Assessment of Intrasubject Parallelism in Ex Vivo Bioassay Using Two One-Sided Tolerance Limits

Hideaki Uehara*¹, Kazuko Satoh*², Nagisa Komokata*², Yohei Tokita*¹, Mitsue Nishiyama*², Maiko Iida*², Junko Ishikawa*², Kazuo Ogawa*² and Masahiro Yamamoto*²

*¹ Kampo Scientific Strategies Division, Tsumura & Co., Tokyo 107-8521, Japan
*² TSUMURA Research Laboratories, Kampo Scientific Strategies Division, Tsumura & Co., Ibaraki 300-1192, Japan
e-mail: uehara_hideaki@mail.tsumura.co.jp

In this article, we propose an alternative criterion for evaluating parallelism in parallel-line assays at the subject level. Unlike another recently proposed criterion (“ISP”; Uehara et al., 2016), this is a single criterion concerning the quantile of intrasubject slope ratio, which is straightforward to interpret. Two one-sided tolerance limits for a linear combination of the control and test slopes are developed as metrics for the evaluation of discrepancy from the ideal intrasubject parallelism. The proposed metrics are exemplified using data from two ex vivo parallel-line experiments, and their characteristics are investigated through Monte Carlo simulations.

Key words: relative potency, parallel-line assay, intrasubject parallelism, two one-sided tests, tolerance limit, variance component.

1. Background
1.1 Assessment of intrasubject parallelism in parallel-line assay

The relative potency (RP), or the “potency ratio” of an unknown preparation of drug, is a pharmacological concept, defined as “the ratio of equipotent doses of the standard preparation and the unknown preparation under the conditions of the assay” (Chapter 5.3 of the European Pharmacopoeia, 2008). This metric is based on the assumption of similarity of the dose-response characteristics (Finney, 1978). This means that the dose-response curves for the test and the standard substance must have similar shapes despite the differences in their values on the abscissa.

Conventionally, the similarity between substances has been assessed via hypothesis testing of the model parameters, although the similarity assumption has implications for individual drug dosages and administration. Demonstrating similarity at the subject level is not a simple task because of the limited number of measurements that are available per individual, even in ex vivo
Considering the importance of RP evaluation for drug studies, there is a need for a practical method for quantifying the intrasubject similarity. Unfortunately, there seems to be no established approach for this, and none are proposed in the current revised drafts of the USP General Chapters (United States Pharmacopeial Convention, 2010a, 2010b, 2010c).

Recently, Uehara et al. proposed a metric for the evaluation of intrasubject parallelism in parallel-line assay (“ISP”), given as $E(d^2_i)/\text{Var}(\hat{b}_i | b_i)$ (Uehara et al., 2016). Here $d_i$ is the difference of dose-response slopes between substances in subject $i$, $b_i$ is the vector of subject specific parameters (intercept and slope for each substance), $b_i$ is the subject specific slope under the control treatment, and $\hat{b}_i$ is its estimate given by the least square method. Their basic idea was to translate the aggregated criterion for individual bioequivalence, which was proposed by the United States government (the “IBC”; Food and Drug Administration, 2001), for use in a similarity evaluation in bioassay. The test drug is judged to have intrasubject parallelism relative to the control when the ISP is demonstrated to be less than a predetermined upper limit, $\Theta$. In practice, the upper confidence limit of linearized criterion $\eta = E(d^2_i) - \text{Var}(\hat{b}_i | b_i) \times \Theta$ is used for the assessment of assay data, whose negative value can be interpreted as evidence of intrasubject parallelism. This method can offer a reasonable power for determining whether there is a deviation from parallelism at the subject level. However, the interpretation of the results is ambiguous because an aggregated measure of the mean and variance regarding the intrasubject slope difference is used. Therefore, if we want a rigorous criterion for intrasubject parallelism, we need to develop alternative metrics that are more straightforward and interpretable than the ISP.

Thanks to the history of individual bioequivalence assessment, again we could find the prototype of alternative methods. Here we focused on the approach taken by Esinhart and Chinchilli (1994a and 1994b) as well as by Brown et al. (1997), who considered the proportion of patients in which the inter-formulation difference of bioavailability stayed within a prescribed acceptance range. To implement this criterion they used the technique for tolerance interval estimation (see, e.g., Krishnamoorthy and Mathew, 2009).

Following these forerunners, in this paper we consider a criterion of intrasubject parallelism based on the intrasubject slope ratio quantile. When the prescribed pair (upper and lower) of quantile points falls within the acceptance range, overall intrasubject parallelism is declared (Figure 1).

To avoid calculating the tolerance limit in terms of the slope ratio, in the case that both the denominator and numerator are unobserved random parameters, we employed an approximation technique which combines the denominator and the numerator linearly (Zhang et al., 2009). Given the positive denominator, we can evaluate the accordance of results to the acceptance condition by examining the sign of the derived variable’s quantile (see Section 2.2.1 below).
Fig. 1. Intrasubject parallelism criterion based on the slope ratio distribution. The subject is deemed to have parallelism when its slope ratio falls within the acceptance range, \((\kappa_L, \kappa_U)\), which is indicated by the two vertical solid lines. When both the lower and upper \(\gamma/2\) quantile points fall in this interval, the proportion of subjects retaining intrasubject parallelism exceeds \(100(1 - \gamma)\) percent. (The two dashed lines indicate the upper and lower \(\gamma/2\) quantile points).

deduced the one-sided confidence limit of the quantile (i.e., the one-sided tolerance limit) of the derived variable by using a variation of the modified large-sample theory technique described by Zou and Donner (2008).

1.2 Motivating example

To provide context for this issue, in Table 1 we displayed the results from two ex vivo parallel-line experiments. Each experiment was designed to evaluate the RP of a test substance B against control substance A using 24 animals, following the standardized experimental procedures. In experiment 1, for each substance, we designated a common sequence of doses at 5 levels around the initial estimate of (geometrically averaged) EC\(_{50}\) for the control substance (A). Each dose level was separated from adjacent doses by a common ratio (approximately \(2^{1/2}\)). At sacrifice, 10 specimens were extracted from each animal and were randomly allocated to one of the 10 planned treatments (dose levels of 1 to 5 for each of the tests and the control substances). Then the treatments were administered in a random order, to acquire 10 responses under 10 different conditions per block (subject animal). Experiment 2 was conducted following essentially the same protocol as experiment 1, except that the dose sequences were updated using the relative potency estimates obtained from experiment 1.
Table 1. Estimated parameters from the ex vivo parallel-line assay dataset

| Block ID | Control Substance (A) | Test Substance (B) | Control Substance (A) | Test Substance (B) |
|----------|-----------------------|--------------------|-----------------------|--------------------|
|          | Intercept  | Slope         | Intercept  | Slope         | Intercept  | Slope         | Intercept  | Slope         |
| 1        | 1.5704    | 0.9273        | 1.5434    | 0.5740        | 1.6087    | 0.8465        | 1.6715    | 0.5376        |
| 2        | 1.5905    | 0.8770        | 1.4659    | 0.6804        | 1.7288    | 0.7511        | 1.7300    | 0.5971        |
| 3        | 1.6869    | 0.6501        | 1.6248    | 0.8111        | 1.7343    | 0.9018        | 1.7641    | 0.8185        |
| 4        | 1.6596    | 1.0057        | 1.6448    | 0.9814        | 1.7394    | 1.2183        | 1.8162    | 0.5440        |
| 5        | 1.5997    | 1.0269        | 1.5100    | 0.8041        | 1.6639    | 0.8283        | 1.5834    | 1.2318        |
| 6        | 1.7491    | 0.7502        | 1.6528    | 1.1433        | 1.5978    | 1.0875        | 1.6166    | 0.9942        |
| 7        | 1.8208    | 0.1082        | 1.8110    | 0.4929        | 1.5302    | 1.0199        | 1.7021    | 0.7997        |
| 8        | 1.5856    | 1.0164        | 1.5013    | 1.4864        | 1.7596    | 0.9607        | 1.7262    | 0.8330        |
| 9        | 1.5726    | 0.8815        | 1.6345    | 0.7944        | 1.5652    | 0.9920        | 1.5620    | 0.9915        |
| 10       | 1.7007    | 1.4634        | 1.7934    | 0.9826        | 1.6258    | 0.8599        | 1.6961    | 0.5989        |
| 11       | 1.5125    | 1.2584        | 1.5487    | 0.9869        | 1.6912    | 1.0284        | 1.7378    | 0.9985        |
| 12       | 1.7819    | 0.6492        | 1.7128    | 0.8776        | 1.5810    | 0.6989        | 1.5133    | 1.3687        |
| 13       | 1.5674    | 0.1005        | 1.7168    | 0.2642        | 1.6952    | 0.7087        | 1.7445    | 0.9805        |
| 14       | 1.7276    | 0.6468        | 1.6550    | 1.2547        | 1.7911    | 1.0104        | 1.7309    | 0.8988        |
| 15       | 1.6102    | 0.7365        | 1.6552    | 0.8454        | 1.6902    | 1.0737        | 1.7071    | 0.7064        |
| 16       | 1.6187    | 1.1380        | 1.6789    | 0.6000        | 1.7389    | 0.9015        | 1.7301    | 0.6740        |
| 17       | 1.6183    | 1.1787        | 1.6725    | 0.4809        | 1.7786    | 0.6490        | 1.7256    | 0.7810        |
| 18       | 1.5226    | 1.2064        | 1.3973    | 0.9842        | 1.7247    | 0.6924        | 1.7955    | 1.0714        |
| 19       | 1.7884    | 1.0057        | 1.6694    | 0.6290        | 1.7408    | 0.7952        | 1.8408    | 0.6526        |
| 20       | 1.6654    | 1.2233        | 1.6464    | 1.3412        | 1.7375    | 0.7192        | 1.6819    | 1.3060        |
| 21       | 1.5714    | 1.1570        | 1.6780    | 1.2058        | 1.7279    | 1.0976        | 1.7894    | 1.0381        |
| 22       | 1.7092    | 1.0498        | 1.6256    | 0.9735        | 1.5210    | 1.2934        | 1.5673    | 1.3611        |
| 23       | 1.6900    | 1.1171        | 1.7384    | 1.0571        | 1.6083    | 0.6806        | 1.6901    | 0.7031        |
| 24       | 1.5057    | 1.0819        | 1.4917    | 1.0754        | 1.6213    | 0.9963        | 1.6750    | 1.1158        |

In each row of the table, we show the simple linear regression coefficients calculated per substance for each block separately, to summarize the intrasubject dose-response relationship. The overall average and variance of the control slope estimates were 0.9181 (95 percent confidence interval: 0.8420 to 0.9941) and 0.0685. A rough estimate for the coefficient of variation of the control slope was 0.285. The overall average and variance of the slope difference estimates were −0.0237 (95 percent confidence interval: −0.1151 to 0.0677) and 0.0991. The residual error variances under the control and test treatments were 0.0124 and 0.0102.

2. Method

In this section, we describe the details of our proposed criterion and the corresponding estimation methods. First, notation is defined in 2.1, and then in 2.2, we present our proposed method, which is based on the one-sided tolerance limit concept. In 2.3, we briefly discuss the problem of integrating multiple assays.
2.1 Notation for a parallel-line assay

Hereinafter we consider a parallel-line assay conducted as an ex vivo randomized-block experiment, where \(n\) animal subjects are used as the experimental blocks, to evaluate the test substance against the control substance.

The statistical analysis is based on the following linear model:

\[
Y_{i,j,k} = D_i + j \cdot a_i + (b_i + j \cdot d_i) \cdot X_{i,j,k} + \varepsilon_{i,j,k}, \quad \varepsilon_{i,j,k} \sim N(0, \sigma_j^2), \quad \text{i.i.d.} \tag{1}
\]

Here \(Y_{i,j,k}\) is the observed response in subject \(i\) under treatment \(j\) (control treatment: \(j = 0\); test treatment: \(j = 1\)) at dose level \(k\) \((k = 1, \ldots, K_{i,j})\). \(D_i\) is the expected response in subject \(i\) at the mean control dose, \(a_i\) is the test treatment effect (intercept difference) in subject \(i\), \(b_i\) is the control slope in terms of the log-dose response by subject, \(d_i\) is the intrasubject slope difference in subject \(i\), and \(\varepsilon_{i,j,k}\) is the observation error in subject \(i\) under treatment \(j\) at dose level \(k\), which is normally distributed with variance \(\sigma_j^2\). \(X_{i,j,k}\) is the log-dose centralized by subtracting the intrasubject mean. \(W_{i,j,k}\) is the log-transformed value of the administered dose in subject \(i\) under treatment \(j\) at dose level \(k\). By definition \(X_{i,j,k} = W_{i,j,k} - \overline{W}_{i,j}^\bullet\), \(\overline{W}_{i,j}^\bullet = \sum_{k=1}^{K_{i,j}} W_{i,j,k} / K_{i,j}\), and intrasubject parallelism is expressed as \(d_i = 0\).

Equation (1) assumes homoskedasticity between dose levels for each study substance. If this is not the case, some suitable variable transformation needs to be sought before applying the following methods.

After sorting the within-subject measurements in order of substance and dose level, the responses from subject \(i\) are given as

\[
Y_i = \left( \begin{array}{c} Y_{i,0} \\ Y_{i,1} \end{array} \right), \quad \text{where} \quad Y_{i,0} = \left( \begin{array}{c} Y_{i,0,1} \\ \vdots \\ Y_{i,0,K_{i,0}} \end{array} \right) \quad \text{and} \quad Y_{i,1} = \left( \begin{array}{c} Y_{i,1,1} \\ \vdots \\ Y_{i,1,K_{i,1}} \end{array} \right).
\]

Then equation (1) can be rewritten as \(Y_i = X_i g_i + \varepsilon\), where

\[
X_i = \left( \begin{array}{c} X_{i,0} \\ X_{i,1} \end{array} \right) = \left( \begin{array}{cccc} j_{K_{i,0}} & 0 & x_{i,0} & 0_{K_{i,0}} \\ j_{K_{i,1}} & j_{K_{i,1}} & x_{i,1} & x_{i,1} \end{array} \right), \quad x_{i,j} = \left( \begin{array}{c} X_{i,j,1} \\ \vdots \\ X_{i,j,K_{i,j}} \end{array} \right),
\]

\[
g_i = \left( \begin{array}{c} a_i \\ b_i \end{array} \right), \quad a_i = \left( \begin{array}{c} D_i \\ a_i \end{array} \right), \quad b_i = \left( \begin{array}{c} b_i \\ d_i \end{array} \right), \quad \varepsilon_i = \left( \begin{array}{c} \varepsilon_{i,0} \\ \varepsilon_{i,1} \end{array} \right), \quad \varepsilon_{i,j} = \left( \begin{array}{c} \varepsilon_{i,j,1} \\ \vdots \\ \varepsilon_{i,j,K_{i,j}} \end{array} \right),
\]

and \(\varepsilon_i \sim N \left( 0_{K_{i,0}+K_{i,1}}, \quad \text{diag} \left( \sigma_0^2 \cdot I_{K_{i,0}}, \quad 0_{K_{i,0} \cdot K_{i,1}}, \quad \sigma_1^2 \cdot I_{K_{i,1}} \right) \right) \). \tag{2}

In (2), \(0_K\) and \(j_K\) are column vectors of length \(K\) consisting of zeros and ones, respectively, \(I_K\) is the identity matrix of order \(K\), and \(0_{K,L}\) is the \(K\)-by-\(L\) zero matrix. The least squares
estimator for the regression parameter for each subject is \( \hat{g}_i = \left( \hat{a}_i \hat{b}_i \right) = \left( X_i^T X_i \right)^{-1} X_i^T Y_i \), where

\[
\hat{a}_i = \left( \hat{D}_i \right) \quad \text{and} \quad \hat{b}_i = \left( \hat{c}_i \right). 
\]

This is distributed as

\[
\hat{g}_i \bigg|_{g_i} \sim N \left( g_i, \begin{pmatrix} A_i & 0_{2,2} \\ 0_{2,2} & B_i \end{pmatrix} \right),
\]

where

\[
\begin{align*}
A_i &= \begin{pmatrix} \frac{\sigma^2_0}{K_{i,0}} & -\frac{\sigma^2_0}{K_{i,0}} \\
-\frac{\sigma^2_0}{K_{i,0}} & \frac{\sigma^2_0}{K_{i,1}} + \frac{\sigma^2_1}{K_{i,1}} \end{pmatrix} \\
B_i &= \begin{pmatrix} -\frac{\sigma^2_0}{x^{'i}_0x_{i,0}} & -\frac{\sigma^2_0}{x^{'i}_0x_{i,0}} \\
-\frac{\sigma^2_0}{x^{'i}_0x_{i,0}} & \frac{\sigma^2_0}{x^{'i}_0x_{i,0}} + \frac{\sigma^2_1}{x^{'i}_1x_{i,1}} \end{pmatrix}.
\end{align*}
\]

The unbiased error variance estimators are

\[
s_j^2 = SSE_j / \nu_j \sim (\nu_j / \nu) \cdot \chi^2_{\nu_j} \quad (j = 0, 1), \quad \text{where} \quad \nu_j = \sum_{i=1}^{n} (K_{i,j} - 2)
\]

and \( SSE_j = \sum_{i=1}^{n} (Y_{i,j} - X_{i,j} \hat{g}_i)'(Y_{i,j} - X_{i,j} \hat{g}_i) \). (4)

When a common dose sequence is used for each substance (\( x_{i,0} = x_0 \) and \( x_{i,1} = x_1 \)), the variance in (3) becomes

\[
A_i = A = \begin{pmatrix} \frac{\sigma^2_0}{K_0} & -\frac{\sigma^2_0}{K_0} \\
-\frac{\sigma^2_0}{K_0} & \frac{\sigma^2_0}{K_0} + \frac{\sigma^2_1}{K_1} \end{pmatrix} \quad \text{and} \quad B_i = B = \begin{pmatrix} -\frac{\sigma^2_0}{x^{'0}_0x_{0,0}} & -\frac{\sigma^2_0}{x^{'0}_0x_{0,0}} \\
-\frac{\sigma^2_0}{x^{'0}_0x_{0,0}} & \frac{\sigma^2_0}{x^{'0}_0x_{0,0}} + \frac{\sigma^2_1}{x^{'1}_1x_{1,1}} \end{pmatrix}. \quad (5)
\]

Here \( K_0 \) and \( K_1 \) are the lengths of \( x_0 \) and \( x_1 \), respectively.

We further assume that is distributed as

\[
b_i \sim N(\beta, \Sigma), \quad \text{where} \quad \beta = \begin{pmatrix} \beta \\ \delta \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma^2_0 & \sigma_{bd} \\ \sigma_{bd} & \sigma^2_d \end{pmatrix} \quad .
\]

The vector of all the responses is denoted \( Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \).

2.2 Proposed criterion of intrasubject slope ratio

As mentioned in 1.1, we propose a criterion of intrasubject parallelism which requires the prescribed pair of quantile points (lower and upper \( \gamma / 2 \) quantile) of slope ratio to fall within the acceptance range, which can be expressed as \( \Pr\{\kappa_L \leq SR_i \leq \kappa_U\} \geq 1 - \gamma \) (see Figure 1). This is sufficient to declare that for more than 100(1 - \( \gamma \)) percent of subjects, the two slopes are sufficiently parallel (i.e., the slope ratio is within the acceptance region).

We consider the tail probabilities in the distribution of slope ratio
The aforementioned criterion may be broken down as
\[ \Pr\{\kappa_L \leq SR_i \} \geq 1 - \gamma/2 \quad \text{and} \quad \Pr\{SR_i \leq \kappa_U \} \geq 1 - \gamma/2, \]
where the interval \((\kappa_L, \kappa_U)\) gives the acceptance range of slope ratio.

From these two inequalities, we can deduce
\[ \Pr\{\kappa_L \leq SR_i \leq \kappa_U \} = \Pr\{SR_i \leq \kappa_U \} - \Pr\{SR_i \leq \kappa_L \} \geq 1 - \gamma. \]

Thus the relations in (8) constitute the sufficient condition for declaring the two slopes to be parallel enough for at least \(100 \cdot (1 - \gamma)\) percent of the subjects.

To take account the fluctuation caused by the sampling error, below we employ the one-sided tolerance limit technique (Novick et al., 2009a; see also Krishnamoorthy and Mathew, 2009) to develop a procedure for parallelism evaluation according to the inequalities in (8) under the assumptions in (6).

### 2.2.1 Approximate linearized inequalities

If the probability of a negative control slope for the subject is negligible, as is expected in any established parallel-line assay system with sufficient precision, the probabilities in (8) can be approximated by the following linearization:
\[ \Pr\{\kappa_L \leq SR_i \} \approx \Pr\{0 \leq l_i \} \quad \text{and} \quad \Pr\{SR_i \leq \kappa_U \} \approx \Pr\{u_i \leq 0 \}. \]

Zhang et al. (2009) examined the use of this approximation for the tolerance interval of the ratio of normal variables, and they concluded that the approximation is quite accurate if the coefficient of variation (CV) of the denominator variable \(b_i\) in our case) is no more than 0.3. We believe that this assumption can be satisfied in most of the established parallel-line assay systems.

Corresponding to the pair of linearized variables \((l_i, u_i)\), we consider two one-sided tolerance limits (TOSTL) with content \((1 - \gamma/2)\) at a confidence level of \((1 - \alpha)\), designated as \((L(Y), U(Y))\), which should satisfy
\[ \Pr\{\Pr(L(Y) < l_i \mid Y) \geq 1 - \gamma/2\} = 1 - \alpha \quad \text{and} \quad \Pr\{\Pr(u_i < U(Y) \mid Y) \geq 1 - \gamma/2\} = 1 - \alpha. \]

If \(0 \leq L(Y)\) and \(U(Y) \leq 0\) simultaneously, we can reject the following pair of null hypotheses:
\[ H_{0l}: \Pr(0 \leq l_i) < 1 - \gamma/2, \quad H_{0u}: \Pr(u_i \leq 0) < 1 - \gamma/2. \]

Here the type 1 error rate is kept under \(\alpha\) because this is an example of an intersection-union test (Dmitrienko and D’Agostino, 2013). Then, at the approximate confidence level of \((1 - \alpha)\), we can conclude that

\[ Jpn J Biomet Vol. 37, No. 2, 2016 \]
\[ \Pr(k_L \leq SR_i \leq k_U) \approx \Pr(u_i \leq 0) - \Pr(l_i < 0) \geq 1 - \gamma. \] (11)

### 2.2.2 Estimation of one-sided tolerance limit

In this subsection, we develop the estimator for \( L(Y) \), the lower one-sided tolerance limit. The estimator for \( U(Y) \) can be deduced similarly. For the sake of simplicity, here we consider experiments without missing data.

Denoting \( \lambda = E(l_i) = c_L \cdot \beta \) and \( \sigma_l^2 = \text{Var}(l_i) = c_L' \cdot \Sigma \cdot c_L, \) \( L(Y) \) is identical to the lower \((1 - \alpha)\) confidence limit for \( \lambda - Z_{1-\gamma/2} \cdot \sigma_l \), where \( Z_\beta \) is the \( \beta \)-quantile of the standard normal distribution.

Without missing data, the variance of the linear statistics \( \hat{l}_i = c_L' \cdot \hat{b}_i \) is given by (5) as \[ \text{Var}(\hat{l}_i | \beta) = \sigma_l^2 + \sigma_h^2. \] Here \( \sigma^2_h = \frac{n}{n-1} \cdot \frac{\sigma_0^2}{x_0'x_0} + \frac{\sigma_1^2}{x_1'x_1}, \) which can be estimated from \( s_h^2 = \frac{\sigma_0^2}{x_0'x_0} + \frac{s_1^2}{x_1'x_1}. \) The corrected sum of squares of \( \hat{l}_i \) is distributed as

\[ SS_l = \sum_{i=1}^{n} (\hat{l}_i - \bar{l}_s)^2 \sim (\sigma_l^2 + \sigma_h^2) \cdot \chi^2_{n-1}. \] (12)

Thus the unbiased estimator of \( \sigma_l^2 \) is given by

\[ \hat{\sigma}_l^2 = s_l^2 - s_h^2 \quad \text{where} \quad s_l^2 = \frac{SS_l}{n - 1}. \] (13)

Under the normality assumption in (1), \( \hat{\sigma}_l^2 \) is the sum of the independent statistics, and the \((1 - \gamma/2, 1 - \alpha)\) lower tolerance limit of \( l_i \) is given by the method of variance estimates recovery (MOVER; Zou and Donner, 2008) as

\[ L(Y) = \bar{l}_s - \left[ Z_{1-\gamma/2} \sqrt{[\hat{\sigma}_l^2]_+} + \sqrt{\frac{s_h^2}{n-1,1-\alpha} \cdot \frac{s_l^2}{n} + Z_{1-\gamma/2}^2 \left( \sqrt{[\hat{\sigma}_l^2]_+} - \sqrt{Q(\alpha)} \right)^2} \right]. \] (14)

Here \([x]_+ = \max(0, x)\), and the upper confidence limit for \( \sigma_l^2 \) is denoted by \( Q(\alpha') \). Note that whenever \( \hat{\sigma}_l^2 > 0 \), \( L(Y) \) given by (14) is smaller than the lower \((1 - \alpha)\) confidence limit of \( \lambda \), which turns out to be a valid lower tolerance limit when \( \sigma_l^2 = 0 \). In such cases, \( L(Y) \) tends to have a conservative nature, typically when the assay has a high reproducibility in the slope.

Defining \( \tau = \sigma_l^2 / \sigma_h^2 \), we have \( SS_l \sim \sigma_h^2 \cdot (1 + \tau) \cdot \chi^2_{n-1} \). Approximating the distribution of \( s_h^2 \) by Satterthwaite’s method, we have

\[ s_h^2 \sim \frac{\sigma_h^2}{\nu} \cdot \chi^2_{\nu}, \quad \text{where} \quad \nu = \frac{\left( \frac{\sigma_l^2}{x_0'x_0} + \frac{\sigma_1^2}{x_1'x_1} \right)^2}{\frac{1}{\nu_0} \cdot \left( \frac{\sigma_0^2}{x_0'x_0} \right)^2 + \frac{1}{\nu_1} \cdot \left( \frac{\sigma_1^2}{x_1'x_1} \right)^2}. \] (15)

Thus the approximate upper confidence limit of \( \tau \) with a \( 100(1 - \alpha) \) percent confidence level is given as \( \tau_{\alpha} = \left[ \frac{s_h^2}{s_l^2} F_{n-1,\nu}^{\alpha} \right]_{+} \), where \( F_{n-1,\nu}^{\alpha} \) is the \( \alpha \)-quantile of the F distribution with degrees of freedom \( n - 1 \) and \( \nu \). In practice, we replace \( \nu \) with \( \hat{\nu} \), which is obtained by substituting \( s_0^2 \) for \( \sigma_0^2 \) and \( s_1^2 \) for \( \sigma_1^2 \) in (15).
Following Hartung and Knapp (2000), we employ

\[ \hat{Q}^{(l)} = \tau_U^2 \cdot s_{\hat{h}}^2. \] (16)

The \((1 - \gamma/2, 1 - \alpha)\) upper tolerance limit of \(u_i\) is obtained by replacing \(l\) with \(u\), and \(L\) with \(U\) in (12) through (16), ensuring a positive sign for the square-bracketed term in (14).

### 2.2.3 Evaluation under a homogeneous design with missing data

Unfortunately, even when a study is well designed, loss of data can occur unexpectedly, and as a consequence, we must analyze unbalanced data. If all the measurements were performed concurrently (e.g., at the sacrifice), then the data can be assumed to be “missing completely at random” (Little and Rubin, 2002), and we can handle the data as if they were taken from a study that was designed to be unbalanced. In this case, we may approximate the distribution of \(s_l^2\) in (13) and (14) by applying a method for unbalanced one-way random effect models (Thomas and Hultquist, 1978; see also Appendix 2 of Uehara et al., 2016) as \(s_l^2 \sim \frac{\sigma_l^2 + \sigma_{hm}^2}{n-1} \cdot \chi_{n-1}^2\), where \(\sigma_{hm}^2 = \sum_{i=1}^{n} \frac{\sigma_i^2}{n} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{k_i^2 \cdot \sigma_i^2}{x_i^0 \cdot x_i^0} + \frac{\sigma_i^2}{x_i^0 \cdot x_i^1} \right)\). Correspondingly, the variance estimator \(\hat{s}_l^2\) in (13) and (14) should be replaced by \(\hat{\sigma}_{lm}^2 = s_l^2 - s_{hm}^2\), where

\[
s_{hm}^2 = \sum_{i=1}^{n} \frac{s_i^2}{n} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{k_i^2 \cdot s_i^2}{x_i^0 \cdot x_i^0} + \frac{s_i^2}{x_i^0 \cdot x_i^1} \right).
\]

As in the previous subsection, we use the upper confidence limit of \(\tau\) to calculate \(Q^{(l)}\) via (16). Defining \(m_i = \sigma_l^2 / \sigma_{\hat{h}}^2\) and \(w_{i(\tau)} = \frac{m_i}{1 + m_i \cdot \tau}\), Wald (1940) showed that

\[
SS_{lw}(\tau) = \sum_{i=1}^{n} \frac{w_{i(\tau)} \cdot \left( \hat{l}_i - \sum_{i=1}^{n} w_{i(\tau)} \cdot \hat{l}_i / \sum_{i=1}^{n} w_{i(\tau)} \right)^2}{\sigma_{\hat{h}}^2 \cdot \chi_{n-1}^2} \sim \sigma_{\hat{h}}^2 \cdot \chi_{n-1}^2.
\] (17)

Thus the approximate upper confidence limit of \(\tau\) with a \(100(1 - \alpha)\) percent confidence level (\(\tau_U^\alpha\)) is determined by solving

\[
F(\tau) = \frac{SS_{lw}(\tau)}{s_{\hat{h}}^2 \cdot (n-1)} = F_{n-1, \nu}^\alpha.
\] (18)

As before, we replace \(\nu\), \(\sigma_{\hat{h}}^2\), and \(\sigma_{hm}^2\) in (17) and (18) with \(\hat{\nu}\), \(s_{\hat{h}}^2\), and \(s_{hm}^2\). Because \(F(\tau)\) is monotonically decreasing with \(\tau\) (Wald, 1940), this has a unique solution \(\tau_U^\alpha\) whenever \(F(0) \geq F_{n-1, \nu}^\alpha\). For the case \(F(0) < F_{n-1, \nu}^\alpha\), \(\tau_U^\alpha\) has to be set to zero.

Note that the unbalanced nature of the design jeopardizes the independence between \(\hat{l}_i\) and \(s_{l_i}^2\), and therefore the prerequisite for the MOVER method is not strictly fulfilled. For any given dataset, the appropriateness of this method will vary according to the pattern or rate of missing data, so it should be examined via Monte Carlo simulation before use.

### 2.3 Integrated analysis of multiple homogeneous assays

When conducting multiple assays, heterogeneity between assays may emerge in the data, the evaluation of intrasubject parallelism needs to reflect both the between- and the within-assay variation. For example \(L(Y)\) can be similarly defined as before, the lower \((1 - \alpha)\) confidence limit for \(\lambda - Z_{1-\gamma/2} \cdot \sigma_l\), but here \(\sigma_l\) stands for the variation between subjects coming from Jpn J Biomet Vol. 37, No. 2, 2016.
From (20) and (21), the estimator (13) becomes

\[ \hat{L}_i^{(r)} = c_L \cdot b_i^{(r)}, \quad \bar{L}^{(r)} = \frac{1}{n} \cdot \sum_{i=1}^{n} \hat{L}_i^{(r)}, \quad \text{and} \quad \bar{L} = \frac{1}{R} \cdot \sum_{r=1}^{R} \bar{L}^{(r)}. \]  

(19)

The unbiased error variance estimators are given by

\[ s_j^2 = \frac{\text{SSE}_j}{\nu_j} \sim \frac{(\sigma_j^2)^2}{\nu_j} \cdot \chi_{\nu_j}^2, \quad (j = 0, 1) \quad \text{where} \quad \nu_j = n \cdot R \cdot (K_j - 2) \]

and

\[ \text{SSE}_j = \sum_{r=1}^{R} \sum_{i=1}^{n} (Y_{i,j}^{(r)} - X_{i,j}^{(r)}b_i^{(r)})^2 = (Y_{i,j}^{(r)} - X_{i,j}^{(r)}\hat{b}_i^{(r)}). \]  

(20)

The between- and within-assay corrected sums of squares of \( \hat{L}_i^{(r)} \) are

\[ \text{SS}_{I(B)} = \sum_{r=1}^{R} (\bar{L}_i^{(r)} - \bar{L})^2 \sim \text{Var}(\bar{L}_i^{(r)}) \cdot \chi_{R-1}^2 \quad \text{and} \]

\[ \text{SS}_{I(W)} = \sum_{r=1}^{R} \sum_{i=1}^{n} (\hat{L}_i^{(r)} - \bar{L}^{(r)})^2 \sim \text{Var}(\hat{L}_i^{(r)} | \mathbf{b}^{(r)}) \cdot \chi_{R(n-1)}^2. \]

(21)

We have

\[ \text{Var}(\hat{L}_i^{(r)} | \mathbf{b}^{(r)}) = \sigma_{I(W)}^2 + \frac{\kappa^2}{x_i'x_0} + \frac{\sigma_x^2}{x_i'x_1}. \]  

(22)

\[ \text{Var}(\bar{L}^{(r)}) = \text{Var}(\mathbb{E}(\hat{L}_i^{(r)} | \mathbf{b}^{(r)})) + \mathbb{E}\{\text{Var}(\hat{L}_i^{(r)} | \mathbf{b}^{(r)})\} \]

\[ = \sigma_{I(B)}^2 + \frac{1}{n} \cdot \left( \sigma_{I(W)}^2 + \frac{\kappa^2}{x_0'x_0} + \frac{\sigma_x^2}{x_1'x_1} \right). \]  

(23)

For the explicit formulas of the variance terms in the general unbalanced case, see Appendix.

From (20) and (21), the estimator (13) becomes

\[ \hat{\sigma}_I^2 = \frac{\text{SS}_{I(B)}}{R - 1} + \frac{\text{SS}_{I(W)}}{R \cdot n} - s_h^2. \]  

(24)

Here \( s_h^2 = \frac{\kappa^2}{x_0'x_0} + \frac{\sigma_x^2}{x_1'x_1} \), and \( L(\mathbf{Y}) \) is given by

\[ L(\mathbf{Y}) = \bar{L} \left[ z_{1-\gamma/2} \sqrt{\hat{\sigma}_I^2} + \sqrt{t_{R-1,1-\alpha}^2 \cdot \frac{\text{SS}_{I(B)}}{R(R-1)} + z_{1-\gamma/2}^2 \left( \sqrt{\hat{\sigma}_I^2} - \sqrt{Q(J)} \right)^2} \right] \]  

(25)

We consider estimating \( Q(J) \) using a quantile of the F distribution. From the first and second moments of the distribution, we have the approximation

\[ \frac{\text{SS}_{I(B)}}{R - 1} + \frac{\text{SS}_{I(W)}}{R \cdot n} \sim \frac{\sigma_{I(B)}^2 + \sigma_{I(W)}^2 + s_h^2}{\omega} \cdot \chi_{\omega}^2 \quad \text{where} \]

\[ \omega = \frac{1}{R - 1} \cdot \left( \sigma_{I(B)}^2 \cdot \frac{\sigma_{I(W)}^2}{n} + \frac{s_h^2}{R \cdot n} \right) + \frac{n - 1}{R \cdot n^2} \cdot \left( \sigma_{I(W)}^2 + s_h^2 \right)^2. \]  

(26)
A crude estimate of \( \omega \) is given by Satterthwaite’s method:
\[
\hat{\omega} = \frac{\left\{ \frac{SS_l(B)}{R-1} + \frac{SS_l(W)}{R \cdot n} \right\}^2}{\left\{ \frac{SS_l(B)}{R-1} \right\}^2 + \frac{\left\{ SS_l(W) \right\}^2}{R^3 \cdot n^2 \cdot (n-1)}}.
\]
\[(27)\]

As in 2.2.2, we estimate \( Q(l) \) using the upper confidence limit of \( \tau = \sigma_l^2 / \sigma_h^2 \):
\[
\hat{Q}(l) = \tau_U^{\alpha} \cdot s_h^2 \quad \text{where} \quad \tau_U^{\alpha} = \left[ \frac{SS_l(B)}{R-1} + \frac{SS_l(W)}{R \cdot n} \right]_{\frac{\sigma_l^2}{\sigma_h^2} \cdot \Gamma_{\nu_0,\nu_1} - 1}.
\]
\[(28)\]

If one intends to evaluate the intrasubject parallelism via an experiment, it is necessary to decide at the design stage whether to conduct a single homogeneous study with a sufficient number of subjects or a series of assays long enough for an integrated analysis. Comparing (26) to (15), it can be seen that \( \omega \) is determined by \( n \) and \( R \), whereas \( \nu \) is determined by \( \nu_0 \) and \( \nu_1 \). Thus an integrated analysis can be advantageous if the number of measurements per subject is limited and increasing the number of animals is relatively easy.

3. Analysis of Ex Vivo Parallel-line Assay Data

We applied our proposed procedures to the dataset presented in Section 1.2, which contained one missing observation from a total of 480. To demonstrate the use of our method, we apply the formulae given in 2.2.2 disregarding both the differences between the experiments and the missing data. The computations were performed using R version 3.2.3 (R Core Team, 2015), by means of the lmList function from the nlme package (Pinheiro et al., 2015).

From Table 1 and the summary statistics given in Section 1.2, the moment estimate of \( \sigma_b^2 \) was 0.0685 – 0.0124/0.2265 ≈ 0.0139, which gives a CV of 0.128, which is deemed acceptable for the use of the approximation in equation (9). The moment estimate of \( \sigma_d^2 \) was 0.0991 – (0.0124 + 0.0102)/0.2265 ≈ −0.0007, which implies that the intersubject homogeneity in terms of slope difference was negligible.

At \( \kappa_L = 2/3 \) and \( \kappa_U = 3/2 \), \( \bar{l} = 0.2823 \), \( s_l^2 = 0.0757 \), \( \bar{u} = -0.4828 \), and \( s_u^2 = 0.1629 \). Based on these statistics, at \( \alpha = 0.05 \) and \( \gamma = 0.2 \), our procedure gave \( L = 0.0018 \) and \( U = -0.1174 \). Thus the intrasubject parallelism was confirmed.

4. Simulation

To examine the operational characteristics of our proposed method, we conducted a Monte Carlo simulation of a five-dose parallel-line assay, which is the analog of the aforementioned example study, with 10,000 iterations.

Here we focused on [1] the coverage rate of true quantile points and [2] the overall power of the procedure to declare the intrasubject parallelism. In [1], we evaluate the coverage probability of \( E(l_i) - Z_{1-\gamma/2} \cdot \sigma_l \) by its one-sided confidence limit \( L(Y) \), as well as the coverage probability Jpn J Biomet Vol. 37, No. 2, 2016.
of \( E(u_i) + Z_{1-\gamma/2} \cdot \sigma_u \) by \( U(Y) \).

The computations were performed using R version 3.2.3 (R Core Team, 2015), by means of the lmList function from the nlme package (Pinheiro et al., 2015).

4.1 Simulation parameters

We chose a design similar to the example above, with a common dose ratio of \( 2^{1/2} \) and the following conditions:

- Centralized log-dose: \( x = \log_{10} 2 \cdot (-1, -0.5, 0, 0.5, 1) (x' x \approx 0.2265) \)
- Control slope: \( \beta = 1 \)
- Mean slope difference: \( \delta \in \{0, 0.15 \cdot \beta, 0.30 \cdot \beta\} \)
- Tolerable range of slope ratio: \( (\kappa_L, \kappa_U) = (2/3, 3/2) \)
- Min. content between TOSTL: \( 1 - \gamma = 0.8 \)
- Number of subjects: \( n = 12, 24, \) or \( 48 \)
- Variance components: \( \Sigma = \begin{pmatrix} \sigma_b^2 & \rho \cdot \sigma_b \cdot \sigma_d \\ \rho \cdot \sigma_b \cdot \sigma_d & \sigma_d^2 \end{pmatrix} \), \( \rho = -0.2 \)
- Error Variance: \( \sigma_b^2 = \sigma_1^2 = \sigma_2^2 = 0.01 \)

With \( \sigma_d = c \cdot |\beta| / Z_{1-\gamma/2} \) and \( \delta = 0 \), the interval \((-c \cdot |\beta|, c \cdot |\beta|)\) has a content of \( 1 - \gamma \); i.e., \( 100 \cdot (1 - \gamma) \) percent of subjects have their slope differences within the interval. For situations in which the CV of intersubject slope variation rises to 0.3, the acceptance limit given by Zhang et al. (2009) is examined at \( \sigma_b^2 = (0.3 \cdot \beta)^2 \).

4.2 Results

The empirical TOSTL coverage rates of the true quantile are summarized in Table 2. When there was no intersubject variation at all in terms of slope, the simulated results exhibited more than 99 percent of coverage, demonstrating the conservatism of the proposed procedures under this exceptional situation. The rates approached the nominal value if there was any heterogeneity in terms of slope.

The empirical TOSTL power of the intrasubject parallelism is summarized in Figure 2. An increase in the mean slope difference rapidly diminished the power, which implies that this method is applicable in practice only when the mean slopes are roughly equal. An increase in the variance of the slope parameter (\( \sigma_b^2 \)) or intrasubject slope difference (\( \sigma_d^2 \)) also decreased the power.

Under homogeneity of slopes (\( \sigma_b^2 = \sigma_d^2 = 0 \)), the conservatism of our procedure was evident, as shown in Table 2. Here an ideal estimate of \( Q^{(l)} \) should take value of zero, which makes the lower (or upper) tolerance limit \( L(Y) \) (or \( U(Y) \)) given by the formula (14) conform to the lower (or upper) confidence limit of \( \lambda = E(l_i) \) (or \( \nu = E(u_i) \)). However, as long as the model assumes heterogeneity of slopes (\( \sigma_b^2 > 0 \) or \( \sigma_d^2 > 0 \)), \( Q^{(l)} \) (the upper confidence limit of \( \sigma_b^2 \)) tends
Assessment of Intrasubject Parallelism in Ex Vivo Bioassay Using TOSTL

Table 2. Empirical TOSTL coverage rates in the simulation study

| $\sigma_b^2$ | $\sigma_d^2$ | N   | $\delta = 0.0$ | $\delta = 0.15 \cdot \beta$ | $\delta = 0.30 \cdot \beta$ |
|-------------|-------------|-----|----------------|-------------------------------|-------------------------------|
|             |             |     | $L(Y)$ | $U(Y)$ | $L(Y)$ | $U(Y)$ | $L(Y)$ | $U(Y)$ | $L(Y)$ | $U(Y)$ |
| 0           | 0           | 12  | 0.9899 | 0.9916 | 0.9891 | 0.9916 | 0.9931 | 0.9904 |
|             |             | 24  | 0.9931 | 0.9934 | 0.9931 | 0.9935 | 0.9928 | 0.9932 |
|             |             | 48  | 0.9942 | 0.9946 | 0.9951 | 0.9948 | 0.9933 | 0.9941 |
| (0.15 \cdot \beta)^2 | (0.15 \cdot \beta)^2 | 12  | 0.9525 | 0.9624 | 0.9512 | 0.9595 | 0.9511 | 0.9637 |
|             |             | 24  | 0.9461 | 0.9579 | 0.9481 | 0.9560 | 0.9462 | 0.9561 |
|             |             | 48  | 0.9483 | 0.9574 | 0.9471 | 0.9553 | 0.9471 | 0.9512 |
| (0.3 \cdot \beta)^2 | (0.3 \cdot \beta)^2 | 12  | 0.9716 | 0.9736 | 0.9756 | 0.9752 | 0.9734 | 0.9723 |
|             |             | 24  | 0.9689 | 0.9704 | 0.9701 | 0.9691 | 0.9712 | 0.9715 |
|             |             | 48  | 0.9656 | 0.9653 | 0.9656 | 0.9643 | 0.9619 | 0.9648 |

$L(Y)$ coverage rate: $\Pr\{L(Y) < E(l_i) - Z_{1-\gamma/2} \cdot \sigma_l\}$

$U(Y)$ coverage rate: $\Pr\{U(Y) > E(u_i) + Z_{1-\gamma/2} \cdot \sigma_u\}$

to be positive even under intrasubject parallelism, especially if the sample size per subject is small (see Table 2 of Hartung and Knapp, 2000). In consequence, the resulting tolerance limit became more distant from the mean, inflating the coverage rate in Table 2. Similar consequences are observed when using the tolerance-limit approach for a mixed-effect model with negligible variance components (e.g., see the first line of Table 3 in Krishnamoorthy and Peng, 2014).

At any rate, the above is an ironic consequence, because the lack of subject-substance interaction is a necessary condition for the validity of the relative potency evaluation, in the strict sense. To put it the other way round, however, a positive TOSTL result is strong evidence for intrasubject parallelism, which is suitable for a rigorous assessment of similarity between substances.

5. Discussion

Despite its essential importance in bioassays, assessment of intrasubject similarity between dose-response curves is not common practice. Significance testing of the second-order interaction between subject, substances, and dose, which does not offer any explicit assurance in terms of the validity of the RP estimation at the subject level, is often employed as a conventional preliminary. In an attempt to solve this problem, Uehara et al. (2016) developed the ISP, an analogue of Jpn J Biomet Vol. 37, No. 2, 2016
Fig. 2. Simulated TOSTL power against number of subjects for different intrasubject distances $\delta$. (a) Fixed control slope and differences between slopes fixed; (b) control slope fixed and differences between slopes varied ($\sigma_d = 0.15 \cdot \beta / Z_{1-\gamma/2}$); (c) control slope varied ($\sigma_b = 0.15 \cdot \beta$) and slope difference fixed; (d) both control slope and slope difference varied ($\sigma_b = 0.15 \cdot \beta$ and $\sigma_d = 0.15 \cdot \beta / Z_{1-\gamma/2}$).

the aggregated IBE criterion proposed by the FDA. The ISP can offer a higher power than a disaggregated procedure, but this higher power can also result in ambiguity in the interpretation of results.

In this paper, we derived alternative procedures via tolerance limit estimation (Brown, Iyer and Wang, 1997; Esinhart and Chinchilli, 1994a, 1994b; also see Zhang et al., 2009 and Section 6.5 in Krishnamoorthy and Mathew, 2009) to evaluate the intrasubject slope differences, assuming heterogeneity of the slope and the slope difference. The simulation study showed that this approach can work well under the presumed heterogeneity, but can be conservative otherwise. In one sense, however, this conservatism can be seen to be suitable for a confirmatory analysis where a rigorous evaluation is required (e.g., the assay before switching the initial standard at the developmental stage to the revised standard for the product quality control).
Our example presented in Sections 1.2 and 3 can be deemed as a case of moderate heterogeneity in control slopes accompanied with a fixed, negligible slope difference. The aforementioned simulation suggests that here we may have suffered a conservatism of the procedure (see results for $\sigma_b = 0.15 \cdot \beta$, $\sigma_d = 0$, and $n = 48$ in Table 2).

In practice, complete homogeneity of slopes is unlikely to occur. In particular, the homogeneity of the control slope cannot be assured effortlessly. Experimenters are often obliged to reduce the number of animals used, and often put the lowest/highest dose of the experiment at the borders of the linear range, because otherwise the study design loses statistical power because of the small effect size. However, the intersubject variation in EC$_{50}$ lets a certain portion of animals deviate from the linear range at its lowest or highest dose level, and the resulting nonlinearity causes the apparent heterogeneity of the control slopes. Also, the slope difference may fluctuate if the preparation method for the test substance is not well established. In such cases, the coverage rate of our TOSTL procedure is more likely to retain the nominal value.

Nevertheless, the most disappointing aspect of the proposed method is the conservatism under “perfect” intrasubject parallelism (i.e., $\sigma_d^2 = 0$). In 4.2, we mentioned our suspicion that the erroneous assumption of variance component structure may be to blame, but we cannot deny the possible inappropriateness of our approach for the quantile inference. In this paper, we deduced the quantile of the random effect distribution under the conventional linear mixed effect model framework. However, what if we employ the quantile itself as the primary parameter of modeling, in place of the mean or the intersubject variance? This may require an entire revision of the model framework, but in our future work we may have to examine the possibility of this alternative approach.

The proposed methods employ several assumptions and, needless to say, their appropriateness for a given dataset must be examined before they are applied. In equation (1), we assumed that the errors were normally distributed and that there was homoscedasticity among the animals and different dose levels of each substance. However, this can be uncertain when the assessment of a test substance is still in the early stages, and so it is important to further examine the robustness of these methods. We expect that the series of articles by Novick et al. (2009a, 2009b, 2009c) will provide a good model for this kind of evaluation.

Note that the example dataset contained the results from two experiments. Strictly speaking, this does not accord with the proposed method. We applied our method to the data in a retrospective manner, for the purpose of demonstrating the utility of the proposed method in practice.

In this paper, we also considered the integration of multiple assays conducted under a common design. This may seem useful for aiding project planning, but may require a substantial number of experiments. The use of integrated analysis needs a careful examination, taking the expected assay precision for each candidate design into consideration. Note that assay integration
can also be an issue when assays are heterogeneously designed. For such cases, we need to further extend our proposed method; our current thought is that we may consider the use of the GLIMMIX procedure in SAS/STAT 9.2, which offers several methods (based on the profile or estimated likelihood and the Wald statistic) for estimating the confidence limits of variance component parameters. The TOSTL deduced by combining these confidence limits and $\bar{t}_\bullet$ (or $\bar{u}_\bullet$) via the MOVER method seems worth examining, although this approach disregards the mutual correlations between the parameters.

Our initial conclusion is that the proposed tolerance limit approach can offer more interpretable results than the aggregated ISP criterion. Although it tends to be more conservative, this can be seen to be suitable for a confirmatory analysis where a rigorous evaluation is required in terms of the intrasubject parallelism. The method for evaluating intrasubject parallelism should be chosen according to the nature and the purpose of the assay.

Acknowledgements

We thank the two referees for their productive comments.

REFERENCES

Brown E.B., Iyer H.K. and Wang C.M. (1997). Tolerance intervals for assessing individual bioequivalence. *Statistics in Medicine*, 16, 803–820.

Dmitrienko A. and D'Agostino R.Sr. (2013). Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*, 32, 5172–5218.

Esinhart J.D. and Chinchilli V.M. (1994a). Sample size considerations for assessing individual bioequivalence based on the method of tolerance interval, *International Journal of Clinical Pharmacology and Therapeutics*, 32, 26–32.

Esinhart J.D. and Chinchilli V.M. (1994b). Extension to the use of tolerance intervals for the assessment of individual bioequivalence, *Journal of Biopharmaceutical Statistics*, 4, 39–52.

Finney D.J. (1978). *Statistical methods in biological assay*, 3rd ed. Charles Griffin and Company Ltd., England.

Food and Drug Administration (2001). *Guidance for industry: Statistical approaches to establishing bioequivalence*.

Graybill F. and Wang C.M. (1980). Confidence intervals on nonnegative linear combinations of variances, *Journal of the American Statistical Association*, 75, 869–873.

Hartung J. and Knapp G. (2000). Confidence intervals for the between group variance in the unbalanced one-way random effects model of analysis of variance. *Journal of Statistical Computation and Simulation*, 65, 311–323.
Howe W.G. (1974). Approximate confidence limits on the mean of X+Y where X and Y are two
tabled independent random variables. *Journal of the American Statistical Association*, 69,
789–794.

Hyslop T., Hsuan F. and Holder D.J. (2000). A small sample confidence interval approach to
assess individual bioequivalence, *Statistics in Medicine*, 19, 2885–2897.

Krishnamoorthy K. and Mathew T. (2009). *Statistical tolerance region: theory, applications, and
computation*. Wiley.

Krishnamoorthy K. and Peng J. (2014). Approximate one-sided tolerance limits in random effects
model and in some mixed models and comparisons, *Journal of Statistical Computation and
Simulation*, 85, 1651–1666.

Novick S. et al. (2009a). A two one-sided parametric tolerance interval test for control of delivered
dose uniformity. Part 1—Characteristics of FDA proposed test. AAPS PharmSciTech, 10,
820–828.

Novick S. et al. (2009b). A two one-sided parametric tolerance interval test for control of delivered
dose uniformity. Part 2—Effect of changing parameters. AAPS PharmSciTech, 10, 841–849.

Novick S. et al. (2009c). A two one-sided parametric tolerance interval test for control of delivered
dose uniformity. Part 3—Investigation of robustness to deviations from normality. AAPS
PharmSciTech, 10, 829–840.

Little and Rubin (2002). *Statistical analysis with missing data*, 2nd ed. Wiley., New Jersey.

Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2015). nlme: Linear and
Nonlinear Mixed Effects Models. R package version 3.1-122, <URL: http://CRAN.R-
project.org/package=nlme>.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation
for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SAS Institute Inc. SAS/STAT 9.2 *User’s Guide*, 2nd Edition. 2013.

The European Pharmacopeia, 2008.

The United States Pharmacopeial Convention, USP 1032 (2010a).

The United States Pharmacopeial Convention, USP 1033 (2010b).

The United States Pharmacopeial Convention, USP 1034 (2010c).

Thomas J.D. and Hultquist R.A. (1978). Interval estimation for the unbalanced case of the
one-way random effect model. *The Annals of Statistics*, 6, 582–587.

Ting N., Burdick R.K., Graybill F.A., Jeyaratnam S. and Lu T.F.C. (1990). Confidence intervals
on linear combinations of variance components that are unrestricted in sign. *Journal of
Statistical Computation and Simulation*, 35, 135–143.

Jpn J Biomet Vol. 37, No. 2, 2016
Uehara H., Satoh K., Komokata N., Tokita Y., Nishiyama M., Iida M., Ishikawa J., Ogawa K. and Yamamoto M. (2016). Combinability of animal data in relative potency estimations. \textit{Japanese Journal of Biometrics}, \textbf{37}, 45–65.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. \textit{The Annals of Mathematical Statistics}, \textbf{11}, 96–100.

Zhang L., Mathew T., Yang H., Krishnamoorthy K. and Cho I. (2009). Tolerance limits for a ratio of normal random variables. \textit{Journal of Biopharmaceutical Statistics}, \textbf{20}, 172–184.

Zou G.Y. and Donner A. (2008). Construction of confidence limits about effect measures: a general approach. \textit{Statistics in Medicine}, \textbf{27}, 1693–1702.

\textbf{Appendix}

Variance of the per-subject linear statistics which appeared in this article

From the formulae in Section 2, we have

\[
\text{Var}(\hat{l}_i^{(r)} | b^{(r)}) = \sigma_0^2 \frac{n_0^2}{x'_{i(r),0}x_{i(r),0}} + \sigma_1^2 \frac{n_1^2}{x'_{i(r),1}x_{i(r),1}}, \tag{A.1}
\]

\[
\text{Var}(\bar{l}_i^{(r)} | b^{(r)}) = \frac{\sigma_0^2}{n_{(r)}} + \frac{1}{n_{(r)}^2} \sum_{i=1}^{n_{(r)}} \left\{ \kappa_L^2 \frac{\sigma_0^2}{x'_{i(r),0}x_{i(r),0}} + \sigma_1^2 \frac{\sigma_1^2}{x'_{i(r),1}x_{i(r),1}} \right\}, \tag{A.2}
\]

\[
\text{Var}(\bar{u}_i^{(r)}) = \text{Var}\{E(\bar{l}_i^{(r)} | b^{(r)})\} + E\{\text{Var}(\bar{l}_i^{(r)} | b^{(r)})\}
= \sigma_0^2 + \frac{1}{R} \sum_{r=1}^{R} \left[ \frac{\sigma_0^2}{n_{(r)}} + \frac{1}{n_{(r)}^2} \sum_{i=1}^{n_{(r)}} \left\{ \kappa_L^2 \frac{\sigma_0^2}{x'_{i(r),0}x_{i(r),0}} + \sigma_1^2 \frac{\sigma_1^2}{x'_{i(r),1}x_{i(r),1}} \right\} \right]. \tag{A.3}
\]

Applying the aforementioned method for an unbalanced one-way random effect model (Thomas and Hultquist, 1978; see also Appendix 2 of Uehara et al., 2016), it can be seen that $SS_l^{(B)}$ and $SS_l^{2(W)}$ are approximately chi-squared distributed according to (20). In unbalanced cases, however, their mutual independence is lost.

The formulae for $\hat{u}_i^{(r)}$ can be obtained from (A.1) to (A.3), replacing $l$ with $u$ and $L$ with $U$. The formulae for $\hat{d}_i^{(r)}$ can be derived by dropping $\kappa_L^2$ and replacing $l$ with $d$. 

\textbf{Jpn J Biomet Vol. 37, No. 2, 2016}