MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
Vol. 12, No. 4, pp. 503–534, 2002

TESTS FOR INTER-SUBJECT AND TOTAL VARIABILITIES UNDER CROSSOVER DESIGNS

Yonghee Lee,[1,2,*] Jun Shao,[1] Shein-Chung Chow,[2] and Hansheng Wang[2]

[1]University of Wisconsin—Madison, Madison, Wisconsin
[2]StatPlus, Inc., Heston Hall, Suite 206, 1790 Yardley-Langhorne Road, Yardley, PA 19067

ABSTRACT

In this paper, we consider statistical tests for inter-subject and total variabilities between treatments under crossover designs. Since estimators of variance components for inter-subject variability and total variability in crossover design are not independent, the usual F-test cannot be applied. Alternatively, we propose a test based on the concept of the extension of the modified large sample method to compare inter-subject variability and total variability between treatments under a $2 \times 2m$ replicated crossover design. An asymptotic power of the proposed test is derived. A sensitivity analysis is performed based on the asymptotic power to determine how the power changes with respect to various parameters such as inter-subject correlation and intra-class correlation. Also the two methods for sample size calculation for testing total variability under 2×4 crossover design are discussed. The method based on the Fisher–Cornish inversion shows better performance than the method based on the normal approximation. Several simulation studies were conducted to investigate the finite sample performance of the proposed test. Our simulation results show that the proposed test can control type I error satisfactorily.

Key Words: Modified large sample; Replicated crossover design; Power; Sensitivity analysis; Sample size calculation

*Corresponding author. E-mail: ylee@statplusinc.com

DOI: 10.1081/BIP-120016233
1054-3406 (Print); 1520-5711 (Online)

INTRODUCTION

In clinical research, besides comparing mean values of various study endpoints, it is also of interest to compare their variability. When the mean values of the primary endpoints from two treatments are close, the one with smaller variability may be considered as clinically superior than the other. Hence, testing noninferiority/superiority of variability becomes important. Also two formulations of drug with similar means may fail to establish bioequivalence due to huge difference in terms of variabilities.[1–3] In such a situation, it is of interest to test equivalence in terms of variability between treatments.

In clinical research, variabilities are usually classified into three categories. They are, namely, intra-subject variability, inter-subject variability, and total variability. Intra-subject variability refers to the variability observed from repeated measurements on the same subject under the same experimental condition. The sources for intra-subject variability is multifold. Two important sources is biological variability and measurement error. Theoretically, the intra-subject variability can be eliminated by repeating the experiment infinite times on one subject under the same condition. In practice, no experiment can be carried out infinite times, which can only be approximated by a large number of replicated experiments. However, even if we could repeat infinite number of experiments on every subject and take the average of replicates, which is called subject-specific mean value, we may still observe difference among the subject-specific mean values. This type of variability is completely due to the heterogeneity between subjects, which is referred to as the inter-subject variability. Assuming the intra- and inter-subject variabilities are independent with each other, the variability observed from a group of subjects is defined as total variability, which can be obtained as the sum of the intra-subject variability and the inter-subject variability from the usual analysis of variance (ANOVA) model.

In practice, it is of interest to test equality, noninferiority/superiority, and equivalence between treatments in terms of the three types of variabilities (i.e., intra-, inter-, and total variability). The problem of comparing intra-subject variabilities under a crossover design is well established based on an F-test.[4] However, how to test for inter-subject and total variabilities is still a challenging problem in clinical practice. Unlike the F-test for comparing intra-subject variabilities, there are two statistical problems in testing inter-subject and total variabilities under a crossover design. First, the unbiased estimators of the inter-subject and total variabilities are usually not chi-square distributed. Thus, the usual F-test is not applicable. Second, the estimators of variance components of the inter-subject and total variability under a crossover design are usually not independent. In practice, although the modified large sample (MLS) method is providing a confidence interval for a linear combination of variance components,[5–7] the dependency among estimators makes it difficult to apply the MLS method directly when comparing inter-subject and total variabilities under a crossover design. Wang[8] developed a test for total variability under

a 2×2 crossover design by combining the UMP invariant test for the case of independent sample with one-sided tests.

In this paper, a test procedure based on the extension of the MLS method for construction of an asymptotically correct confidence bounds or intervals for a linear combination of variance components under different treatments is proposed. The hypothesis of interest can be tested by comparing these confidence bounds or intervals with appropriate reference value. The proposed test is easy to implement as it is derived from the standard moment estimators of variance components. Without loss of generality, we will focus on a $2 \times 2m$ replicated crossover design, where m is number of replicates. The idea can be easily extended to other higher-order crossover designs. In addition, an explicit form of the asymptotic power function of the test is derived for the sensitivity analysis. It is shown that the asymptotic power depends upon the parameters of the ratio of variability, intra-class correlation, the correlation of inter-subject random effects between treatment and the sample size. A sensitivity analysis was performed to examine how the power of the test changes as the parameter changes.

In addition, two methods of sample size calculation for testing total variability under 2×4 crossover design are discussed. First, we will discuss the method based on the normal approximation which is the most popular in practice. Since the method based on the normal approximation becomes too conservative in some cases, the alternative method will be developed and it is based on the Fisher–Cornish inversion of the test statistics.[9] The method based on the Fisher–Cornish inversion shows better performance than the method based on the normal approximation in the simulation study.

In the next section, the statistical model and an extension of the MLS method will be introduced. In the third section, the statistical procedure for testing total variability based on a $2 \times 2m$ replicated crossover design will be developed. In the fourth section, we extend the proposed method for comparing inter-subject variabilities. The asymptotic power of the tests are derived for sensitivity analysis in the fifth section. In the sixth section, two methods for sample size calculation are discussed for the test of total variability under 2×4 crossover design. Extensive simulation studies were carried out to evaluate the finite sample performance of the proposed testing procedures in the seventh section. A brief discussion is given in the eighth section.

EXTENSION OF THE MODIFIED LARGE SAMPLE METHOD

The MLS method is a useful tool for construction of a confidence bound for a linear combination of variance components. To illustrate the basic idea of the MLS method, we consider a linear combination of two variance components. Suppose σ_x^2 and σ_y^2 are two variance components of interest. Suppose we are interested in

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

constructing a $(1-\alpha)100\%$ confidence bound for the parameter

$$\eta = a\sigma_x^2 + b\sigma_y^2, \tag{1}$$

where a and b are the fixed real numbers and α is any real number in $(0, 0.5)$. Let s_x^2 and s_y^2 be independent unbiased estimators of σ_x^2 and σ_y^2, respectively. Assume that $n_1 s_x^2/\sigma_x^2$ and $n_2 s_y^2/\sigma_y^2$ follow a χ^2 distribution with n_1 and n_2 degrees of freedom, respectively. If both a and b are positive, the MLS $(1-\alpha)100\%$ lower and upper confidence bounds for η can be obtained as follows

$$\hat{\eta}_{\mathrm{L}} = as_x^2 + bs_y^2 - \sqrt{a^2 s_x^4 \left(\frac{n_1}{\chi^2_{1-\alpha/2,n_1}} - 1\right)^2 + b^2 s_y^4 \left(\frac{n_2}{\chi^2_{1-\alpha/2,n_2}} - 1\right)^2},$$

$$\hat{\eta}_{\mathrm{U}} = as_x^2 + bs_y^2 + \sqrt{a^2 s_x^4 \left(\frac{n_1}{\chi^2_{\alpha/2,n_1}} - 1\right)^2 + b^2 s_y^4 \left(\frac{n_2}{\chi^2_{\alpha/2,n_2}} - 1\right)^2},$$

where $\chi^2_{\alpha,n}$ is the α-level quantile of χ^2 distribution with n degrees of freedom. When a and b in Eq. (1) have different signs, without loss of generality we assume $a < 0 < b$. In this case, the MLS $(1-\alpha)100\%$ lower and upper confidence bounds for η are given by

$$\hat{\eta}_{\mathrm{L}} = as_x^2 + bs_y^2 - \sqrt{a^2 s_x^4 \left(\frac{n_1}{\chi^2_{\alpha/2,n_1}} - 1\right)^2 + b^2 s_y^4 \left(\frac{n_2}{\chi^2_{1-\alpha/2,n_2}} - 1\right)^2},$$

$$\hat{\eta}_{\mathrm{U}} = as_x^2 + bs_y^2 + \sqrt{a^2 s_x^4 \left(\frac{n_1}{\chi^2_{1-\alpha/2,n_1}} - 1\right)^2 + b^2 s_y^4 \left(\frac{n_2}{\chi^2_{\alpha/2,n_2}} - 1\right)^2}.$$

Note that the MLS $(1-\alpha)100\%$ confidence bounds are exact if either $\sigma_x^2 = 0$ or $\sigma_y^2 = 0$. Furthermore, it gives an asymptotic coverage probability of $1-\alpha$ when $n_1 \to \infty$ and $n_2 \to \infty$.

Under a crossover design, however, it can be easily shown that the estimators for the total variability or the inter-subject variability are not independent. Thus, neither the usual F-test nor the MLS method is appropriate for testing the equality or equivalence in variability between treatments. However, since the MLS method can be extended when estimators of variance components are not independent, we may adopt to use the extension of the MLS method to assess the equality or equivalence in total or inter-subject variability between treatments.

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

Let $(x_i, y_i)^t$, $i = 1, \ldots, n$ be the random vectors from a bivariate normal distribution with mean $(0, 0)^t$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}, \tag{2}$$

where $\mathbf{x}^t$ is a transpose of vector $\mathbf{x}$. The covariance matrix Σ can be estimated by the sample covariance matrix $\hat{\Sigma}$ such that

$$\hat{\Sigma} = \begin{pmatrix} \frac{1}{n}\sum_i x_i^2 & \frac{1}{n}\sum_i x_i y_i \\ \frac{1}{n}\sum_i x_i y_i & \frac{1}{n}\sum_i y_i^2 \end{pmatrix} \equiv \begin{pmatrix} s_x^2 & s_{xy}^2 \\ s_{xy}^2 & s_y^2 \end{pmatrix}.$$

Suppose we are interested in constructing a $(1 - \alpha)100\%$ upper confidence bound for a linear combination of variance components $\eta = a\sigma_x^2 + b\sigma_y^2$. Let

$$\Theta = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

Also, let λ_1 and λ_2 be the eigenvalues of $\Theta\Sigma$, where Σ is given in Eq. (2). Explicit forms of λ_1 and λ_2 can be obtained as follows

$$\lambda_i = \frac{a\sigma_x^2 + b\sigma_y^2 \pm \sqrt{(a\sigma_x^2 - b\sigma_y^2)^2 + 4ab\rho^2\sigma_x^2\sigma_y^2}}{2} \quad \text{for } i = 1, 2.$$

Let $\mathcal{L}(x)$ denotes the law of random variable x. The following lemma establishes the distribution of the estimator $\hat{\eta} = as_x^2 + bs_y^2$.

Lemma 1.

$$\mathcal{L}(as_x^2 + bs_y^2) = \mathcal{L}(\lambda_1 U_1/n + \lambda_2 U_2/n),$$

where U_1 and U_2 are independent chi-square random variables with n degrees of freedom.

The above lemma indicates that the linear combination of two correlated variance components is actually distributed as a weighted linear combination of two independent chi-square random variables. Assume that $a < 0 < b$ and define sample version of eigenvalues $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as

$$\hat{\lambda}_i = \frac{as_x^2 + bs_y^2 \pm \sqrt{(as_x^2 - bs_y^2)^2 + 4abs_{xy}^4}}{2} \quad \text{for } i = 1, 2.$$

Note that $\hat{\eta} = as_x^2 + bs_y^2 \equiv \hat{\lambda}_1 + \hat{\lambda}_2$. Without loss of generality, we assume that $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$. Then, a $(1 - \alpha)100\%$ lower and upper confidence bounds for

$\eta = a\sigma_x^2 + b\sigma_y^2$ based on the extension of the MLS method are given by

$$\hat{\eta}_{\mathrm{L}} = as_x^2 + bs_y^2 - \sqrt{\hat{\lambda}_1^2\left(\frac{n}{\chi^2_{\alpha/2,n}} - 1\right)^2 + \hat{\lambda}_2^2\left(\frac{n}{\chi^2_{1-\alpha/2,n}} - 1\right)^2},$$

$$\hat{\eta}_{\mathrm{U}} = as_x^2 + bs_y^2 + \sqrt{\hat{\lambda}_1^2\left(\frac{n}{\chi^2_{1-\alpha/2,n}} - 1\right)^2 + \hat{\lambda}_2^2\left(\frac{n}{\chi^2_{\alpha/2,n}} - 1\right)^2}.$$

The only difference of the extension of the MLS method is that it uses $\hat{\lambda}_1$ and $\hat{\lambda}_2$ (the eigenvalues of $\Theta\hat{\Sigma}$ which include the estimator s_{xy}^2 of the covariance) instead of s_x^2 and s_y^2 only. Based on the extension of the MLS method, tests for comparing total variabilities between treatments will be derived. More details regarding the extension of MLS method can be found in Ref. [10].

TEST FOR TOTAL VARIABILITY UNDER A $2 \times 2m$ REPLICATED CROSSOVER DESIGN

Without loss of generality, consider a $2 \times 2m$ replicated crossover design comparing two treatments. For convenience sake, we will refer to the two treatments as a test formulation and a reference formulation. Under a $2 \times 2m$ replicated crossover design, in each sequence, each subject receives the test formulations m times and the reference formulations m times at different dosing periods. When $m = 1$, the $2 \times 2m$ replicated crossover design reduces to the standard two-sequence, two-period (2×2) crossover design. On the other hand, when $m = 2$, the $2 \times 2m$ replicated crossover design becomes 2×4 crossover design, which is recommended by FDA for assessment of population/individual bioequivalence.[3] Note that when $m = 1$, the standard 2×2 crossover design allows a test for total variability. In this section, we will focus on statistical inference for $m > 1$. All of the formulas and estimators can be applied to the case when $m = 1$ with a minor modification.

Statistical Model

Let y_{ijkl} be the response observed from the jth ($j = 1, \ldots, n_i$) subject in the ith ($i = 1, 2$) sequence under the lth ($l = 1, \ldots, m$) replicate of the kth ($k = T, R$) treatment. The following mixed effects model is usually considered

$$y_{ijkl} = \mu_k + \gamma_{ikl} + S_{ijk} + \epsilon_{ijkl}, \tag{3}$$

where μ_k is the treatment effect for formulation k, γ_{ikl} is the fixed effect of the lth replicate on the kth treatment in the ith sequence. Note that γ_{ikl} are the fixed

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

period, sequence, and interaction effects, which satisfy the constraint

$$\sum_{i=1}^{2}\sum_{l=1}^{m}\gamma_{ikl}=0 \quad \text{for each } k=T,R.$$

This parameterization for fixed effects are proposed by Chinchilli and Esinhart[4] and also used in FDA[3] for testing bioequivalence. $(S_{ijT}, S_{ijR})^t$ is the random effect of the jth subject in the ith sequence, which are independent and identically distributed as a bivariate normal random vector with mean $(0,0)^t$ and covariance matrix

$$\Sigma_{\mathrm{B}}=\begin{pmatrix}\sigma^2_{\mathrm{BT}} & \rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}}\\ \rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}} & \sigma^2_{\mathrm{BR}}\end{pmatrix}.$$

Note that σ^2_{BT} and σ^2_{BR} are inter-subject variances for the test formulation and the reference formulation, respectively. ϵ_{ijkl}'s are independent random variables from the normal distribution with mean 0 and variance σ^2_{WT} or σ^2_{WR} (intra-subject variances) depending on the formulation applied to the subject. For the sake of convenience, we define the intra-subject covariance matrix Σ_{W} as

$$\Sigma_{\mathrm{W}}=\begin{pmatrix}\sigma^2_{\mathrm{WT}} & 0\\ 0 & \sigma^2_{\mathrm{WR}}\end{pmatrix}.$$

$(S_{ijT}, S_{ijR})^t$ and ϵ_{ijkl} are assumed independent. The total variability σ^2_{TT} and σ^2_{TR} for the test formulation and the reference formulation are then given by

$$\sigma^2_{\mathrm{TT}}=\sigma^2_{\mathrm{BT}}+\sigma^2_{\mathrm{WT}} \quad \text{and} \quad \sigma^2_{\mathrm{TR}}=\sigma^2_{\mathrm{BR}}+\sigma^2_{\mathrm{WR}}.$$

More details regarding the designs and inferences under crossover designs can be found in Ref. [4].

To obtain estimates of the variance components, we use an orthogonal transformation, which is also considered by Chinchilli and Esinhart.[4] Consider the random vector $\mathbf{y}^t_{ij}=(\mathbf{y}^t_{ijT}, \mathbf{y}^t_{ijR})$ where $\mathbf{y}^t_{ijT}$ and $\mathbf{y}^t_{ijR}$ are given by

$$\mathbf{y}^t_{ijT}=(\mathbf{y}_{ijT1}, y_{ijT2}, \ldots, \mathbf{y}_{ijTm}) \quad \text{and} \quad \mathbf{y}^t_{ijR}=(y_{ijR1}, y_{ijR2}, \ldots, y_{ijRm}).$$

Define $\mathbf{y}^*_{ijk}=\mathbf{P}^t\mathbf{y}_{ijk}$, $\mathbf{P}$ is an $m\times m$ orthogonal transformation defined as

$$\mathbf{P}^t=\begin{pmatrix}\frac{\mathbf{1}}{m}1^t_m\\ \mathbf{p}^t_2\\ \vdots\\ \mathbf{p}^t_m\end{pmatrix}$$

and $\mathbf{P}^t\mathbf{P}=(1/m)\mathbf{I}_m$. Then, the first element of the random vector $\mathbf{y}^*_{ijk}$ is y_{ijk}, which

is the average of the responses from the jth subject in the ith sequence with the kth treatment. Now we define two random vectors $\mathbf{x}_i^t = (\mathbf{x}_{iT}^t, \mathbf{x}_{iR}^t)$ and $\mathbf{z}_i^t = (\mathbf{z}_{iT}^t, \mathbf{z}_{iR}^t)$ for each sequence as follows

$$\mathbf{x}_{iT}^t = (y_{i1T1}^*, y_{i2T1}^*, \ldots, y_{in_iT1}^*) = (\bar{y}_{i1T}, \ldots, \bar{y}_{in_iT}),$$

$$\mathbf{x}_{iR}^t = (y_{i1R1}^*, y_{i2R1}^*, \ldots, y_{in_iR1}^*) = (\bar{y}_{i1R}, \ldots, \bar{y}_{in_iR}),$$

$$\mathbf{z}_{iT}^t = (y_{i1T2}^*, \ldots, y_{in_iT2}^*, y_{i1T3}^*, \ldots, y_{in_iT3}^*, \ldots, y_{i1Tm}^*, \ldots, y_{in_iTm}^*),$$

$$\mathbf{z}_{iR}^t = (y_{i1R2}^*, \ldots, y_{in_iR2}^*, y_{i1R3}^*, \ldots, y_{in_iR3}^*, \ldots, y_{i1Rm}^*, \ldots, y_{in_iRm}^*).$$

Thus, sample covariance matrices ($\hat{\Sigma}_{\mathrm{M}}$ and $\hat{\Sigma}_{\mathrm{W}}$) of the $\mathbf{x}_i$'s and $\mathbf{z}_i$'s are given by

$$\hat{\Sigma}_{\mathrm{M}} = \begin{pmatrix} s_{\mathrm{MT}}^2 & s_{\mathrm{MTR}}^2 \\ s_{\mathrm{MTR}}^2 & s_{\mathrm{MR}}^2 \end{pmatrix}, \quad \hat{\Sigma}_{\mathrm{W}} = \begin{pmatrix} s_{\mathrm{WT}}^2 & 0 \\ 0 & s_{\mathrm{WR}}^2 \end{pmatrix},$$

where

$$s_{\mathrm{MT}}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\bar{y}_{ijT\cdot} - \bar{y}_{i\cdot T\cdot})^2,$$

$$s_{\mathrm{MR}}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\bar{y}_{ijR\cdot} - \bar{y}_{i\cdot R\cdot})^2,$$

$$s_{\mathrm{MTR}}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\bar{y}_{ijT\cdot} - \bar{y}_{i\cdot T\cdot})(\bar{y}_{ijR\cdot} - \bar{y}_{i\cdot R\cdot}),$$

$$s_{\mathrm{WT}}^2 = \frac{m}{(n_1 + n_2 - 2)(m-1)} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \sum_{l=2}^{m} (y_{ijTl}^* - \bar{y}_{i\cdot Tl}^*)^2,$$

$$s_{\mathrm{WR}}^2 = \frac{m}{(n_1 + n_2 - 2)(m-1)} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \sum_{l=2}^{m} (y_{ijRl}^* - \bar{y}_{i\cdot Rl}^*)^2.$$

Note that $\hat{\Sigma}_{\mathrm{M}}$ is an unbiased estimator of Σ_{M} such that

$$\Sigma_{\mathrm{M}} = \Sigma_{\mathrm{B}} + \Sigma_{\mathrm{W}/m} = \begin{pmatrix} \sigma_{\mathrm{BT}}^2 + \sigma_{\mathrm{WT}}^2/m & \rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}} \\ \rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}} & \sigma_{\mathrm{BR}}^2 + \sigma_{\mathrm{WR}}^2/m \end{pmatrix},$$

which is the covariance matrix of $(\bar{y}_{ijT\cdot}, \bar{y}_{ijR\cdot})$. Also, $\hat{\Sigma}_{\mathrm{W}}$ is an unbiased estimator of Σ_{W}.

Suppose that the parameter of the following interest is a linear combination of total variances for the test and reference formulation

$$\eta = \theta_1 \sigma_{\mathrm{TT}}^2 + \theta_2 \sigma_{\mathrm{TR}}^2, \tag{4}$$

where θ_1 and θ_2 are some real numbers. An unbiased estimator of η can be obtained as

$$\hat{\eta} = \theta_1 s_{\mathrm{MT}}^2 + \theta_2 s_{\mathrm{MR}}^2 + \frac{m-1}{m}(\theta_1 s_{\mathrm{WT}}^2 + \theta_2 s_{\mathrm{WR}}^2). \tag{5}$$

Define matrix $\mathbf{A}_i$ and $\mathbf{B}_i$ such that $\mathbf{A}_i = \mathbf{\Theta} \otimes \mathbf{J}_{n_i}$ and $\mathbf{B}_i = \mathbf{\Theta} \otimes \mathbf{I}_{m-1} \otimes \mathbf{J}_{n_i}$ where $\mathbf{\Theta}$ is given by

$$\mathbf{\Theta} = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}.$$

Then, the two quadratic forms Q_1' and Q_2' can be obtained as

$$Q'_1 = \sum_{i=1}^{2} \mathbf{x}_i^t \mathbf{A}_i \mathbf{x}_i = (n_1 + n_2 - 2)(\theta_1 s_{\mathrm{MT}}^2 + \theta_2 s_{\mathrm{MR}}^2),$$

$$Q'_2 = \sum_{i=1}^{2} \mathbf{z}_i^t \mathbf{B}_i \mathbf{z}_i = (n_1 + n_2 - 2)(m-1)(\theta_1 s_{\mathrm{WT}}^2 + \theta_2 s_{\mathrm{WR}}^2)/m.$$

Note that Q_1' is a linear combination of the between-subject sum of squares and Q_2' is a linear combination of the within-subject sum of squares. The unbiased estimator $\hat{\eta}$ of the parameter in Eq. (4) can be written in terms of the two quadratic forms as follows

$$\hat{\eta} = \frac{Q_1'}{n_1 + n_2 - 2} + \frac{Q_2'}{(n_1 + n_2 - 2)}. \tag{6}$$

Under normality assumption and independence and by the orthogonal transformation, the following lemma regarding the distribution of estimator in Eq. (6) follows. The proof is straightforward and hence omitted.

Lemma 2. *Under the model* (3)*, the distribution of the estimator in Eq.* (6) *have an equivalent form such that*

$$\mathcal{L}(\hat{\eta}) = \mathcal{L}(Q_{11} + Q_{12} + Q_{21} + Q_{22}),$$

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

where the distribution of Q's are given as

$$\mathcal{L}(Q_{11}) = \lambda_{11}\frac{\chi^2(n_1+n_2-2)}{n_2+n_2-2}, \quad \mathcal{L}(Q_{12}) = \lambda_{12}\frac{\chi^2(n_1+n_2-2)}{n_2+n_2-2},$$

$$\mathcal{L}(Q_{21}) = \lambda_{21}\frac{\chi^2((n_1+n_2-2)(m-1))}{(n_2+n_2-2)(m-1)},$$

$$\mathcal{L}(Q_{22}) = \lambda_{22}\frac{\chi^2((n_1+n_2-2)(m-1))}{(n_2+n_2-2)(m-1)},$$

where λ_{11} *and* λ_{12} *are the eigenvalues of* $\Theta\Sigma_{\mathrm{M}}$ *and*

$$\lambda_{21} = (m-1)\theta_1\sigma^2_{\mathrm{WT}}/m, \quad \lambda_{22} = (m-1)\theta_2\sigma^2_{\mathrm{WR}}/m.$$

Furthermore, Q_{11}, Q_{12}, Q_{21}, *and* Q_{22} *are independent.*

Now we will consider the various tests for total variability by using the extension of the MLS method.

Test for Equality and Noninferiority/Superiority

For testing equality, the following hypotheses are usually considered

$$H_0 : \frac{\sigma^2_{\mathrm{TT}}}{\sigma^2_{\mathrm{TR}}} = 1 \quad \text{vs.} \quad H_{\mathrm{a}} : \frac{\sigma^2_{\mathrm{TT}}}{\sigma^2_{\mathrm{TR}}} \neq 1,$$

which is equivalent to test

$$H_0 : \sigma^2_{\mathrm{TT}} - \sigma^2_{\mathrm{TR}} = 0 \quad \text{vs.} \quad H_{\mathrm{a}} : \sigma^2_{\mathrm{TT}} - \sigma^2_{\mathrm{TR}} \neq 0.$$

The parameter of interest $\eta = \sigma^2_{\mathrm{TT}} - \sigma^2_{\mathrm{TR}}$ in Eq. (4) can be defined with $\theta_1 = 1$ and $\theta_2 = -1$. Then, Lemma 2 can be used to construct a $(1-\alpha)100\%$ confidence interval for η by using the MLS method. Since θ_1 and θ_2 have the opposite signs, without loss of generality, we assume that $\hat{\lambda}_{11} < 0 < \hat{\lambda}_{12}$, which are the eigenvalues of $\Theta\hat{\Sigma}_{\mathrm{M}}$. Also, assume that $\hat{\lambda}_{21} = (m-1)\theta_2 s^2_{\mathrm{WR}}/m < 0$ and $\hat{\lambda}_{22} = (m-1)\theta_1 s^2_{\mathrm{WT}}/m > 0$.

A $(1-\alpha)100\%$ confidence interval of η is given by $(\hat{\eta}_{\mathrm{L}}, \hat{\eta}_{\mathrm{U}})$, where

$$\hat{\eta}_{\mathrm{L}} = \hat{\eta} - \sqrt{\Delta_{\mathrm{L}}}, \quad \hat{\eta}_{\mathrm{U}} = \hat{\eta} + \sqrt{\Delta_{\mathrm{U}}}, \tag{7}$$

and

$$\Delta_{\mathrm{L}} = \hat{\lambda}^2_{11}\left(\frac{n_{\mathrm{s}}}{\chi^2(\alpha/2, n_{\mathrm{s}})} - 1\right)^2 + \hat{\lambda}^2_{21}\left(\frac{n_{\mathrm{sp}}}{\chi^2(\alpha/2, n_{\mathrm{sp}})} - 1\right)^2$$
$$+ \hat{\lambda}^2_{12}\left(\frac{n_{\mathrm{s}}}{\chi^2(1-\alpha/2, n_{\mathrm{s}})} - 1\right)^2 + \hat{\lambda}^2_{22}\left(\frac{n_{\mathrm{sp}}}{\chi^2(1-\alpha/2, n_{\mathrm{sp}})} - 1\right)^2, \tag{8}$$

$$\Delta_U = \hat{\lambda}_{11}^2\left(\frac{n_s}{\chi^2(1-\alpha/2,n_s)}-1\right)^2+\hat{\lambda}_{21}^2\left(\frac{n_{sp}}{\chi^2(1-\alpha/2,n_{sp})}-1\right)^2$$
$$+\hat{\lambda}_{12}^2\left(\frac{n_s}{\chi^2(\alpha/2,n_s)}-1\right)^2+\hat{\lambda}_{22}^2\left(\frac{n_{sp}}{\chi^2(\alpha/2,n_{sp})}-1\right)^2,$$

and $n_s = n_1 + n_2 - 2$, $n_{sp} = (n_1 + n_2 - 2)(m-1)$. We then reject the null hypothesis if $0 \notin (\hat{\eta}_L, \hat{\eta}_U)$.

For testing noninferiority/superiority in total variability, we can consider the one-sided test with the following hypotheses

$$H_0 : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \geq \delta \quad \text{vs.} \quad H_a : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} < \delta,$$

for some $\delta > 0$. This test can be easily performed by using the upper confidence bound of the parameter $\eta = \sigma_{TT}^2 - \delta\sigma_{TR}^2$.

Test for Equivalence

In order to establish equivalence in variability between two treatments, the following hypotheses are of particular interest

$$H_0 : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \notin (\delta_1, \delta_2) \quad \text{vs.} \quad H_a : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \in (\delta_1, \delta_2),$$

where $\delta_1 < 1 < \delta_2$ are given real numbers. The above hypotheses can be decomposed into the following two one-sided hypotheses

$$H_{01} : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \geq \delta_2 \quad \text{vs.} \quad H_{a1} : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} < \delta_2$$

and

$$H_{02} : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \leq \delta_1 \quad \text{vs.} \quad H_{a2} : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} > \delta_1.$$

These are equivalent to the following two one-sided hypotheses

$$H_{01} : \sigma_{TT}^2 - \delta_2\sigma_{TR}^2 \geq 0 \quad \text{vs.} \quad H_{a1} : \sigma_{TT}^2 - \delta_2\sigma_{TR}^2 < 0 \tag{9}$$

and

$$H_{02} : \sigma_{TT}^2 - \delta_1\sigma_{TR}^2 \leq 0 \quad \text{vs.} \quad H_{a2} : \sigma_{TT}^2 - \delta_1\sigma_{TR}^2 > 0. \tag{10}$$

Equivalence in total variability between two treatments can be established if both of the above two hypotheses are rejected at the α level of significance. In order to test the hypotheses of Eq. (9), we can use the $(1-\alpha)100\%$ confidence upper

bound for $\eta = \sigma_{TT}^2 - \delta_2\sigma_{TR}^2$. Similarly, the $(1-\alpha)100\%$ confidence lower bound for $\eta = \sigma_{TT}^2 - \delta_1\sigma_{TR}^2$ can be used to test the hypotheses in Eq. (10).

TESTS FOR INTER-SUBJECT VARIABILITY UNDER A $2 \times 2m$ REPLICATED CROSSOVER DESIGN

Suppose following linear combination of inter-subject variances is of interest

$$\eta = \theta_1\sigma_{BT}^2 + \theta_2\sigma_{BR}^2. \tag{11}$$

Under a $2 \times 2m$ replicated crossover design with $m \geq 2$, η can be estimated by its moment estimator follows as

$$\hat{\eta} = \theta_1 s_{MT}^2 + \theta_2 s_{MR}^2 - \frac{1}{m}(\theta_1 s_{WT}^2 + \theta_2 s_{WR}^2). \tag{12}$$

Note that the inter-subject variability cannot be assessed under the standard 2×2 crossover design, under which estimates of total variabilities are available. Since the estimator of the linear combination of inter-subject variances given in Eq. (12) is exactly the same as that of total variance in Eq. (5) except that the coefficient for estimators of intra-subject variance, the extension of the MLS method can be similarly applied to obtain a test for inter-subject variability with a minor modification.

For testing equality, similarly, the following hypotheses are usually considered

$$H_0 : \frac{\sigma_{BT}^2}{\sigma_{BR}^2} = 1 \quad \text{vs.} \quad H_a : \frac{\sigma_{BT}^2}{\sigma_{BR}^2} \neq 1,$$

which is equivalent to

$$H_0 : \sigma_{BT}^2 - \sigma_{BR}^2 = 0 \quad \text{vs.} \quad H_a : \sigma_{BT}^2 - \sigma_{BR}^2 \neq 0.$$

Let $\theta_1 = 1$ and $\theta_2 = -1$. Without loss of generality, assume $\hat{\lambda}_{11} < 0 < \hat{\lambda}_{12}$, which are the eigenvalues of $\Theta\hat{\Sigma}_M$. Also, assume that $\hat{\lambda}_{21} = -\theta_1 s_{WT}^2/m < 0$ and $\hat{\lambda}_{22} = -\theta_2 s_{WR}^2/m > 0$. Then, a $(1-\alpha)100\%$ confidence interval of η is the same as those given in Eqs. (7) and (8). We will reject the null hypothesis if $0 \notin (\hat{\eta}_L, \hat{\eta}_U)$. Tests for noninferiority/superiority can be similarly obtained.

Similarly, consider the following hypotheses for establishment of bioequivalence in variability

$$H_0 : \frac{\sigma_{BT}^2}{\sigma_{BR}^2} \notin (\delta_1, \delta_2) \quad \text{vs.} \quad H_a : \frac{\sigma_{BT}^2}{\sigma_{BR}^2} \in (\delta_1, \delta_2),$$

where $\delta_1 < 1 < \delta_2$ are given real numbers. The above hypotheses can be decomposed into the following two one-sided hypotheses

$$H_{01} : \sigma_{BT}^2 - \delta_2\sigma_{BR}^2 \geq 0 \quad \text{vs.} \quad H_{a1} : \sigma_{BT}^2 - \delta_2\sigma_{BR}^2 < 0 \tag{13}$$

Table 1. Sample Size Selected Using Eqs. (20) and (22) and the Corresponding Power of the Test Based on 10,000 Simulation; $\rho = 0.8$

ρ	r_T	r_R	δ	P^*	n^*	P^{**}	n^{**}
0.80	0.30	0.30	0.75	0.8194	70	0.8194	70
			0.50	0.8563	26	0.8280	24
			0.25	0.9156	14	0.8644	12
	0.30	0.50	0.75	0.8051	72	0.8052	72
			0.50	0.8389	26	0.8221	25
			0.25	0.8745	13	0.8160	11
	0.30	0.70	0.75	0.7931	77	0.7958	78
			0.50	0.8212	27	0.8040	26
			0.25	0.8483	13	0.8162	12
	0.50	0.30	0.75	0.8289	77	0.8270	75
			0.50	0.8687	29	0.8307	26
			0.25	0.9105	15	0.8443	12
	0.50	0.50	0.75	0.8080	74	0.8126	74
			0.50	0.8488	27	0.8119	25
			0.25	0.8869	14	0.8403	12
	0.50	0.70	0.75	0.8075	76	0.8066	76
			0.50	0.8275	27	0.8120	26
			0.25	0.8548	13	0.8204	12
	0.70	0.30	0.75	0.8335	86	0.8220	83
			0.50	0.8654	32	0.8452	29
			0.25	0.9272	17	0.8346	13
	0.70	0.50	0.75	0.8145	79	0.8143	78
			0.50	0.8589	29	0.8097	26
			0.25	0.9034	15	0.8145	12
	0.70	0.70	0.75	0.8169	77	0.8074	76
			0.50	0.8223	27	0.8154	26
			0.25	0.8815	14	0.8137	12

and

$$H_{02} : \sigma_{BT}^2 - \delta_1 \sigma_{BR}^2 \leq 0 \quad \text{vs.} \quad H_{a2} : \sigma_{BT}^2 - \delta_1 \sigma_{BR}^2 > 0. \tag{14}$$

Equivalence in inter-subject variability between two treatments can be established if both of the above two hypotheses are rejected at the α level of significance.

SENSITIVITY ANALYSIS BASED ON ASYMPTOTIC POWER

In this section we will consider the power of proposed test for inter-subject variability and total variability. Since the exact power of the proposed test is not trackable, an asymptotic argument will be used assuming that the numbers of subjects in each sequence are sufficiently large, i.e., $n_i \to \infty$ for all $i = 1, 2$, and

Table 2. Sample Size Selected Using Eqs. (20) and (22) and the Corresponding Power of the Test Based on 10,000 Simulation; $\rho = 1.0$

ρ	r_T	r_R	δ	P^*	n^*	P^{**}	n^{**}
1.00	0.30	0.30	0.75	0.8370	44	0.8280	43
			0.50	0.8797	18	0.8500	16
			0.25	0.9483	11	0.8586	8
	0.30	0.50	0.75	0.8151	53	0.8098	53
			0.50	0.8482	20	0.8120	18
			0.25	0.9075	11	0.8361	9
	0.30	0.70	0.75	0.8035	66	0.8029	67
			0.50	0.8108	23	0.7926	22
			0.25	0.8661	12	0.7959	10
	0.50	0.30	0.75	0.8365	58	0.8222	56
			0.50	0.8842	23	0.8406	20
			0.25	0.9372	13	0.8675	10
	0.50	0.50	0.75	0.8162	61	0.8160	60
			0.50	0.8567	23	0.8217	21
			0.25	0.8970	12	0.8385	10
	0.50	0.70	0.75	0.8117	68	0.8098	68
			0.50	0.8271	24	0.8020	23
			0.25	0.8631	12	0.8286	11
	0.70	0.30	0.75	0.8291	74	0.8258	71
			0.50	0.8677	28	0.8359	25
			0.25	0.9254	15	0.8583	12
	0.70	0.50	0.75	0.8290	71	0.8130	69
			0.50	0.8521	26	0.8209	24
			0.25	0.9070	14	0.8214	11
	0.70	0.70	0.75	0.8056	72	0.8076	71
			0.50	0.8346	26	0.8017	24
			0.25	0.8765	13	0.8093	11

the number of replicates m is fixed. For simplicity, we assume that the number of subjects in all sequences are same ($n = n_1 = n_2$).

We first consider the power of the test for total variability with the following one-sided hypotheses

$$H_0 : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \geq 1 \quad \text{vs.} \quad H_a : \frac{\sigma_{TT}^2}{\sigma_{TR}^2} < 1. \tag{15}$$

For evaluation of the power of the test under the alternative hypothesis, the following lemma is useful.

Lemma 3. *Let $\chi^2_{\alpha,n}$ denote the αth quantile of a standard chi-square random variable with degrees of freedom n. It follows that*

$$\sqrt{\frac{n}{2}}\left(\frac{n}{\chi^2_{\alpha,n}} - 1\right) \to -z_\alpha, \quad as \quad n \to \infty,$$

where $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of a standard normal random variable.

Table 3. Type I Error Rate for Testing Equality in Total Variability Under a 2×4 Replicated Crossover Design

	Parameters		$n = n_1 = n_2$					Parameters		$n = n_1 = n_2$			
ρ	r_T	r_R	5	10	20	30	ρ	r_T	r_R	5	10	20	30
0.0	0.9	0.9	0.0232	0.0346	0.0394	0.0456	0.6	0.9	0.9	0.0212	0.0350	0.0374	0.0414
	0.9	0.7	0.0282	0.0374	0.0370	0.0440		0.9	0.7	0.0270	0.0336	0.0402	0.0418
	0.9	0.5	0.0274	0.0384	0.0442	0.0426		0.9	0.5	0.0302	0.0376	0.0418	0.0408
	0.9	0.3	0.0332	0.0466	0.0478	0.0424		0.9	0.3	0.0336	0.0424	0.0536	0.0504
	0.9	0.1	0.0406	0.0446	0.0486	0.0490		0.9	0.1	0.0378	0.0468	0.0442	0.0476
	0.7	0.7	0.0260	0.0372	0.0394	0.0426		0.7	0.7	0.0294	0.0372	0.0510	0.0390
	0.7	0.5	0.0284	0.0436	0.0418	0.0526		0.7	0.5	0.0292	0.0408	0.0424	0.0488
	0.7	0.3	0.0342	0.0410	0.0444	0.0506		0.7	0.3	0.0278	0.0480	0.0462	0.0458
	0.7	0.1	0.0436	0.0462	0.0446	0.0518		0.7	0.1	0.0426	0.0452	0.0458	0.0498
	0.5	0.5	0.0352	0.0410	0.0450	0.0472		0.5	0.5	0.0352	0.0406	0.0440	0.0426
	0.5	0.3	0.0376	0.0532	0.0420	0.0496		0.5	0.3	0.0376	0.0430	0.0440	0.0474
	0.5	0.1	0.0488	0.0424	0.0468	0.0484		0.5	0.1	0.0456	0.0494	0.0478	0.0522
	0.3	0.3	0.0414	0.0500	0.0454	0.0500		0.3	0.3	0.0460	0.0472	0.0518	0.0506
	0.3	0.1	0.0442	0.0548	0.0494	0.0518		0.3	0.1	0.0476	0.0560	0.0486	0.0530
	0.1	0.1	0.0470	0.0488	0.0484	0.0506		0.1	0.1	0.0586	0.0468	0.0492	0.0512
0.2	0.9	0.9	0.0242	0.0330	0.0386	0.0486	0.8	0.9	0.9	0.0248	0.0312	0.0412	0.0432
	0.9	0.7	0.0258	0.0384	0.0362	0.0408		0.9	0.7	0.0232	0.0370	0.0356	0.0414
	0.9	0.5	0.0310	0.0336	0.0404	0.0468		0.9	0.5	0.0256	0.0408	0.0388	0.0442
	0.9	0.3	0.0338	0.0444	0.0462	0.0458		0.9	0.3	0.0344	0.0448	0.0460	0.0482
	0.9	0.1	0.0366	0.0470	0.0458	0.0432		0.9	0.1	0.0372	0.0434	0.0508	0.0400
	0.7	0.7	0.0284	0.0340	0.0386	0.0498		0.7	0.7	0.0270	0.0342	0.0408	0.0424
	0.7	0.5	0.0338	0.0430	0.0510	0.0440		0.7	0.5	0.0288	0.0438	0.0392	0.0466

(continued)

Table 3. Continued

	Parameters		$n = n_1 = n_2$					Parameters		$n = n_1 = n_2$			
ρ	r_T	r_R	5	10	20	30	ρ	r_T	r_R	5	10	20	30
	0.7	0.3	0.0342	0.0440	0.0458	0.0552		0.7	0.3	0.0330	0.0426	0.0482	0.0516
	0.7	0.1	0.0398	0.0468	0.0492	0.0502		0.7	0.1	0.0460	0.0460	0.0474	0.0474
	0.5	0.5	0.0348	0.0394	0.0422	0.0414		0.5	0.5	0.0326	0.0428	0.0468	0.0466
	0.5	0.3	0.0372	0.0504	0.0458	0.0476		0.5	0.3	0.0364	0.0430	0.0468	0.0476
	0.5	0.1	0.0490	0.0536	0.0474	0.0552		0.5	0.1	0.0458	0.0446	0.0482	0.0456
	0.3	0.3	0.0474	0.0498	0.0500	0.0446		0.3	0.3	0.0384	0.0470	0.0484	0.0438
	0.3	0.1	0.0504	0.0520	0.0520	0.0518		0.3	0.1	0.0462	0.0474	0.0452	0.0506
	0.1	0.1	0.0542	0.0484	0.0506	0.0512		0.1	0.1	0.0592	0.0578	0.0514	0.0486
0.4	0.9	0.9	0.0264	0.0322	0.0408	0.0434	1.0	0.9	0.9	0.0278	0.0350	0.0342	0.0392
	0.9	0.7	0.0234	0.0366	0.0388	0.0462		0.9	0.7	0.0250	0.0360	0.0434	0.0472
	0.9	0.5	0.0282	0.0376	0.0410	0.0404		0.9	0.5	0.0284	0.0388	0.0378	0.0436
	0.9	0.3	0.0340	0.0428	0.0474	0.0422		0.9	0.3	0.0322	0.0398	0.0452	0.0482
	0.9	0.1	0.0434	0.0432	0.0454	0.0486		0.9	0.1	0.0400	0.0456	0.0490	0.0478
	0.7	0.7	0.0274	0.0390	0.0436	0.0440		0.7	0.7	0.0244	0.0368	0.0436	0.0398
	0.7	0.5	0.0312	0.0356	0.0430	0.0422		0.7	0.5	0.0282	0.0430	0.0422	0.0398
	0.7	0.3	0.0430	0.0510	0.0510	0.0476		0.7	0.3	0.0366	0.0406	0.0426	0.0510
	0.7	0.1	0.0426	0.0474	0.0464	0.0466		0.7	0.1	0.0336	0.0478	0.0482	0.0484
	0.5	0.5	0.0338	0.0438	0.0476	0.0498		0.5	0.5	0.0324	0.0428	0.0466	0.0442
	0.5	0.3	0.0348	0.0490	0.0516	0.0476		0.5	0.3	0.0320	0.0420	0.0442	0.0486
	0.5	0.1	0.0438	0.0518	0.0486	0.0516		0.5	0.1	0.0424	0.0474	0.0506	0.0470
	0.3	0.3	0.0496	0.0482	0.0526	0.0482		0.3	0.3	0.0374	0.0448	0.0478	0.0470
	0.3	0.1	0.0522	0.0494	0.0506	0.0560		0.3	0.1	0.0434	0.0476	0.0474	0.0454
	0.1	0.1	0.0560	0.0528	0.0532	0.0524		0.1	0.1	0.0428	0.0514	0.0500	0.0486

Table 4. Type I Error Rate for Testing Equality in Inter-subject Variability Under a 2×4 Replicated Crossover Design

	Parameters		$n = n_1 = n_2$					Parameters		$n = n_1 = n_2$			
ρ	r_T	r_R	5	10	20	30	ρ	r_T	r_R	5	10	20	30
0.0	0.9	0.9	0.0274	0.0330	0.0390	0.0384	0.6	0.9	0.9	0.0254	0.0286	0.0380	0.0402
	0.9	0.7	0.0356	0.0432	0.0456	0.0486		0.9	0.7	0.0290	0.0432	0.0456	0.0456
	0.9	0.5	0.0382	0.0438	0.0440	0.0466		0.9	0.5	0.0338	0.0394	0.0428	0.0474
	0.9	0.3	0.0386	0.0464	0.0464	0.0478		0.9	0.3	0.0408	0.0392	0.0478	0.0438
	0.9	0.1	0.0420	0.0462	0.0462	0.0478		0.9	0.1	0.0448	0.0424	0.0452	0.0494
	0.7	0.7	0.0324	0.0382	0.0408	0.0440		0.7	0.7	0.0298	0.0390	0.0406	0.0416
	0.7	0.5	0.0312	0.0382	0.0362	0.0456		0.7	0.5	0.0340	0.0394	0.0428	0.0410
	0.7	0.3	0.0344	0.0394	0.0448	0.0502		0.7	0.3	0.0340	0.0478	0.0432	0.0468
	0.7	0.1	0.0412	0.0372	0.0434	0.0484		0.7	0.1	0.0404	0.0438	0.0462	0.0492
	0.5	0.5	0.0376	0.0408	0.0502	0.0480		0.5	0.5	0.0312	0.0382	0.0480	0.0456
	0.5	0.3	0.0376	0.0502	0.0436	0.0472		0.5	0.3	0.0384	0.0414	0.0456	0.0502
	0.5	0.1	0.0448	0.0468	0.0440	0.0512		0.5	0.1	0.0426	0.0498	0.0504	0.0500
	0.3	0.3	0.0414	0.0472	0.0488	0.0532		0.3	0.3	0.0456	0.0460	0.0516	0.0508
	0.3	0.1	0.0468	0.0504	0.0456	0.0476		0.3	0.1	0.0512	0.0570	0.0518	0.0520
	0.1	0.1	0.0478	0.0494	0.0480	0.0516		0.1	0.1	0.0580	0.0484	0.0484	0.0520
0.2	0.9	0.9	0.0226	0.0330	0.0388	0.0426	0.8	0.9	0.9	0.0230	0.0350	0.0382	0.0412
	0.9	0.7	0.0338	0.0476	0.0428	0.0446		0.9	0.7	0.0310	0.0436	0.0402	0.0476
	0.9	0.5	0.0442	0.0438	0.0440	0.0482		0.9	0.5	0.0344	0.0434	0.0442	0.0436
	0.9	0.3	0.0470	0.0452	0.0504	0.0452		0.9	0.3	0.0426	0.0432	0.0518	0.0470
	0.9	0.1	0.0438	0.0438	0.0486	0.0492		0.9	0.1	0.0470	0.0450	0.0450	0.0510
	0.7	0.7	0.0246	0.0380	0.0418	0.0412		0.7	0.7	0.0204	0.0348	0.0438	0.0460
	0.7	0.5	0.0334	0.0404	0.0422	0.0460		0.7	0.5	0.0354	0.0444	0.0460	0.0472
	0.7	0.3	0.0372	0.0374	0.0466	0.0450		0.7	0.3	0.0424	0.0416	0.0442	0.0476

(continued)

Table 4. Continued

	Parameters		$n = n_1 = n_2$					Parameters		$n = n_1 = n_2$			
ρ	r_T	r_R	5	10	20	30	ρ	r_T	r_R	5	10	20	30
	0.7	0.1	0.0380	0.0532	0.0426	0.0506		0.7	0.1	0.0396	0.0478	0.0474	0.0526
	0.5	0.5	0.0340	0.0420	0.0468	0.0412		0.5	0.5	0.0348	0.0428	0.0444	0.0488
	0.5	0.3	0.0398	0.0446	0.0492	0.0452		0.5	0.3	0.0388	0.0446	0.0464	0.0476
	0.5	0.1	0.0442	0.0536	0.0466	0.0552		0.5	0.1	0.0394	0.0444	0.0482	0.0458
	0.3	0.3	0.0436	0.0472	0.0488	0.0492		0.3	0.3	0.0364	0.0420	0.0474	0.0436
	0.3	0.1	0.0476	0.0506	0.0534	0.0522		0.3	0.1	0.0474	0.0482	0.0490	0.0470
	0.1	0.1	0.0556	0.0512	0.0510	0.0516		0.1	0.1	0.0582	0.0590	0.0508	0.0516
0.4	0.9	0.9	0.0264	0.0344	0.0356	0.0406	1.0	0.9	0.9	0.0202	0.0358	0.0386	0.0364
	0.9	0.7	0.0332	0.0406	0.0400	0.0492		0.9	0.7	0.0344	0.0396	0.0456	0.0474
	0.9	0.5	0.0374	0.0454	0.0518	0.0480		0.9	0.5	0.0350	0.0444	0.0442	0.0480
	0.9	0.3	0.0388	0.0444	0.0510	0.0468		0.9	0.3	0.0422	0.0484	0.0452	0.0544
	0.9	0.1	0.0434	0.0464	0.0458	0.0476		0.9	0.1	0.0420	0.0368	0.0490	0.0464
	0.7	0.7	0.0262	0.0358	0.0496	0.0436		0.7	0.7	0.0280	0.0370	0.0380	0.0416
	0.7	0.5	0.0316	0.0408	0.0472	0.0426		0.7	0.5	0.0276	0.0398	0.0414	0.0428
	0.7	0.3	0.0420	0.0476	0.0502	0.0480		0.7	0.3	0.0380	0.0376	0.0388	0.0424
	0.7	0.1	0.0404	0.0488	0.0490	0.0496		0.7	0.1	0.0422	0.0482	0.0492	0.0442
	0.5	0.5	0.0306	0.0428	0.0482	0.0456		0.5	0.5	0.0306	0.0372	0.0430	0.0448
	0.5	0.3	0.0406	0.0466	0.0498	0.0478		0.5	0.3	0.0312	0.0404	0.0496	0.0458
	0.5	0.1	0.0454	0.0506	0.0490	0.0466		0.5	0.1	0.0448	0.0456	0.0474	0.0490
	0.3	0.3	0.0410	0.0492	0.0540	0.0510		0.3	0.3	0.0332	0.0466	0.0424	0.0506
	0.3	0.1	0.0498	0.0498	0.0464	0.0524		0.3	0.1	0.0360	0.0446	0.0524	0.0486
	0.1	0.1	0.0568	0.0552	0.0530	0.0508		0.1	0.1	0.0472	0.0492	0.0504	0.0542

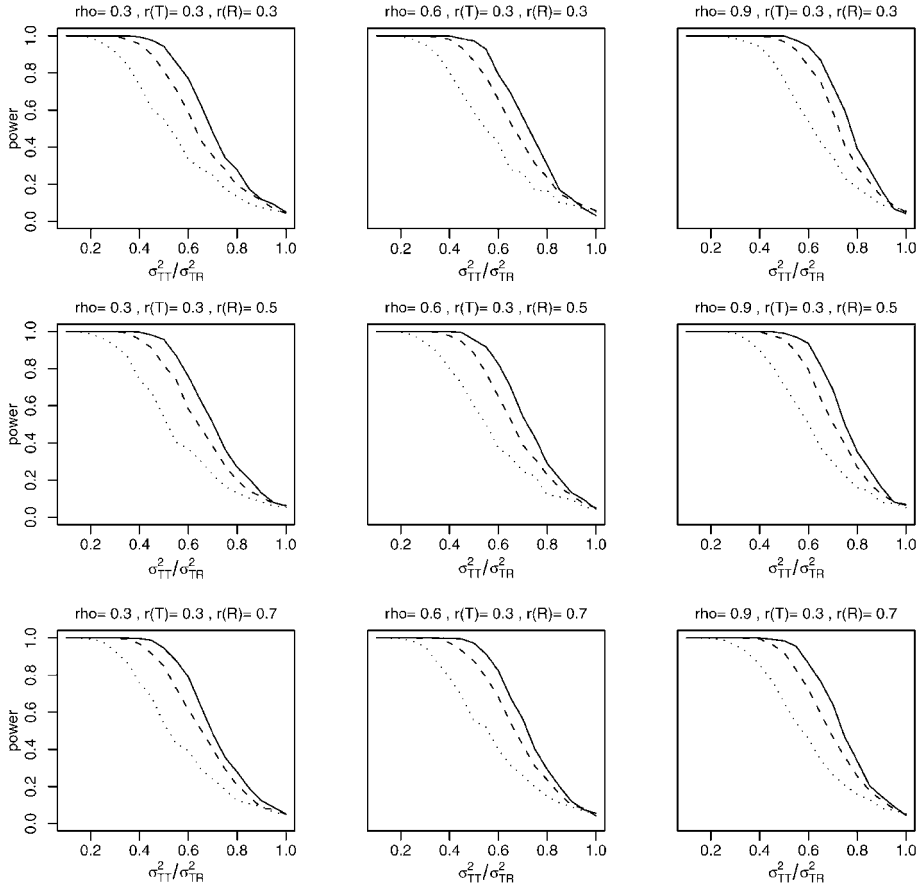

Figure 1. Power of the one-sided test for total variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T)$ = intra corr. of test; $r(R)$ = intra corr. of reference).

Proof is straightforward and therefore omitted.

The asymptotic variance of estimator $\hat{\eta}$ in Eq. (5) with $\theta_1 = 1$ and $\theta_2 = -1$ is given by

$$\tau^2 = 2\left(\sigma^4_{BT} + \frac{2}{m}\sigma^2_{BT}\sigma^2_{WT} + \frac{1}{m}\sigma^4_{WT} + \sigma^4_{BR} + \frac{2}{m}\sigma^2_{BR}\sigma^2_{WR} + \frac{1}{m}\sigma^4_{WR} - 2\rho^2\sigma^2_{BT}\sigma^2_{BR}\right). \tag{16}$$

We assume $\delta = \sigma^2_{TT}/\sigma^2_{TR} < 1$ and $\eta = \sigma^2_{TT} - \sigma^2_{TR} < 0$. If n is sufficiently large,

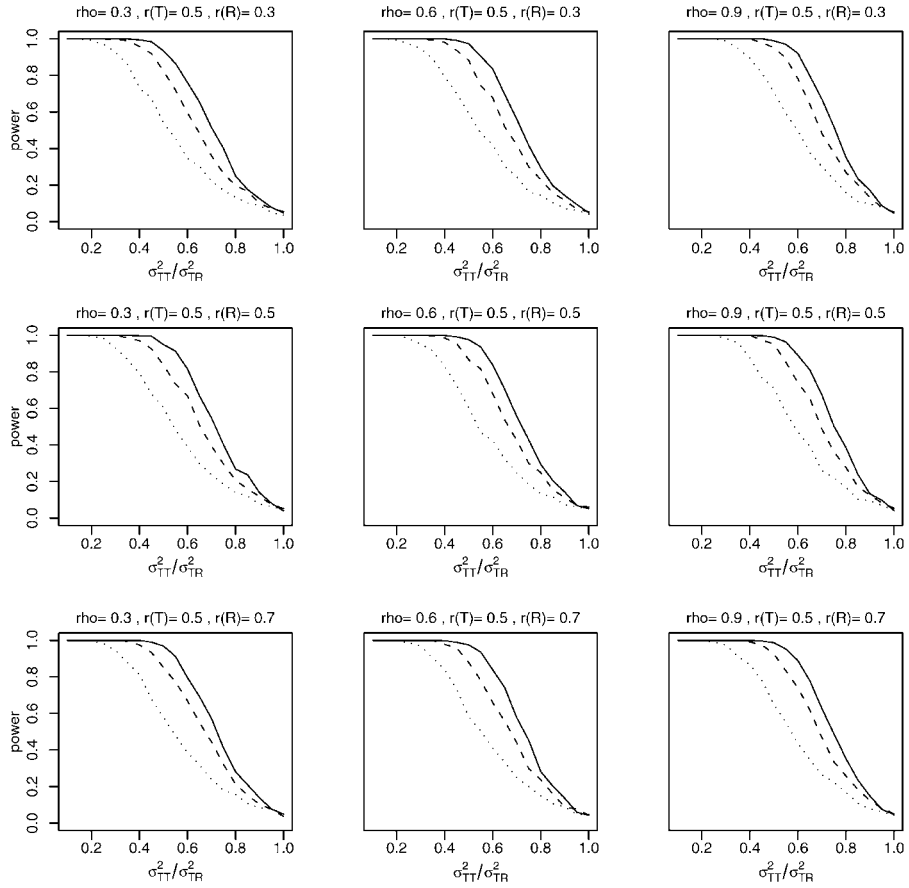

Figure 2. Power of the one-sided test for total variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T)$ = intra corr. of test; $r(R)$ = intra corr. of reference).

then the power of test for total variability can be approximated by

$$P_{\mathrm{T}} = P(\hat{\eta}_{\mathrm{U}} < 0) = P\left(\frac{\sqrt{2n-2}(\hat{\eta}-\eta)}{\tau} < -\frac{\sqrt{2n-2}(\sqrt{\Delta_{\mathrm{U}}}+\eta)}{\tau}\right)$$

$$\approx P\left(Z < z_\alpha - \sqrt{2n}\frac{\eta}{\tau}\right)$$

by Lemma 3 and the central limit theorem (CLT). Define the intra-class correlation for the test and reference formulation such that

$$r_{\mathrm{T}} = \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{BT}} + \sigma^2_{\mathrm{WT}}} \quad \text{and} \quad r_{\mathrm{R}} = \frac{\sigma^2_{\mathrm{WR}}}{\sigma^2_{\mathrm{BR}} + \sigma^2_{\mathrm{WR}}}. \tag{17}$$

Then, the power of test for total variability can be written as a function of four

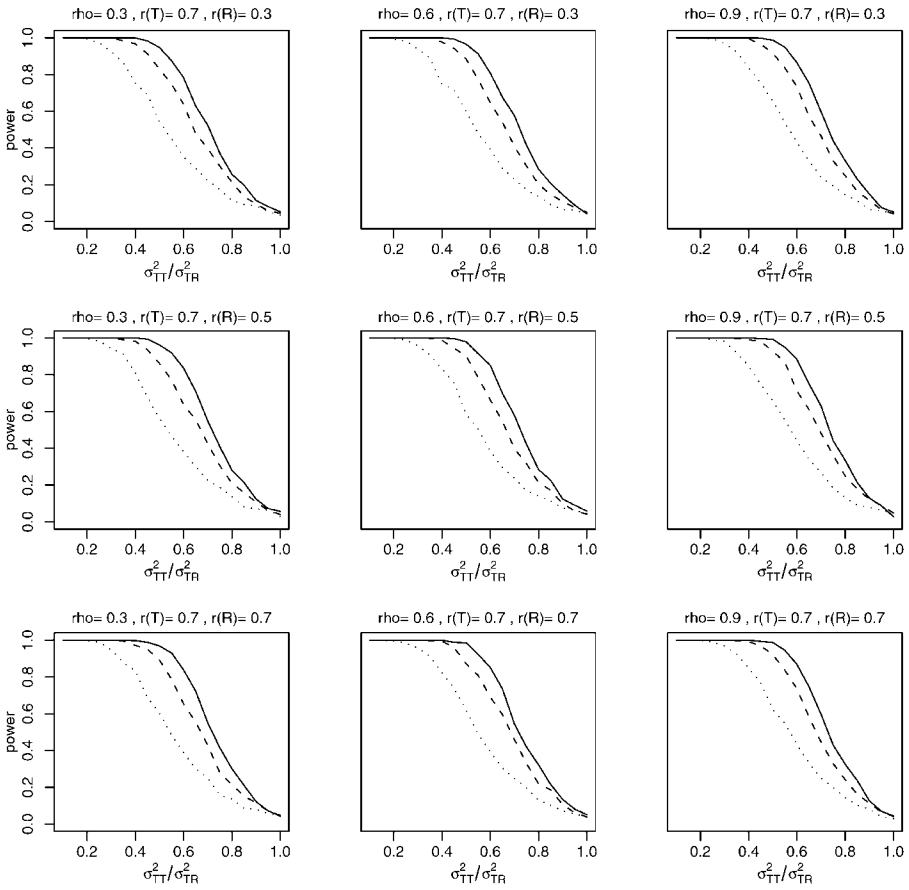

Figure 3. Power of the one-sided test for total variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T)$ = intra corr. of test; $r(R)$ = intra corr. of reference).

parameters (δ, r_T, r_R, ρ) and n as

$$P_T = P(Z < z_\alpha - \sqrt{n}K), \tag{18}$$

where $K = K(\delta, r_T, r_R, \rho)$ is given by

$$K = (\delta - 1)\left[\delta^2(1 - r_T)^2 + \frac{2}{m}\delta^2 r_T(1 - r_T) + \frac{1}{m}\delta^2 r_T^2 \right.$$

$$\left. +(1 - r_R)^2 + \frac{2}{m}r_R(1 - r_R) + \frac{1}{m}r_R^2 - 2\rho^2\delta(1 - r_T)(1 - r_R)\right]^{-1/2}$$

If we assume that $r = r_T = r_R$, then the constant K in Eq. (18) can be further

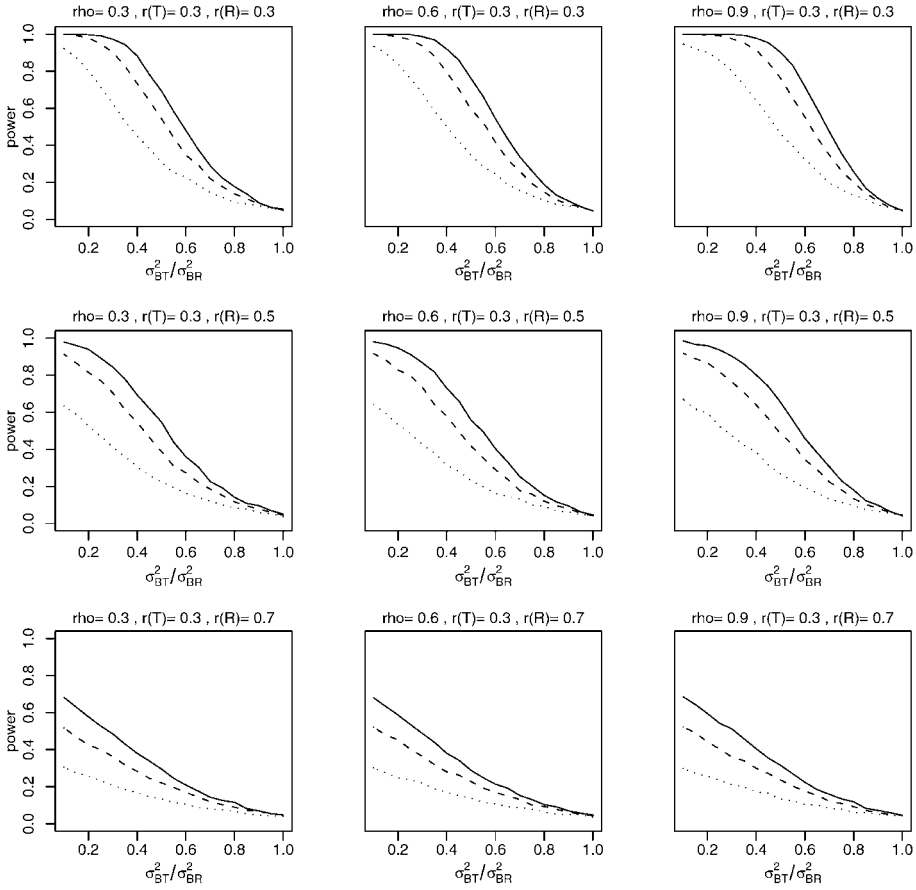

Figure 4. Power of the one-sided test for inter-subject variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T) =$ intra corr. of test; $r(R) =$ intra corr. of reference).

simplified as

$$K = (\delta - 1)\left[(1 - r)^2\left(\frac{m-1}{m}(\delta^2 + 1) - 2\rho^2\delta\right) + \frac{1}{m}(\delta^2 + 1)\right]^{-1/2}.$$

It can be easily seen that the power of the test is an increasing function as (i) n increases, (ii) δ decreases, and (iii) ρ increases. Since K is a bounded function of the intra-class correlation, i.e., $0 < K <$ const. for all $0 \le r \le 1$, the power of the test for total variability is not sensitive to the change in intra-class correlation. On the other hand, suppose we consider the test for inter-subject variability with the following hypothesis

$$H_0 : \frac{\sigma^2_{BT}}{\sigma^2_{BR}} \ge 1 \quad \text{vs.} \quad H_a : \frac{\sigma^2_{BT}}{\sigma^2_{BR}} < 1. \tag{19}$$

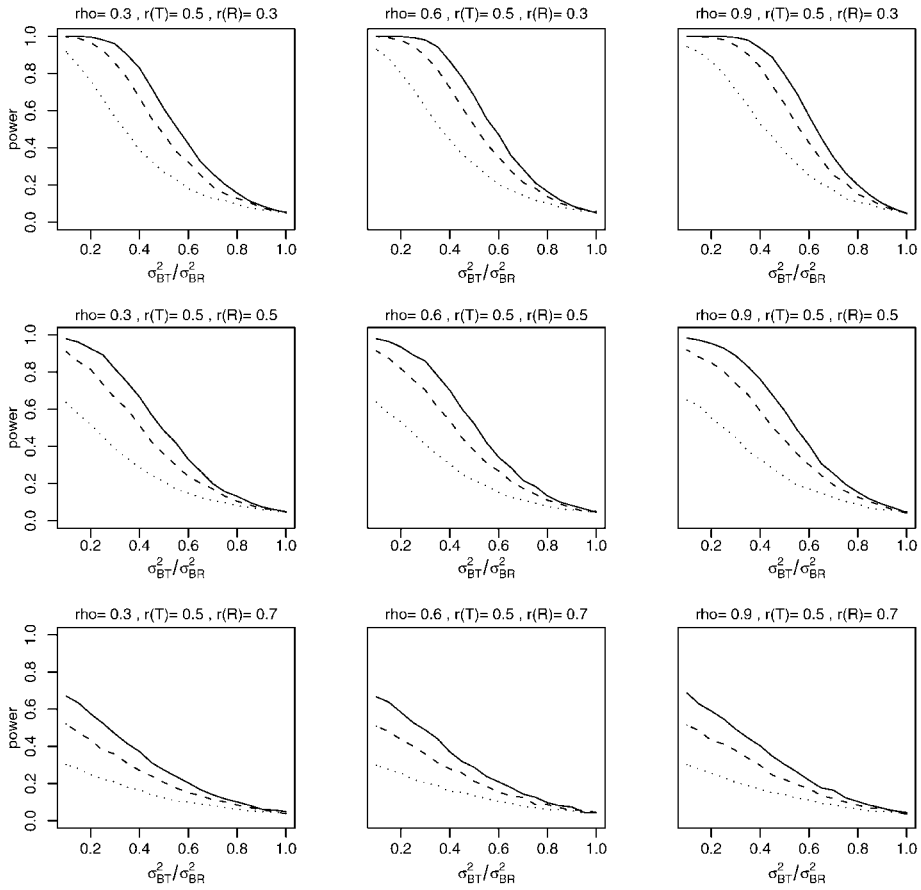

Figure 5. Power of the one-sided test for inter-subject variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T) =$ intra corr. of test; $r(R) =$ intra corr. of reference).

Assume that $\delta = \sigma^2_{\mathrm{BT}}/\sigma^2_{\mathrm{BR}} < 1$ and $\eta = \sigma^2_{\mathrm{BT}} - \sigma^2_{\mathrm{BR}} < 0$. Then, the power function of the test for inter-subject variability can be approximated by the probability given in Eq. (18) where K is given by

$$K = (\delta - 1)\left[\frac{(1+\delta^2)}{m}\left\{\frac{1}{m-1}\left(\frac{r}{1-r}\right)^2 + 2\left(\frac{r}{1-r}\right) + 1\right\} - 2\rho^2\delta\right]^{-1/2},$$

where $r = r_{\mathrm{T}} = r_{\mathrm{R}}$. The power of the test for inter-subject variability has the same properties as the test for total variability except when r is close to 1. If the intra-class correlation r converges to 1, K converges to 0 in the order of exponential. Thus, it is difficult to obtain a sufficient power even when n is large. This might be partially explained by the fact that the inter-subject variability cannot be assessed separately and it is always confounded with intra-subject variability in between-subject sum of square. Hence, if the contribution of the inter-subject variability

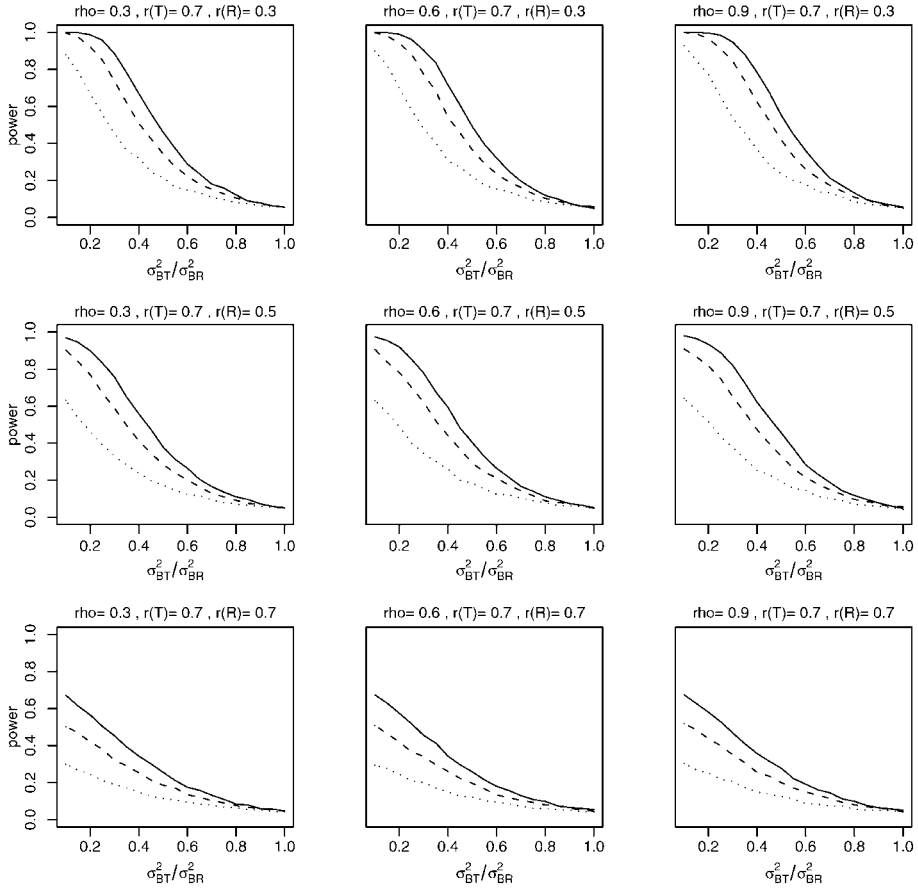

Figure 6. Power of the one-sided test for inter-subject variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T) =$ intra corr. of test; $r(R) =$ intra corr. of reference).

toward total variability is very small, it is very hard to obtain the estimate of the inter-subject variability with reasonable precision by the estimation method, which is based on subtracting the within-subject sum of squares from the between-subject sum of squares. The test for total variability, however, does not have this problem since total variability is assessed by adding the within-subject and the between-subject sum of squares.

SAMPLE SIZE CALCULATION

In this section we will consider sample size calculation for the one side test of total variability under 2×4 crossover design. An extension of the sample size calculation to $2 \times 2m$ design and also the test of inter-subject variability can be obtained by using similar argument. First, the sample size calculation based on

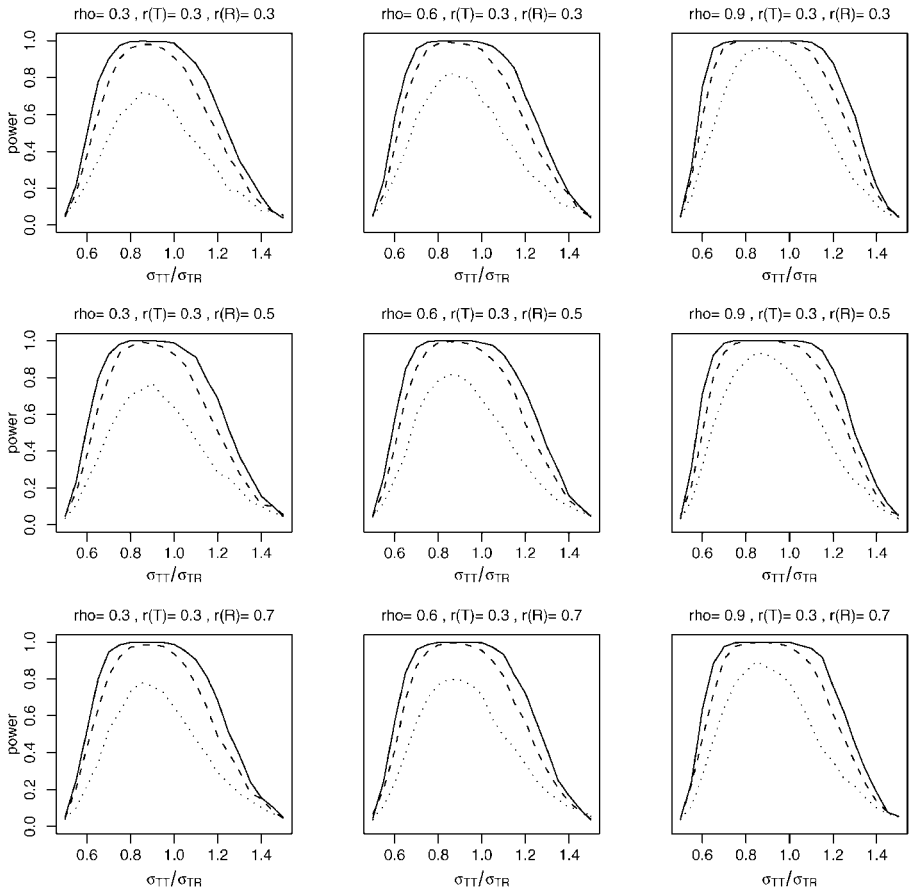

Figure 7. Power of equivalence test for total variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T)$ = intra corr. of test; $r(R)$ = intra corr. of reference).

the normal approximation will be discussed. Furthermore, the method based on Fisher–Cornish expansion will be discussed.

Consider the hypothesis (15) and assume $\delta = \sigma_{\mathrm{TT}}^2/\sigma_{\mathrm{TR}}^2 < 1$ and $\eta = \sigma_{\mathrm{TT}}^2 - \sigma_{\mathrm{TR}}^2 < 0$. By using the same argument for the power function in the previous section, the sample size n needed for achieving a power of $1 - \beta$ can be obtained by solving the following equation in term of n:

$$z_\alpha - \sqrt{2n}\frac{\eta}{\tau} = z_\beta \tag{20}$$

and it gives

$$n^* = \frac{\tau^2(z_\alpha + z_\beta)^2}{2\eta^2} + 1, \tag{21}$$

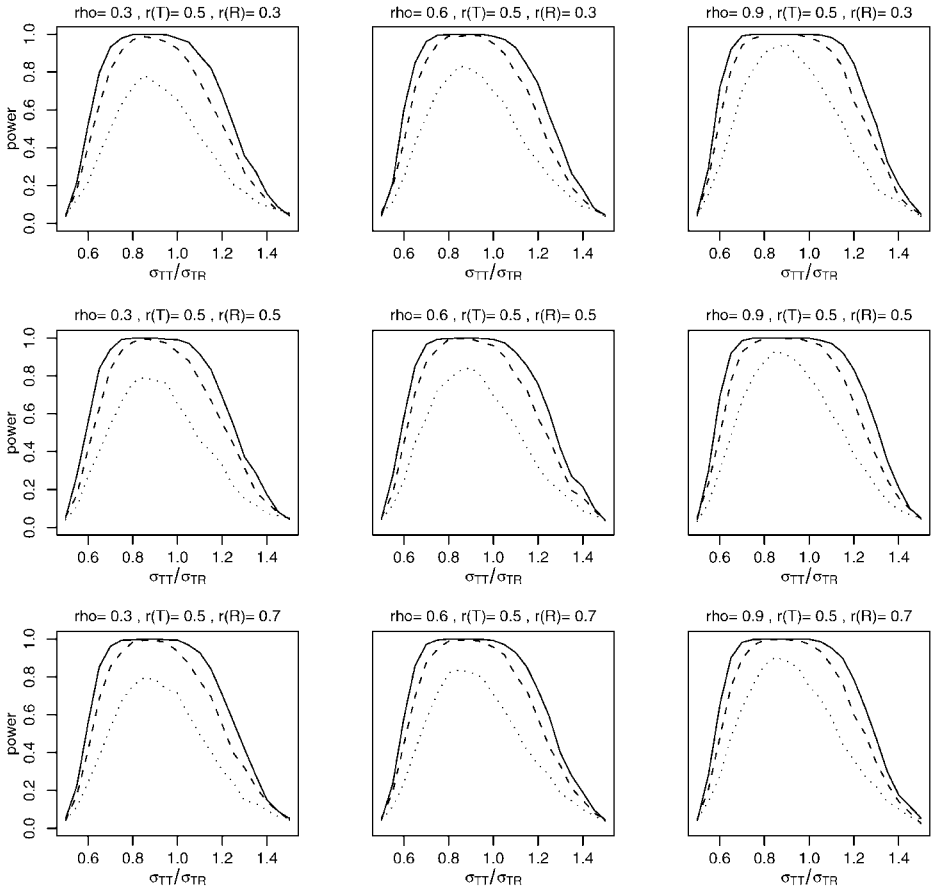

Figure 8. Power of equivalence test for total variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T)$ = intra corr. of test; $r(R)$ = intra corr. of reference).

where τ^2 is given in Eq. (16) with $m = 2$. Here, the reason for adding 1 in Eq. (21) is recovering 1 degree of freedom from the normal approximation.

The sample size from the normal approximation can be refined by using the Fisher–Cornish inversion. By Lee and Shao[9] the exact quantile of $\sqrt{2n}(\hat{\eta} - \eta)/\tau$ has the asymptotic expansion such that

$$x_\gamma \cong z_\gamma - n^{-1/2}\left[\frac{\lambda_{12}^2 + \lambda_{22}^2 - \lambda_{11}^2 - \lambda_{21}^2}{\lambda_{12}^2 + \lambda_{22}^2 + \lambda_{11}^2 + \lambda_{21}^2}\right]\left\{\frac{4z_\gamma^2 + 2}{3\sqrt{2}}\right\},$$

where $P(\sqrt{2n}(\hat{\eta} - \eta)/\tau > x_\gamma) = \gamma$ for all γ in any compact subset of $(0, 1)$ with an error term $O_p(n^{-1})$. Here, the eigenvalues λs are defined in Lemma 2 with $m = 2$. If z_α and z_β in Eq. (20) are replaced with the x_α and x_β, we can get

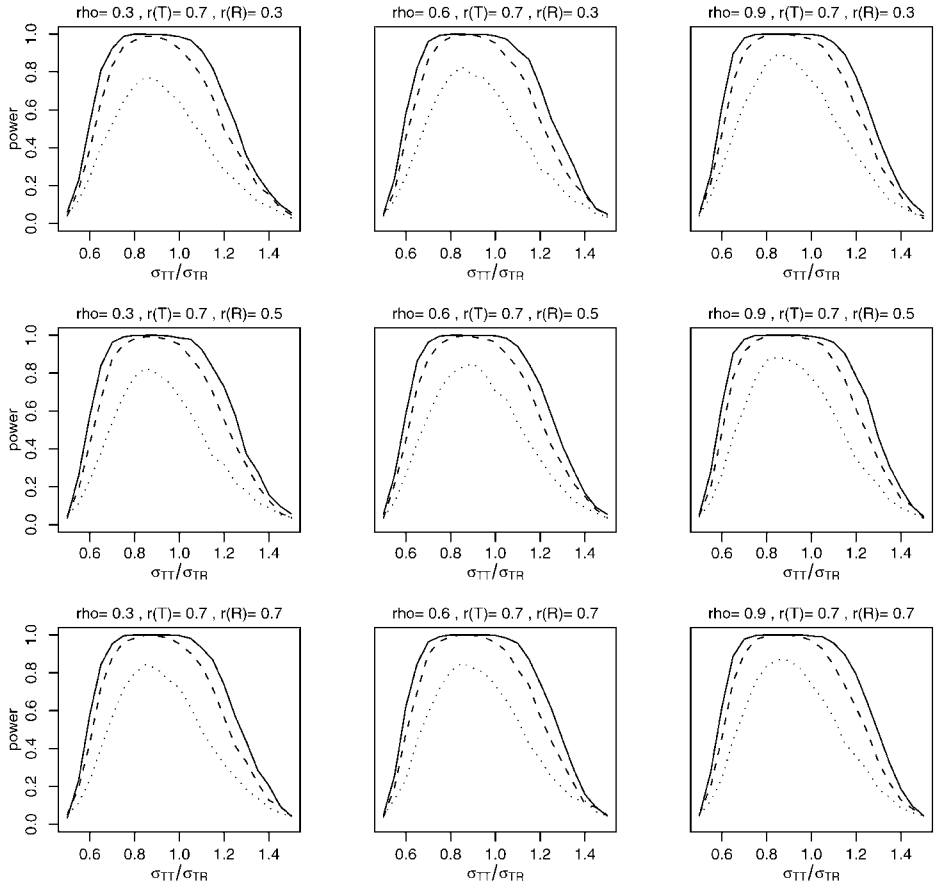

Figure 9. Power of equivalence test for total variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T)$ = intra corr. of test; $r(R)$ = intra corr. of reference).

the second order equation for n:

$$x_\alpha - \sqrt{2n}\frac{\eta}{\tau} = x_\beta. \tag{22}$$

Let n_1^{**} and n_2^{**} be two solutions for Eq. (22). Define $n^{**} = \max(n_1^{**}, n_2^{**}) + 1$ and it is the sample size needed for achieving a power of $1 - \beta$ based on Fisher–Cornish inversion. Note that this n^{**} can be easily obtained since Eq. (22) is the second order polynomial with respect to $\sqrt{n}$. However, the solution may not exit when n is very small.

To compare two methods for the sample size calculation for testing total variability under 2×4 crossover design, a small simulation study is considered. We consider the type I error probability is 0.05 and type II error probability is 0.2 (i.e., the power is 0.8). Some combinations of intra-class correlations ($r_k = 0.3, 0.5, 0.7$) for $k = T, R$ and inter-subject correlation ($\rho = 0.8, 1.0$) are

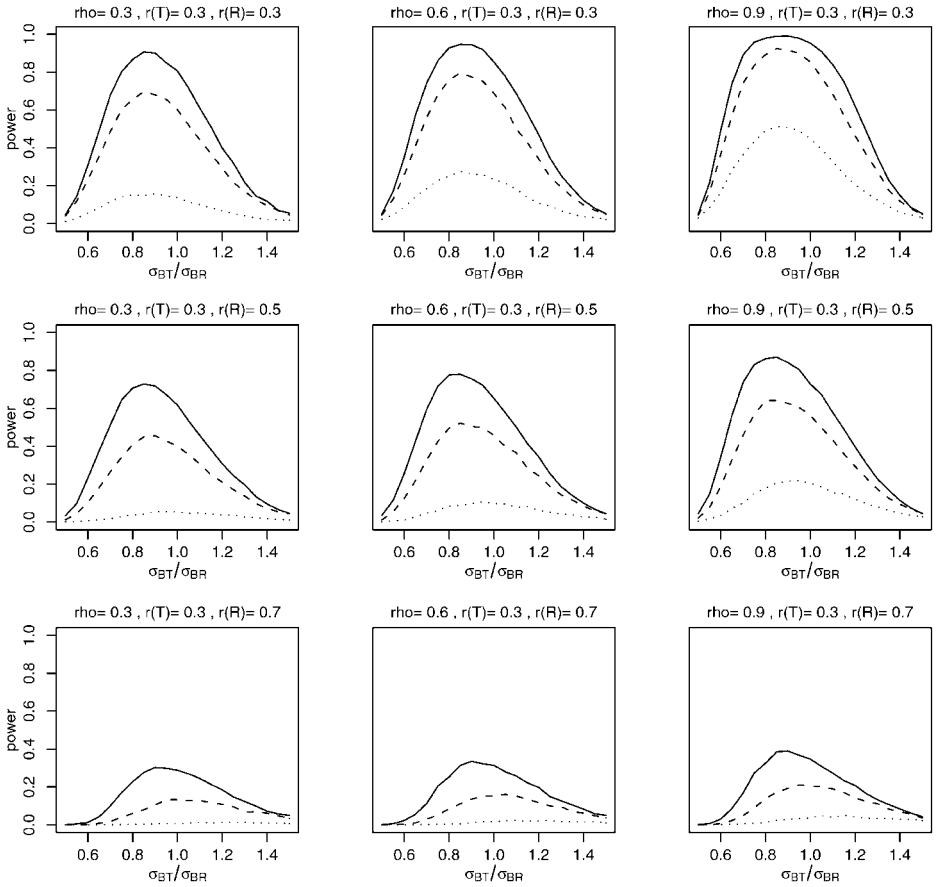

Figure 10. Power of equivalence test for inter-subject variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T) =$ intra corr. of test; $r(R) =$ intra corr. of reference).

considered. Also three values for $\delta = \sigma_{\mathrm{TT}}^2/\sigma_{\mathrm{TR}}^2$ are considered ($\delta = 0.75, 0.5, 0.25$) for the alternative hypothesis. When all the parameters are given, two sample size n^* and n^{**} are calculated and the corresponding powers P^* and P^{**} are approximated by 10,000 simulation. The result are given in Tables 1 and 2. The sample size based on the normal approximation become more conservative as δ is decreasing. However, the sample size based on the Fisher–Cornish inversion shows better performance in sense that the simulated power is very close to the nominal power 80% for all values of δ.

SIMULATION STUDIES

Several Monte Carlo simulations were performed to investigate the performance of the extension of the MLS method for testing inter-subject and

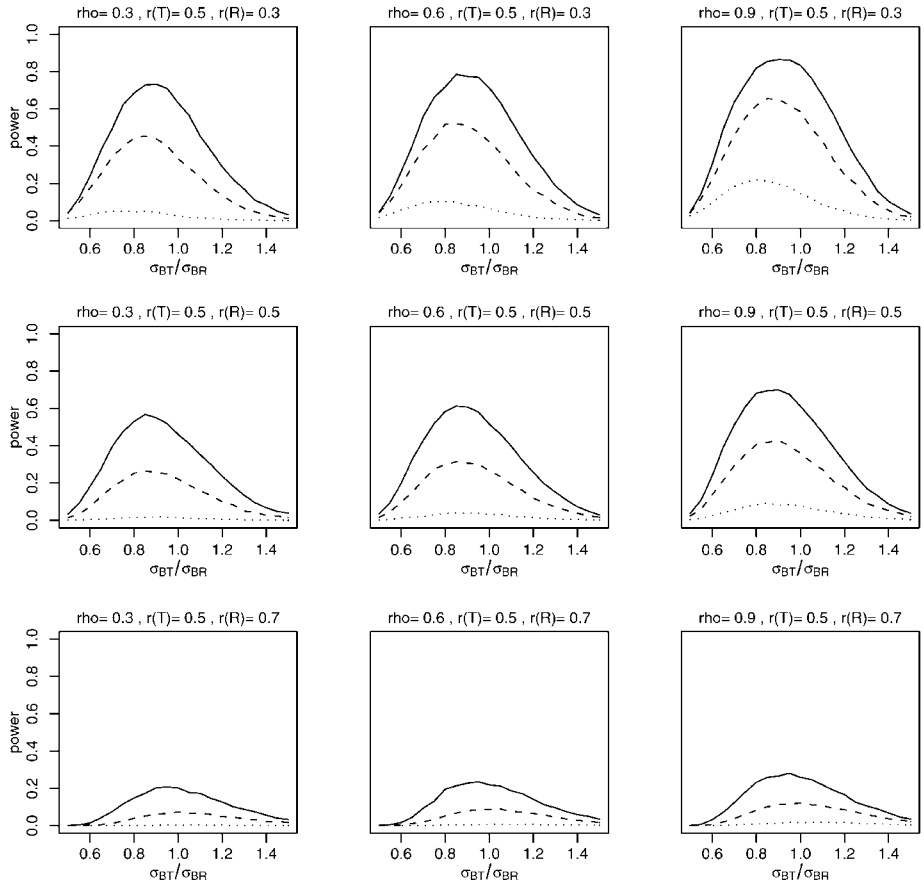

Figure 11. Power of equivalence test for inter-subject variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T) =$ intra corr. of test; $r(R) =$ intra corr. of reference).

total variabilities. For illustration purpose, we only consider the 2×4 replicated crossover design as recommended by the FDA in these simulations. In each simulation 5000, simulation runs were considered.

The first simulation was conducted to study the type I error probability of the proposed test for equality of inter-subject variability and total variability. The two-sided hypotheses for equality were considered with significant level $\alpha = 0.05$. Various combinations of intra-class correlations (r_T, r_R) and inter-subject correlation (ρ) were considered. In this simulation, six different intra-class correlations (0.1, 0.3, 0.5, 0.7, and 0.9) and six inter-subject correlations (0.0, 0.2, 0.4, 0.6, 0.8, and 1.0) were considered. Tables 3 and 4 summarize the simulation results of type I probabilities of the proposed test for total variability and inter-subject variability, respectively. The results show that type I error probabilities are well controlled under all the parameters considered. As sample size increases, type I error probability converges to the level $\alpha = 0.05$ very fast.

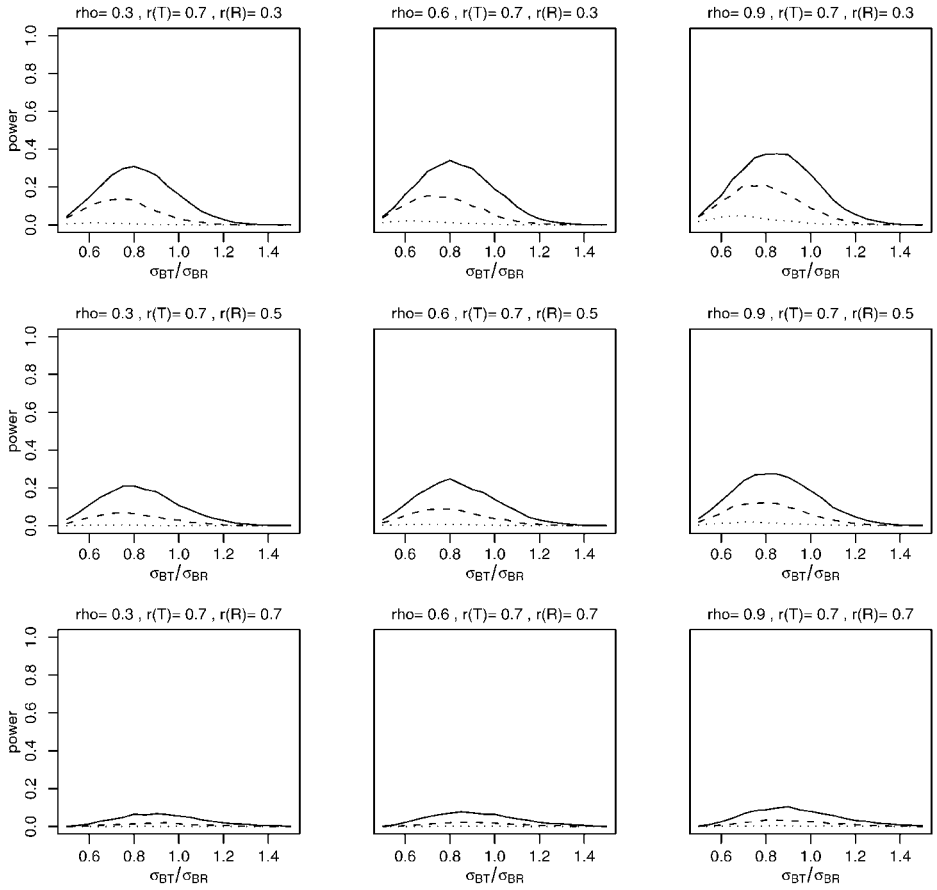

Figure 12. Power of equivalence test for inter-subject variability under a 2×4 replicated crossover design (dotted line: $n = 10$; dashed line: $n = 20$; solid line: $n = 30$; $r(T) =$ intra corr. of test; $r(R) =$ intra corr. of reference).

The second simulation was performed for the investigation of the power of the one-sided test. The hypotheses given in Eqs. (15) and (19) were considered. The power functions in various combinations of parameters are plotted in Figs. 1–3 for total variability and Figs. 4–6 for inter-subject variability, respectively. The power functions of the one-sided test for total variability show that the power increases rapidly as the sample size increases. Note that the power increases as the correlation ρ in the inter-subject covariance increases. Also, the power of the test for total variability is not sensitive to the intra-class correlation. However, the power function of the one-sided test for inter-subject variability is sensitive to the intra-class correlation. As both intra-class correlations approach to 1.0, the power is low even when n is large. However, when both intra-class correlations are less than 0.5, the test has a reasonably high power.

The third simulation considered the power for the test of equivalence in total and inter-subject variability. The hypotheses considered in this simulation were as follows

$$H_0 : \frac{\sigma_{\mathrm{TT}}}{\sigma_{\mathrm{TR}}} \notin (0.5, 1.5) \quad \text{vs.} \quad H_{\mathrm{a}} : \frac{\sigma_{\mathrm{TT}}}{\sigma_{\mathrm{TR}}} \in (0.5, 1.5)$$

and

$$H_0 : \frac{\sigma_{\mathrm{BT}}}{\sigma_{\mathrm{BR}}} \notin (0.5, 1.5) \quad \text{vs.} \quad H_{\mathrm{a}} : \frac{\sigma_{\mathrm{BT}}}{\sigma_{\mathrm{BR}}} \in (0.5, 1.5).$$

Note that the hypotheses are described as standard deviation instead of variance. In this simulation, we considered the criteria (50, 150%) for equivalence of variability instead of (80, 120%), which is the criteria for equivalence of mean. The power functions of the test for equivalence are given in Figs. 7–9 and Figs. 10–12 for total and inter-subject variability, respectively. For total variability, like the one-sided test, the power also increases rapidly as the sample size increases. The power increases as the correlation ρ in the inter-subject covariance increases. For inter-subject variability, the test for equivalence suffers from a very low power when both intra-class correlations are greater than 0.5.

In conclusion the test for total variability under the 2×4 crossover design based on the concept of the extension of the MLS method performs well in small sample in sense that it controls the type I error rate well. In addition, the proposed test has a reasonably high power in most situations under study. The test for inter-subject variability shows well-maintained type I error. However, the test suffers from a low power when the intra-class correlation is large, specially in the test of equivalence.

DISCUSSION

In this paper, we propose test for comparing inter-subject variability and total variability based on the concept of the extension of the MLS method under general crossover designs. The proposed test has an asymptotically correct type I error since it uses information of the correlation between the two treatments. Also the proposed test is very easy to implement since the test statistics are based on the moment estimator of variance components and the quantiles of central chi-square distribution.

The proposed test for total variability shows a very robust power function, which is not sensitive to the changes in parameters. On the other hand, the proposed test for inter-subject variability does have a reasonably high powers when the intra-class correlations are less then 0.5, it is very sensitive when the intra-class correlations are high. Hence, it should be used with caution. In practice, this sensitivity is inevitable since the inter-subject variability is always

confounded with the intra-subject variability in the usual ANOVA. The development of robust estimation methods for inter-subject variability requires further research.

In addition, the two methods for sample size calculation for testing total variability under 2×4 crossover design are discussed. The method based on the Fisher–Cornish inversion shows better performance than the method based on the normal approximation.

REFERENCES

1. FDA. *In Vivo Bioequivalence Studies Based on Population and Individual Bioequivalence Approaches*; Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 1999.
2. FDA. *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administrated Drug Products—General Considerations*; Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2000.
3. FDA. *Statistical Approaches to Establishing Bioequivalence*; Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2001.
4. Chinchilli, V.M.; Esinhart, J.D. Design and Analysis of Intra-subject Variability in Cross-Over Design. Stat. Med. **1996**, *15*, 1619–1634.
5. Graybill, F.A.; Wang, C.-M. Confidence Intervals on Nonnegative Linear Combinations of Variances. J. Am. Stat. Assoc. **1980**, *75*, 869–873.
6. Ting, N.; Burdick, R.K.; Graybill, F.A.; Jeyaratnam, S.; Lu, T.-F.C. Confidence Intervals on Linear Combinations of Variance Components That Are Unrestricted in Sign. J. Stat. Comput. Simulat. **1990**, *35*, 135–143.
7. Hyslop, T.; Hsuan, F.; Holder, D.J. A Small Sample Confidence Interval Approach to Assess Individual Bioequivalence. Stat. Med. **2000**, *19*, 2885–2897.
8. Wang, W. On Equivalence of Two Variances of a Bivariate Normal Vector. J. Stat. Plann. Inference **1999**, *81*, 279–292.
9. Lee, Y. Confidence Intervals and Tests on a Linear Combination of Variance Components When Estimators Are Dependent. Ph.D. Thesis, University of Wisonsin at Madison, 2002.
10. Lee, Y.; Shao, J.; Chow, S.-C. Confidence Interval for Linear Combinations of Variance Components When Estimators of Variance Components Are Dependent: An Extension of Modified Large Sample Method with Applications. **2002**, submitted for publication.