A New Procedure of One-Sided Test in Clinical Trials with Multiple Endpoints

Yoshiomi Nakazuru*1,*2, Takashi Sozu*3, Chikuma Hamada*4 and Isao Yoshimura*4
*1 Clinical Statistics, Pfizer Japan Inc.
*2 Department of Mathematical Information Science, Graduate School of Science, Tokyo University of Science
*3 Department of Biostatistics, Kyoto University School of Public Health
*4 Department of Management Science, Faculty of Engineering, Tokyo University of Science

This paper considers clinical trials with multiple endpoints, in which the efficacy of a test treatment is confirmed only when the superiority of the test treatment to control is evidenced in at least one endpoint and non-inferiority is observed in the remaining endpoints. Perlman and Wu (2004) proposed a one-sided testing procedure that was adaptable to this type of trials. This paper proposes a modification of this procedure, in which the likelihood ratio test is replaced with another test similar to that proposed by Tang et al. (1989). The performance of the proposed procedure was examined through theoretical consideration and Monte Carlo simulations assuming normality and homoscedasticity. The simulation study demonstrated that the power of the proposed procedure was higher than that of the procedure proposed by Perlman and Wu; in this procedure, type I error rates are maintained within nominal significance levels unless primary endpoints are highly correlated.

Key words: Clinical trial; Multiple endpoints; Multivariate normal distribution; Non-inferiority; Superiority.

1. Introduction

In confirmatory clinical trials, it is advisable to select only one primary endpoint for evaluating the efficacy of a test treatment (Lewis et al., 1999). However, in reality, there are many cases where multiple primary endpoints, say p endpoints, are required as exemplified by Offen et al. (2007). This paper deals with a case where the efficacy of the test treatment was confirmed only when the superiority of the test treatment to control was evidenced in at least one endpoint and non-inferiority was observed in the remaining endpoints.

For the above mentioned type of trials, Perlman and Wu (2004) proposed a testing procedure (Perlman-Wu procedure) that was modified from the testing procedure proposed by Bloch et al. (2001) (Bloch et al. procedure). In this modification, Hotelling’s $T^2$ statistic was replaced...
with a one-sided likelihood ratio statistic. Since the superiority of the Perlman-Wu procedure to the procedures previously proposed in the literature was confirmed (Perlman and Wu, 2004) under the assumption of normality and homoscedasticity, we re-examined the characteristics of the Perlman-Wu procedure and found scope for improvement. Thereafter, we devised a new procedure (the proposed IUT) that may possibly be superior to the Perlman-Wu procedure and examined its performance. In the text that follows, we will explain the devised procedure and discuss the results of the examination.

2. Notations and Hypotheses

Consider a clinical trial with 2 treatment groups comprising \( n_1 \) and \( n_2 \) subjects, respectively. Without loss of generality, assume that the first group corresponds to a test treatment and the second to the control treatment and that the efficacy is recognized when the response in the former is greater than the latter.

Let \( Y_{ijk} \) denote the response of the \( k \)th subject to the \( i \)th treatment at the \( j \)th endpoint \((i = 1, 2; j = 1, \ldots, p; k = 1, \ldots, n_i)\), \( \bar{Y}_{ij} \) the sample average of the number of \( j \)th endpoints to the \( i \)th treatment, and set \( X = (X_1, \ldots, X_p)^t \) with \( X_j = (\bar{Y}_{1j} - \bar{Y}_{2j}) \), where the superscript \( t \) denotes transpose.

Assume that \( Y_{ik} = (Y_{i1k}, \ldots, Y_{ipk})^t \) \((i = 1, 2; k = 1, \ldots, n_i)\) denotes independently and normally distributed random vectors with mean vector \( \eta_i = (\eta_{i1}, \ldots, \eta_{ip})^t \) and covariance matrix \( \Sigma \), where \( \Sigma = \{\sigma_{ij}\} \) is positive definite. Let \( \mu = \eta_1 - \eta_2 = (\mu_1, \ldots, \mu_p)^t \); then, \( X \) is normally distributed with mean vector \( \mu \) and covariance matrix \((n_1^{-1} + n_2^{-1})\Sigma\).

Let \( \hat{\Sigma} \) denote the pooled sample covariance matrix, then \( S \equiv (n_1 + n_2 - 2)\hat{\Sigma} \) is distributed as a Wishart distribution \( W_p(\Sigma, n_1 + n_2 - 2) \) when \( n_1 + n_2 > p + 2 \). We assume that this condition would be applicable throughout this paper.

Consider a pair of null and alternative hypotheses expressed by

\[
H_0: \left\{ \max_{1 \leq j \leq p} \mu_j \leq 0 \right\} \cup \left\{ \min_{1 \leq j \leq p} (\mu_j + \epsilon_j) \leq 0 \right\} \text{ versus } H_1: \text{not } H_0, \tag{1}
\]

where \( \epsilon \) denotes prespecified positive constants. By defining subhypotheses “\( H_0^{(0)}: \max_{1 \leq j \leq p} \mu_j \leq 0 \)” and “\( H_0^{(j)}: \mu_j \leq -\epsilon_j \) \((j = 1, \ldots, p)\)”, \( H_0 \) is expressed as shown in Equation (2) which is adaptable to an intersection-union test (IUT) (Berger, 1982).

\[
H_0 \equiv H_0^{(0)} \cup \left\{ H_0^{(1)} \cup \cdots \cup H_0^{(p)} \right\} \tag{2}
\]

Figure 1 shows the sets of \( \mu \) that correspond to the respective hypotheses when \( p = 2 \) and \( \epsilon_1 = \epsilon_2 = 0.5 \).

For the convenience of explanation, we further define two subhypotheses: “\( H_{00}: \mu_j = 0 \) \((j = 1, \ldots, p)\)” and “\( H_{11}: \mu_j \geq 0 \) \((j = 1, \ldots, p)\) and \( \max_{1 \leq j \leq p} \mu_j > 0 \)”.

Jpn J Biomet Vol. 35, No. 1, 2014
3. Proposal of a New Procedure

3.1 Perlman-Wu procedure

Perlman-Wu procedure is an IUT incorporating size-$\alpha$ one-sided $t$-tests for testing $H_0^{(j)}$ ($j = 1, 2, \ldots, p$) and a size-$\alpha$ one-sided likelihood ratio test (LRT) proposed by Perlman (1969) for testing $H_0^{(0)}$. It rejects $H_0$ if and only if

$$\|X - \pi_S(X; N^p)\|_S^2 > c^*_\alpha$$

and

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} (X_j + \epsilon_j) / \sqrt{\hat{\sigma}_{jj}} > t_\alpha \quad \text{for } j = 1, \ldots, p. \tag{4}$$

Here $N^p \equiv \{ (\mu_1, \ldots, \mu_p) \mid \max_{1 \leq j \leq p} \mu_j \leq 0 \}$ is the non-positive orthant in $\mathbb{R}^p$ ($p$-dimensional Euclidean space), $\|x\|_S^2 \equiv x^T S^{-1} x$ is the Euclidean norm determined by $S$, $\pi_S(X; N^p)$ is the projection of $X$ onto $N^p$ with respect to this norm, $\hat{\sigma}_{jj}$ is the $j$th diagonal element of $\hat{\Sigma}$, $t_\alpha$ is the upper $\alpha$ quantile of $t$-distribution with the degrees of freedom of $\hat{\Sigma}$, and $c^*_\alpha$ is the critical value of size-$\alpha$ one-sided LRT for $H_0^{(0)}$ determined by

$$\alpha = \frac{1}{2} Pr \left[ \frac{\chi^2_{p-1}}{\chi^2_{n_1 + n_2 - p}} > c^*_\alpha \right] + \frac{1}{2} Pr \left[ \frac{\chi^2_p}{\chi^2_{n_1 + n_2 - p - 1}} > c^*_\alpha \right], \tag{5}$$

where $\chi^2_q$ is a chi-square variable with $q$ degrees of freedom.

Perlman and Wu (2004) illustrated the advantage of the Perlman-Wu procedure over the Bloch et al. procedure showing Figure 2, which indicates the difference in rejection regions between the Perlman-Wu (solid line) and Bloch et al. (broken line) procedures. The difference essentially resulted from the fact that $T^2$ test was not one-sided while the Equation (3) was one-sided. For example, when the correlation coefficient is extremely negative, as shown in Figure 2 (c), a part of sample space in the negative quadrant is included in the rejection region of the Bloch et al. procedure, while such a region disappears in the Perlman-Wu procedure.

However, one issue needs to be addressed in the practical applications of the Perlman-Wu procedure. The critical value provided by Equation (5) corresponds to the supremum of possible choice of type I error rates among various covariance matrices, which automatically makes the Jpn J Biomet Vol. 35, No. 1, 2014
Perlman-Wu procedure less powerful.

We think that this procedure might be improved if Equation (3) in the Perlman-Wu procedure was replaced with an approximate testing procedure. To execute this idea, we utilized a hint drawn from Glimm et al. (2002), which extended the approximate likelihood ratio test (ALRT) in Tang et al. (1989). We briefly explain Tang et al.’s and Glimm et al.’s ALRTs in succeeding subsections which are followed by the introduction of the proposed procedures.

3.2 Tang et al.’s ALRT

When $\Sigma$ is known, there is a positive definite matrix $A$ such that $A^tA = \Sigma^{-1}$; then, the statistic $u \equiv (u_1, \ldots, u_p)^t = \sqrt{\frac{n_1n_2}{n_1+n_2}}AX$ is distributed as a $p$-variate normal distribution with mean $\sqrt{\frac{n_1n_2}{n_1+n_2}}A\mu$ and covariance matrix $I$ (the identity matrix).

Note that $A$ is not uniquely determined as discussed by Glimm et al. (2002), but for simplicity, we propose to use the set of eigen vectors multiplied by the square root of corresponding eigenvalue to represent $A$ henceforth in this paper.

With this transformation, Tang et al. (1989) proposed a one-sided ALRT of significance level $\alpha$ for accepting $H_{11}$ as the alternative to $H_{00}$; it rejected $H_{00}$ if and only if

$$\bar{u}^2 \equiv \sum_{j=1}^{p} \max(u_j,0)^2 > c,$$

where $c$ is determined by

$$\sum_{j=0}^{p} \frac{p!}{j!(p-j)!} \frac{1}{2^p} Pr (\chi_j^2 > c) = \alpha.$$

$\chi_j^2$ denotes the $\chi^2$ distribution with $j$ degrees of freedom, and $\chi^2(0)$ is defined as the constant zero.

One issue to be addressed in the practical application of Tang et al.’s ALRT is that the covariance matrix is assumed to be known and, consequently, must be replaced with its estimate along with the critical value.
3.3 Glimm et al.’s ALRT

Glimm et al. (2002) addressed this issue by introducing another one-sided ALRT (Glimm et al.’s ALRT) based on the sum of the squares of deviation. We briefly describe Glimm et al.’s ALRT with a minor modification, namely replacing the sum of squares with the sample covariance matrix, in order to facilitate the comparison with Tang et al.’s ALRT.

Define the statistic $G$ as (Srivastava, 2002)

$$G = \frac{1}{n_1 + n_2 - 1} \left( \sum_{k=1}^{n_1} (Y_{1k} - \hat{\mu})(Y_{1k} - \hat{\mu})^t + \sum_{k=1}^{n_2} (Y_{2k} - \hat{\mu})(Y_{2k} - \hat{\mu})^t \right),$$

(8)

where $\hat{\mu} = (n_1 \bar{Y}_1 + n_2 \bar{Y}_2)/(n_1 + n_2)$. Since we assumed $n_1 + n_2 > p + 2$, $G$ is positive definite with probability 1, and there exists a positive definite matrix $A$ such that $A^tA = G^{-1}$.

Let $u \equiv (u_1, \ldots, u_p)^t = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} A X$. On the basis of this transformation, Glimm et al. (2002) proposed a one-sided ALRT of significance level $\alpha$, which rejected $H_{00}$ if and only if

$$\bar{u}^2 \equiv \sum_{j=1}^{p} \max(u_j,0)^2 > c,$$

(9)

where $c$ is determined by

$$\frac{1}{2^p} \sum_{j=0}^{p-1} \frac{p!}{j!(p-j)!} B(\frac{p-j}{2}, \frac{n_1 + n_2 - 1}{2} - 1) \int_0^{\frac{c}{n_1 + n_2 - 1} - 1} r^{\frac{p-j}{2} - 1} (1 - r) \left(\frac{n_1 + n_2 - 1 - p + j}{2}\right)^{\frac{p-j}{2} - 1} dr + \frac{1}{2^p} = 1 - \alpha,$$

(10)

where $B(\bullet, \bullet)$ is a beta function defined with gamma functions as $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

3.4 Disadvantageous property of Glimm et al.’s ALRT

Glimm et al.’s ALRT has a disadvantageous property, in common with Tang et al.’s ALRT, for testing $H_{00}$ versus $H_{11}$ when correlation coefficients among the endpoints are positive. Glimm et al.’s ALRT is not irrational if the test statistic is constructed in the sample and parameter spaces for $X$ and $\mu$ as may be realized in the case of zero correlation. In fact, it was constructed in the sample space for $u$, where the positive orthant for transformed parameters $\sqrt{\frac{n_1 n_2}{n_1 + n_2}} A \mu$ does not correspond to the positive orthant in the space for $\mu$ (Silvapulle, 1997). There is a slight discrepancy between the one-sidedness realized in Glimm et al.’s ALRT and the intended one-sidedness for testing $H_{00}$ versus $H_{11}$.

Let us explain this issue using a simple numerical example (we provide the explanation of more general cases in Appendices A and B). Assume that $p = 2$, $n_1 = n_2 = 100$, and the estimated standard deviations and correlation coefficient are $s_1 = s_2 = 1$ and $r = 0.8$, respectively.

Then $G$ and $A$ are as follows:

$$G = \begin{pmatrix} 1.00 & 0.80 \\ 0.80 & 1.00 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 1.49 & -0.75 \\ -0.75 & 1.49 \end{pmatrix}$$

(11)

Figure 3 (a) illustrates the relationship between the points and curves in the spaces for $X$. 

Jpn J Biomet Vol. 35, No. 1, 2014
(left) and \( u \) (right) under the condition that \( s_1 = 1, \ s_2 = 1, \) and \( r = 0.8 \) are given. Since the transformation is linear, lines in \( X \)-space are transformed to lines in \( u \)-space and a circle with the center at the origin \((0,0)\) is transformed to an ellipse with the center at the origin, and vice versa.

The curve that satisfies \( \bar{u}^2 = c \) in \( u \)-space is composed of a quarter arc and two lines smoothly connected with the arc as is shown in Figure 3 (a) (right). When this curve in \( u \)-space is transformed back into \( X \)-space, the trajectory is a part of an ellipse and two lines connected smoothly with the ellipse as is shown in Figure 3 (a) (left).

The rejection region of Gimm et al.’s ALRT is outside (upper right side) of this curve in \( u \)-space, which is equivalently the corresponding outside area in \( X \)-space. As is seen in Figure 3 (a) (left), a part of rejection region may be inside the third quadrant, which is disadvantageous for testing not \( H_{00} \) but \( H_0^{(0)} \) in our intention.

Note that this disadvantageous property disappears when \( r \) is negative. Moreover, this property is maintained even when \( p > 2 \), as is explained in Appendices A and B for the case of intraclass correlation structure.

### 3.5 Proposed ALRT

We modify the transformation matrix and the test statistic in Glimm et al.’s ALRT in order to improve the above-mentioned disadvantage of their method. The detail of the modification is as follows.

Let \( B \) be the matrix substituting the off-diagonal elements of \( A \) with their absolute values, e.g.,

\[
B = \begin{pmatrix}
1.49 & 0.75 \\
0.75 & 1.49
\end{pmatrix}
\quad \text{for} \quad A = \begin{pmatrix}
1.49 & -0.75 \\
-0.75 & 1.49
\end{pmatrix}.
\]  

Consider the two transformations such that

\[
u_A \equiv (u_{A1}, \ldots, u_{Ap})^t = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} AX \quad \text{and} \]

\[
u_B \equiv (u_{B1}, \ldots, u_{Bp})^t = \left( \frac{\det A}{\det B} \right)^{2/p} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} BX.
\]

In these transformations, the term including the ratio of determinants is a correction factor for equating the scale of enlargement, although it is unity for \( p = 2 \). The meaning of the transformation by \( B \) is illustrated in Figure 3 (b). It represents the correspondence between \( X \)-space and \( u \)-space induced by the transformation matrix \( B \) under the same condition as Figure 3 (a), i.e., the correlation coefficient between the two variables is positive. Using these transformations, we propose a “modified ALRT” that rejects \( H_0^{(0)} \) if and only if

\[
\min(\bar{u}_A^2, \bar{u}_B^2) > c,
\]

where \( \bar{u}_A^2 \) and \( \bar{u}_B^2 \) are defined by Equations (16) and (17), respectively, and \( c \) has the same value as the one determined by Equation (10).
A New Procedure of One-Sided Test in Clinical Trials with Multiple Endpoints

(a) Correspondence of $X$-space (left) and $u$-space (right) transformed by the matrix $A = \begin{pmatrix} 1.49 & -0.75 \\ -0.75 & 1.49 \end{pmatrix}$

(b) Correspondence of $X$-space (left) and $u$-space (right) transformed by the matrix $B = \begin{pmatrix} 1.49 & 0.75 \\ 0.75 & 1.49 \end{pmatrix}$

(c) The boundary of the rejection region of the proposed ALRT given $A$
Point $Q_i$ corresponds to Point $P_i$ ($i = A1, A2, A3, B1, B2, B3$) and Line $M_j$ corresponds to Line $L_j$ ($j = A2, A3, B2, B3$). Rejection region of Gimm et al.'s ALRT lies outside the curve in Figure 3 (a) (left). Rejection region of the proposed ALRT lies outside the curve in Figure 3 (c).

Fig. 3. Relationship between $X$-space and $u$-space given $A$. 
\[ \hat{u}_A^2 = \sum_{j=1}^{p} \max(u_{A_j}, 0)^2 \] (16)
\[ \hat{u}_B^2 = \sum_{j=1}^{p} \max(u_{B_j}, 0)^2 \] (17)

The reason why this newly proposed ALRT may be more advantageous than Glimm et al.’s ALRT in terms of one-sidedness can also be explained with the above example. Since the proposed ALRT rejects \( H_0^{(0)} \) when the both of \( \hat{u}_A^2 \) and \( \hat{u}_B^2 \) are greater than the constant \( c \), the rejection region in \( X \)-space is the joint region of two areas lying outside the curve in Figure 3 (a) (left) and the curve in Figure 3 (b) (left), i.e., the area lying outside (upper right side) the curve in Figure 3 (c), whereas the rejection region of Glimm et al.’s ALRT is the area lying outside the curve in Figure 3 (a) (left). The joint rejection area for the proposed ALRT does not include the disadvantageous region shown as the outside area of the curve in Figure 3 (a) (left).

Although it is difficult to visualize the relationship of \( X \) and \( u \) for \( p > 2 \), the relationship between the proposed ALRT and Glimm et al.’s ALRT in three or higher dimensions is similar to that in the two-dimensional case, as explained in Appendix B.

### 3.6 Proposed IUT

This paper proposes a new IUT (proposed IUT) that replaces Perlman’s LRT in the Perlman-Wu procedure with the proposed ALRT for testing \( H_0 \) versus \( H_1 \). More precisely, the proposed IUT rejects \( H_0 \) if and only if

\[ \min(\hat{u}_A^2, \hat{u}_B^2) > c \] (18)
\[ \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( X_j + \epsilon_j \right) / \sqrt{\tilde{\sigma}_{jj}} > t_\alpha \quad \text{for } j = 1, \ldots, p. \] (19)

Even when \( p = 2 \), the test statistic of the proposed IUT is a function of five statistics comprising the estimates of two means, two standard deviations, and one correlation coefficient. Consequently, the rejection region cannot be expressed in two-dimensional space. However, in order to visualize the rejection region of the proposed IUT, we chose some distributions lying on the boundary of \( H_0 \) in the parameter space and generated a number of random samples that obeyed these distributions. By applying the proposed IUT to these sample data, we obtained approximately 5% of rejected data, which denoted the rejection region.

We distributed them on the two-dimensional \( X \)-space shown in Figure 4, where three figures—left, center, and right—are plotted for given \( r = -0.8, r = 0, \) and \( r = 0.8 \), respectively. In other words, five dimensional surface of the boundary is sliced by the three cases of estimated correlation coefficient \( r \) and, further, projected on the two dimensional \( X \)-space presenting the variability due to estimated standard deviations by random points.

If \( r \) is continuously moved from \(-1\) through \(+1\), figures for rejection region like these three figures are smoothly connected to make the boundary area of the whole rejection region.

---

Jpn J Biomet Vol. 35, No. 1, 2014
Given estimated correlation coefficient \( r \), the lower boundary of the rejection region is given by randomly generated rejected points distributed on \( X \)-space as shown in these figures.

**Fig. 4.** Rejection region of the proposed IUT for \( p = 2 \).

According to Figure 4, the characteristic property of the proposed IUT is remarkably reflected in that of the proposed ALRT for large values of \( r \) that are likely to be produced under large values of the population correlation coefficient \( \rho \).

When the standard deviations are different between two variables \( X_1 \) and \( X_2 \), the rejection regions is distorted depending on the ratio of the scales of two coordinates, while essential feature of the proposed IUT is maintained.

For \( p > 2 \), the visual illustration of the characteristic property of the proposed IUT is impossible. However, when an intraclass correlation structure is assumed, a similar argument to the above is possible as shown in Appendix B.

### 3.7 Numerical comparison of the proposed IUT with the Perlman-Wu procedure

We performed Monte Carlo simulation experiments to compare the performance of the proposed IUT with that of the Perlman-Wu procedure, focusing our attention on type I error rates and powers as the criteria for evaluation.

In the simulation, the level of significance was set at 0.05. In addition, we assumed that \( \epsilon_1 = \epsilon_2 = 0.5 \) and that the covariance matrices are given by

\[
\Sigma = \begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix}
\]

for \( p = 2 \), and

\[
\Sigma =
\begin{pmatrix}
1 & \rho & \rho \\
\rho & 1 & \rho \\
\rho & \rho & 1
\end{pmatrix}
\]

for \( p = 3 \). Using the Statistical Analysis System (SAS) program, the generation of simulated data was repeated 100,000 times for evaluating type I error rates and 10,000 times for evaluating powers.

Although we examined numerous parameter settings, only the result of simulations pertaining to typical cases, where the characteristic property of the proposed IUT is represented, are presented in Tables 1 and 2 for \( p = 2 \) and Tables 3 and 4 for \( p = 3 \). In these tables, the sample size is set as \( n = n_1 = n_2 \), and only the cases with \( n = 30 \) and \( n = 100 \) are shown because the Jpn J Biomet Vol. 35, No. 1, 2014
Table 1. Estimated type I error rates for \( p = 2 \).

| \( \rho \) | \( n = 30 \) | \( \mu \) | \( n = 100 \) | \( \mu \) |
|-------|----------------|----------------|----------------|----------------|
|       | Method         | \((-0.50, 0)\) | \((-0.25, 0)\) | \( (0, 0) \) | \((-0.50, 0)\) | \((-0.25, 0)\) | \( (0, 0) \) |
| \   |                |                |                |                |                |
| 0.8  | Proposed       | 0.029          | 0.047          | 0.046          | 0.012          | 0.051          | 0.047          | 0.019          |
|      | Perlman        | 0.009          | 0.011          | 0.019          | 0.009          | 0.012          | 0.019          | 0.019          |
| 0.6  | Proposed       | 0.011          | 0.024          | 0.044          | 0.006          | 0.025          | 0.047          | 0.020          |
|      | Perlman        | 0.005          | 0.010          | 0.022          | 0.005          | 0.011          | 0.024          | 0.024          |
| 0.4  | Proposed       | 0.005          | 0.016          | 0.042          | 0.003          | 0.016          | 0.044          | 0.026          |
|      | Perlman        | 0.003          | 0.009          | 0.024          | 0.003          | 0.010          | 0.026          | 0.026          |

\( \sigma_{11} = \sigma_{22} = 1, \, \epsilon_1 = \epsilon_2 = 0.5, \, \) and \( \alpha = 0.05 \).

Table 2. Estimated powers for \( p = 2 \).

| \( \rho \) | \( n = 30 \) | \( \mu \) | \( \mu \) | \( n = 100 \) | \( \mu \) | \( \mu \) |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|
|       | Method         | \((-0.25, 0.50)\) | \( (0.50) \) | \( (0.50, 0.50) \) | \((-0.25, 0.25)\) | \( (0.25) \) | \( (0.25, 0.25) \) |
| \   |                |                |                |                |                |                |                |
| 0.8  | Proposed       | 0.239          | 0.523          | 0.573          | 0.343          | 0.561          | 0.522          | 0.284          | 0.301          | 0.400          |
|      | Perlman        | 0.197          | 0.332          | 0.455          | 0.284          | 0.301          | 0.400          | 0.284          | 0.301          | 0.400          |
| 0.6  | Proposed       | 0.198          | 0.407          | 0.616          | 0.279          | 0.444          | 0.559          | 0.244          | 0.308          | 0.455          |
|      | Perlman        | 0.159          | 0.302          | 0.517          | 0.244          | 0.308          | 0.455          | 0.244          | 0.308          | 0.455          |
| 0.4  | Proposed       | 0.165          | 0.350          | 0.667          | 0.256          | 0.398          | 0.610          | 0.226          | 0.313          | 0.521          |
|      | Perlman        | 0.136          | 0.281          | 0.575          | 0.226          | 0.313          | 0.521          | 0.226          | 0.313          | 0.521          |

\( \sigma_{11} = \sigma_{22} = 1, \, \epsilon_1 = \epsilon_2 = 0.5, \, \) and \( \alpha = 0.05 \).

influence of the increases of sample size is negligible for any \( n \) greater than 100. That is, \( n = 30 \) represents the small sample size, while \( n = 100 \) represents a large sample size. Concerning the correlation coefficient \( \rho \), only large values cause serious \( \alpha \)-violation, i.e., the inflation of type I error rates beyond the nominal significance level. Therefore, only the cases where \( \rho \) greater than 0.4 are listed in the tables.

According to Table 1 for \( p = 2 \), type I error rates in the proposed IUT were below the nominal significance level, except when \( \rho = 0.8 \). The large powers listed in Table 2 should be
regarded as reflecting the result of the substitution of LRT in the Perlman-Wu procedure with the proposed ALRT.

The result for cases when $p = 3$ is similar to the case of $p = 2$ as shown in Tables 3 and 4, although the restriction on $\rho$ for controlling type I error rates below the nominal significance level is more stringent, i.e., it should ideally be less than 0.6.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{\(n = 30\)} & \textbf{\(\rho\)} & \textbf{Method} & \textbf{\(\mu\)} \\
\hline
& 0.8 & Proposed & 0.057 \ (−0.30, 0, 0) & \multirow{2}{*}{0.045} \ (−0.10, 0, 0) & \multirow{2}{*}{0.046} \\
& & Perlman & 0.006 & & \\
\hline
& 0.6 & Proposed & 0.028 \ (−0.20, 0, 0) & \multirow{2}{*}{0.034} \ (−0.10, 0, 0) & \multirow{2}{*}{0.041} \\
& & Perlman & 0.007 & & \\
\hline
& 0.4 & Proposed & 0.017 \ (−0.10, 0, 0) & \multirow{2}{*}{0.028} \ (0, 0, 0) & \multirow{2}{*}{0.036} \\
& & Perlman & 0.006 & & \\
\hline
\end{tabular}
\caption{Estimated type I error rates for \(p = 3\).}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{\(n = 30\)} & \textbf{\(\rho\)} & \textbf{Method} & \textbf{\(\mu\)} \\
\hline
& 0.8 & Proposed & 0.118 \ (−0.20, 0.50, 0.50) & \multirow{2}{*}{0.060} \ (−0.10, 0.50, 0.50) & \multirow{2}{*}{0.047} \\
& & Perlman & 0.007 & & \\
\hline
& 0.6 & Proposed & 0.058 \ (−0.10, 0.50, 0.50) & \multirow{2}{*}{0.043} \ (0, 0.50, 0.50) & \multirow{2}{*}{0.046} \\
& & Perlman & 0.008 & & \\
\hline
& 0.4 & Proposed & 0.031 \ (0, 0.50, 0.50) & \multirow{2}{*}{0.036} \ (0.50, 0.50, 0.50) & \multirow{2}{*}{0.045} \\
& & Perlman & 0.008 & & \\
\hline
\end{tabular}
\caption{Estimated powers for \(p = 3\).}
\end{table}
4. Least Favorable Configuration (LFC) for $\alpha$-violation

As is seen in Tables 1 and 3, the proposed IUT involves a serious issue of $\alpha$-violation. We investigated what the least favorable configuration (LFC) of parameters is and what condition will actualize $\alpha$-violation.

4.1 Theoretical investigation on the simplest case

As the simplest case, we assume that the covariance matrix is known and that it has an intraclass correlation structure with unit variances, i.e., all correlation coefficients are the same. This assumption is justified when the asymptotic theory is applicable because of large sample size and multiple endpoints are evenly correlated.

Consider the case where $p = 2$, $n_1 = n_2 (= n) = 100$, and $\epsilon_1 = \epsilon_2 = 0.5$. The critical value $c$ is 4.2188 from Equation (10), and the transformation matrices $A$ and $B$ are given by Equation (12) as a special case of those shown in Appendices A and B. Under this condition, the rejection region is symmetrical with respect to a line that is equiangular to both axes, such as the one shown in Figure 5. Consequently without loss of generality, we can concentrate on the upper left area.

When $\rho$ is considerably high, there is a sharp dent (or nuzzle) in the rejection region indicated by the point $P_{C2}$ in Figure 5, which appears to greatly contribute to the inflation of type I error rates. Since the probability distribution on this sample space is symmetrical with respect to the equiangular line to both axes, the least favorable distribution to type I inflation must be the one

![Fig. 5. Illustration of the method for identifying the theoretical LFC.](image-url)
Table 5. Theoretical and estimated LFCs with estimated type I error rates at LFCs.

| $\rho$ | 0.70 | 0.80 | 0.90 | 0.95 |
|--------|------|------|------|------|
| $\mu_1$ | Theoretical LFC | -0.23 | -0.20 | -0.16 | -0.12 |
|        | Estimated LFC    | -     | -0.19 | -0.18 | -0.15 |

Type I error rates

|                  | At theoretical LFC | At estimated LFC |
|------------------|---------------------|------------------|
| Theoretical LFC  | 0.036               | 0.060            |
| Estimated LFC    | 0.057               | 0.106            |

The simulation data is generated 100,000 times for $p = 2$, $n_1 = n_2 = 100$, and $\alpha = 0.05$. For $\rho = 0.7$ estimated type I error rates increase with $\mu_1$ with the maximum 0.046 at $\mu_1 = 0$. With $\mu$ on the horizontal axis, i.e., the upper boundary of the null hypothesis $H_0^{(o)}$. Moreover, the value of $\mu$ is obtained by projecting the sharp dent along the equiangular line to the axes, as is illustrated in Figure 5. The thus obtained values of the first coordinate $\mu_1$ of LFC are listed in the 2nd row of Table 5. According to this theory, the dependency of type I error rates on $\mu_1$ must be unimodal with LFC at the position of mode.

Although clear theoretical investigation is difficult for more general cases as explained in appendix C, there does not appear to be much doubt that LFCs are located on axes within the region of $H_0^{(o)}$, and that the unimodality holds on these axes if $\alpha$-violation occurs.

4.2 Numerical confirmation of the theoretical results

In order to confirm the validity of the above theory, we conducted a simulation experiment with 100,000 times of data generations under the same setting as the above. Figure 6 (upper) illustrates the dependency of type I error rates on $\mu_1$ as given by the results of simulation. It provides an evidence on the unimodality of type I error rates when $\alpha$-violation occurs. The values of $\mu_1$ which produced the maximum of type I error rates in the simulation, are presented in the 3rd row of Table 5 with the corresponding type I error rate in the 5th row. The estimated type I error rates at theoretical values of $\mu_1$ are in the 4th row of the table. According to Table 5, the theoretically identified LFCs are quite close to the values of $\mu_1$ obtained by the simulation experiment, and type I error rates are almost the same between the theoretical and estimated $\mu_1$s. Although identifying theoretical LFCs for $p > 2$ is difficult, the unimodality on the boundary $(\mu = (\mu_1, 0, 0)^T)$ of $H_0$ in $\mu$ space is shown on the basis of simulation experiment, as illustrated in Figure 6 (lower).

5. Discussion

Controlling type I error rates is the most serious issue in the practical application of the proposed IUT. Based on the theoretical investigation and simulation experiment, it can be stated with certainty that the LFC is located on the axes in $\mu$-space and that type I error rates are unimodal on these axes. For small sample sizes, the $\alpha$-violation is not so serious because $t$-tests in the proposed IUT work well. When the sample size is considerable—e.g., as large as 100—the
maximum type I error rates increases with the correlation coefficients and \( p \).

According to Offen et al. (2007) and Sankoh et al. (1999), the correlation coefficients among multiple endpoints in clinical trials approximately equal 0.4 and fall in the range between 0.2 and 0.8. In practical situations, such information regarding the correlation coefficients as well as the

\[
\begin{align*}
\text{(a) } p = 2 & \\
\text{(b) } p = 3
\end{align*}
\]
necessary sample size \( n \) are available before the clinical trial is designed. Also, in clinical trials, \( p \) is generally small as per the recommendation of “Statistical Principles for Clinical Trials” (Lewis et al., 1999). We have shown that when \( p \) is small and \( \rho \) is not so extreme, type I error rate is controlled below the nominal significant level.

If there is no sufficient information regarding the correlation coefficients before the beginning of the study, it is necessary to examine the type I error rates under various values of the correlation. The type I error rates can be assessed by conducting simulation under \( H_0 \). In particular, when \( p = 2 \), one example might be to pre-specify the following in the analysis plan: first, calculate the confidence interval of the obtained correlation coefficient.; second, simulate type I error rate under \( H_0 \) with assumed correlation equal to the lower and upper limits of the confidence interval.; and, third, if type I error rate falls below the nominal significance level, then the proposed IUT can be used.

If type I error rate falls above the nominal significance level, the proposed approach cannot be used directly. One approach in such a case is to adjust the critical value so as to control type I error rates below the nominal significance level. Such method is a topic of future research.

Therefore, we believe that the proposed IUT can be practically used under certain conditions with respect to the correlation coefficients and the number of endpoints \( p \).

Acknowledgements

This study was conducted when all authors belonged to the Biomedical Statistics Course, Graduate School of Engineering, Tokyo University of Science. We are grateful to Professor Takashi Seo of Department of Graduate School of Mathematical Information Science, Faculty of Science, Tokyo University of Science, for his advice on the theoretical aspect of this study. The authors also wish to thank the referee and editor for comments that improved the presentation of this article. This research was partially supported by Grant-in-Aid Grant-in-Aid for Young Scientists (B), 25870332, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

REFERENCES

Berger RL. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics, 24*, 295–300.

Bloch DA., Lai TL., and Tubert-Bitter P. (2001). One-sided tests in clinical trials with multiple endpoints. *Biometrics, 57*, 1039–1047.

Glimm E., Srivastava M., and Lauter J. (2002). Multivariate tests of normal mean vectors with restricted alternatives. *Communications in Statistics, B 31*, 589–604.
Appendix

A. Explicit expression of the transformation matrices for intraclass correlation structure

As the simplest case, assume $p = 2$ and $G = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \quad (-1 < r < 1)$. Then, $G$ is decomposed as $G = H\Lambda H^t$ with the orthogonal matrix $H$ defined below and, consequently, $G^{-1} = H\Lambda^{-1}H^t$, where

\[
\Lambda = \begin{pmatrix} 1 - r & 0 \\ 0 & 1 + r \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} .
\]  

(A.1)

Defining the matrix $\Lambda^{-1/2}$ as

\[
\Lambda^{-1/2} = \begin{pmatrix} 1/\sqrt{1 - r} & 0 \\ 0 & 1/\sqrt{1 + r} \end{pmatrix},
\]
A New Procedure of One-Sided Test in Clinical Trials with Multiple Endpoints

\[ A = H \Lambda^{-1/2} H^t = \begin{pmatrix} 
\frac{1}{2\sqrt{1-r}} + \frac{1}{2\sqrt{1+r}} & -\frac{1}{2\sqrt{1-r}} + \frac{1}{2\sqrt{1+r}} \\
-\frac{1}{2\sqrt{1-r}} + \frac{1}{2\sqrt{1+r}} & \frac{1}{2\sqrt{1-r}} + \frac{1}{2\sqrt{1+r}} 
\end{pmatrix} \quad (A.2) \]

is symmetric and \( A^t A = G^{-1} \).

If \( r > 0 \), the off-diagonal elements of \( A \) are negative and, consequently,

\[ B = \begin{pmatrix} 
\frac{1}{2\sqrt{1-r}} + \frac{1}{2\sqrt{1+r}} & \frac{1}{2\sqrt{1-r}} - \frac{1}{2\sqrt{1+r}} \\
\frac{1}{2\sqrt{1-r}} - \frac{1}{2\sqrt{1+r}} & \frac{1}{2\sqrt{1-r}} + \frac{1}{2\sqrt{1+r}} 
\end{pmatrix}. \quad (A.3) \]

Otherwise \( r < 0 \), \( B = A \) and the rejection region of the proposed ALRT coincides with that of Glimm et al.’s.

As a more general case, assume \( p > 2 \) and that \( G \) has an intraclass correlation structure such that

\[ G = \begin{pmatrix} 
1 & r & \cdots & r \\
r & 1 & \cdots & r \\
\vdots & \vdots & \ddots & \vdots \\
r & r & \cdots & 1 
\end{pmatrix} \quad (-1/(p-1) < r < 1). \quad (A.4) \]

Let \( \Lambda^{-1/2} \) and \( H \) be a diagonal matrix and an orthogonal matrix, respectively, such that

\[ \Lambda^{-1/2} = \begin{pmatrix} 
1 & 0 & \cdots & 0 \\
0 & \sqrt{1-r} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sqrt{1+r} 
\end{pmatrix} \quad (A.5) \]

and

\[ H = \begin{pmatrix} 
h_{11} & \cdots & h_{1(p-1)} & 1/\sqrt{p} \\
h_{21} & \cdots & h_{2(p-1)} & 1/\sqrt{p} \\
\vdots & \vdots & \vdots & \vdots \\
h_{p1} & \cdots & h_{p(p-1)} & 1/\sqrt{p} 
\end{pmatrix}. \quad (A.6) \]

Next, \( A = H \Lambda^{-1/2} H^t \) is written as (Siotani, 1985)

\[ A = \begin{pmatrix} 
p - 1 & 1 & \cdots & 1 \\
-\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} & -\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} & \cdots & -\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} \\
-\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} & -\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} & \cdots & -\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} \\
\vdots & \vdots & \vdots & \vdots \\
-\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} & -\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} & \cdots & -\frac{1}{p\sqrt{1-r}} + \frac{1}{p\sqrt{1+(p-1)r}} 
\end{pmatrix}. \quad (A.7) \]
and satisfies $A^TA = G^{-1}$. The off-diagonal elements of $A$ are negative if $r > 0$ and, consequently, the off-diagonal elements of $B$ are obtained from those of $A$ by inverting the sign.

When $G$ is represented as

$$
G = \begin{pmatrix}
    s_1 & 0 & \cdots & 0 \\
    0 & s_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & s_p
\end{pmatrix}
\begin{pmatrix}
    1 & r & \cdots & r \\
    r & 1 & \cdots & r \\
    \vdots & \vdots & \ddots & \vdots \\
    r & r & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
    s_1 & 0 & \cdots & 0 \\
    0 & s_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & s_p
\end{pmatrix}, \quad (A.8)
$$

an argument similar to the above one is possible for $Y$s defined by Equation (A.9) instead of $X$s.

$$
\begin{pmatrix}
    Y_1 \\
    Y_2 \\
    \vdots \\
    Y_p
\end{pmatrix}
= \begin{pmatrix}
    1/s_1 & 0 & \cdots & 0 \\
    0 & 1/s_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & 1/s_p
\end{pmatrix}
\begin{pmatrix}
    X_1 \\
    X_2 \\
    \vdots \\
    X_p
\end{pmatrix}, \quad (A.9)
$$

This signifies that the above argument is not very particular.

Theoretical considerations of more general cases apart from the intraclass correlation structure is difficult. However, when all correlation coefficients are positive, numerical studies demonstrated that the off-diagonal elements of $A$ are likely to be negative.

**B. Relation between the rejection region of the proposed ALRT and that of Glimm et al.’s ALRT**

Assume that $G$ is approximated by the population covariance matrix $\Sigma$ with the intraclass correlation structure represented by Equation (B.1) and unit variances given the large sample size. We assume $0 < \rho < 1$ because the proposed and Glimm et al.’s ALRTs are almost the same if $\rho \leq 0$.

$$
G = \begin{pmatrix}
    1 & \rho & \cdots & \rho \\
    \rho & 1 & \cdots & \rho \\
    \vdots & \vdots & \ddots & \vdots \\
    \rho & \rho & \cdots & 1
\end{pmatrix}, \quad (B.1)
$$

For $p > 2$, the relationship between the rejection regions of $X$-space and $u$-space cannot be visualized. However, when an intraclass correlation structure is assumed, the rejection region of Glimm et al.’s ALRT in $X$-space consists of the top of a $p$-dimensional ellipsoid with a $p$-dimensional conic skirt smoothly connected with the top because

$$
\frac{n_1n_2}{n_1 + n_2} X^t A X = c
$$

constructs an ellipsoid. The rejection region of the proposed procedure is the intersection of two rejection regions with boundaries of a flat ellipsoid and a sharp ellipsoid, respectively.

Intuitively, the boundary of the rejection region is like a bowler with a cannon ball on the top penetrating a discoid. The major axis of the symmetrical form of the bowler is the line equiangular to all the axes in $X$-space. The brim of the bowler is almost perpendicular to the
major axis and guards the rejection region against going down to negative area of $X$-space, which is an advantageous property of the proposed ALRT.

Although the detail is left out, Monte Carlo experiments can provide evidences for the proposed ALRT to have the property similar to the above when all correlation coefficients hold considerably large positive values.

### C. Least favorable configuration for $\alpha$-violation in the proposed IUT

The most serious issue concerning the proposed IUT is the $\alpha$-violation that occurs at some configurations of parameters. The least favorable configuration (LFC) of the parameter must be the one corresponding to the least favorable area in the rejection region, which is the nearest area to $\mu$ in $H^{(0)}_0$ overlaid on $X$-space.

For $p = 2$, it is the area around the dent $P_{C_2}$ sketched in Figure 5. Since the principal axis of the ellipsoid of concentration in the normal distribution is equiangular from both axes, the value of $\mu$ that yields the mode of type I error rates can be mathematically estimated from the location of the dent. It must be the projected point—the coordinate of which is represented by $(\mu_1, 0)$ without loss of generality—on the $\mu_1$-axis overlaid on the $X_1$-axis from the dent along the equiangular line.

Since the point $(0, \sqrt{c})$ in $u$-space is transformed back to $X$-space as two symmetric points $P_{A2} = (a_1, a_2)$ and $P_{B2} = (-a_1, a_2)$ by matrices $A^{-1}$ and $B^{-1}$, respectively, as shown in Figure 5, the dent point $P_{C_2} = (X_1, X_2)$ is identified as the solution of the simultaneous equation of a line and an ellipse such as Equation (C.1). The values of $\mu_1$ of LFC is calculated as $\mu_1 = X_1 - X_2$.

The theoretical LFCs listed in the 2nd row of Table 5 are calculated with this logic.

$$\begin{align*}
\frac{n_1n_2}{n_1+n_2} \left\{ \left( \frac{1}{2(1-\rho)} + \frac{1}{2(1+\rho)} \right) X_1^2 + 2 \left( -\frac{1}{2(1-\rho)} + \frac{1}{2(1+\rho)} \right) X_1 X_2 \right. \\
+ \left( \frac{1}{2(1-\rho)} + \frac{1}{2(1+\rho)} \right) X_2^2 \right\} = c \\
\sqrt{\frac{n_1n_2}{n_1+n_2}} \left\{ \left( \frac{1}{2\sqrt{1-\rho}} - \frac{1}{2\sqrt{1+\rho}} \right) X_1 + \left( \frac{1}{2\sqrt{1-\rho}} + \frac{1}{2\sqrt{1+\rho}} \right) X_2 \right\} = \sqrt{c}
\end{align*}$$

(C.1)

The unimodality of type I error rates on the $\mu_1$-axis is assumed from the fact that the boundary of the rejection region is moved back from the dent point.

Although it is difficult to present mathematical formulas of LFC for $p > 2$, the conjecture that it lies on an axis of the parameter space expressed as, e.g., $(\mu_1, 0, 0)$ and that type I error rates are unimodal on this axis is thought to be true, since it corresponds to, as it were, the pleat of the bowler.

Jpn J Biomet Vol. 35, No. 1, 2014