Comparing Variabilities in Clinical Research

Yonghee Lee
University of Wisconsin–Madison, Madison, Wisconsin, U.S.A.

Hansheng Wang
Peking University, Beijing, People's Republic of China

Shein-Chung Chow
Millennium Pharmaceuticals, Inc., Cambridge, Massachusetts, U.S.A.

INTRODUCTION

In most clinical trials comparing a test drug and a control (e.g., a placebo control or an active control), treatment effect is usually established by comparing mean response change from baseline of some primary study endpoints assuming that their corresponding variabilities are comparable. In practice, however, variabilities associated with the test drug and the control could be very different. When the variability of the test drug is much larger than that of the reference drug, safety of the test drug could be a concern. Thus, in addition to comparing mean responses between treatments, it is also of interest to compare the variabilities associated with the responses between treatments.

In practice, the variabilities associated with the response are usually classified into two categories, namely, the intrasubject (or within-subject) variability and the intersubject (or between-subjects) variability. Intrasubject variability refers to the variability observed from repeated measurements from the same subject under the same experimental condition. On the other hand, intersubject variability is the variability due to the heterogeneity among subjects. The total variability is simply the sum of the intra- and intersubject variabilities. In practice, it is of interest to test for equality, noninferiority/superiority, and similarity (or equivalence) between treatments in terms of the intrasubject, intersubject, and/or total variabilities. Under a replicated parallel design, the usual F-test can be performed to compare intrasubject variabilities. Under a replicated crossover design, Chinchilli and Esinhart[1] proposed the F-test using an orthogonal transformation to test for difference in intrasubject variability between treatments. Under replicated parallel designs, however, intersubject and total variabilities between two treatments cannot be compared by the usual F-test because the distribution of estimators does not follow a chi-square distribution. In these cases, the modified large sample (MLS) method is suggested (see, e.g., Refs. [2–5]). The MLS method is shown to be superior to the other approximation methods in practice. Under a replicated crossover design, the MLS method is not appropriate to compare intersubject and total variabilities because estimators of variance components are not independent. Lee et al.[6] extend the MLS method when estimators of variance components are dependent. In addition, Lee et al.[7] study a test for comparing intersubject and total variabilities under replicated crossover designs based on the extension of the MLS method.

In practice, instead of comparing intrasubject variabilities, it may be of interest to compare intrasubject coefficient of variation (CV). In recent years, the use of intrasubject CV has become popular when assessing the reproducibility, similarity, and stability of a drug product when the drug product is repeatedly administered over a period of time. In this entry, we will discuss two statistical methods by Chow and Tse[8] and Quan and Shih.[9]

OVERVIEW

In the next four sections, statistical tests for equality, superiority/noninferiority, and similarity under a crossover design or a parallel design in clinical trials are derived for comparing intrasubject variability, intrasubject CV, intersubject variabilities, and total variabilities, respectively. The section ''Sample Size Calculation'' provides formulas for sample size calculation corresponding to respective statistical tests. Some concluding remarks are given in the last section. Throughout this article, for the sake of convenience, we will denote test formulation (drug) and reference formulation (drug) by T and R, respectively.

COMPARING INTRASUBJECT VARIABILITIES

To assess intrasubject variability, replicates from the same subject are necessarily obtained. For this purpose,

Encyclopedia of Biopharmaceutical Statistics
DOI: 10.1081/E-EBS 120017301

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

replicated crossover designs or parallel group designs with replicates are commonly employed. In what follows, statistical tests for comparing intrasubject variabilities under a parallel design with replicates and a replicated crossover design (e.g., a $2 \times 2m$ replicated crossover design) will be studied.

Parallel Design with Replicates

Let x_{ijk} be the observation of the kth replicate ($k = 1, \ldots, m$) of the jth subject ($j = 1, \ldots, n_i$) from the ith treatment ($i = \mathrm{T}, \mathrm{R}$). The following linear mixed effects model is usually considered:

$$x_{ijk} = \mu_i + S_{ij} + e_{ijk} \tag{1}$$

where μ_i is the treatment effect, S_{ij} is the random effect due to the jth subject in the ith treatment group, and e_{ijk} is the intrasubject variability under the ith treatment. It is assumed that for a fixed i, S_{ij} is independent and identically distributed as a normal random variable with mean 0 and variance $\sigma^2_{\mathrm{B}i}$, and e_{ijk}, $k = 1, \ldots, m$ is independent and identically distributed as a normal random variable with mean 0 and variance $\sigma^2_{\mathrm{W}i}$. Under this model, an unbiased estimator for intrasubject variance $\sigma^2_{\mathrm{W}i}$ is given by

$$\hat{\sigma}^2_{\mathrm{W}i} = \frac{1}{n_i(m-1)} \sum_{j=1}^{n_i} \sum_{k=1}^{m} (x_{ijk} - \bar{x}_{ij\cdot})^2 \tag{2}$$

where

$$\bar{x}_{ij\cdot} = \frac{1}{m} \sum_{k=1}^{m} x_{ijk} \tag{3}$$

It can be seen that $n_i(m-1)\hat{\sigma}^2_{\mathrm{W}i}/\sigma^2_{\mathrm{W}i}$ is distributed as a χ^2-distribution with $n_i(m-1)$ degrees of freedom. Note that σ^2_{WT} and σ^2_{WR} are intrasubject variances for test and reference formulation, respectively. Also, σ^2_{BT} and σ^2_{BR} are the intersubject variances for test and reference formulation, respectively.

Test for equality

In practice, it is often of interest to test whether two drug products have the same intrasubject variability. The following hypotheses are then of interest:

$$\mathrm{H}_0 : \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{WR}}} = 1 \quad \text{vs.} \quad \mathrm{H}_\mathrm{a} : \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{WR}}} \neq 1 \tag{4}$$

A commonly used test statistic for testing the above hypotheses is given by

$$T = \frac{\hat{\sigma}^2_{\mathrm{WT}}}{\hat{\sigma}^2_{\mathrm{WR}}}$$

Under the null hypothesis, T is distributed as an F distribution with $n_\mathrm{T}(m-1)$ and $n_\mathrm{R}(m-1)$ degrees of freedom. Hence, we reject the null hypothesis at the α level of significance if

$$T > F_{\alpha/2, n_\mathrm{T}(m-1), n_\mathrm{R}(m-1)}$$

or

$$T < F_{1-\alpha/2, n_\mathrm{T}(m-1), n_\mathrm{R}(m-1)}$$

where $F_{\alpha/2, n_\mathrm{T}(m-1), n_\mathrm{R}(m-1)}$ is the upper $(\alpha/2) \times 100\%$ quantile of an F distribution with $n_\mathrm{T}(m-1)$ and $n_\mathrm{R}(m-1)$ degrees of freedom.

Test for noninferiority/superiority

The problem of testing noninferiority and superiority can be unified by the following hypotheses

$$\mathrm{H}_0 : \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{WR}}} \geq \delta \quad \text{vs.} \quad \mathrm{H}_\mathrm{a} : \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{WR}}} < \delta \tag{5}$$

When $\delta < 1$, the rejection of the null hypothesis indicates the superiority of the test product over the reference in terms of the intrasubject variability. When $\delta > 1$, the rejection of the null hypothesis indicates the noninferiority of the test product over the reference. The test statistic is given by

$$T = \frac{\hat{\sigma}^2_{\mathrm{WT}}}{\delta\hat{\sigma}^2_{\mathrm{WR}}}$$

Under the null hypothesis, T is distributed as an F random variable with $n_\mathrm{T}(m-1)$ and $n_\mathrm{R}(m-1)$ degrees of freedom. Hence, we reject the null hypothesis at the α level of significance if

$$T < F_{1-\alpha, n_\mathrm{T}(m-1), n_\mathrm{R}(m-1)}$$

Test for similarity

For testing similarity, the following hypotheses are usually considered:

$$\mathrm{H}_0 : \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{WR}}} \notin \left(\frac{1}{\delta}, \delta\right) \quad \text{vs.} \quad \mathrm{H}_\mathrm{a} : \frac{\sigma^2_{\mathrm{WT}}}{\sigma^2_{\mathrm{WR}}} \in \left(\frac{1}{\delta}, \delta\right) \tag{6}$$

where $\delta > 1$ is the similarity limit. The above hypotheses can be decomposed into the following two one-sided hypotheses:

$$H_{01}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} \geq \delta \quad \text{vs.} \quad H_{a1}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} < \delta$$

$$H_{02}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} \leq \frac{1}{\delta} \quad \text{vs.} \quad H_{a2}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} > \frac{1}{\delta} \tag{7}$$

These two one-sided hypotheses can be tested by the following two test statistics

$$T_1 = \frac{\hat{\sigma}_{WT}^2}{\delta\hat{\sigma}_{WR}^2} \quad \text{and} \quad T_2 = \frac{\delta\hat{\sigma}_{WT}^2}{\hat{\sigma}_{WR}^2}$$

We then reject the null hypothesis and conclude similarity at the α level of significance if

$$T_1 < F_{1-\alpha,n_T(m-1),n_R(m-1)} \quad \text{and} \quad T_2 > F_{\alpha,n_T(m-1),n_R(m-1)}$$

Replicated Crossover Design

Compared with the parallel design with replicates, the merit of a crossover design is the ability to make comparisons of treatment effect within subjects. In this section, without loss of generality, consider a $2 \times 2m$ replicated crossover design comparing two treatments. 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 a 2×4 crossover design, which is recommended by the United States Food and Drug Administration (FDA) for assessment of population/individual bioequivalence.[10]

Suppose that n_1 subjects are assigned to the first sequence and n_2 subjects are assigned to the second sequence. Let x_{ijkl} be the observation from the jth subject ($j = 1, \ldots, n_i$) in the ith sequence ($i = 1, 2$) under the lth replicate ($l = 1, \ldots, m$) of the kth treatment ($k = T, R$). As indicated in Ref. [1], the following linear mixed effects model can be considered to describe a $2 \times 2m$ replicated crossover design:

$$x_{ijkl} = \mu_k + \gamma_{ikl} + S_{ijk} + \epsilon_{ijkl} \tag{8}$$

where μ_k is the treatment effect for formulation k, γ_{ikl} is the fixed effect of the lth replicate on treatment k in the ith sequence with constraint

$$\sum_{i=1}^{2}\sum_{l=1}^{m}\gamma_{ikl} = 0$$

$(S_{ijT}, S_{ijR})'$ are the random effects of the jth subject in the ith sequence, which are independent and identically distributed as a bivariate normal random vector with mean $(0,0)'$ and covariance matrix

$$\sum\nolimits_B = \begin{pmatrix} \sigma_{BT}^2 & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{BR}^2 \end{pmatrix}$$

Note that σ_{BT}^2 and σ_{BR}^2 are intersubject variances under 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_{WT}^2 or σ_{WR}^2, which are intrasubject variabilities under the test formulation and the reference formulation, respectively. It is assumed that $(S_{ijT}, S_{ijR})'$ and ϵ_{ijkl} are independent.

To obtain estimators of intrasubject variances, it is a common practice to use an orthogonal transformation, which is considered by Chinchilli and Esinhart.[1] A new random variable z_{ijkl} can be obtained by using the orthogonal transformation

$$\mathbf{z}_{ijk} = \mathbf{P}'\mathbf{x}_{ijk} \tag{9}$$

where

$$\mathbf{x}'_{ijk} = (x_{ijk1}, x_{ijk2}, \ldots, x_{ijkm}),$$

$$\mathbf{z}'_{ijk} = (z_{ijk1}, z_{ijk2}, \ldots, z_{ijkm})$$

and $\mathbf{P}$ is an $m \times m$ orthogonal transformation under which $\mathbf{P}'\mathbf{P}$ is an $m \times m$ diagonal matrix. The first column of $\mathbf{P}$ is usually defined by the vector $1/m(1, 1, \ldots, 1)'$ to obtain $z_{ijk1} = \bar{x}_{ijk}$ and the other columns can be defined to satisfy the orthogonality of $\mathbf{P}$ and $\text{Var}(z_{ijkl}) = \sigma_{Wi}^2$ for $l = 2, \ldots, m$. For example, in the 2×4 crossover design, the new random variable z_{ijkl} can be defined as

$$z_{ijk1} = \frac{x_{ijk1} + x_{ijk2}}{2} \equiv \bar{x}_{ijk.} \quad \text{and} \quad z_{ijk2} = \frac{x_{ijk1} - x_{ijk2}}{\sqrt{2}}$$

Now, the estimator of intrasubject variance can be defined as

$$\hat{\sigma}_{WT}^2 = \frac{1}{(n_1 + n_2 - 2)(m-1)}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\sum_{l=2}^{m}(z_{ijTl} - \bar{z}_{i.Tl})^2$$

$$\hat{\sigma}_{WR}^2 = \frac{1}{(n_1 + n_2 - 2)(m-1)}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\sum_{l=2}^{m}(z_{ijRl} - \bar{z}_{i.Rl})^2 \tag{10}$$

where

$$\bar{z}_{i.kl} = \frac{1}{n_i}\sum_{j=1}^{n_i} z_{ijkl}$$

It should be noted that $\hat{\sigma}^2_{WT}$ and $\hat{\sigma}^2_{WR}$ are independent and unbiased estimators for σ^2_{WT} and σ^2_{WR}, respectively.

Test for equality

Under the null hypothesis of Eq. 4, a test statistic

$$T = \frac{\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}}$$

is distributed as an F random variable with d and d degrees of freedom, where $d = (n1 + n_2 - 2)(m-1)$. Hence, we reject the null hypothesis at the α level of significance if

$$T > F_{\alpha/2,d,d} \quad \text{or} \quad T < F_{1-\alpha/2,d,d}$$

Test for noninferiority/superiority

Under the null hypothesis of Eq. 5, $T = \hat{\sigma}^2_{WT}/\delta\hat{\sigma}^2_{WT}$is distributed as an F random variable with d and d degrees of freedom. Hence, we reject the null hypothesis at the α level of significance if

$$T < F_{1-\alpha,d,d}$$

Test for similarity

Under the two one-sided hypotheses as discussed in Eq. 7, the following two test statistics

$$T_1 = \frac{\hat{\sigma}^2_{WT}}{\delta\hat{\sigma}^2_{WR}} \quad \text{and} \quad T_2 = \frac{\delta\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}}$$

are used. We then reject the null hypothesis and conclude similarity at the α level of significance if

$$T_1 < F_{1-\alpha,d,d} \quad \text{and} \quad T_2 > F_{\alpha,d,d}$$

COMPARING INTRASUBJECT COEFFICIENTS OF VARIATION

In addition to comparing intrasubject variabilities, it is often of interest to compare intrasubject CV, which is a relative standard deviation adjusted by the mean. In recent years, the use of intrasubject CV has become increasingly popular. For example, the FDA defines highly variable drug products based on their intrasubject CVs. A drug product is said to be a highly variable drug if its intrasubject CV is greater than 30%. The intrasubject CV is also used as a measure for reproducibility of blood levels (or blood concentration–time curves) of a given formulation when the formulation is repeatedly administered at different dosing periods. In addition, the information regarding the intrasubject CV of a reference product is usually used for performing power analysis for sample size calculation in bioavailability and bioequivalence studies. In practice, two methods are commonly used for comparing intrasubject CVs. One is proposed by Chow and Tse,[8] which is referred to as conditional random effects model. The other one is suggested by Quan and Shih,[9] which is a simple one-way random effects model. In what follows, we will briefly introduce these two methods.

Simple Random Effects Model

Quan and Shih[9] developed a method to estimate the intrasubject CV based on the simple mixed effects model Eq. 1 under which the intrasubject variability is a constant for all subjects in the same treatment. An intuitive unbiased estimator for μ_i is then given by

$$\hat{\mu}_i = \frac{1}{n_i m}\sum_{j=1}^{n_i}\sum_{k=1}^{m} x_{ijk}$$

Hence, an estimator of the intrasubject CV can be obtained by, for i = T, R,

$$\widehat{CV}_i = \frac{\hat{\sigma}_{Wi}}{\hat{\mu}_i}$$

By Taylor's expansion, it follows that

$$\begin{aligned}\widehat{CV}_i - CV_i &= \frac{\hat{\sigma}_{Wi}}{\hat{\mu}_i} - \frac{\sigma_{Wi}}{\mu_i} \\ &\approx \frac{1}{2\mu_i\sigma_{Wi}}(\hat{\sigma}^2_{Wi} - \sigma^2_{Wi}) - \frac{\sigma_{Wi}}{\mu_i^2}(\hat{\mu}_i - \mu_i)\end{aligned}$$

Hence, by central limit theorem (CLT), $\widehat{CV}_i$ is asymptotically distributed as a normal random variable with mean CV_i and variance σ_i^{*2}/n_i, where

$$\begin{aligned}\sigma_i^{*2} &= \frac{\sigma^2_{Wi}}{2m\mu_i^2} + \frac{\sigma^4_{Wi}}{\mu_i^4} \\ &= \frac{1}{2m}CV_i^2 + CV_i^4\end{aligned}$$

Thus, an intuitive estimator of σ_i^{*2} is given by

$$\hat{\sigma}_i^{*2} = \frac{1}{2m}\widehat{\mathrm{CV}}_i^2 + \widehat{\mathrm{CV}}_i^4$$

Test for equality

The following hypotheses are usually considered for testing equality in intrasubject CVs

$$\mathrm{H}_0: \mathrm{CV_T} = \mathrm{CV_R} \quad \text{vs.} \quad \mathrm{H_a}: \mathrm{CV_T} \neq \mathrm{CV_R} \tag{11}$$

Under the null hypothesis, the test statistic

$$T = \frac{\widehat{\mathrm{CV}}_\mathrm{T} - \widehat{\mathrm{CV}}_\mathrm{R}}{\sqrt{\hat{\sigma}_\mathrm{T}^{*2}/n_\mathrm{T} + \hat{\sigma}_\mathrm{R}^{*2}/n_\mathrm{R}}}$$

is asymptotically distributed as a standard normal random variable. Hence, we reject the null hypothesis at the α level of significance if $|T| > z_{\alpha/2}$.

Test for noninferiority/superiority

Similarly, the problem of testing noninferiority and superiority can be unified by the following hypotheses:

$$\begin{aligned}&\mathrm{H}_0: \mathrm{CV_T} - \mathrm{CV_R} \geq \delta \quad \text{vs.}\\ &\mathrm{H_a}: \mathrm{CV_T} - \mathrm{CV_R} < \delta\end{aligned} \tag{12}$$

where δ is the noninferiority/superiority margin. When $\delta > 0$, the rejection of the null hypothesis indicates the noninferiority of the test formulation over the reference formulation. When $\delta < 0$, the rejection of the null hypothesis indicates the superiority of the test formulation over the reference formulation.

Under the null hypothesis, the test statistic

$$T = \frac{\widehat{\mathrm{CV}}_\mathrm{T} - \widehat{\mathrm{CV}}_\mathrm{R} - \delta}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}}$$

is asymptotically distributed as a standard normal random variable. Hence, we reject the null hypothesis at the α level of significance if $T < -z\alpha$.

Test for similarity

For testing similarity, the following hypotheses are usually considered:

$$\begin{aligned}&\mathrm{H}_0: |\mathrm{CV_T} - \mathrm{CV_R}| \geq \delta \quad \text{vs.}\\ &\mathrm{H_a}: |\mathrm{CV_T} - \mathrm{CV_R}| < \delta\end{aligned} \tag{13}$$

The two drug products are concluded similar to each other if the null hypothesis is rejected at a given significance level. The null hypothesis is rejected at the α level of significance if

$$\frac{\widehat{\mathrm{CV}}_\mathrm{T} - \widehat{\mathrm{CV}}_\mathrm{R} + \delta}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}} > z_\alpha \quad \text{and}$$
$$\frac{\widehat{\mathrm{CV}}_\mathrm{T} - \widehat{\mathrm{CV}}_\mathrm{R} - \delta}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}} < -z_\alpha$$

Conditional Random Effects Model

In practice, the variability of the observed response often increases as the mean increases. In many cases, the standard deviation of the intrasubject variability is approximately proportional to the mean value. To best describe this type of data, Chow and Tse[8] proposed the following conditional random effects model.

$$x_{ijk} = A_{ij} + A_{ij}e_{ijk} \tag{14}$$

where x_{ijk} is the observation from the kth replicate ($k = 1, \ldots, m$) of the jth subject ($j = 1, \ldots, n_i$) from the ith treatment ($i = \mathrm{T}, \mathrm{R}$) and A_{ij} is the random effect due to the jth subject in the ith treatment. It is assumed that A_{ij} is normally distributed as a normal random variable with mean μ_i and variance $\sigma_{\mathrm{B}i}^2$ and e_{ijk} is normally distributed as a normal random variable with mean 0 and variance $\sigma_{\mathrm{W}i}^2$. For a given subject with a fixed A_{ij}, x_{ijk} is normally distributed as a normal random variable with mean A_{ij} and variance $A_{ij}^2\sigma_{\mathrm{W}i}^2$. Hence, the CV for this subject is given by

$$\mathrm{CV} = \frac{|A_{ij}\sigma_{\mathrm{W}i}|}{|A_{ij}|} = \sigma_{\mathrm{W}i}$$

As can be seen, the conditional random effects model assumes the CV is a constant across subjects.

Define

$$\bar{x}_{i\cdot\cdot} = \frac{1}{nm}\sum_{j=1}^{n_i}\sum_{k=1}^{m} x_{ijk}$$

$$M_{i1} = \frac{m}{n_i - 1}\sum_{j=1}^{n_i}(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot})^2$$

$$M_{i2} = \frac{1}{n_i(m_i - 1)}\sum_{j=1}^{n_i}\sum_{k=1}^{m}(x_{ijk} - \bar{x}_{ij\cdot})^2$$

It can be verified that

$$\begin{aligned}E(\bar{x}_{i\cdot\cdot}) &= \mu_i\\ E(M_{i1}) &= (\mu_i^2 + \sigma_{\mathrm{BT}}^2)\sigma_{\mathrm{W}i}^2 + m\sigma_{\mathrm{B}i}^2 = \tau_{i1}^2\\ E(M_{i2}) &= (\mu_i^2 + \sigma_{\mathrm{BT}}^2)\sigma_{\mathrm{W}i}^2 = \tau_{i2}^2\end{aligned}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

It follows that

$$\mathrm{CV}_i = \sigma_{\mathrm{W}i} = \sqrt{\frac{E(M_{i2})}{E^2(\bar{x}_{i\cdot\cdot}) + (M_{i1} - M_{i2})}}$$

Hence, an estimator for CV_i is given by

$$\widehat{\mathrm{CV}_i} = \sqrt{\frac{M_{i2}}{\bar{x}_{i\cdot\cdot}^2 + (M_{i1} - M_{i2})/m}}$$

By Taylor's expansion,

$$\begin{aligned}\widehat{\mathrm{CV}_i} - \mathrm{CV}_i &\approx \frac{1}{\tau_{i1}\tau_{i2}}(M_{i2} - \tau_{i2}^2) - \frac{\mu_i\tau_{i2}}{\tau_{i1}^3}(\bar{x}_{i\cdot\cdot} - \mu_i) \\ &\quad - \frac{\tau_{i2}}{2m\tau_{i1}^3}(M_{i1} - M_{i2} - (\tau_{i1}^2 - \tau_{i2}^2)) \\ &= k_0(\bar{x}_{i\cdot\cdot} - \mu_i) + k_1(M_{i1} - \tau_{i1}^2) \\ &\quad + k_2(M_{i2} - \tau_{i2}^2)\end{aligned}$$

where

$$k_0 = -\frac{\mu_i\tau_{i2}}{\tau_{i1}^3}$$

$$k_1 = -\frac{\tau_{i2}}{2m\tau_{i1}^3}$$

$$k_2 = \left(\frac{1}{\tau_{i1}\tau_{i2}} + \frac{\tau_{i2}}{2m\tau_{i1}^3}\right)$$

As a result, $\widehat{\mathrm{CV}_i}$'s distribution can be approximated by a normal random variable with mean CV_i and variance $\sigma_{\mathrm{i}}^{*2}/n_{\mathrm{i}}$, where

$$\sigma_i^{*2} = \mathrm{var}\left[k_0\bar{x}_{ij\cdot} + mk_1(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot})^2 + \frac{k_2}{m-1}\sum_{k=1}^{m}(x_{ijk} - \bar{x}_{ij\cdot})^2\right]$$

An intuitive estimator for σ_i^{*2} is the sample variance, denoted by $\hat{\sigma}_i^{*2}$, of

$$k_0\bar{x}_{ij\cdot} + mk_1(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot})^2 + \frac{k_2}{m-1}\sum_{k=1}^{m}(x_{ijk} - \bar{x}_{ij\cdot})^2$$

$$j = 1, \cdots, n_i$$

Test for equality

Under the null hypothesis in Eq. 11, the test statistic

$$T = \frac{\widehat{\mathrm{CV}_\mathrm{T}} - \widehat{\mathrm{CV}_\mathrm{R}}}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}}$$

is asymptotically distributed as a standard normal random variable. Hence, we reject the null hypothesis at the α level of significance if $|T| > z_{\alpha/2}$.

Test for noninferiority/superiority

Under the null hypothesis of Eq. 12, the test statistic

$$T = \frac{\widehat{\mathrm{CV}_\mathrm{T}} - \widehat{\mathrm{CV}_\mathrm{R}} - \delta}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}}$$

is asymptotically distributed as a standard normal random variable. Hence, we reject the null hypothesis at the α level of significance if $T < -z_\alpha$.

Test for similarity

For testing similarity, consider the hypotheses

$$\mathrm{H}_0: |\mathrm{CV}_\mathrm{T} - \mathrm{CV}_\mathrm{R}| \geq \delta \quad \text{vs.} \quad \mathrm{H}_\mathrm{a}: |\mathrm{CV}_\mathrm{T} - \mathrm{CV}_\mathrm{R}| < \delta$$

The two drug products are concluded similar to each other if the null hypothesis is rejected at a given significance level. The null hypothesis is rejected at α level of significance if

$$\frac{\widehat{\mathrm{CV}_\mathrm{T}} - \widehat{\mathrm{CV}_\mathrm{R}} + \delta}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}} > z_\alpha \quad \text{vs.}$$

$$\frac{\widehat{\mathrm{CV}_\mathrm{T}} - \widehat{\mathrm{CV}_\mathrm{R}} - \delta}{\sqrt{\sigma_\mathrm{T}^{*2}/n_\mathrm{T} + \sigma_\mathrm{R}^{*2}/n_\mathrm{R}}} < -z_\alpha$$

COMPARING INTERSUBJECT VARIABILITIES

In addition to comparing intrasubject variabilities or intrasubject CVs, it is also of interest to compare intersubject variabilities. In practice, it is not uncommon that clinical results may not be reproducible from subject to subject within the target population or from subjects within the target population to subjects within a similar but slightly different population due to the intersubject variability. How to test a difference in intersubject and total variability between two treatments is a challenging problem to clinical scientists, especially biostatisticians, because of the following reasons. First, unbiased estimators of the intersubject and total variabilities are usually not chi-square distributed under both parallel and crossover design with replicates. Second, the estimators for the intersubject and total variabilities under different treat-

ments are usually not independent under a crossover design. As a result, unlike tests for comparing intrasubject variabilities, the standard F-test is not applicable. Tests for comparing intersubject variabilities under a parallel design can be performed by using the MLS method (see Refs. [2–5]). As indicated earlier, the MLS method is superior to many other approximation methods. Under crossover designs, however, the MLS method cannot be directly applied because estimators of variance components are not independent. Lee et al.[6] proposed an extension of the MLS method when estimators of variance components are not independent. In addition, tests for comparing intersubject and total variabilities under crossover designs are studied by Lee et al.[7] Note that the MLS method by Hyslop et al.[5] is recommended as a statistical test for individual bioequivalence by the FDA.[10]

PARALLEL DESIGN WITH REPLICATES

Under model Eq. 1, for $i = \mathrm{T}, \mathrm{R}$, define

$$s_{\mathrm{B}i}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot})^2 \tag{15}$$

where $\bar{x}_{ij}$ is given in Eq. 3 and

$$\bar{x}_{i\cdot\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{x}_{ij\cdot}$$

Note that $E({s_{\mathrm{B}i}}^2) = \sigma_{\mathrm{B}i}^2 + \sigma_{\mathrm{W}i}^2/m$. Therefore, the unbiased estimators for the intersubject variances are given by

$$\hat{\sigma}_{\mathrm{B}i}^2 = s_{\mathrm{B}i}^2 - \frac{1}{m}\hat{\sigma}_{\mathrm{W}i}^2$$

where $\hat{\sigma}_{\mathrm{W}i}^2$ is defined in Eq. 2.

Test for equality

For testing equality in intersubject variability, the following hypotheses are usually considered

$$\mathrm{H}_0 : \frac{\sigma_{\mathrm{BT}}^2}{\sigma_{\mathrm{BR}}^2} = 1 \quad \text{vs.} \quad \mathrm{H}_\mathrm{a} : \frac{\sigma_{\mathrm{BT}}^2}{\sigma_{\mathrm{BR}}^2} \neq 1 \tag{16}$$

Testing the above hypotheses is equivalent to testing the following hypotheses

$$\mathrm{H}_0 : \sigma_{\mathrm{BT}}^2 - \sigma_{\mathrm{BR}}^2 = 0 \quad \text{vs.} \quad \mathrm{H}_\mathrm{a} : \sigma_{\mathrm{BT}}^2 - \sigma_{\mathrm{BR}}^2 \neq 0 \tag{17}$$

Let $\eta = \sigma_{\mathrm{BT}}^2 - \sigma_{\mathrm{BR}}^2$. An intuitive estimator of η is given by

$$\hat{\eta} = \hat{\sigma}_{\mathrm{BT}}^2 - \hat{\sigma}_{\mathrm{BR}}^2$$

It follows that

$$\begin{aligned}\hat{\eta} &= \hat{\sigma}_{\mathrm{BR}}^2 - \hat{\sigma}_{\mathrm{BT}}^2 \\ &= s_{\mathrm{BT}}^2 - s_{\mathrm{BR}}^2 - \hat{\sigma}_{\mathrm{WT}}^2/m + \hat{\sigma}_{\mathrm{WR}}^2/m\end{aligned}$$

A $(1-\alpha)\times 100\%$ MLS confidence interval of η is given by

$$(\hat{\eta}_\mathrm{L}, \hat{\eta}_\mathrm{U})$$

where

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

and

$$\begin{aligned}\Delta_\mathrm{L} = {} & s_{\mathrm{BT}}^4 \left(1 - \frac{n_\mathrm{T}-1}{\chi_{\alpha/2,n_\mathrm{T}-1}^2}\right)^2 + s_{\mathrm{BR}}^4 \left(1 - \frac{n_\mathrm{R}-1}{\chi_{1-\alpha/2,n_\mathrm{R}-1}^2}\right)^2 \\ & + \frac{\hat{\sigma}_{\mathrm{WT}}^4}{m^2} \left(1 - \frac{n_\mathrm{T}(m-1)}{\chi_{1-\alpha/2,n_\mathrm{T}(m-1)}^2}\right)^2 \\ & + \frac{\hat{\sigma}_{\mathrm{WR}}^4}{m^2} \left(1 - \frac{n_\mathrm{R}(m-1)}{\chi_{\alpha/2,n_\mathrm{R}(m-1)}^2}\right)^2 \\ \Delta_\mathrm{U} = {} & s_{\mathrm{BT}}^4 \left(1 - \frac{n_\mathrm{T}-1}{\chi_{1-\alpha/2,n_\mathrm{T}-1}^2}\right)^2 + s_{\mathrm{BR}}^4 \left(1 - \frac{n_\mathrm{R}-1}{\chi_{\alpha/2,n_\mathrm{R}-1}^2}\right)^2 \\ & + \frac{\hat{\sigma}_{\mathrm{WT}}^4}{m^2} \left(1 - \frac{n_\mathrm{T}(m-1)}{\chi_{\alpha/2,n_\mathrm{T}(m-1)}^2}\right)^2 \\ & + \frac{\hat{\sigma}_{\mathrm{WR}}^4}{m^2} \left(1 - \frac{n_\mathrm{R}(m-1)}{\chi_{1-\alpha/2,n_\mathrm{R}(m-1)}^2}\right)^2\end{aligned}$$

We reject the null hypothesis at the α level of significance if $0 \notin (\hat{\eta}_\mathrm{L}, \hat{\eta}_\mathrm{U})$.

Test for noninferiority/superiority

Similar to testing intrasubject variabilities, the problem of testing noninferiority/superiority can be unified by the following hypotheses

$$\mathrm{H}_0 : \frac{\sigma_{\mathrm{BT}}^2}{\sigma_{\mathrm{BR}}^2} \geq \delta \quad \text{vs.} \quad \mathrm{H}_\mathrm{a} : \frac{\sigma_{\mathrm{BT}}^2}{\sigma_{\mathrm{BR}}^2} < \delta \tag{18}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Testing the above hypotheses is equivalent to testing the following hypotheses

$$\begin{aligned} &H_0: \sigma_{BT}^2 - \delta\sigma_{BR}^2 \geq 0 \quad \text{vs.} \\ &H_1: \sigma_{BT}^2 - \delta\sigma_{BT}^2 < 0 \end{aligned} \tag{19}$$

Define

$$\eta = \sigma_{BT}^2 - \delta\sigma_{BR}^2$$

For a given significance level α, similarly, the $(1-\alpha)\times 100\%$th MLS upper confidence bound of η can be constructed as

$$\hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

where Δ_U is given by

$$\begin{aligned} \Delta_U = {} & s_{BT}^4\left(1 - \frac{n_T - 1}{\chi^2_{1-\alpha,n_T-1}}\right)^2 + \delta^2 s_{BR}^4\left(1 - \frac{n_R - 1}{\chi^2_{\alpha,n_R-1}}\right)^2 \\ & + \frac{\hat{\sigma}_{WT}^4}{m^2}\left(1 - \frac{n_T(m-1)}{\chi^2_{\alpha,n_T(m-1)}}\right)^2 \\ & + \frac{\delta^2\hat{\sigma}_{WR}^4}{m^2}\left(1 - \frac{n_R(m-1)}{\chi^2_{1-\alpha,n_R(m-1)}}\right)^2 \end{aligned}$$

We then reject the null hypothesis at the α level of significance if $\hat{\eta}_U < 0$.

Test for similarity

Similarly, consider the following hypotheses for establishment of similarity in variability:

$$H_0: \frac{\sigma_{BT}^2}{\sigma_{BR}^2} \notin \left(\frac{1}{\delta}, \delta\right) \quad \text{vs.} \quad H_a: \frac{\sigma_{BT}^2}{\sigma_{BR}^2} \in \left(\frac{1}{\delta}, \delta\right) \tag{20}$$

where $\delta > 1$ is the similarity limit. The above hypotheses can be decomposed into the following two one-sided hypotheses:

$$\begin{aligned} &H_{01}: \sigma_{BT}^2 - \delta\sigma_{BR}^2 \geq 0 \quad \text{vs.} \quad H_{a1}: \sigma_{BT}^2 - \delta\sigma_{BR}^2 < 0 \\ &H_{02}: \delta\sigma_{BT}^2 - \sigma_{BR}^2 \leq 0 \quad \text{vs.} \quad H_{a2}: \delta\sigma_{BT}^2 - \sigma_{BR}^2 > 0 \end{aligned} \tag{21}$$

Similarity in intersubject variability between two treatments can be established if both of the above two ypotheses are rejected at the α level of significance. The test can be performed by constructing the $(1-\alpha)\times 100\%$ upper confidence bound of $\eta_1 = \sigma_{BT}^2 - \delta\sigma_{BR}^2$ and the $(1-\alpha)\times 100\%$ lower confidence bound of $\eta_2 = \delta\sigma_{BT}^2 - \sigma_{BR}^2$ by the MLS method. For example, the $(1-\alpha)\times 100\%$ lower confidence bound of η_2 is given by

$$\hat{\eta}_{2L} = \hat{\eta}_2 - \sqrt{\Delta_L}$$

where

$$\begin{aligned} \Delta_L = {} & \delta^2 s_{BT}^4\left(1 - \frac{n_T - 1}{\chi^2_{\alpha,n_T-1}}\right)^2 + s_{BR}^4\left(1 - \frac{n_R - 1}{\chi^2_{1-\alpha,n_R-1}}\right)^2 \\ & + \frac{\delta^2\hat{\sigma}_{WT}^4}{m^2}\left(1 - \frac{n_T(m-1)}{\chi^2_{1-\alpha,n_T(m-1)}}\right)^2 \\ & + \frac{\hat{\sigma}_{WR}^4}{m^2}\left(1 - \frac{n_R(m-1)}{\chi^2_{\alpha,n_R(m-1)}}\right)^2 \end{aligned}$$

Replicated Crossover Design

Under model Eq. 8 and new random variable $z_{ijk1} \equiv \bar{x}_{ijk}$ in Eq. 9, the intersubject sums of squares are defined by

$$\begin{aligned} s_{BT}^2 &= \frac{1}{n_1 + n_2 - 2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\bar{x}_{ijT.} - \bar{x}_{i.T.})^2 \\ s_{BR}^2 &= \frac{1}{n_1 + n_2 - 2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\bar{x}_{ijR.} - \bar{x}_{i.R.})^2 \end{aligned} \tag{22}$$

where

$$\bar{x}_{i.k.} = \frac{1}{n_i}\sum_{j=1}^{n_i}\bar{x}_{ijk.}$$

Note that $E(s_{Bk}^2) = \sigma_{Bk}^2 + \sigma_{Wk}^2/m$ for $k = T, R$. Therefore, the unbiased estimators for the intersubject variance are given by

$$\hat{\sigma}_{BT}^2 = s_{BT}^2 - \frac{1}{m}\hat{\sigma}_{WT}^2 \tag{23}$$

$$\hat{\sigma}_{BR}^2 = s_{BR}^2 - \frac{1}{m}\hat{\sigma}_{WR}^2 \tag{24}$$

where $\hat{\sigma}_{WT}^2$ and $\hat{\sigma}_{WR}^2$ are defined in Eq. 10.

Test for equality

For testing the equality in intersubject variability, the hypotheses in Eq. 17 are considered. Let $\eta = \sigma_{BT}^2 - \sigma_{BR}^2$. An intuitive estimator of η is given by

$$\hat{\eta} = \hat{\sigma}_{BT}^2 - \hat{\sigma}_{BR}^2$$

where $\hat{\sigma}^2_{BT}$ and $\hat{\sigma}^2_{BR}$ are given in Eqs. 23 and 24, respectively. It follows that

$$\begin{aligned}\hat{\eta} &= \hat{\sigma}^2_{BR} - \hat{\sigma}^2_{BT} \\ &= s^2_{BT} - s^2_{BR} - \frac{1}{m}\hat{\sigma}^2_{WT} + \frac{1}{m}\hat{\sigma}^2_{WR}\end{aligned}$$

Mean vector for the jth subject in ith sequence $(\bar{x}_{ijT}, \bar{x}_{ijR.})'$ has a bivariate normal distribution with covariance matrix given by

$$\mathbf{\Omega}_B = \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT}/m & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} + \sigma^2_{WR}/m \end{pmatrix} \tag{25}$$

The unbiased estimator of covariance matrix $\mathbf{\Omega}_B$ can be obtained as

$$\hat{\Omega}_B = \begin{pmatrix} s^2_{BT} & s^2_{BTR} \\ s^2_{BTR} & s^2_{BR} \end{pmatrix} \tag{26}$$

where

$$s^2_{BTR} = \frac{1}{n_1 + n_2 - 2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\bar{x}_{ijT.} - \bar{x}_{i.T.})(\bar{x}_{ijR.} - \bar{x}_{i.R.})$$

which is the sample covariance between $\bar{x}_{ijT}$ and $\bar{x}_{ijR}$. Let λ_i, $i = 1, 2$ be the two eigenvalues of the matrix $\Theta\Omega_B$where

$$\Theta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{27}$$

Hence, the population eigenvalues λ_i, $i = 1, 2$ can be estimated by the sample eigenvalues of $\Theta\hat{\Omega}_B$ and the explicit form of estimators is given as

$$\hat{\lambda}_i = \frac{s^2_{BT} - s^2_{BR} \pm \sqrt{(s^2_{BT} + s^2_{BR}) - 4s^4_{BTR}}}{2}$$

for $i = 1, 2$

Without loss of generality, it can be assumed that $\hat{\lambda}_1$. By the extension of the MLS method in Ref. [7], a $(1-\alpha)\times 100\%$ MLS confidence interval of η is given by

$$(\hat{\eta}_L, \hat{\eta}_U)$$

where

$$\hat{\eta}_L = \hat{\eta} - \sqrt{\Delta_L}, \quad \hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

and

$$\begin{aligned}\Delta_L &= \hat{\lambda}_1^2\left(1 - \frac{n_s - 1}{\chi^2_{\alpha/2,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1 - \frac{n_s - 1}{\chi^2_{1-\alpha/2,n_s-1}}\right)^2 \\ &\quad + \frac{\hat{\sigma}^4_{WT}}{m^2}\left(1 - \frac{n_s(m-1)}{\chi^2_{\alpha/2,n_s(m-1)}}\right)^2 \\ &\quad + \frac{\hat{\sigma}^4_{WR}}{m^2}\left(1 - \frac{n_s(m-1)}{\chi^2_{1-\alpha/2,n_s(m-1)}}\right)^2 \\ \Delta_U &= \hat{\lambda}_1^2\left(1 - \frac{n_s - 1}{\chi^2_{1-\alpha/2,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1 - \frac{n_s - 1}{\chi^2_{\alpha/2,n_s-1}}\right)^2 \\ &\quad + \frac{\hat{\sigma}^4_{WT}}{m^2}\left(1 - \frac{n_s(m-1)}{\chi^2_{1-\alpha/2,n_s(m-1)}}\right)^2 \\ &\quad + \frac{\hat{\sigma}^4_{WR}}{m^2}\left(1 - \frac{n_s(m-1)}{\chi^2_{\alpha/2,n_s(m-1)}}\right)^2\end{aligned}$$

where $n_s = n_1 + n_2 - 2$. Then, we reject the null hypothesis at the α level of significance if $0 \notin (\hat{\eta}_L, \hat{\eta}_U)$.

Test for noninferiority/superiority

Similar to testing intrasubject variabilities, the problem of testing noninferiority/superiority can be unified by the hypotheses described in Eq. 19. For a given significance level of α, the $(1-\alpha)\times 100\%$ MLS upper confidence bound of $\eta = \sigma^2_{BT} - \delta\sigma^2_{BR}$ can be constructed as

$$\hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

where Δ_U is given by

$$\begin{aligned}\Delta_U &= \hat{\lambda}_1^2\left(1 - \frac{n_s - 1}{\chi^2_{1-\alpha,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1 - \frac{n_s - 1}{\chi^2_{\alpha,n_s-1}}\right)^2 \\ &\quad + \frac{\hat{\sigma}^4_{WT}}{m^2}\left(1 - \frac{n_s(m-1)}{\chi^2_{1-\alpha,n_s(m-1)}}\right)^2 \\ &\quad + \frac{\delta^2\hat{\sigma}^4_{WR}}{m^2}\left(1 - \frac{n_s(m-1)}{\chi^2_{\alpha,n_s(m-1)}}\right)^2\end{aligned}$$

where $n_s = n_1 + n_2 - 2$ and $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ are the eigenvalues of $\Theta\hat{\Omega}_B$ with

$$\Theta = \begin{pmatrix} 1 & 0 \\ 0 & -\delta \end{pmatrix} \tag{28}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Test for similarity

Consider the hypotheses in Eq. 21 for establishment of similarity in intersubject variability. Similarity between two treatments can be established if both of the two hypotheses in Eq. 21 are rejected at the α level of significance. The test can be performed by constructing the $(1-\alpha)\times 100\%$ upper confidence bound of $\eta_1 = \sigma_{\mathrm{BT}}^2 - \delta\sigma_{\mathrm{BR}}^2$ and the $(1-\alpha)\times 100\%$ lower confidence bound of $\eta_2 = \delta\sigma_{\mathrm{BT}}^2 - \sigma_{\mathrm{BR}}^2$ by the MLS method. For example, the $(1-\alpha)\times 100\%$ lower confidence bound of η_2 is given by

$$\hat{\eta}_{2\mathrm{L}} = \hat{\eta}_2 - \sqrt{\Delta_{\mathrm{L}}}$$

where

$$\Delta_{\mathrm{L}} = \hat{\lambda}_1^2\left(1-\frac{n_s-1}{\chi^2_{\alpha,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_s-1}{\chi^2_{1-\alpha,n_s-1}}\right)^2 + \frac{\delta^2\hat{\sigma}_{\mathrm{WT}}^4}{m^2} \times\left(1-\frac{n_s(m-1)}{\chi^2_{\alpha,n_s(m-1)}}\right)^2 + \frac{\hat{\sigma}_{\mathrm{WR}}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{1-\alpha,n_s(m-1)}}\right)^2$$

where $n_s = n_1+n_2-2$ and $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ are the eigenvalues of $\Theta\hat{\Omega}_{\mathrm{B}}$ with

$$\Theta = \begin{pmatrix}\delta & 0\\ 0 & -1\end{pmatrix} \tag{29}$$

COMPARING TOTAL VARIABILITIES

In practice, it may be also of interest to compare total variabilities between drug products. Total variability is related to the drug's prescribability,[10] which is the main concept for population bioequivalence. Because the total variability can be estimated under parallel/crossover designs both with and without replicates, we will consider both designs with and without replicates.

Parallel Designs Without Replicates

The following model can be used for a parallel design without replicates

$$x_{ij} = \mu_i + \epsilon_{ij}$$

where x_{ij} is the observation from the jth subject in the ith treatment group. Also we assume that the random variable ϵ_{ij} has normal distribution with mean 0 and variance $\sigma_{\mathrm{T}i}^2$ for i = T, R. Hence, the total variability can be estimated by

$$\hat{\sigma}_{\mathrm{T}i}^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2$$

where

$$\bar{x}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

Test for equality

For testing equality in total variability, the following hypotheses are considered:

$$\mathrm{H}_0: \frac{\sigma_{\mathrm{TT}}^2}{\sigma_{\mathrm{TR}}^2} = 1 \quad \text{vs.} \quad \mathrm{H}_{\mathrm{a}}: \frac{\sigma_{\mathrm{TT}}^2}{\sigma_{\mathrm{TR}}^2} \neq 1 \tag{30}$$

Under the null hypothesis, the test statistic

$$T = \frac{\hat{\sigma}_{\mathrm{TT}}^2}{\hat{\sigma}_{\mathrm{TR}}^2}$$

is distributed as an F random variable with $n_{\mathrm{T}}-1$ and $n_{\mathrm{R}}-1$ degrees of freedom. Hence, we reject the null hypothesis at the α level of significance if

$$T > F_{\alpha/2,n_{\mathrm{T}},n_{\mathrm{R}}}$$

or

$$T < F_{1-\alpha/2},n_{\mathrm{T}},n_{\mathrm{R}}$$

Test for noninferiority/superiority

Similarly, the problem of testing noninferiority and superiority can be unified by the following hypotheses

$$\mathrm{H}_0: \frac{\sigma_{\mathrm{TT}}^2}{\sigma_{\mathrm{TR}}^2} \geq \delta \quad \text{vs.} \quad \mathrm{H}_{\mathrm{a}}: \frac{\sigma_{\mathrm{TT}}^2}{\sigma_{\mathrm{TR}}^2} < \delta \tag{31}$$

The test statistic is given by

$$T = \frac{s_{\mathrm{TT}}^2}{\delta s_{\mathrm{TR}}^2}$$

Under the null hypothesis, T is distributed as an F random variable with $n_{\mathrm{T}}-1$ and $n_{\mathrm{R}}-1$ degrees of freedom. Hence, we reject the null hypothesis at the α level of significance if

$$T < F_{1-\alpha,n_{\mathrm{R}},n_{\mathrm{T}}}$$

Test for similarity

For testing similarity, the hypotheses of interest are given by

$$\mathrm{H}_0: \frac{\sigma_{\mathrm{TT}}^2}{\sigma_{\mathrm{TR}}^2} \notin \left(\frac{1}{\delta},\delta\right) \quad \text{vs.} \quad \mathrm{H}_{\mathrm{a}}: \frac{\sigma_{\mathrm{TT}}^2}{\sigma_{\mathrm{TR}}^2} \in \left(\frac{1}{\delta},\delta\right) \tag{32}$$

where $\delta > 1$ is the similarity limit. The above hypotheses can be decomposed into the following two one-sided hypotheses

$$H_{01}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \geq \delta \quad \text{vs.} \quad H_{a1}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} < \delta$$
$$H_{02}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \leq \frac{1}{\delta} \quad \text{vs.} \quad H_{a2}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} > \frac{1}{\delta} \tag{33}$$

These two hypotheses can be tested by the following two test statistics

$$T_1 = \frac{s_{TT}^2}{\delta s_{TR}^2} \quad \text{and} \quad T_2 = \frac{\delta s_{TT}^2}{s_{TR}^2}$$

We then reject the null hypothesis and conclude similarity at the α level of significance if

$$T_1 < F_{1-\alpha,n_T,n_R} \quad \text{and} \quad T_2 > F_{\alpha,n_T,n_R}$$

Parallel Design with Replicates

In practice, a parallel design with replicates can also be used to assess total variability. Although total variabilities can be assessed by the design without replicates, there is a merit of the design with replicates. Because intersubject and intrasubject variabilities can be assessed under a parallel design with replicates, it can serve more than just one purpose. If we consider model Eq. 1, an unbiased estimator for total variabilities is given by

$$\hat{\sigma}_{Ti}^2 = s_{Bi}^2 + \frac{m-1}{m}\hat{\sigma}_{Wi}^2$$

where ${s_{Bi}}^2$ is defined in Eq. 15 and σ_{Wi}^2 is defined in Eq. 2. Let $\eta = \sigma_{TT}^2 - \sigma_{TR}^2$. Then, a natural estimator for η is given by

$$\hat{\eta} = \hat{\sigma}_{TT}^2 - \hat{\sigma}_{TR}^2$$

Test for equality

For testing the equality in total variability, consider the hypotheses in Eq. 30, which is equivalent to testing the following hypotheses:

$$H_0: \sigma_{TT}^2 - \sigma_{TR}^2 = 0 \quad \text{vs.} \quad H_a: \sigma_{TT}^2 - \sigma_{TR}^2 \neq 0 \tag{34}$$

A $(1-\alpha)\times 100\%$ MLS confidence interval of η is given by

$$(\hat{\eta}_L, \hat{\eta}_U)$$

where

$$\hat{\eta}_L = \hat{\eta} - \sqrt{\Delta_L}, \quad \hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

and

$$\Delta_L = s_{BT}^4\left(1 - \frac{n_T - 1}{\chi^2_{\alpha/2,n_T-1}}\right)^2 + s_{BR}^4\left(1 - \frac{n_R - 1}{\chi^2_{1-\alpha/2,n_R-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WT}^4}{m^2}\left(1 - \frac{n_T(m-1)}{\chi^2_{1-\alpha/2,n_T(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1 - \frac{n_R(m-1)}{\chi^2_{\alpha/2,n_R(m-1)}}\right)^2$$

$$\Delta_U = s_{BT}^4\left(1 - \frac{n_T - 1}{\chi^2_{1-\alpha/2,n_T-1}}\right)^2 + s_{BR}^4\left(1 - \frac{n_R - 1}{\chi^2_{\alpha/2,n_R-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WT}^4}{m^2}\left(1 - \frac{n_T(m-1)}{\chi^2_{\alpha/2,n_T(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1 - \frac{n_R(m-1)}{\chi^2_{1-\alpha/2,n_R(m-1)}}\right)^2$$

We reject the null hypothesis at the α level of significance if $0 \notin (\hat{\eta}_L, \hat{\eta}_U)$.

Test for noninferiority/superiority

Similar to testing intrasubject variabilities, the problem of testing noninferiority/superiority can be unified by the hypotheses described in Eq. 31. Testing the hypotheses in Eq. 31 is equivalent to testing the following hypotheses:

$$H_0: \sigma_{TT}^2 - \delta\sigma_{TR}^2 \geq 0 \quad \text{vs.}$$
$$H_a: \sigma_{TT}^2 - \delta\sigma_{TR}^2 < 0 \tag{35}$$

Let $\eta = \sigma_{TT}^2 - \delta\sigma_{TR}^2$. For a given significance level of α, similarly, the $(1-\alpha)100\%$ upper confidence bound of η can be constructed as

$$\hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

where $\hat{\eta} = \hat{\sigma}_{TT}^2 - \delta\hat{\sigma}_{TR}^2$ and Δ_U is given by

$$\Delta_U = s_{BT}^4\left(1 - \frac{n_T - 1}{\chi^2_{1-\alpha,n_T-1}}\right)^2 + \delta^2 s_{BR}^4\left(1 - \frac{n_R - 1}{\chi^2_{\alpha/2,n_R-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WT}^4}{m^2}\left(1 - \frac{n_T(m-1)}{\chi^2_{\alpha,n_T(m-1)}}\right)^2 + \frac{\delta^2(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1 - \frac{n_R(m-1)}{\chi^2_{1-\alpha,n_R(m-1)}}\right)^2$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

We then reject the null hypothesis at the α level of significance if $\hat{\eta}_U < 0$.

Test for similarity

For testing similarity, the hypotheses of interest are described in Eq. 33. We then reject the null hypothesis and conclude similarity at the α level of significance if both hypotheses in Eq. 33 are rejected with significance level α. The test can be performed by constructing the $(1-\alpha)\times 100\%$ upper confidence bound of $\eta_1 = \sigma_{TT}^2 - \delta\sigma_{TR}^2$ and the $(1-\alpha)\times 100\%$ lower confidence bound of $\eta_2 = \delta\sigma_{TT}^2 - \sigma_{TR}^2$ by the MLS method. For example, the $(1-\alpha)\times 100\%$ lower confidence bound of η_2 is given by

$$\hat{\eta}_{2L} = \hat{\eta}_2 - \sqrt{\Delta_L}$$

where

$$\Delta_L = \delta^2 s_{BT}^4\left(1-\frac{n_T-1}{\chi^2_{\alpha,n_T-1}}\right)^2 + s_{BR}^4\left(1-\frac{n_R-1}{\chi^2_{1-\alpha,n_R-1}}\right)^2 + \frac{(m-1)^2\delta^2\hat{\sigma}_{WT}^4}{m^2}\left(1-\frac{n_T(m-1)}{\chi^2_{1-\alpha,n_T(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1-\frac{n_R(m-1)}{\chi^2_{\alpha,n_R(m-1)}}\right)^2$$

The Standard 2×2 Crossover Design

For the standard 2×2 crossover design, model Eq. 8 is still useful if the index for replicates l is omitted (i.e., $m = 1$). Unbiased estimators for the total variabilities are given by

$$\hat{\sigma}_{TT}^2 = s_{BT}^2 = \frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ijT}-\bar{x}_{i\cdot T})^2$$

$$\hat{\sigma}_{TR}^2 = s_{BR}^2 = \frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ijR}-\bar{x}_{i\cdot R})^2$$

where

$$\bar{x}_{i\cdot T} = \frac{1}{n_i}\sum_{j=1}^{n_i}x_{ijT} \quad \text{and} \quad \bar{x}_{i\cdot R} = \frac{1}{n_i}\sum_{j=1}^{n_i}x_{ijR}$$

Test for equality

For testing the equality in total variability, consider the hypotheses in Eq. 34. Let $\eta = \sigma_{TT}^2 - \sigma_{TR}^2$. An intuitive estimator of η is given by

$$\hat{\eta} = \hat{\sigma}_{TT}^2 - \hat{\sigma}_{TR}^2$$

Let

$$s_{BTR}^2 = \frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ijT}-\bar{x}_{i\cdot T})(x_{ijR}-\bar{x}_{i\cdot R})$$

Let $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ be the two eigenvalues of the matrix $\Theta\hat{\Omega}_B$ where $\hat{\Omega}_B$ is defined in Eq. 26 and Θ is given in Eq. 27. A $(1-\alpha)\times 100\%$ MLS confidence interval of η is given by

$$(\hat{\eta}_L, \hat{\eta}_U)$$

where

$$\hat{\eta}_L = \hat{\eta} - \sqrt{\Delta_L}, \quad \hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

and

$$\Delta_L = \hat{\lambda}_1^2\left(1-\frac{n_1+n_2-2}{\chi^2_{1-\alpha/2,n_1+n_2-2}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_1+n_2-2}{\chi^2_{\alpha/2,n_1+n_2-2}}\right)^2$$

$$\Delta_U = \hat{\lambda}_1^2\left(1-\frac{n_1+n_2-2}{\chi^2_{\alpha/2,n_1+n_2-2}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_1+n_2-2}{\chi^2_{1-\alpha/2,n_1+n_2-2}}\right)^2$$

Test for noninferiority/superiority

Similar to testing intrasubject variabilities, consider the hypotheses in Eq. 35. For a given significance level of α, similarly, the $(1-\alpha)\times 100\%$ upper confidence bound of $\eta = \sigma_{TT}^2 - \delta\,\sigma_{TR}^2$ can be constructed as

$$\hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

where Δ_U is given by

$$\Delta_U = \hat{\lambda}_1^2\left(\frac{n_1+n_2-2}{\chi^2_{\alpha,n_1+n_2-2}}-1\right)^2 + \hat{\lambda}_2^2\left(\frac{n_1+n_2-2}{\chi^2_{1-\alpha,n_1+n_2-2}}-1\right)^2$$

and $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ be the two eigenvalues of the matrix $\Theta\hat{\Omega}_B$ where Θ is given in Eq. 28. We then reject the null hypothesis at the α level of significance if $\eta_U < 0$.

Test for similarity

For testing similarity, the hypotheses of interest are described in Eq. 33. We then reject the null hypothesis and conclude similarity at the α level of significance if both hypotheses in Eq. 33 are rejected with significance

level α. The test can be performed by constructing the $(1-\alpha)\times 100\%$ upper confidence bound of $\eta_1 = \sigma_{TT}^2 - \delta\sigma_{TR}^2$ and the $(1-\alpha)\times 100\%$ lower confidence bound of $\eta_2 = \sigma_{TT}^2 - \delta\sigma_{TR}^2$ by the MLS method. For example, the $(1-\alpha)\times 100\%$ lower confidence bound of η_2 is given by

$$\hat{\eta}_{2L} = \hat{\eta}_2 + \sqrt{\Delta_L}$$

where

$$\Delta_L = \hat{\lambda}_1^2\left(1-\frac{n_1+n_2-2}{\chi^2_{1-\alpha,n_1+n_2-2}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_1+n_2-2}{\chi^2_{\alpha,n_1+n_2-2}}\right)^2$$

where $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ are the two eigenvalues of the matrix $\Theta\hat{\Omega}_B$ where $\mathbf{\Theta}$ is given in Eq. 29.

Replicated $2\times 2m$ Crossover Design

We can use a similar argument for the test of intersubject variabilities under model Eq. 8 with the estimators, for $k = T, R$,

$$\hat{\sigma}_{Tk}^2 = s_{Bk}^2 + \frac{m-1}{m}\hat{\sigma}_{Wk}^2$$

where $\hat{s}_{Bk}^2$ and $\hat{\sigma}_{Wk}^2$ are defined in Eq. 22 and Eq. 10, respectively.

Test for equality

For testing the equality in total variability, consider the hypotheses in Eq. 34. An intuitive estimator of $\eta = \sigma_{TT}^2 - \sigma_{TR}^2$ is given by $\hat{\eta} = \hat{\sigma}_{TT}^2 - \hat{\sigma}_{TR}^2$. Let $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ be the two eigenvalues of the matrix $\Theta\hat{\Omega}_B$ where $\mathbf{\Theta}$ is given in Eq. 27. By Lee et al.,[7] a $(1-\alpha)\times 100\%$ confidence interval of η is given by

$$(\hat{\eta}_L, \hat{\eta}_U)$$

where

$$\hat{\eta}_L = \hat{\eta} - \sqrt{\Delta_L}, \quad \hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

and

$$\Delta_L = \hat{\lambda}_1^2\left(1-\frac{n_s-1}{\chi^2_{1-\alpha/2,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_s-1}{\chi^2_{\alpha/2,n_s-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WT}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{\alpha/2,n_s(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{1-\alpha/2,n_s(m-1)}}\right)^2$$

$$\Delta_U = \hat{\lambda}_1^2\left(1-\frac{n_s-1}{\chi^2_{\alpha/2,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_s-1}{\chi^2_{1-\alpha/2,n_s-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WT}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{1-\alpha/2,n_s(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{\alpha/2,n_s(m-1)}}\right)^2$$

We reject the null hypothesis at the α level of significance if $0 \notin (\hat{\eta}_L, \hat{\eta}_U)$.

Test for noninferiority/superiority

Consider a test of the hypotheses in Eq. 35. Let $\eta = \sigma_{TT}^2 - \delta\sigma_{TR}^2$ For a given significance level of α, similarly, the $(1-\alpha)\times 100\%$ upper confidence bound of η can be constructed as

$$\hat{\eta}_U = \hat{\eta} + \sqrt{\Delta_U}$$

where $\hat{\eta} = \hat{\sigma}_{TT}^2 - \delta\sigma_{TR}^2$ and Δ_U is given by

$$\Delta_U = \hat{\lambda}_1^2\left(1-\frac{n_s-1}{\chi^2_{\alpha,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_s-1}{\chi^2_{1-\alpha,n_s-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WT}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{1-\alpha,n_s(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}_{WR}^4}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{\alpha,n_s(m-1)}}\right)^2$$

where $n_s = n_1+n_2-2$ and $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ are the eigenvalues of $\Theta\hat{\Omega}_B$ with $\mathbf{\Theta}$ given in Eq. 28. We then reject the null hypothesis at the α level of significance if $\hat{\eta}_U$.

Test for similarity

For testing similarity, the hypotheses of interest are described in Eq. 33. We then reject the null hypothesis and conclude similarity at the α level of significance if both hypotheses in Eq. 33 are rejected with significance level α. The test can be performed by constructing the $(1-\alpha)\times 100\%$ upper confidence bound of $\eta_1 = \sigma_{TT}^2 - \delta\sigma_{TR}^2$ and the $(1-\alpha)\times 100\%$ lower confidence bound of $\eta_2 = \delta\sigma_{TT}^2 - \sigma_{TR}^2$ by the MLS method. For example, the $(1-\alpha)\times 100\%$ lower confidence bound of η_2 is given by

$$\hat{\eta}_{2L} = \hat{\eta}_2 - \sqrt{\Delta_L}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

where

$$\Delta_L = \hat{\lambda}_1^2\left(1-\frac{n_s-1}{\chi^2_{1-\alpha,n_s-1}}\right)^2 + \hat{\lambda}_2^2\left(1-\frac{n_s-1}{\chi^2_{\alpha,n_s-1}}\right)^2 + \frac{(m-1)^2\hat{\sigma}^4_{WT}}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{\alpha,n_s(m-1)}}\right)^2 + \frac{(m-1)^2\hat{\sigma}^4_{WR}}{m^2}\left(1-\frac{n_s(m-1)}{\chi^2_{1-\alpha,n_s(m-1)}}\right)^2$$

where $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ be the two eigenvalues of the matrix $\Theta\hat{\Omega}_B$ where Θ is given in Eq. 29.

SAMPLE SIZE CALCULATION

For clinical scientists/biostatisticians, a great concern is how many subjects should be recruited in the clinical trial to achieve the desired power of statistical test, which is used to show equality, superiority, or similarity of the test formulation. If an explicit power function of a test is not available or difficult to obtain, sample size calculation depends on asymptotic power, which is usually obtained by using the central limit theorem. Also, some assumption about unknown parameters should be made to calculate sample size. These assumptions may come from previous or related clinical studies that can provide reasonable estimates for unknown parameters. If information about parameters is not available, it is common to assume the worst-case scenario so that sample size is big enough to assure that a clinical study will not fail to achieve its objective. In practice, however, it is difficult to obtain a large sample size because a clinical study is restricted by time and budget. Here, we introduce some statistical techniques to calculate sample size for comparing variabilities to achieve a desired power under some assumption about unknown parameters.

Sample Size for *F*-Test: Comparing Intrasubject Variabilities

If a test is based on F distribution, the explicit power function can be obtained so that fairly accurate sample size can be provided. Also, for most cases, the target sample size is available as an implicit solution of the equation so that no explicit formula for sample size is available. However, the target sample size can be obtained by numerical methods such as the grid search method or Newton–Rhapson algorithm.

First, consider the hypotheses in Eq. 4 for test for equality of intrasubject variabilities under a parallel design with replicates in ''Parallel Design with Replicates'' under the section ''Comparing Intrasubject Variabilities.'' Under the alternative hypothesis, without loss of generality, we assume that $\sigma^2_{WT} < \sigma^2_{WR}$. The power of the test can then be approximated by

$$\begin{aligned}1-\beta &= P(T < F_{1-\alpha/2,n_T(m-1),n_R(m-1)})\\ &= P(1/T > F_{\alpha/2,n_R(m-1),n_T(m-1)})\\ &= P\left(\frac{\hat{\sigma}^2_{WR}/\sigma^2_{WR}}{\hat{\sigma}^2_{WT}/\sigma^2_{WT}} > \frac{\sigma^2_{WT}}{\sigma^2_{WR}}F_{\alpha/2,n_R(m-1),n_T(m-1)}\right)\\ &= P\left(F_{(n_R(m-1),n_T(m-1))} > \frac{\sigma^2_{WT}}{\sigma^2_{WR}}F_{\alpha/2,n_R(m-1),n_T(m-1)}\right)\end{aligned}$$

where $F_{(a,b)}$ denotes an F random variable with a and b degrees of freedom. Under the assumption that $n = n_R = n_T$ and with a fixed value for m, σ^2_{WT}, and σ^2_{WR}, the sample size needed to achieve a desired power of $1-\beta$ can be obtained by solving the following equation for n:

$$\frac{\sigma^2_{WT}}{\sigma^2_{WR}} = \frac{F_{1-\beta,n(m-1),n(m-1)}}{F_{\alpha/2,n(m-1),n(m-1)}}$$

Next, consider the hypotheses in Eq. 6 for the test for similarity of intrasubject variabilities under a parallel design with replicates. Assuming that $n = n_T = n_R$, under the alternative hypothesis that $\sigma_{WT}{}^2 = \sigma WR^2$, the power of the test can be approximated by

$$\begin{aligned}1-\beta &= P\left(\frac{F_{\alpha,n(m-1),n(m-1)}}{\delta} < \frac{\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}} < \delta F_{1-\alpha,n(m-1),n(m-1)}\right)\\ &= P\left(\frac{1}{F_{\alpha,n(m-1),n(m-1)}\delta} < \frac{\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}} < \delta F_{1-\alpha,n(m-1),n(m-1)}\right)\\ &= 1-2P\left(\frac{\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}} > \delta F_{1-\alpha,n(m-1),n(m-1)}\right)\\ &= 1-2P\left(F_{(n(m-1),n(m-1))} > \delta F_{1-\alpha,n(m-1),n(m-1)}\right)\end{aligned}$$

Thus, the sample size required for achieving a desired power of $1-\beta$ can be obtained by solving the following equation for n

$$\delta = \frac{F_{\beta/2,n(m-1),n(m-1)}}{F_{1-\alpha,n(m-1),n(m-1)}}$$

Note that under the alternative hypothesis, it is also possible that $\sigma^2_{WT} \neq \sigma^2_{WR}$. Without loss of generality, we assume that $\sigma^2_{WT}/\sigma^2_{WR} = r \in (1, \delta)$. In this case, the power of the above test can be approximated by

$$\begin{aligned} 1-\beta &= P\left(\frac{F_{\alpha,n(m-1),n(m-1)}}{\delta} < \frac{\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}} < \delta F_{1-\alpha,n(m-1),n(m-1)} \right) \\ &\approx P\left(\frac{\hat{\sigma}^2_{WT}}{\hat{\sigma}^2_{WR}} < \delta F_{1-\alpha,n(m-1),n(m-1)} \right) \\ &= P\left(\frac{\hat{\sigma}^2_{WT}}{r\hat{\sigma}^2_{WR}} < \frac{\delta}{r} F_{1-\alpha,n(m-1),n(m-1)} \right) \\ &= P\left(F_{(n(m-1),n(m-1))} < \frac{\delta}{r} F_{1-\alpha,n(m-1),n(m-1)} \right) \end{aligned}$$

Hence, the sample size needed to achieve the desired power of $1-\beta$ can be obtained by solving the following equation for n:

$$\frac{\delta}{r} = \frac{F_{\beta,n(m-1),n(m-1)}}{F_{1-\alpha,n(m-1),n(m-1)}}$$

Sample Size for *Z*-Test: Comparing Intrasubject Coefficients of Variation

If a test is already based on the normal approximation of the test statistic, then sample size can be easily obtained by simple calculation of power function, which is based on the normal distribution. In this case, the explicit form of target sample size is usually available.

Here, we consider the test for comparing intrasubject CVs under the section ''Simple Random Effects Model.'' First, consider the hypothesis in Eq. 11 for the test of equality of intrasubject CVs. Under the alternative hypothesis, without loss of generality, it is assumed that $CV_T > CV_R$. The distribution of test statistics can be approximated by a normal distribution with unit variance and mean

$$\frac{CV_T - CV_R}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}}$$

Thus, the power is given by

$$\begin{aligned} 1-\beta &= P(|T| > z_{\alpha/2}) \approx P(T > z_{\alpha/2}) \\ &= P\left(N(0,1) > z_{\alpha/2} - \frac{CV_T - CV_R}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} \right) \end{aligned}$$

Under the assumption that $n = n_1 = n_2$, the sample size needed to have a power of $1-\beta$ can be obtained by solving the following equation

$$z_{\alpha/2} - \frac{CV_T - CV_R}{\sqrt{\sigma^{*2}_T/n + \sigma^{*2}_R/n}} = -z_\beta$$

This leads to

$$n = \frac{(\sigma^{*2}_T + \sigma^{*2}_R)(z_{\alpha/2} + z_\beta)^2}{(CV_T - CV_R)^2}$$

Next, consider the hypothesis in Eq. 13 for the test of similarity of intrasubject CVs under simple random effects model. Under the alternative hypothesis that $CV_T = CV_R$, the power of the test procedure is given by

$$\begin{aligned} 1-\beta &= P\left(z_\alpha - \frac{\delta}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} < N(0,1) < \frac{\delta}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} - z_\alpha \right) \\ &= 1 - 2P\left(N(0,1) > \frac{\delta}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} - z_\alpha \right) \end{aligned}$$

Hence, under the assumption that $n = n_1 = n_2$, the sample size needed to achieve $1-\beta$ power at the α level of significance can be obtained by solving the following equation

$$\frac{\delta}{\sqrt{\sigma^2_1/n + \sigma^2_2/n}} - z_\alpha = z_{\beta/2}$$

This leads to

$$n = \frac{(z_\alpha + z_{\beta/2})^2(\sigma^2_1 + \sigma^2_2)}{\delta^2}$$

Under the alternative, without loss of generality, we assume that $0 < CV_T - CV_R < \delta$. Then, the power of the test procedure can be approximated by

$$\begin{aligned} 1-\beta &= P\left(z_\alpha - \frac{\delta + |CV_T - CV_R|}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} < N(0,1) < \frac{\delta - |CV_T - CV_R|}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} - z_\alpha \right) \\ &\approx P\left(N(0,1) < \frac{\delta - |CV_T - CV_R|}{\sqrt{\sigma^{*2}_T/n_T + \sigma^{*2}_R/n_R}} - z_\alpha \right) \end{aligned}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Hence, under the assumption that $n = n_1 = n_2$, the sample size needed to achieve $1-\beta$ power at the α level of significance can be obtained by solving the following equation

$$\frac{\delta - |\mathrm{CV_T} - \mathrm{CV_R}|}{\sqrt{\sigma_T^{*2}/n_T + \sigma_R^{*2}/n_R}} - z_\alpha = z_\beta$$

This gives

$$n = \frac{(z_\alpha + z_\beta)^2(\sigma_T^{*2} + \sigma_R^{*2})}{(\delta - |\mathrm{CV_T} - \mathrm{CV_R}|)^2}$$

Sample Size for Modified Large Sample Method: Comparing Intersubject Variabilities

If a test is based on the MLS method, the explicit power function is difficult to calculate so that approximated power function should be obtained by using CLT to find the target sample size. The crucial step for normal approximation is the calculation of the asymptotic variance of a test statistic. The asymptotic variance of test statistics based on the MLS method can be easily calculated because estimators of variance components are usually distributed as chi-square distribution. When estimators of variance components are not independent, the asymptotic variance can be easily obtained via the eigenvalues of $\mathbf{\Theta\Omega}_B$ where $\mathbf{\Omega}_B$ is a 2×2 intersubject covariance matrix and $\mathbf{\Theta}$ is some 2×2 matrix.

We will consider the test of intersubject variabilities in ''Parallel Design Replicates'' under the section ''Comparing Intersubject Variabilities.'' Consider the hypothesis in Eq. 11 for the test of equality of intersubject variabilities. Under the alternative hypothesis, without loss of generality, we assume that $\sigma_{BR}^2 \sigma_{BT}^2$ and $n = n_T = n_R$. Thus, the power of the test procedure can be approximated by

$$P\left(N\left(\frac{\sqrt{n}(\sigma_{BR}^2 - \sigma_{BT}^2)}{\sigma^*}, 1\right) > z_{\alpha/2}\right)$$

where asymptotic variance σ^* of the test statistic is given by

$$\sigma^{*2} = 2\left[\left(\sigma_{BT}^2 + \frac{\sigma_{WT}^2}{m}\right)^2 + \left(\sigma_{BR}^2 + \frac{\sigma_{WR}^2}{m}\right)^2 + \frac{\sigma_{WT}^4}{m^2(m-1)} + \frac{\sigma_{WR}^4}{m^2(m-1)}\right]$$

Because a parallel design is used, the estimators of variance components are independent so that asymptotic variance σ^{*2} can be obtained by using variance of chi-square random variable. As a result, the sample size needed to achieve the desired power of $1-\beta$ at the α level of significance can be obtained by solving the following equation

$$z_{\alpha/2} - \frac{\sqrt{n}(\sigma_{BT}^2 - \sigma_{BR}^2)}{\sigma^*} = -z_\beta$$

This leads to

$$n = \frac{\sigma^{*2}(z_{\alpha/2} + z_\beta)^2}{(\sigma_{BT}^2 - \sigma_{BR}^2)^2}$$

Now consider the test of equality of total variabilities under the section ''Standard 2×2 Crossover Design.'' Here, the test is based on the MLS method and estimators of variance components are not independent. Under the alternative hypothesis and letting $n = n_1 + n_2 - 2$, without loss of generality, we assume $\sigma_{TR}^2 > \sigma_{TT}^2$. The power of the above test can be determined by

$$P\left(N\left(\frac{\sqrt{n}(\sigma_{TT}^2 - \sigma_{TR}^2)}{\sigma^*}, 1\right) > z_{\alpha/2}\right)$$

where

$$\sigma^{*2} = 2(\lambda_1^2 + \lambda_2^2) \equiv 2(\sigma_{TT}^4 + \sigma_{TR}^- 2\rho^2\sigma_{BT}^2\sigma_{BR}^2)$$

where let $\lambda_1 < 0 < \lambda_2$ are the two eigenvalues of the matrix $\mathbf{\Theta\Omega}_B$ where $\mathbf{\Omega}_B$ is defined in Eq. 25 and $\mathbf{\Theta}$ is given in Eq. 27.

Hence, the sample size needed to achieve the power of $1-\beta$ at the α level of significance can be obtained by solving the following equation

$$z_{\alpha/2} = -\frac{\sqrt{n}(\sigma_{TT}^2 - \sigma_{TR}^2)}{\sigma^*} = -z_\beta$$

which implies that

$$n = \frac{\sigma^{*_2}(z_{\alpha/2} + z_\beta)^2}{(\sigma_{TT}^2 - \sigma_{TR}^2)^2}$$

CONCLUSION

For comparing intrasubject variabilities and/or intrasubject CVs between treatment groups, replicates from the same subject are essential regardless of the study design. In clinical research, data are often log-transformed before the analysis. It should be noted that the intrasubject standard deviation of log-transformed data is approximately equal to the intrasubject CV of the untransformed (raw) data. As a result, it is suggested that intrasubject

variability be used when analyzing log-transformed data and the intrasubject CV be considered when analyzing untransformed data.

In recent years, the assessment of reproducibility in terms of intrasubject variability or intrasubject CV in clinical research has received much attention. Chow[11] defined reproducibility of a study drug as a collective term that encompasses consistency, similarity, and stability (control) within the therapeutic index (or window) of a subject's clinical status (e.g., clinical response of some primary study endpoint, blood levels, or blood concentration–time curve) when the study drug is repeatedly administered at different dosing periods under the same experimental condition. Reproducibility of clinical results observed from a clinical study can be quantitated through the evaluation of the so-called reproducibility probability, which will be briefly introduced.[12]

For assessment of intersubject variability and/or total variability, Chow and Tse[13] indicated that the usual analysis of variance model could lead to negative estimates of the variance components, especially the intersubject variance component. In addition, the sum of the best estimates of the intrasubject variance and the intersubject variance may not lead to the best estimate for the total variance. Chow and Shao[14] proposed an estimation procedure for variance components, which cannot only avoid negative estimates but can also provide a better estimate than the maximum likelihood estimates. For estimation of total variance, Chow and Tse[13] proposed a method as an alternative to the sum of estimates of individual variance components. These ideas could be applied to provide a better estimate of sample sizes for studies intended for comparing variabilities between treatment groups.

REFERENCES

1. Chinchilli, V.M.; Esinhart, J.D. Design and analysis of intra-subject variability in cross-over design. Stat. Med. **1996**, *15*, 1619–1634.
2. Howe, W.G. Approximate confidence limits on the mean of $x+y$ where x and y are two tabled independent random variables. J. Am. Stat. Assoc. **1974**, *69*, 789–794.
3. Graybill, F.A.; Wang, C.-M. Confidence intervals on nonnegative linear combinations of variances. J. Am. Stat. Assoc. **1980**, *75*, 869–873.
4. 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. Simul. **1990**, *35*, 135–143.
5. Hyslop, T.; Hsuan, F.; Holder, D.J. A small sample confidence interval approach to assess individual bioequivalence. Stat. Med. **2000**, *19*, 2885–2897.
6. Lee, Y.; Shao, J.; Chow, S.-C. *Confidence Intervals for Linear Combinations of Variance Components when Estimators of Variance Components are Dependent: An Extension of the Modified Large Sample Method with Application*; 2002, submitted.
7. Lee, Y.; Shao, J.; Chow, S.-C.; Wang, H. Test for intersubject and total variabilities under crossover design. J. Biopharm. Stat. **2002**, *12*.
8. Chow, S.-C.; Tse, S.-K. A related problem in bioavailability/bioequivalence studies—Estimation of the intrasubject variability with a common Cv. Biom. J. J. Math. Methods Biosci. **1990**, *32*, 597–607.
9. Quan, H.; Shih, W.J. Assessing reproducibility by the within-subject coefficient of variation with random effects models. Biometrics **1996**, *52*, 1195–1203.
10. FDA. *Guidance for Industry: Statistical Approaches to Establishing Bioequivalence*; Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2001.
11. Chow, S.C. *Statistical Methods for Assessment of Reproducibility*; Presented at Critical Assessment Review Advisory Board Meeting, Pfizer-Aventis: NewYork, NY, 1995.
12. Chow, S.-C.; Shao, J. *Statistics in Drug Research—Methodology and Recent Development*; Marcel Dekker, Inc: New York, NY, 2001.
13. Chow, S.-C.; Tse, S.-K. On the estimation of total variability in assay validation. Stat. Med. **1991**, *10*, 1543–1553.
14. Chow, S.-C.; Shao, J. A new procedure for the estimation of variance components. Stat. Probab. Lett. **1988**, *6*, 349–355.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016