In a relative potency assessment, it is necessary to make assumptions about the similarities between substances and their dose-response profile. For example, in a parallel line bioassay which uses the dose-response data within the linear response range, we need to demonstrate that the dose-response slopes of the study substances are approximately parallel. When using multiple animals for testing, it is also crucial to confirm that this parallelism exists not only for the averages but also within each animal (Uehara et al. 2016a).

Meanwhile, when applying a linear mixed effect model to the analysis of parallel line assays, the between-substance difference of the slopes can be treated as a random effect. Thus, under a balanced assay design, we can derive an efficient score test to assess the quantile of the slope difference (McCulloch et al. 2008), which enables us to determine whether the majority of animals have their slope difference within the acceptable range.

We applied this approach to the assessment of intrasubject parallelism with the intention of ameliorating the conservatism of our previous method (Uehara et al. 2016b). We present an example that uses the proposed method, along with the results of simulation studies.

**Key words**: relative potency, parallel-line assay, intrasubject parallelism, efficient score test, random effects quantile.

1. **Introduction**

   It is common to use bioassays to compare the pharmaceutical activities of analog substances (Finney 1978). Typical examples are the slope-ratio assay, which compares the responses at lower doses, and the parallel curve assay, which compares the location of dose-response curves on the abscissa.

   These experiments are conducted to estimate the relative potency (RP), i.e., “the ratio of
equipotent doses of the standard preparation and the unknown preparation under the conditions of the assay” (cited from Chapter 5.3 of the European Pharmacopoeia 2008). 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.

Therefore, in the parallel line bioassay, which uses dose-response data within the linear response range, we need to demonstrate the parallelism between the dose-response slopes. When using multiple animals, it is also crucial to retain parallelism not only in the averages across subjects but also within each animal. Regarding this, Uehara et al. (2016a) stated that “if, for any subject, the two dose-response curves cross, there the log-RP estimate does not make sense, and thus the estimate obtained by averaging all subjects becomes meaningless.” So far, the issue of intrasubject parallelism has not attracted much attention, and the number of proposals to solve this problem is limited.

Meanwhile, when applying the linear mixed effect model (McCulloch et al. 2008) to the analysis of parallel line assays, the between-substance difference of slopes can be treated as a random effect and, under a balanced assay design, we can derive an efficient score test (Cox and Hinkley 1974) regarding the quantile of the slope difference. Using this procedure, we can determine whether the majority of animals have their slope difference within the acceptable range.

We applied this approach to the assessment of intrasubject parallelism with the intention of ameliorating the conservatism of our previous method (Uehara et al. 2016b). In this approach, the authors proposed the use of two one-sided tolerance limits for the intrasubject slope differences. Unfortunately, this moment-based method could not appropriately address the situation in which the slope differences did not vary among subject animals. Therefore, we considered an alternative likelihood-based approach to better model the homogeneity of slope differences. Below we describe our proposed method in detail and present an example of its use.

2. Method

2.1 Assay design and intrasubject parallelism criterion

Below we consider a balanced parallel line assay conducted as an ex vivo experiment. The experiment is designed to evaluate the RP of the test substance against that of the control, using $m$ animals in total. There are $K$ levels of doses ($k = 1, \ldots, K$) for substance $j$ ($j = 0$ for the standard, $j = 1$ for the test), with a common dose ratio separating the adjacent levels. The dose levels are located around the initial estimate of EC$_{50}$ for each substance. At sacrifice, $2 \cdot K$ specimens are extracted from each animal and are randomly allocated to one of the $2 \cdot K$ planned treatments (the dose levels 1 to $K$ for each substance). Thus, after administration of the two study substances (the test and the control), $2 \cdot K$ responses under $2 \cdot K$ different conditions are acquired per block (subject animal).

Note that we cannot choose a very large number of dose levels $K$, because in practice the
maximum number of specimens per subject (or organ) is limited. Instead, we may augment the number of animals, although this will increase the cost of the experiment.

In accordance with Uehara et al. (2016b), we employ the following criterion as the sufficient condition for intrasubject parallelism:

If the designated acceptance interval contains the majority of slope differences, this establishes the intrasubject parallelism.

2.2 Statistical model

The structure of the data is described by the following model:

\[ Y_{i,j,k} = D_i + a_i + (b_i + d_i) \cdot X_k + \varepsilon_{i,j,k}, \quad \varepsilon_{i,j,k} \sim N(0, \sigma_j^2) \]  

(1)

Here the random effects parameters \( D_i, a_i, b_i, d_i \) determine the dose-response profile of subject \( i \). More concretely, \( D_i \) stands for the average response to the standard substance, \( a_i \) for the mean difference between drugs, \( b_i \) for the slope of the standard, and \( d_i \) for the difference between the slopes.

The centralized log-dose \( X_k \) is given by \( X_k = W_{j,k} - \frac{1}{K} \sum_{l=1}^{K} W_{j,l} \), where \( W_{j,k} \) is the actual log-dose at level \( k \) of substance \( j \). We can write the model above using vector and matrix notation:

\[ Y_i = X b_i + \varepsilon_i, \quad \text{where} \quad X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix} = \begin{pmatrix} j_K & 0_K & x \\ j_K & x & 0_K \end{pmatrix}, \quad b_i = \begin{pmatrix} D_i \\ a_i \\ b_i \\ d_i \end{pmatrix}, \]  

(2)

\[ x = \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix} = \left( I_K - \frac{1}{K} J_K \right) w_j, \quad w_j = \begin{pmatrix} W_{j,1} \\ \vdots \\ W_{j,K} \end{pmatrix}, \]  

(3)

\[ \varepsilon_i = \begin{pmatrix} \varepsilon_{i,0} \\ \varepsilon_{i,1} \end{pmatrix} \sim N \left( \begin{pmatrix} 0_{2K} \\ \sigma_0^2 \sigma_1^2 \end{pmatrix} \otimes I_K \right), \quad \text{and} \quad \varepsilon_{i,j} = \begin{pmatrix} \varepsilon_{i,j,1} \\ \vdots \\ \varepsilon_{i,j,K} \end{pmatrix} \]  

(4)

We denote the identity matrix of order \( K \) as \( I_K \), the square matrix of order \( K \) whose elements are all 1 as \( J_K \), the column vector of order \( K \) whose elements are all 1 as \( j_K \), and \( J_K = j_K j_K' \). Also, note that \( j_K' x = 0 \).

According to the convention of linear mixed effect models, we also assume the following:
\[ b_i \sim N(\beta, \Sigma_b), \text{ where } \beta = \begin{pmatrix} \alpha \\ \beta \\ \delta \end{pmatrix}, \Sigma_b = \begin{pmatrix} \sigma^2_D & \sigma_{Da} & \sigma_{Db} & \sigma_{Dd} \\ \sigma_{Da} & \sigma^2_a & \sigma_{ab} & \sigma_{ad} \\ \sigma_{Db} & \sigma_{ab} & \sigma^2_b & \sigma_{bd} \\ \sigma_{Dd} & \sigma_{ad} & \sigma_{bd} & \sigma^2_d \end{pmatrix}. \] 

(5)

### 2.3 Deduction of the proposed efficient score test

#### 2.3.1 Reduction of random effect parameters

Under a balanced design, we can reduce the number of random effect parameters by measuring the difference in response between substances at each dose level:

\[ Y_{i,\text{diff}} = (\mathbf{I}_K - \mathbf{J}_K) \begin{pmatrix} Y_i \\ x \end{pmatrix} \begin{pmatrix} a_i \\ d_i \end{pmatrix} + \varepsilon_{i,\text{diff}}, \text{ where } \varepsilon_{i,\text{diff}} = \varepsilon_{i,1} - \varepsilon_{i,0} \sim N(\mathbf{0}_K, \sigma^2 \cdot \mathbf{I}_K) \text{ and } \sigma^2 = \sigma^2_0 + \sigma^2_1. \] 

Furthermore, we may remove \( a_i \) by centralizing the aforementioned paired differences:

\[ Y_{i,C} = (\mathbf{I}_K - \frac{1}{K} \mathbf{J}_K) Y_{i,\text{diff}} = d_i \cdot x + \left( \mathbf{I}_K - \frac{1}{K} \mathbf{J}_K \right) \varepsilon_{i,\text{diff}}. \] 

(7)

Note that

\[ \text{Var}(Y_{i,C}) = \Sigma_C = \left( \mathbf{I}_K - \frac{1}{K} \mathbf{J}_K \right) \left( \sigma^2 \cdot \mathbf{I}_K + \sigma^2_a \cdot xx' \right) \left( \mathbf{I}_K - \frac{1}{K} \mathbf{J}_K \right). \] 

\[ \therefore \text{rank}(\Sigma_C) = K - 1 < K. \] 

(8)

To remove the collinearity in (7), we drop the dose level \( K \):

\[ Y_{i,D} = \mathbf{A}' Y_{i,\text{diff}} = \begin{pmatrix} Y_{i,\text{diff}} - \frac{1}{K} \sum_{l=1}^{K} Y_{i,\text{diff}}^l \\ \vdots \\ Y_{i,\text{diff}} - \frac{1}{K} \sum_{l=1}^{K} Y_{i,\text{diff}}^l \end{pmatrix} \]

\[ \mathbf{A} = \begin{pmatrix} \mathbf{I}_{K-1} \\ 0_{K-1} \end{pmatrix} - \frac{1}{K} \mathbf{J}_{K-1} \mathbf{J}_{K-1}' \] 

(9)

#### 2.3.2 Likelihood function

Using the derived variable \( Y_{i,D} \) within each animal, the likelihood function of \( \theta = (\delta, \sigma^2, \sigma^2_a)' \) is given as follows:

\[ L(\theta; Y) = \prod_{i=1}^{m} L(\theta; Y_{i,D}) \]

\[ = \frac{1}{(2 \cdot \pi)^{K \cdot m} |\Sigma_D|^{m/2}} \cdot \exp \left( -\frac{1}{2} \sum_{i=1}^{m} (Y_{i,D}^l - \delta \cdot \mathbf{A}' x)' \Sigma_D^{-1} (Y_{i,D}^l - \delta \cdot \mathbf{A}' x) \right) \]

\[ = \frac{1}{(2 \cdot \pi)^{K \cdot m} |\Sigma_D|^{m/2}} \cdot \exp \left( -\frac{1}{2} \sum_{i=1}^{m} (Y_{i,\text{diff}} - \delta \cdot x)' \mathbf{A} \Sigma_D^{-1} \mathbf{A}' (Y_{i,\text{diff}} - \delta \cdot x) \right) \]
Evaluation of Intrasubject Parallelism in Balanced Ex Vivo Bioassay

\[ \frac{1}{(2 \cdot \pi)^{K_m/2} |\Sigma_p|^{m/2}} \cdot \exp \left( -\frac{Q}{2 \cdot \sigma^2} \right). \]  

We can decompose (see Appendix A) the quadratic form \( Q \) as

\[ Q = \text{SSE} + \frac{\sigma^2}{\lambda} \cdot \text{SSX} + \sum_{i=1}^{m} \left( \frac{Y_{i}^{\text{diff}} - \delta}{x^{'x}} \right)^2, \]  

where \( Y_{i}^{\text{diff}} = \frac{1}{m} \sum_{i=1}^{m} Y_{i}^{\text{diff}} \). 

\[ \lambda = \sigma^2 + \sigma_{\delta}^2 \cdot x^{'x} \quad \text{and} \quad \text{SSX} = \sum_{i=1}^{m} \left( x^{'}(Y_{i}^{\text{diff}} - \hat{Y}_{i}^{\text{diff}}) \right)^2 \]  

Note that the expected values of the sums of squares in (11) are \( E(\text{SSE}) = m \cdot (K - 2) \cdot \sigma^2 \) and \( E(\text{SSX}) = (m - 1) \cdot x^{'x} \cdot \lambda. \) As \( x^{'}Y_{i}^{\text{diff}} = (x^{'x}) \cdot d_0 + x^{'}e_{i}^{\text{diff}} \sim N((x^{'x}) \cdot \delta, (x^{'x}) \cdot \lambda), \) we have \( \hat{\delta} = x^{'}Y_{i}^{\text{diff}} / x^{'x} \sim N(\delta, \frac{\lambda}{m \cdot x^{'x}}). \)

### 2.3.3 Efficient score vector, and Hessian and Fisher information matrices

We can calculate the elements of the score vector as the first derivatives of the log-likelihood function:

\[ \frac{\partial l}{\partial \theta} = m \cdot x^{'x} \lambda, \quad \frac{\partial l}{\partial \sigma^2} = \frac{m \cdot (K - 2)}{2 \cdot \sigma^2} - m + \frac{\text{SSE}}{2 \cdot \lambda} + \frac{\text{SSX}}{2 \cdot \lambda^2} \cdot \frac{\partial l}{\partial \lambda}, \]  

\[ \frac{\partial l}{\partial \sigma_{\delta}^2} = \frac{m \cdot (x^{'x})^2}{2 \cdot \lambda^2} \cdot \frac{\partial l}{\partial \lambda}. \]  

The second derivatives then give the Hessian matrix (see Appendix B), and the expected sum of squares gives the following Fisher information matrix:

\[ I(\theta) = m \cdot \begin{pmatrix} \frac{x^{'x}}{\lambda} & 0 & 0 \\ 0 & \frac{K - 2}{2 \cdot \sigma^2} + \frac{1}{2 \cdot \lambda^2} \cdot \frac{x^{'x}}{\lambda} & 0 \\ 0 & 0 & \frac{(x^{'x})^2}{2 \cdot \lambda^2} \end{pmatrix}. \]  

#### 2.3.4 Efficient score test for the quantile

For the inference of the 100 \( \cdot q \% \) quantile of the slope difference \( q = \delta + Z_g \cdot \sigma_d \), we rewrite the previous results using the new parameter vector \( \xi = (q, \sigma^2, \sigma_d)' \). Using the chain rule, we have

\[ \frac{\partial l(\xi; Y)}{\partial q} = \frac{\partial l(\theta; Y)}{\partial \delta} + \frac{2 \cdot \sigma_d}{Z_g} \cdot \frac{\partial l(\theta; Y)}{\partial \sigma_d} \quad \text{and} \quad \frac{\partial l(\xi; Y)}{\partial \sigma_d} = -Z_g \cdot \frac{\partial l(\theta; Y)}{\partial \delta} + \frac{2 \cdot \sigma_d}{Z_g} \cdot \frac{\partial l(\theta; Y)}{\partial \sigma_d}. \]

Therefore,

\[ \frac{\partial l(\xi; Y)}{\partial \xi} = \frac{\partial l(\theta; Y)}{\partial \theta}' \quad \text{where} \quad J = \frac{\partial \theta'}{\partial \xi} = \begin{pmatrix} 1 & 0 & 2 \cdot \sigma_d/Z_g \\ 0 & 1 & 0 \\ -Z_g & 0 & 2 \cdot \sigma_d \end{pmatrix}. \]
The test statistic for the null hypothesis, $H_0$: $q = \kappa$, is given as
\[
T = \frac{s_q}{\sqrt{\text{var}(s_q \mid H_0)}} \quad \text{where} \quad s_q = \left. \frac{\partial l(\xi, Y)}{\partial q} \right|_{\xi = \tilde{\xi}_{H_0}} \cdot \left. \frac{\partial l(\xi, Y)}{\partial \sigma_d} \right|_{\xi = \tilde{\xi}_{H_0}} = 0 \quad \text{and} \quad \tilde{\xi}_{H_0} = \arg \max_{\xi} l(\xi, Y).
\]

Here $\text{var}(s_q \mid H_0)$ is the conditional variance of $s_q$ under the null hypothesis. We can calculate this as $\text{var}(s_q \mid H_0) = m \cdot (i_{1,1} - i_{1,2} \tilde{\xi}_{2,1} \tilde{\xi}_{2,2})$ if the Fisher information matrix $I(\xi) = m \cdot \begin{pmatrix} i_{1,1} & i_{1,2} \\ i_{1,2} & i_{2,2} \end{pmatrix}$ is non-singular (Anderson 1984). Alternatively, if the estimate of $\sigma_d$ is zero and the estimates of $q$ and $\delta$ are identical, then we use $I_{1,1} = m \cdot i_{1,1}$ as $\text{var}(s_q \mid H_0)$.

### 2.3.5 Proposed procedure for intrasubject parallelism

Given the acceptance interval $(\kappa_L, \kappa_U)$, we can restate the intrasubject parallelism criterion (Subsection 2.1) as $\Pr(\kappa_L < d_i \leq \kappa_U) > 1 - \gamma$. We can demonstrate this using the following two tests, both with a type 1 error rate of $\alpha$:

- [L] $H^L_0$: $q_{1/2} = \kappa_L$ vs. $H^L_1$: $q_{1/2} > \kappa_L$
- [U] $H^U_0$: $q_{1-\gamma/2} = \kappa_U$ vs. $H^U_1$: $q_{1-\gamma/2} < \kappa_U$

If both tests reject the null hypotheses, according to the alternative hypotheses $H^L_1$ and $H^U_1$, we can claim $\Pr(\kappa_L < d_i \leq \kappa_U) = \Pr(d_i \leq \kappa_U) - \Pr(d_i \leq \kappa_L) > 1 - \gamma$, which satisfies the intrasubject parallelism criterion above (Figure 1).

Because this procedure is a sort of intersection-union test, it keeps the overall type 1 error below $\alpha$ (Dmitrienko and D’Agostino 2013).

### 3. Example

We applied our proposed method to the assay results in Table 1 of Uehara et al. (2016b), which presented the results of a parallel line assay conducted using 48 animals in total ($m = 48$). They employed five dose levels ($K = 5$), with a dose ratio of around $\sqrt{2}$, which gives

\[
x = (-\log_{10}(2), \ -\log_{10}(2)/2, \ 0, \ \log_{10}(2)/2, \ \log_{10}(2))' = (-0.3010, \ -0.1505, \ 0, \ 0.1505, \ 0.3010)'.
\]

The mean slope difference $\delta = x' \tilde{Y}_{\cdot j} / x' x$ was calculated to be $-0.0237$. The $SSX$ was
Evaluation of Intrasubject Parallelism in Balanced Ex Vivo Bioassay

Fig. 1. Intrasubject parallelism criteria based on the slope difference distribution. Parallelism is deemed to exist between substances when the slope difference falls within the acceptance range, $(\kappa_L, \kappa_U)$, indicated by the two vertical solid lines. When both the lower and upper $\gamma/2$ quantile points of the slope difference distribution 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 $\gamma/2$ quantiles)

calculated to be 0.2405 using the following formula:

$$SSX = \sum_{i=1}^{m} \left\{ x'(Y_i^{\text{diff}} - \bar{Y}^{\text{diff}}) \right\}^2 = \sum_{i=1}^{m} \left\{ x' \cdot \left( \frac{x'Y_i^{\text{diff}}}{x'x} - \frac{x'\bar{Y}^{\text{diff}}}{x'x} \right) \right\}^2$$

$$= \sum_{i=1}^{m} \left( x' \cdot (\Delta \text{slope}_i - \bar{\Delta \text{slope}}) \right)^2,$$

where $\Delta \text{slope}_i = \text{slope}_i^B - \text{slope}_i^A$ and $\bar{\Delta \text{slope}} = \frac{1}{m} \cdot \sum_{i=1}^{m} \Delta \text{slope}_i$.

Here $\text{slope}_i^A$ and $\text{slope}_i^B$ are the standard and test slopes, respectively, observed in subject $i$. As Uehara et al. (2016b) did not include the value of the $SSE$, we used their error variance estimates instead, 0.0124 for control substance A and 0.0102 for test substance B (which can be seen in the final line of Subsection 1.2 in the aforementioned literature), to estimate the $SSE$ as $SSE = 48 \cdot (5 - 2) \cdot (0.0124 + 0.0102) = 3.2544$.

For the likelihood maximization that restricted the variance components to non-negative values, we used the `nlminb` function from the R statistical programming language (R Core Team 2017). The maximum likelihood estimates based on these statistics were $\hat{\delta} = -0.0237$, $\hat{\sigma}^2 = 0.0225$, and $\hat{\sigma}_d^2 = 0.0000$.

Jpn J Biomet Vol.39, No.1, 2018
Applying our proposed method with the minimum content probability \(1 - \gamma = 0.8\) and the acceptance interval \((\kappa_L, \kappa_U) = (-0.5, 0.5)\), we obtained the following results:

[L] Testing \(H^L_0: q_{\gamma/2} = \kappa_L = -0.5\)

Under the null hypothesis, the maximum likelihood estimates were \(\hat{\sigma}^2 = 0.0211\) and \(\hat{\sigma}_d = 0.2971\), which gave a test statistic of \(T = 2.615\) \((P = 0.0044)\).

[U] Testing \(H^U_0: q_{0.9} = \kappa_U = 0.5\)

Under the null hypothesis, the maximum likelihood estimates were \(\hat{\sigma}^2 = 0.0211\) and \(\hat{\sigma}_d = 0.3220\), which gave a test statistic of \(T = -2.835\) \((P = 0.0023)\).

In summary, the rejection of both null hypotheses demonstrated the intrasubject parallelism. This result shows that more than 80% of the subject animals had a slope difference between \(-0.5\) and \(0.5\), at the 95% confidence level.

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 example above, with 10,000 iterations. Here we focused on the size of each one-sided test and the overall power of the procedure to determine the intrasubject parallelism. As before, we used R for the computation.

Simulation parameters

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

Centralized log-dose:
\[ x = \log_{10} 2 \cdot (-1, -0.5, 0, 0.5, 1)' \]
\((x'x \approx 0.2265)\),

Mean slope difference:
\(\delta \in \{0, 0.0, 0.3, 0.5\}\),

Acceptance range of slope difference:
\((\kappa_L, \kappa_U) = (-0.5, 0.5)\),

Minimum proportion of acceptable subjects:
\(1 - \gamma = 0.8\),

Number of animals:
\(m = 12, 24, \text{ or } 48\),

Variance of the slope difference:
\(\sigma_d \in \{0, 0.3/Z_{1-\gamma/2}, 0.5/Z_{1-\gamma/2}\}\),

Error variance:
\(\sigma_0^2 = \sigma_1^2 = 0.01\),

Nominal type 1 error rate of each test:
\(\alpha = 0.05\).

In section 3, we estimated the parameters as \(\hat{\delta} \approx 0\), \(\hat{\sigma}^2 \approx 0.02\), and \(\hat{\sigma}_d^2 = 0.0\); thus, this simulation reflects the experimental conditions. As noted above, homogeneity in the slope differences is represented by \(\sigma_d = 0.0\), and the empirical size of each one-sided test will be obtained for \(\delta = 0\) and \(\sigma_d = 0.5/Z_{1-\gamma/2}\).
Evaluation of Intrasubject Parallelism in Balanced Ex Vivo Bioassay

Table 1. Empirical size and power of the proposed tests

| δ   | σ₂ | m   | Lower | Upper | Lower & Upper |
|-----|----|-----|-------|-------|---------------|
| 0.0 | 0.4151 | 0.4048 | 0.2851 |
| 24  | 0.8225 | 0.8210 | 0.7260 |
| 48  | 0.9889 | 0.9897 | 0.9800 |
| 0.3 | 0.1340 | 0.1384 | 0.0630 |
| 24  | 0.2998 | 0.3118 | 0.1656 |
| 48  | 0.5741 | 0.5789 | 0.4028 |
| 0.5 | 0.6206 | 0.6310 | 0.0086 |
| 24  | 0.6174 | 0.6372 | 0.0067 |
| 48  | 0.6489 | 0.6481 | 0.0069 |
| 0.0 | 0.8876 | 0.9015 | 0.0915 |
| 24  | 0.9998 | 1.1457 | 0.1457 |
| 48  | 1.0000 | 2.3400 | 2.3400 |
| 0.3 | 0.5370 | 0.5189 | 0.0188 |
| 24  | 0.9311 | 0.9129 | 0.0129 |
| 48  | 0.9996 | 0.0075 | 0.0075 |
| 0.5 | 0.1852 | 0.0039 | 0.0037 |
| 24  | 0.4618 | 0.0006 | 0.0006 |
| 48  | 0.7797 | 0.0000 | 0.0000 |
| 0.0 | 0.9854 | 0.0275 | 0.0275 |
| 24  | 1.0000 | 0.0268 | 0.0268 |
| 48  | 1.0000 | 0.0264 | 0.0264 |
| 0.3 | 0.8226 | 0.0048 | 0.0048 |
| 24  | 0.9974 | 0.0013 | 0.0013 |
| 48  | 1.0000 | 0.0003 | 0.0003 |
| 0.5 | 0.4075 | 0.0004 | 0.0004 |
| 24  | 0.8338 | 0.0000 | 0.0000 |
| 48  | 0.9911 | 0.0000 | 0.0000 |

Results

In the “Lower” or “Upper” columns of Table 1, the empirical sizes of each one-sided test are shown in the rows with italic font (at δ = 0 and σ₂ = 0.5/Z₁−γ/2). All were less than 0.05, which indicates the conservative nature of the proposed method. The rates approached the nominal value as the number of subjects increased.

The “Lower & Upper” column in Table 1 displays the empirical power of the proposed procedure. The increase in the mean and variance of the slope difference rapidly diminished the power, which reached 80% only when δ = 0, σ₂ = 0, and m = 48. This result implies that in practice the proposed method is applicable only when the slopes are identical and homogeneous. It seems that further work is needed to address the conservatism under small sample sizes.
5. Discussion

Despite its essential importance in bioassays, the assessment of intrasubject similarity between dose-response curves is not usually considered. Significance testing of the second-order interaction between subject, substances, and dose, which does not offer any explicit assurance regarding the validity of the RP estimation at the subject level, is often the only measure used to test the validity of the experimental assumptions. In an attempt to address this issue, Uehara et al. (2016a) developed the Intrasubject Parallelism Criterion (ISP), an analog of the aggregated Individual Bioequivalence Criterion proposed by the Food and Drug Administration. 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. As an alternative, Uehara et al. (2016b) derived an approach via tolerance limit estimation (Brown, Iyer and Wang 1997) to evaluate the intra-subject slope ratio, assuming heterogeneity of the slope and the slope difference. A simulation study showed that this approach could work well under the presumed heterogeneity in the slope difference, but can be conservative under the assumption of homogeneity even with moderate sample sizes.

As yet another approach, here we introduced an efficient score test for the quantile of the random slope difference, based on the conventional linear mixed effect model, which we used to determine whether the majority of animals have their slope difference within the acceptable range. According to our simulation study, the proposed procedure may be appropriate for moderate sample sizes. For small sample sizes, however, we could not completely solve the issue of conservatism, which suggests that further consideration is required.

The Monte Carlo simulation in Subsection 4 was conducted to aid understanding of the example situation. The scope of this simulation study was narrow; thus, before applying our proposed method to a different situation, an additional more comprehensive simulation study may be necessary.

Other limitations of our approach include our assumption in equation (1) of a Gaussian distribution for the slope differences. We may alleviate this assumption by considering other distributions for the random effect parameters. The error terms were also assumed to be Gaussian, as well as homoscedastic across the animals and the dose levels. In practice, these assumptions may require a variance-stabilizing transformation. Another possible issue is the choice of the acceptance interval, which should be decided according to the precision of the assay system. We need to further consider this aspect of our model as well.

The example dataset originally came from two different experiments (Uehara et al. 2016b). Although the heterogeneity of the slope difference turned out to be negligible, this is not exactly in accord with the proposed method. We applied our method to the data in a retrospective manner, to demonstrate the utility of the proposed method in practice.

In our view, these efficient score based approaches are quite versatile and it is worthwhile
pursuing their possible uses. They could also be useful for deriving the sample size and the confidence interval formulae, which are potential objectives of our future work. As is mentioned in our previous work (Uehara et al. 2016a), we are also interested in the intrasubject similarity in the parallel curve analysis, which requires not only parallelism over the linear response range but also similarities in the responses at both asymptotes. The extension to nonlinear models may be cumbersome, but the need for such models is clear.

Note that standardization is common practice in bioassays. The response range of each sample is determined before the administration of the study substance, which can be used to derive the response variable with homoscedasticity. Although omitted in this paper, it is straightforward to extend our method to utilize the additional measurements as covariates in an ANCOVA-type analysis. This may further improve the precision of estimates, and may contribute to reducing the number of sacrificed animals.

Regarding the inference of the random effect’s quantile, there are further implications. Whenever the treatment effect is deemed to be heterogeneous, there are concerns about qualitative interactions, which are an implicit but ubiquitous issue accompanying the mixed effect models. For example, the ICH E9 guideline (3.2 “Multicentre Trials”) (ICH Harmonised Tri-partite Guideline 1998) briefly considered the use of mixed models to explore the heterogeneity of the treatment effect, in which the center and treatment-by-center effects were considered to be random, provided that there were a sufficiently large number of sites. Here the existence of qualitative interaction implies the non-negligible proportion of clinical sites which are under risk of negative treatment effect. To show the general applicability of the treatment, it should be useful to demonstrate the pre-determined lower quantile of treatment effect being non-negative, where our proposed testing procedure might be useful.

We conclude that the proposed procedure can offer simple and easily interpretable results regarding the intrasubject parallelism. Although our approach is still overly conservative for small sample sizes, it is suitable for a confirmatory analysis when a rigorous evaluation is required.

Acknowledgements

We thank the two referees for their insightful comments.

REFERENCES

Anderson T.W. (1984). An Introduction to Multivariate Statistical Analysis, 2nd ed., Wiley, Canada.

Brown, Iyer & Wang (1997). Tolerance intervals for assessing individual bioequivalence. Statistics in Medicine, 16, 803–820.

Cox D.R., Hinkley D.V. (1974). Theoretical Statistics, Chapman and Hall, UK.
Appendix A

From (9), we have

\[ |A' A| = |I_{K-1}| \left(1 - \frac{1}{K} j'_{K-1} j_{K-1} \right) = \frac{1}{K} \left(A' A\right)^{-1} = I_{K-1} + j_{K-1}, \tag{A.1} \]

\[ A(A' A)^{-1} A' = A(I_{K-1} + j_{K-1}) A' = A(I_{K-1}, -j_{K-1}) = (A, -A j_{K-1}) \]

\[ = \left( \frac{I_{K-1} - \frac{1}{K} j_{K-1}}{\frac{1}{K} j_{K-1}} \right) \right) = I_{K-1} - \frac{1}{K} j_{K} \]. \tag{A.2} \]

Denote \( \text{Var}(Y^D) = \Sigma_D = A' \text{Var}(Y^\delta) A = \sigma^2 A' x x' A + \sigma^2 \cdot A' A. \) Then

\[ |\Sigma_D| = |\sigma^2 A' x x' A + \sigma^2 \cdot A' A| = |\sigma^2 A' A| \cdot (1 + \sigma^2 \cdot x' A(\sigma^2 \cdot A' A)^{-1} A' x) \]

\[ = \sigma^{2(K-1)} \cdot A' A \cdot \left(1 + \frac{\sigma^2}{\sigma^2 \cdot x' A(A' A)^{-1} A' x} \right) \]

\[ = \sigma^{2(K-2)} \cdot \frac{\sigma^2 + \sigma^2 \cdot x' x}{K}, \tag{A.3} \]

\[ \Sigma_D^{-1} = (\sigma^2 \cdot A' x x' A + \sigma^2 \cdot A' A)^{-1} \]

\[ = \sigma^{-2} \cdot \left\{ (A' A)^{-1} - \frac{\sigma^2 \cdot (A' A)^{-1} x' x (A' A)^{-1}}{\sigma^2 + \sigma^2 \cdot x' x} \right\} . \tag{A.4} \]

\[ Q = \sigma^2 \cdot \sum_{i=1}^{m} (Y^{\delta_i} - \delta \cdot x)' A \Sigma_D^{-1} A (Y^{\delta_i} - \delta \cdot x) \]
\[ SSE = \frac{\sigma^2 \cdot SSX}{\lambda \cdot x'} + m \frac{\sigma^2 \cdot x'x}{\lambda} \left( \frac{x'Y_{\text{diff}}}{x'x} - \delta \right)^2. \]
Appendix B

From (10) and (11), we have

\[
\begin{align*}
\frac{\partial^2 l(\theta; Y)}{\partial \delta^2} &= \frac{-m \cdot x'x}{\lambda}, \\
\frac{\partial^2 l(\theta; Y)}{\partial (\sigma^2) \partial \delta} &= \frac{-m \cdot x'x}{\lambda^2} (\delta - \hat{\delta}), \\
\frac{\partial^2 l(\theta; Y)}{\partial (\sigma^2) \partial \hat{\delta}} &= \frac{-m \cdot (x'x)^2}{\lambda^2} (\delta - \hat{\delta}), \\
\frac{\partial^2 l(\theta; Y)}{\partial \sigma^2 \partial \sigma^2} &= \frac{m \cdot (K - 2)}{2 \cdot \sigma^4} + \frac{m}{2 \cdot \lambda^2} - \frac{SSX}{\sigma^6} \frac{SSX}{\lambda^3} \frac{m \cdot x'x}{\lambda^3} \cdot (\hat{\delta} - \delta)^2, \\
\frac{\partial^2 l(\theta; Y)}{\partial (\sigma^2)^2} &= \frac{m \cdot (x'x)^2}{2 \cdot \lambda^2} - \frac{SSX \cdot x'x}{\lambda^4} - \frac{m \cdot (x'x)^3}{\lambda^3} \cdot (\hat{\delta} - \delta)^2, \\
\text{and} \quad \frac{\partial^2 l(\theta; Y)}{\partial (\sigma^2)^2} &= \frac{m \cdot xx}{2 \cdot \lambda^2} - \frac{SSX}{\lambda^3} - \frac{m \cdot (x'x)^2}{\lambda^3} \cdot (\hat{\delta} - \delta)^2. 
\end{align*}
\]