} - \boldsymbol{\tau})' \widehat{\mathbf{\Sigma}}^{-1}_{\mathbf{D}} (\mathbf{D} - \boldsymbol{\tau}) > (K^{-2/n} - 1)(n-2). \tag{21}$$

Using (16) and the fact that H_0 in (9) can be written as the union of all H_{0_m} in (20) we have that

$$\lambda(\mathbf{D}) < K \iff \inf_{\boldsymbol{\tau} \in \bigcup_{m=1}^{\binom{q}{p}} H_{0_m}} (\mathbf{D} - \boldsymbol{\tau})' \widehat{\mathbf{\Sigma}}^{-1}_{\mathbf{D}} (\mathbf{D} - \boldsymbol{\tau}) > (K^{-2/n} - 1)(n-2).$$

The right hand side holds if and only if the right hand side of (21) holds for $m = 1, 2, \ldots, \binom{q}{p}$. Then, using (19) and that the $\mathbf{a}_{m_i}$'s are just the $\mathbf{a}_k$'s we have

$$\lambda(\mathbf{D}) < K \iff |\mathbf{a}'_k D| < \Delta - t_{n-2}(\alpha)(\mathbf{a}'_k \widehat{\mathbf{\Sigma}}_{\mathbf{D}} \mathbf{a}_k)^{1/2}, \text{ for } k = 1, 2, \ldots, q.$$

□

3.3 The confidence set approach

Another way to obtain a test for (9) is to use the confidence set approach. Let $C(Y)$ be a $1 - \alpha$ confidence set for $\boldsymbol{\tau}$. Then, the test for (9) that rejects the null hypothesis if and only if

$$C(\mathbf{Y}) \subset H_A \tag{22}$$

has level α. Consider the confidence set based on the Hotelling's T^2, which in this case is

$$T^2 = (\mathbf{D} - \boldsymbol{\tau})' \widehat{\mathbf{\Sigma}}^{-1}_{\mathbf{D}} (\mathbf{D} - \boldsymbol{\tau}).$$

Then, the set

$$C(\mathbf{Y}) = \{\boldsymbol{\tau} : T^2 \leq C_1^2\}, \tag{23}$$

has coverage probability $1-\alpha$, if

$$C_1^2 = F_{p,n-1-p}(\alpha)((n-2)p)/(n-1-p), \tag{24}$$

where $F_{p,n-1-p}(\alpha)$ is the upper α quantile of the F distribution with p and $n-1-p$ degrees of freedom. Putting the test (22) together with the confidence set (23) yields a test with rejection region

$$|\mathbf{a}_k'\mathbf{D}| < \Delta - C_1(\mathbf{a}_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2},\ k = 1, 2, ..., q \tag{25}$$

That is the intersection-union test (17), despite that the t-percentile, $t_{n-2}(\alpha)$ has been replaced with C_1.

Theorem 3.3:
The test based on the confidence interval approach described above has size

$$\alpha^* = P(T_{n-2} < -C_1), \tag{26}$$

where T_{n-2} denotes a stochastic variable that is t-distributed with $n-2$ degrees of freedom and C_1 is the positive square root of (24).

Proof: Because the confidence interval approach ends up in the intersection-union test with rejection region (25), which despite of C_1 is the same as the intersection-union test (17), the size α^* equals the size of the individual tests as it is for the intersection-union test (17). Thus, we get the size α^* from (15) when $\Delta\mathbf{a_k}'\mathbf{a_{k'}} = \Delta$ and $l \to \infty$. The size equals

$$\begin{aligned} \alpha^* &= P\left((\mathbf{a}_k'\mathbf{D} - \Delta)/(a_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k)^{1/2} < -C_1\right) = \\ &= P(T_{n-2} < -C_1), \end{aligned} \tag{27}$$

where T_{n-2} is t-distributed with $n-2$ degrees of freedom.

□

Using the symmetry of the t-distribution we see that the probability in (27) can be written as

$$P\left((\mathbf{a}_k'\mathbf{D} - \Delta)^2/(a_k'\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k) > C_1^2\right)/2.$$

Thus, another way to express the actual size of the confidence set approach is that of Wang *et al.* (1999), i.e.

$$\alpha^* = P\left(\chi_1^2/\chi_{n-2}^2 > C_1^2\right)/2,$$

where χ^2_1 and χ^2_{n-2} denotes independent chi-squared random variables with 1 respective $n-2$ degrees of freedom. That is

$$\alpha^* = P\left(F_{1,n-2} > (n-2)C_1^2\right)/2,$$

where $F_{1,n-2}$ denotes a F-distributed stochastic variable with 1 respective $n-2$ degrees of freedom.

The actual size decreases fast with the number of parameters but does not depend on the number of restrictions. Table 1 shows the actual size for the confidence set approach when we reject H_0 if and only if the symmetric 95% confidence set lies within H_A. On the other hand, Table 2 shows the appropriate level on the confidence set so that the corresponding test has size 0.05.

Table 1: *The actual size α^* of the test based on the confidence interval approach derived from a 95% confidence set ($\alpha = 0.05$ in (24)), the number of subjects n as well as the number of characteristics p varies.*

n	p=1	p=2	p=3	p=4	p=5	p=10
20	0.025	6.5×10^{-3}	2.0×10^{-3}	6.1×10^{-4}	1.9×10^{-4}	1.4×10^{-7}
30	0.025	6.7×10^{-3}	2.2×10^{-3}	7.6×10^{-4}	2.7×10^{-4}	1.1×10^{-6}
50	0.025	6.9×10^{-3}	2.4×10^{-3}	8.7×10^{-4}	3.3×10^{-4}	3.2×10^{-6}
∞	0.025	7.2×10^{-3}	2.6×10^{-3}	1.0×10^{-3}	4.4×10^{-4}	9.4×10^{-6}

Note that the actual size does not depend on the number of subjects for $p = 1$ whereas it does for $p > 1$. The last row ($n = \infty$), is calculated using that $T_{n-2} \sim N(0,1)$ in (26) and $C_1^2 \to \chi_p(\alpha)$ in (24) when $n \to \infty$. To use the confidence set approach to make a test that has an actual size α^* we have to choose $\alpha > \alpha^*$ in (24). To get size α^* we have from Theorem 3.3 that C_1 has to be equal to the upper α^* quantile of the t-distribution with $n-2$ degrees of freedom, $t_{n-2}(\alpha^*)$. Solving for α in equation (24) when $C_1 = t_{n-2}(\alpha^*)$ gives

$$\alpha = P\left(F_{p,n-1-p} > t^2_{n-2}(\alpha^*)(n-p-1)/(p(n-2))\right), \tag{28}$$

where $F_{p,n-1-p}$ stand for a F distributed stochastic variable with p respective $n-1-p$ degrees of freedom. In the limit as $n \to \infty$, we have that

$$\alpha = P\big(\chi^2_p > z^2(\alpha^*)\big),$$

where χ^2_p denotes a chi-squared random variable with p degrees of freedom and $z(\alpha^*)$ is the upper α^* quantile in the standard normal distribution.

Table 2: *The confidence level* $(1-\alpha)$ *on the confidence set so that the actual size equals 0.05.* α *is calculated from (28) when* α^* *is set to 0.05, the number of characteristics* p *as well as the number of subjects* n *varies.*

n	p=1	p=2	p=3	p=4	p=5	p=10
20	0.90	0.73	0.53	0.35	0.21	3.3×10^{-3}
30	0.90	0.73	0.54	0.37	0.22	5.9×10^{-3}
50	0.90	0.74	0.55	0.38	0.24	8.3×10^{-3}
∞	0.90	0.74	0.56	0.39	0.25	1.2×10^{-2}

4 Test for equality in variability

This section starts with a description on how to test for equivalence in variability by using the distributional results of Guilbaud (1993). However, the TOST or the equivalent confidence interval approach generate a very conservative test. In fact, there are no possibility at all to show equivalence under most common practical situations. We show how to modify the test such that a desired size of the test is obtained. Further on, we use this modified test together with an intersection-union test to test for equality in variability for general bioequivalence restrictions concerning the variability.

4.1 The one dimensional case

If we only measure one characteristic, Y_{ij1} and Y_{ij2} and all other elements in model (1) are one-dimensional. In particular, $\boldsymbol{\varepsilon}_A$ and $\boldsymbol{\varepsilon}_B$ in (3) are univariate normal distributed stochastic variables with zero means and variances that we denote σ_A^2 respective σ_B^2. We want to test if these variances are equal. Guilbaud (1993) defined

$$\gamma = (\sigma_A^2 - \sigma_B^2)/(\sigma_A^2 + \sigma_B^2)$$

as a measure of the difference between the two variances. Because γ is an one to one function of σ_A^2/σ_B^2, making statistical inference about γ is equivalent to make statistical inference about the relative variability, σ_A^2/σ_B^2. Guilbaud (1999) showed how to make various exact statistical inference about γ using that the test statistic defined by

$$T(c) = (\gamma^* - c)/s_*, \tag{29}$$

follows a t-distribution with $n-3$ degrees of freedom for $c = \gamma$. Here, γ^* is an exactly median unbiased estimator of γ defined by

$$\gamma^* = S_{-+}/S_{--},$$

where S_{-+} and S_{--} are the pooled within group corrected sums of squares and cross products corresponding to the within subject sums and crossover (treatment A - treatment B) differences, i.e.

$$S_{-+} = \sum_{j=1}^{2} \sum_{i=1}^{n_j} (Y_{ij}^{-} - \bar{Y^{-}})(Y_{ij}^{+} - \bar{Y^{+}})$$

S_{--} and S_{++} are defined in the similar way. The standard error of the estimator γ^* is denoted by s_* and is given by

$$s_*^2 = \Big(S_{++}/S_{--} - (\gamma^*)^2 \Big) / (n-3). \tag{30}$$

It may occur that γ^* does not belong to the interval $(-1, 1)$. It may also occur that the confidence region for γ is not included within $(-1, 1)$. This is pointed out and discussed how to handle in Guilbaud (1999). However, when testing for bioequivalence, this is no problem because the interesting part is when γ^* is close to zero and when the width of the confidence interval is small.

We can use TOST as well as the symmetric confidence interval approach together with (29) to test for equivalence in variability, i.e. test

$$H_0 : |\gamma| \geq \Delta_\gamma \text{ versus } H_A : |\gamma| < \Delta_\gamma,$$

where Δ_γ is the predefined cut-off number for γ. The confidence interval approach rejects H_0 in favor of H_A if the the entire $1 - 2\alpha$ confidence interval for γ lies within H_A. Using the exact distribution in (29), a symmetric $1 - 2\alpha$ confidence interval for γ is

$$C_\gamma = \Big(\gamma^* - t_\alpha(n-3)s_*, \gamma^* + t_\alpha(n-3)s_* \Big). \tag{31}$$

That is the same as the TOST, which rejects H_0 in favor of H_A if and only if

$$|\gamma^*| < \Delta_\gamma - t_\alpha(n-3)s_*. \tag{32}$$

This test is a level α test, i.e. the probability rejecting H_0 erroneously is at most α. The size is

$$\alpha^* \quad = \quad \sup_{\theta \in H_0} P_\theta \Big(|\gamma^*| < \Delta_\gamma - t_\alpha(n-3)s_* \Big). \tag{33}$$

Note that s_* depends on the estimate γ^*, thus the variance of the estimate can not be made arbitrary small for any parameter sequence in H_0 as it can for estimates of the mean. This implies that the size is less than α, i.e. the test is conservative.

The maximum in (33) is taken on the edge to H_0, i.e. for $\gamma = \Delta_\gamma$. Because the expectation of S_{++} equals $(n-2)(\sigma_A^2 + \sigma_B^2 + 4\text{Var}(\xi_{ij}))$ and $E[\xi_{ij}] = 0$, it is more likely that s_*^2 is small when there are no between subject variability present. Thus, the size is given when $\gamma = \Delta_\gamma$ and $\xi_{ij} = 0, i = 1, 2, \ldots, n_j, j = 1, 2$. However, it is hard (impossible?) to derive the α in (32) that gives size α^* analytically. Instead, we first derive α roughly, then use this rough α as starting point in a repeated simulation procedure to find a better estimate of the α that gives the desired size.

4.1.1 A rough calculation to get closer to the desired size

If we set

$$s_*^2 = \tilde{s}_*^2 = (1 - (\gamma^*)^2)/(n-3)$$

in (33), then it is possible to calculate α^*. Note that, if there are no between subject variation, then S_{--} and S_{++} have the same distribution. This motivates that we can use $\tilde{s}_*^2$ instead of s_*^2 in order to calculate a rough estimate of the actual size.

Using that the probability of rejecting H_0 is maximized when there are no between subject variability and when $\gamma = \Delta_\gamma$ we have that the rough size $\tilde{\alpha}^*$ equals

$$\tilde{\alpha}^* = P_{\Delta_\gamma}\left(|\gamma^*| < \Delta_\gamma - t_\alpha\sqrt{(1-(\gamma^*)^2)/(n-3)}\right), \tag{34}$$

where we have written t_α instead of $t_\alpha(n-3)$. Thus, we will find α such that this probability equals $\tilde{\alpha}^*$. It is clear that $\alpha \geq \tilde{\alpha}^*$ and that the choice depends on Δ_γ and n. Of course, the width of the confidence interval has to be smaller than $2\Delta_\gamma$ if there should be any possibility to reject H_0 at all. This yields the restriction

$$t_\alpha(n-3) < \Delta_\gamma\sqrt{n-3}.$$

Thus, we search for α that satisfies

$$P\left(T_{n-3} > \Delta_\gamma\sqrt{n-3}\right) < \alpha < 0.5, \tag{35}$$

where T_{n-3} is t-distributed with $n-3$ degrees of freedom.

Even though $\tilde{s}_*$ takes its maximum in $\gamma^* = 0$ and decreases with the distance between γ^* and zero, we have that

$$\max\{|\gamma^* - t_\alpha\tilde{s}_*|, |\gamma^* + t_\alpha\tilde{s}_*|\} = \begin{cases} |\gamma^* - t_\alpha\tilde{s}_*|, & \text{if } \gamma^* \leq 0 \\ |\gamma^* + t_\alpha\tilde{s}_*|, & \text{if } \gamma^* > 0 \end{cases} \tag{36}$$

takes its minimum for $\gamma^* = 0$. We also have that both $\gamma^* - t_\alpha\tilde{s}_*$ and $\gamma^* + t_\alpha\tilde{s}_*$ are monotone increasing functions of γ^* in the interval

$$\left(-(1 + t_\alpha^2/(n-3))^{-1}, (1 + t_\alpha^2/(n-3))^{-1}\right),$$

i.e. in the interesting part around zero. The monotonicity together with (36) gives that if we reject H_0 for $\gamma^* = c$ then we also reject H_0 for $\gamma^* \leq |c|$.

As mentioned earlier, the TOST or the equivalent confidence interval approach for testing (44) is a level α test. But it is conservative, in fact it is common that it is impossible to reject

H_0. For example, using $\alpha = 0.05$ in a study containing $n = 25$ subjects. Observing $\gamma^* = 0$ and using $\tilde{s}_*$ instead of s_* in (31) gives the rough confidence interval $(-0.37, 0.37)$. Thus, it is possible to reject H_0 only if $\Delta_\gamma > 0.37$. But $\Delta_\gamma > 0.37$ allows one of the within subject variances to be more than twice as big as the other one, i.e. $\Delta_\gamma = 0.37 \Rightarrow \sigma_A^2/\sigma_B^2 = 2.2$. This is a far to wide bioequivalence region for most experiments. A more reasonable limit is $\sigma_A^2/\sigma_B^2 = 1.2$, which corresponds to $\Delta_\gamma = 0.091$. We now show how to construct a test at size $\tilde{\alpha}^*$ for all cut-off limits by choosing α in (34) appropriately.

We have from (34) that the rough size $\tilde{\alpha}^*$ can be written as

$$P_{\Delta_\gamma}\left(\frac{\gamma^* - \Delta_\gamma}{\sqrt{(1-(\gamma^*)^2)/(n-3)}} < -t_\alpha\right) - P_{\Delta_\gamma}\left(\frac{\gamma^* + \Delta_\gamma}{\sqrt{(1-(\gamma^*)^2)/(n-3)}} < t_\alpha\right) = \tag{37}$$

$$= \alpha - P_{\Delta_\gamma}\left(\frac{\gamma^* + \Delta_\gamma}{\sqrt{(1-(\gamma^*)^2)/(n-3)}} < t_\alpha\right)$$

Rewriting the inequality in the left probability in (37) as $\gamma^* < f(\Delta_\gamma, t_\alpha, n)$ and using that this probability equals α, we can calculate the distribution function of γ^* (under the conditions that $E[\gamma^*] = \Delta_\gamma$ and no between subject variability). We have

$$P_{\Delta_\gamma}\left(\gamma^* < \frac{(n-3)\Delta_\gamma - \sqrt{(n-3)t_\alpha^2 + t_\alpha^4 - (n-3)\Delta_\gamma^2 t_\alpha^2}}{n-3+t_\alpha^2}\right) = \alpha. \tag{38}$$

The right probability in (37) is unknown, but can be calculated from the known probability (38) if we write the inequality in terms of $\gamma^* < g(\Delta_\gamma, t_\alpha, n)$, i.e.

$$P_{\Delta_\gamma}\left(\gamma^* < \frac{-(n-3)\Delta + \sqrt{(n-3)t_\alpha^2 + t_\alpha^4 - (n-3)\Delta_\gamma^2 t_\alpha^2}}{n-3+t_\alpha^2}\right). \tag{39}$$

We now use the known probability (38) to solve the unknown (39) by solving for $t_{\alpha'}$ in the equation

$$f(\Delta_\gamma, t_{\alpha'}, n) = g(\Delta_\gamma, t_\alpha, n). \tag{40}$$

There is no explicit solution to this equation, but it can be solved by numerical methods. However, it does not exist a solution for all t_α, but it does for $t_\alpha < \Delta_\gamma\sqrt{n-3}$. The condition, $t_{\alpha'} < t_\alpha$ guaranties an unique solution.

Thus, the rough size is

$$\tilde{\alpha}^* = \alpha - \alpha',$$

where α' is the corresponding probability to $t_{\alpha'}$.

The appropriate α such that the rough size equals $\tilde{\alpha}^*$ can be found using iterative methods. An easy way to do that is to start with an α close to the lower restriction in (35) and then try new candidates in such a way that we successive enclose the appropriate value in a smaller and smaller interval. We continue the iterations until we get a size sufficient close to $\tilde{\alpha}^*$. Results are shown in Table 3.

Table 3: *The α in (31) and (32) that corresponds to a test at rough size 0.05 ± 10^{-4}. The number of subjects n as well as the bioequivalence limit Δ_γ varies.*

n	Δ_γ=0.05	Δ_γ=0.10	Δ_γ=0.15
20	0.445	0.368	0.298
30	0.424	0.329	0.247
50	0.391	0.274	0.181
100	0.337	0.190	0.100
∞	0.05	0.05	0.05

4.1.2 An iterative simulation procedure that constructs a test at desired size

We now describe an iterative simulation procedure to find the α in (33) that corresponds to a test of size 0.05. Note first that the distribution of γ^* is scale invariant. Thus, the distribution is given by the relation between the variances, i.e. by γ.

1. Choose starting values on: $\tilde{\alpha}$; the number of simulations m; the step length d. Choose also an appropriate cut-off limit δ. We chose $\tilde{\alpha}$ from Table 3; $m = 10000$; $\delta = 0.0005$; and $d = 0.01$.

2. Simulate m studies from the univariate case ($p = 1$) of model (1) when $\gamma = \Delta_\gamma$ and no between subject variability, $\xi_{ij} = 0$, $i = 1, 2, \ldots, n_j$, $j = 1, 2$.

3. Calculate the number of studies (X) out of the m simulated, that rejects H_0 using the test (32). $X \sim Bin(m, p)$, where $p =$ the size corresponding to $\tilde{\alpha}$. Calculate the symmetric 99% approximative confidence interval (c.i.) for p, using normal approximation.

4. If the simulated c.i. is included within $(0.05 - \delta, 0.05 + \delta)$ then

   - Stop the iterations;

- $\hat{\alpha} = \tilde{\alpha}$;

Else if

- 0.05 is included within the c.i. then
  - increase m ($m = 3m$) and decrease the step length d ($d = d/2$).
- 0.05 lies below the c.i. then
  - $\tilde{\alpha} = \tilde{\alpha} - d$.
- 0.05 lies over the c.i. then
  - $\tilde{\alpha} = \tilde{\alpha} + d$.

Goto step 2.

The procedure described above generates $\hat{\alpha}$ as an estimate of the α in (33) that corresponds to size 0.05 and is so accurate that the interval $(0.04995, 0.0505)$ contains the true size with probability ≈ 0.99. The results are shown in Table 4 below.

Table 4: *$\hat{\alpha}$'s from the iterative simulation procedure corresponding to tests at size 0.05. The number of subjects n as well as the bioequivalence limit Δ_γ varies.*

n	Δ_γ=0.05	Δ_γ=0.10	Δ_γ=0.15
20	0.441	0.357	0.278
30	0.421	0.320	0.232
50	0.390	0.270	0.172
100	0.336	0.186	0.098
∞	0.05	0.05	0.05

The last row in Table 4 equals 0.05 because $s_* \to 0$ as $n \to \infty$.

4.2 Multivariate extension

We can transform the multivariate model (1) into an one-dimensional one by taking linear transformations of $\mathbf{Y}_{ij1}$ and $\mathbf{Y}_{ij2}$. If $\mathbf{a}$ is a $p \times 1$ vector, then $\mathbf{a}'\mathbf{Y}_{ij1}$ and $\mathbf{a}'\mathbf{Y}_{ij2}$ are univariate random variables that can be written as

$$(\mathbf{a}'\mathbf{Y}_{ij1}, \mathbf{a}'\mathbf{Y}_{ij2}) = (\mathbf{a}'\boldsymbol{\mu}_{j1}, \mathbf{a}'\boldsymbol{\mu}_{j2}) + (\mathbf{a}'\boldsymbol{\xi}_{ij} + \mathbf{a}'\boldsymbol{\varepsilon}_{ij1}, \mathbf{a}'\boldsymbol{\xi}_{ij} + \mathbf{a}'\boldsymbol{\varepsilon}_{ij2}), \tag{41}$$

where $\mathbf{a}'\boldsymbol{\varepsilon}_{ij1}$ and $\mathbf{a}'\boldsymbol{\varepsilon}_{ij2}$ are normally distributed random variables with zero means and variances depending on the treatment. From (3) we have that $\mathbf{a}'\boldsymbol{\varepsilon}_A \sim N(0, \mathbf{a}'\boldsymbol{\Lambda}_A\mathbf{a})$ and

$\mathbf{a}'\boldsymbol{\varepsilon}_B \sim N(0, \mathbf{a}'\boldsymbol{\Lambda}_B\mathbf{a})$. Thus, we can make exact statistical inference about $\mathbf{a}^{'}\boldsymbol{\Lambda}_A\mathbf{a}/\mathbf{a}^{'}\boldsymbol{\Lambda}_B\mathbf{a}$ or equivalent about

$$\begin{aligned}\gamma(\mathbf{a}) &= (\mathbf{a}^{'}\boldsymbol{\Lambda}_A\mathbf{a} - \mathbf{a}^{'}\boldsymbol{\Lambda}_B\mathbf{a})/(\mathbf{a}^{'}\boldsymbol{\Lambda}_A\mathbf{a} + \mathbf{a}^{'}\boldsymbol{\Lambda}_B\mathbf{a}) = \\ &= (\mathbf{a}^{'}(\boldsymbol{\Lambda}_A - \boldsymbol{\Lambda}_B)\mathbf{a})/(\mathbf{a}^{'}(\boldsymbol{\Lambda}_A + \boldsymbol{\Lambda}_B)\mathbf{a}) \qquad (42)\end{aligned}$$

in the same way as described in section 4.1. Because $\boldsymbol{\Lambda}_A$ and $\boldsymbol{\Lambda}_B$ are covariance matrices, it is assumed that they are positive definite. That is, $\mathbf{a}'\boldsymbol{\Lambda}_A\mathbf{a} > 0$ and $\mathbf{a}'\boldsymbol{\Lambda}_B\mathbf{a} > 0$ hold for all vectors $\mathbf{a} \neq 0$. Thus, the denominator in (42) is > 0 and we have that $-1 \leq \gamma(\mathbf{a}) \leq 1$, for all column vectors $\mathbf{a} \neq 0$. Note also that $\gamma(\mathbf{a})$ is an one to one function of $\mathbf{a}^{'}\boldsymbol{\Lambda}_A\mathbf{a}/\mathbf{a}^{'}\boldsymbol{\Lambda}_B\mathbf{a}$, which may be the measure we are interested in. The choice of the vector $\mathbf{a}$ depends on the comparison we are interested in. If we only want to compare the variances of the first characteristic we choose $\mathbf{a}$ to be the first coordinate vector, i.e. $\mathbf{a} = (1, 0, \ldots, 0)$. Other choices than choosing a coordinate vector includes covariances as well as variances in the comparison. In bioequivalence applications, the choice of $\mathbf{a}$ depends on the bioequivalence restrictions. Thus, by choosing $\mathbf{a}$ we can test for bioequivalence for a bioequivalence criterion that includes both variances of the characteristics and covariances between the characteristics.

For example, if two characteristics are measured ($p = 2$) and $\mathbf{a}' = (a_1, a_2)$, we have that

$$\begin{aligned}\mathbf{a}'\boldsymbol{\Lambda}_A\mathbf{a} &= a_1^2\sigma_{A_{11}}^2 + a_2^2\sigma_{A_{22}}^2 + 2a_1a_2\sigma_{A_{12}} \\ \mathbf{a}'\boldsymbol{\Lambda}_B\mathbf{a} &= a_1^2\sigma_{B_{11}}^2 + a_2^2\sigma_{B_{22}}^2 + 2a_1a_2\sigma_{B_{12}},\end{aligned}$$

where $\sigma_{A_{11}}^2$, $\sigma_{A_{22}}^2$ and $\sigma_{A_{12}}$ are the elements in the 2×2 covariance matrix $\boldsymbol{\Lambda}_A$; and $\sigma_{B_{11}}^2$, $\sigma_{B_{22}}^2$ and $\sigma_{B_{12}}$ are the elements in $\boldsymbol{\Lambda}_B$. If it is more important that the variances in the first characteristic are equal than in the second characteristic we should choose a large value on a_1. We have to take account of that there are different variability in the measurements of the two characteristics when we choose the vector $\mathbf{a}$. These variances are of course unknown but we may a priori have an idea about the relative magnitude between them that can be of use.

We see $\gamma(\mathbf{a})$ as a measure of the difference between the covariance matrices $\boldsymbol{\Lambda}_A$ and $\boldsymbol{\Lambda}_B$. If $\boldsymbol{\Lambda}_A = \boldsymbol{\Lambda}_B$ then $\gamma(\mathbf{a}) = 0$. The estimate of $\gamma(\mathbf{a})$ is derived from the estimate in one dimension,

$$\gamma^*(\mathbf{a}) = \mathbf{a}'S_{+-}\mathbf{a}/\mathbf{a}'S_{--}\mathbf{a} \qquad (43)$$

and the corresponding standard error, $s_*(\mathbf{a})$ is given by,

$$s_*^2(\mathbf{a}) = \left(\mathbf{a}'S_{++}\mathbf{a}/\mathbf{a}'S_{--}\mathbf{a} - \gamma^*(\mathbf{a})^2\right)/(n-3).$$

We can test

$$H_0 : |\gamma(\mathbf{a})| \geq \Delta_\gamma \text{ versus } H_A : |\gamma(\mathbf{a})| < \Delta_\gamma, \qquad (44)$$

by the same technics described in section 4.1. The confidence interval approach rejects H_0 if and only if the $1-2\alpha'$ confidence interval for $\gamma(\mathbf{a})$, $C_{\gamma(\mathbf{a})}$ lies within the equivalence limits $(-\Delta_\gamma, \Delta_\gamma)$. Here,

$$C_{\gamma(\mathbf{a})} = \Big(\gamma^*(\mathbf{a}) - t_{\alpha'}(n-3)s_*(\mathbf{a}), \gamma^*(\mathbf{a}) + t_{\alpha'}(n-3)s_*(\mathbf{a})\Big).$$

This test is of course equivalent with the TOST with rejection region

$$R_{\gamma(\mathbf{a})} = \Big\{|\gamma^*(\mathbf{a})| < \Delta_\gamma - t_{\alpha'}(n-3)s_*(\mathbf{a})\Big\}. \tag{45}$$

If we choose α' by the same procedure as described in section 4.1 this test has size α.

4.3 Intersection-Union test

Adopting the intersection-union method, we can simultaneously test (44) for more than one linear combination. That is a test of

$$H_0 : \max_{1\le k\le q} |\gamma(\mathbf{a}_k)| \ge \Delta_\gamma \text{ versus } H_A : \max_{1\le k\le q} |\gamma(\mathbf{a}_k)| < \Delta_\gamma, \tag{46}$$

where q is the number of linear combinations and Δ_γ is the equivalence limit for the kth comparison. The hypothesis in (46) can be expressed as the union respective intersection of the following hypotheses

$$H_{0_k} : |\gamma(\mathbf{a}_k)| \ge \Delta_\gamma \text{ versus } H_{A_k} : |\gamma(\mathbf{a}_k)| < \Delta_\gamma,$$

i.e. $H_0 = \bigcup_{k=1}^q H_{0_k}$ respective $H_A = \bigcap_{k=1}^q H_{A_k}$. Thus, using the intersection-union method, the rejection region defined by

$$R_\gamma^I = \bigcap_{k=1}^q R_{\gamma(\mathbf{a}_k)} = \bigcap_{k=1}^q \Big\{|\gamma^*(\mathbf{a}_k)| < \Delta_\gamma - t_{\alpha'}(n-3)s_*(\mathbf{a}_k)\Big\} \tag{47}$$

defines a level α test for (46) because each marginal test is a test of size α. Thus, the probability of making an error of type I is at most α. However, the following theorem assures that the intersection-union test (47) has size α if $\mathbf{a}_k$ equals the kth coordinate vector, $k = 1, 2, \ldots, p$.

Theorem 4.1:
The intersection-union test defined by (47) for testing (46) has size α if the vectors $\mathbf{a}_k$, $k = 1, 2, \ldots, q$ are orthogonal.

Proof: It is sufficient to prove this when $\mathbf{a}_k$ equals the kth coordinate vector because transforming the data using the orthogonal vectors generates an equivalent problem.

The theorem follows from theorem 2.2 in Berger and Hsu (1996). Let the parameter sequence θ_l belong to H_{0_1} in such a way that there are no between subject variability; $\gamma(\mathbf{a}_1) = \Delta_\gamma$; and $\gamma(\mathbf{a}_k) = \delta_l$ where $\delta_l \to 0$ as $l \to \infty$ for $k \neq 1$. From the results in section 4.1 we have that

$$\lim_{l\to\infty} P_{\theta_l}(\mathbf{Y} \in R_1) = \alpha.$$

For $k \neq 1$ we have that $\gamma^*(\mathbf{a}_k) \to 0$ with probability 1 as $l \to \infty$. This implies that

$$\lim_{l\to\infty} P_{\theta_l}(\mathbf{Y} \in R_k) = 1,$$

which completes the proof.

□

Of course, the number of vectors q has to be less then or equal to the number of characteristics p.

5 Tests including restrictions on both means and variability

We now combine the results from sections 3 and 4. Thus, we test for bioequivalence when the bioequivalence region contains restrictions on both means and variances. First we use an intersection-union test to test for bioequivalence for a general bioequivalence region. Then we show how to control the size of this test in the special case when the same orthogonal restrictions are made for both means and variability.

5.1 General test

The following test includes restrictions on both means and variablity.

$$H_0 : \max_{1\leq k\leq q} |\mathbf{a}_k'\boldsymbol{\tau}| \geq \Delta \text{ and/or } \max_{1\leq k'\leq q'} |\gamma(\mathbf{b}_{k'})| \geq \Delta_\gamma$$

$$\text{versus} \tag{48}$$

$$H_A : \max_{1\leq k\leq q} |\mathbf{a}_k'\boldsymbol{\tau}| \leq \Delta \text{ and } \max_{1\leq k'\leq q'} |\gamma(\mathbf{b}_{k'})| \leq \Delta_\gamma,$$

where the $\mathbf{a}_k$'s are vectors on length 1 whereas the $\mathbf{b}_{k'}$'s are orthogonal. The null hypothesis can be expressed as the union of the null hypothesis in (9) and (46), whereas the alternative is the intersection of the alternatives in (9) and (46). The intersection-union test for testing (48) rejects H_0 if and only if both (9) and (46) are rejected. That is, using the intersection of (12) and (47), a test with rejection region

$$R^I = R^I_{\boldsymbol{\tau}} \cap R^I_{\gamma}.$$

Suppose that the corresponding sizes are α and α' for the means respective the variances, then this is a level α^* test, $\alpha^* = \max\{\alpha, \alpha'\}$ (Berger and Hsu (1996)). Thus, the probability rejecting H_0 erroneously is at most α^*.

5.2 Restrictions on both means and variances for each characteristic

Suppose $\mathbf{a}_k$ is the kth coordinate vector $k = 1, 2, \ldots, q$, $q \leq p$ and we want to test

$$H_{0_k} : \left\{|\mathbf{a}_k'\boldsymbol{\tau}| \geq \Delta\right\} \cup \left\{|\gamma(\mathbf{a}_k)| \geq \Delta_\gamma\right\} \text{ vs } H_{A_k} : \left\{|\mathbf{a}_k'\boldsymbol{\tau}| \leq \Delta\right\} \cap \left\{|\gamma(\mathbf{a}_k)| \leq \Delta_\gamma\right\}. \tag{49}$$

The intersection-union test of this has rejection region that is the intersection of the marginal rejection regions R_k in (11) and $R_{\gamma(\mathbf{a}_k)}$ in (45), i.e.

$$R^I_k = R_k \cap R_{\gamma(\mathbf{a}_k)}.$$

Now, because $(\mathbf{a}_k'\mathbf{D}, \mathbf{a}_k\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k')$ and the statistic corresponding to (29), $T(\Delta_\gamma) = (\gamma^*(\mathbf{a}_k) - \Delta_\gamma)/s_*(\mathbf{a}_k)$ are independent (Guilbaud (1993)) and that R_k is a function of $(\mathbf{a}_k'\mathbf{D}, \mathbf{a}_k\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{a}_k')$ whereas $R_{\gamma(\mathbf{a}_k)}$ is a function of $T(\Delta_\gamma)$, we have that the probability of rejecting H_{0_k} is

$$P_\theta(\mathbf{Y} \in R^I_k) = P_\theta(\mathbf{Y} \in R_k)P_\theta(\mathbf{Y} \in R_{\gamma(\mathbf{a})}). \tag{50}$$

We have from the proofs of theorem 3.1 and theorem 4.1 that each of the two probabilities in the right hand side of (50) are maximized in the parameter sequence belonging to H_{0_k} defined by, no between subject variation; $\mathbf{a}_k\boldsymbol{\tau} = \Delta$ and $\boldsymbol{\Sigma}_{\mathbf{D}} \to \mathbf{0}$ in such a way that $\gamma(\mathbf{a}_k) = \Delta_\gamma$. It is now evident that the probability of rejecting H_{0_k} erroneously using the intersection-union test is maximized in the limit of this parameter sequence. Thus, the size equals the product of the sizes of the marginal tests. To construct a test of a desired size we choose the marginal sizes so that they multiplies to the desired size.

We now use the intersection-union test once again to construct a test including all relevant characteristics, i.e. we test

$$H_0 : \max_{1\leq k\leq q} |\mathbf{a}_k'\boldsymbol{\tau}| \geq \Delta \text{ and/or } \max_{1\leq k\leq q} |\gamma(\mathbf{a}_k)| \geq \Delta_\gamma$$

versus (51)

$$H_A : \max_{1\le k\le q} |\mathbf{a}_k'\boldsymbol{\tau}| \le \Delta \text{ and } \max_{1\le k\le q} |\gamma(\mathbf{a}_k)| \le \Delta_\gamma.$$

Thus if R_k^I is a rejection region determining a size α test for (49), then the intersection-union test of (51) has rejection region

$$R^I = \bigcap_{k=1}^{q} R_k^I.$$

This test has of course level α but the following theorem assures that it also has size α.

Theorem 5.1:
The intersection-union test R^I for testing (51) has size α.

Proof: The theorem follows easily by the same technique as used in the proofs of theorem 3.1 and theorem 4.1. That is in this case using theorem 2.2 in Berger and Hsu (1996) together with (50) for the parameter sequence in H_{0_k} determined by, no between subject variation; $\mathbf{a}_k\boldsymbol{\tau} = \Delta$; $\mathbf{a}_{k'}\boldsymbol{\tau} = 0$ for $k' \neq k$; and $\boldsymbol{\Sigma}_\mathbf{D} \to \mathbf{0}$ in such a way that $\gamma(\mathbf{a}_k) = \Delta_\gamma$ and $\gamma(\mathbf{a}_{k'}) \to 0$ for $k' \neq k$.

□

It should be noted that the theorem not only holds for the coordinate vectors but even for orthogonal vectors, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_q$, $q \le p$.

References

Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, **11**(4), 283–319.

Chow, S.-C. (1997). *Special issue on bioavailability/bioequivalence, Journal of Biopharmaceutical statistics, Volume 7, Number 1, 1997*. Marcel Dekker, inc.

FDA (2001). Guidance for industry. Statistical approaches to eatablishing bioequivalence. *U.S. department of health and humal services, food and drug administration, center for drug evaluation and research (CDER)*.

Guilbaud, O. (1993). Exact inference about the within-subject variability in 2×2 crossover trials. *Journal of the American Statistical Association*, **88**(423), 939–946.

Guilbaud, O. (1999). Exact comparisons of means and within-subject variances in 2×2 crossover trials. *Drug information Journal*, **33**(2), 455–469.

Jones, B. and Kenward, M. (1989). *Design and analysis of cross-over trials*. Chapman and Hall.

Krzanowski, W. J. (1988). *Principles of multivariate analysis*. Oxford university press.

Okpako, D. T. (1991). *Principles of pharmacology: a tropical approach*. Cambridge university press.

Schuirmann, D. J. (1981). On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. *Biometrics*, **37**, 617.

Wang, W., Hwang, G., and Dasgupta, A. (1999). Statistical tests for multivariate bioequivalence. *Biometrika*, **86**(2), 395–402.

Westlake, W. J. (1981). Response to T.B.L. Kirkwood: bioequivalence testing—a need to rethink. *Biometrics*, **37**, 395–402.