Closed Testing Procedures Based on $\bar{\chi}^2$-Statistics in Multi-Sample Models with Bernoulli Responses under Simple Ordered Restrictions

Taka-aki Shiraiishi and Shin-ichi Matsuda
Department of Systems and Mathematical Science, Nanzan University,
18 Yamazato-cho, Showa-ku, Nagoya 466-8673, Japan
e-mail:marble@nanzan-u.ac.jp

Multiple comparison procedures provide differences among the groups that is of interest. The procedures are used in clinical trials and agricultural fields experiments. We consider multi-sample model with Bernoulli responses under simple ordered restrictions of proportions. Shiraiishi (2014b) proposed closed testing procedures based on maximum values of two-sample test statistics for all pairwise comparisons. The equality of sample sizes is needed in the asymptotic theory of Shiraiishi’s procedures. We propose closed testing procedures based on statistics having asymptotically a $\bar{\chi}^2$-distribution which is appeared in Chernoff (1954). The proposed procedures are applicable for the models with unequal sample sizes. Although single-step multiple comparison procedures are utilized in general, the power of these procedures is low for a large number of groups. The closed testing procedures stated in the present paper are more powerful than the single-step procedures. Simulation studies are performed under the null hypothesis and some alternative hypotheses. In this studies, the proposed procedures show a good performance. We also illustrate applying to a dose-finding trial data with unequal sample sizes.

Key words: Multiple comparison procedure, Binomial distribution, Multi-step procedures, All pairwise comparisons, Multiple comparisons with a control, Asymptotic properties, Unequal sample sizes.

1. Introduction

In many applications of $k$ sample models for comparing treatment effects in clinical trials or experimental studies, individuals are classified as two possible outcomes, “success” and “failure.” Let $p_1, \ldots, p_k$ correspond to the probabilities that individuals at each of $k$ levels of treatment possess “success.” For this setting, Williams (1990) and Agresti and Coull (1996) assumed $k$-sample Bernoulli responses models with the simple ordered restrictions

$$p_1 \leq p_2 \leq \cdots \leq p_k.$$  

Received August 2016. Revised November 2016. Accepted November 2016.
For the $k$-sample models, we suppose specifically that $(X_{i1}, \ldots, X_{in_i})$ is a random sample of size $n_i$ from the $i$-th Bernoulli population with unknown success probability $p_i$ ($i = 1, \ldots, k$). Furthermore, $X_{ij}$'s are assumed to be independent. Under same sample sizes of $n_1 = \cdots = n_k$, Bartholomew (1959) and Shi (1991) constructed tests for null hypothesis $H_0: p_1 = \cdots = p_k$ vs. the ordered alternative $H^A: p_1 \leq p_2 \leq \cdots \leq p_k$ with at least one strict inequality. Even if the null hypothesis $H_0$ is rejected by using their procedures, we only get the conclusion of $p_1 < p_k$, which simply represents the presence of monotonicity, indicating many possible differential patterns among the groups of interest. Multiple comparison procedures can detect more specific differences among the groups that is of interest. Such procedures are more informative and thus more useful for clinical trials and experiments in agriculture or other fields. We consider multiple comparison procedures in $k$ sample models. Shiraishi (2011b) proposed the Tukey-Kramer type procedure based on arcsine transformation for all pairwise comparisons of \{ the null hypothesis $H_{(i,i')}: p_i = p_{i'}$ vs. the alternative $H^A_{(i,i')}: p_i \neq p_{i'}$ \}. Furthermore Shiraishi (2011a) discussed a closed testing procedure which was superior to the Tukey-Kramer type procedure. Shiraishi (2012) proposed multiple comparisons procedures based on $F$ distributions for all proportions of \{ the null hypothesis $H^*_i: p_i = p_{i0}$ vs. the alternative $H^{A*}_i: p_i \neq p_{i0}$ \}. Under the simple ordered restrictions (1) and under same sample sizes, Shiraishi (2014b) proposed the Hayter-type procedure similar to the normal theory of Hayter (1990) for all pairwise comparisons of \{ the null hypothesis $H_{(i,i')}: p_i = p_{i'}$ vs. the alternative $H^A_{(i,i')}: p_i < p_{i'}$ \}.

In the present paper, we propose closed testing procedures as multiple comparison tests for all pairwise comparisons under the simple ordered restrictions (1). The test procedures are based on statistics having asymptotically a $\chi^2$-distribution, which is appeared in Chernoff (1954). Furthermore we propose closed testing procedures for comparisons with a control of \{ the null hypothesis $H_i: p_i = p_1$ vs. the alternative $H^A_i: p_i > p_1$ \}. In section 2, we introduce single-step procedures. In section 3, we give new closed testing procedures as multi-step tests and we introduce previous closed testing procedures. In section 4, by the Monte Carlo simulations, we compare the proposed procedures with the other multiple comparisons tests in the sense of all-pairs power defined in Ramsey (1978). As the result of section 4, we find that the proposed procedures are superior to the other procedures. Especially in all pairwise comparisons, the power of the proposed closed testing procedures is fairly higher than that of the single-step procedures. Although the equality of sample sizes is needed in the previous procedures, the proposed procedures are applicable for the models with unequal sample sizes. In section 5, we give the application to a dose-finding trial data. We illustrate applying the proposed procedure using critical values for unequal sample sizes. We conclude with a discussion in section 6.
2. Condition and Single-Step Procedures

For \( k \) sample models with Bernoulli responses of the previous section, \( X_i = \sum_{j=1}^{n_i} X_{ij} \) has the binomial distribution with parameters \( n_i \) and \( p_i \) and estimators of \( p_i \)'s are given by \( \hat{p}_i = X_i/n_i \) or \((X_i + 0.5)/(n_i + 1)\) \((i = 1, \ldots, k)\). We put \( n = \sum_{i=1}^{k} n_i \) and add the condition (C1).

\[
(C1) \quad \lim_{n \to \infty} n_i/n = \lambda_i > 0 \quad (i = 1, \ldots, k).
\]

Let us put \( \hat{\nu}_i = 2\arcsin(\sqrt{\hat{p}_i}) \) and \( \nu_i = 2\arcsin(\sqrt{p_i}) \). Then using a central limit theorem and Slutsky’s theorem, under (C1), we get

\[
\sqrt{n}(\hat{\nu}_i - \nu_i) \overset{d}{\to} Y_i \sim N\left(0, \frac{1}{\lambda_i}\right),
\]

where \( \overset{d}{\to} \) denotes convergence in law and \( Y \sim N(0, \sigma^2) \) denotes that \( Y \) is distributed according to \( N(0, \sigma^2) \).

2.1 \( \chi^2 \)-Test

The test statistic for the null hypothesis of homogeneity of proportions \( H_0: p_1 = \cdots = p_k \) vs. the alternative \( H_1: p_i \neq p_{i'} \) for some \( 1 \leq i < i' \leq k \) is given by

\[
S = \sum_{i=1}^{k} n_i \left( \hat{\nu}_i - \sum_{j=1}^{k} \lambda_{ni} \hat{\nu}_j \right)^2,
\]

where \( \lambda_{ni} = n_i/n \) \((i = 1, \ldots, k)\). If the null hypothesis \( H_0 \) holds, by using (2) and the Cramér-Wold technique, we get, under (C1), as \( n \to \infty \),

\[
S \overset{d}{\to} \sum_{i=1}^{k} \lambda_i \left( Y_i - \sum_{j=1}^{k} \lambda_j Y_j \right)^2 \sim \chi^2_{k-1},
\]

where \( \chi^2_{k-1} \) denotes a chi-squared distribution with \( k - 1 \) degrees of freedom. Hence, we can reject \( H_0 \) when the value of \( S \) is larger than \( \chi^2_{k-1}(\alpha) \), where \( \chi^2_{k-1}(\alpha) \) stands for the 100\( (1 - \alpha) \) percentile of the chi-squared distribution with \( k - 1 \) degrees of freedom.

2.2 \( \bar{\chi}^2 \)-Test

When the simple ordered restrictions of (1) is satisfied, we consider the null hypothesis \( H_0 \) vs. the alternative \( H^A: p_1 \leq p_2 \leq \cdots \leq p_k \) with at least one strict inequality, which is equivalent to \( H_0: p_1 = p_k \) vs. \( H^A: p_1 < p_k \). We define \( \{\hat{\nu}^*_i \mid i = 1, \ldots, k\} \) by \( \{u_i \mid i = 1, \ldots, k\} \) which minimize \( \sum_{i=1}^{k} \lambda_{ni} (u_i - \hat{\nu}_i)^2 \) under simple ordered restrictions \( u_1 \leq u_2 \leq \cdots \leq u_k \), i.e.,

\[
\sum_{i=1}^{k} \lambda_{ni} (\hat{\nu}^*_i - \hat{\nu}_i)^2 = \min_{u_1 \leq \cdots \leq u_k} \sum_{i=1}^{k} \lambda_{ni} (u_i - \hat{\nu}_i)^2.
\]

\( \hat{\nu}_1^*, \ldots, \hat{\nu}_k^* \) are computed by using the pool-adjacent-violators algorithm stated in Robertson et al. (1988). We put

\begin{align*}
\text{Jpn J Biomet Vol. 37, No. 2, 2016}
\end{align*}
Hayter and Liu (1996) showed that the value of $Y$ in Robertson et al. (1988) can be derived by the recurrence of one-dimensional computational integration. Let $I$ with $L$ degrees of freedom. The recurrence formula of computing $P(L,k;\lambda)$ is written in Robertson et al. (1988). We shall state the recurrence formula along the above notations. Hayter and Liu (1996) showed that the value of

$$P(k,k;\lambda) = P(Y_1 < Y_2 < \cdots < Y_k)$$

(4)

can be derived by the recurrence of one-dimensional computational integration. Let $I_1^d, I_2^d, \ldots, I_L^d$ be a partition of $\{1, 2, \ldots, k\}$ satisfying the following property (P1)

(P1) Each $I_s^d$ is a nonempty set composed of consecutive integers or an integer. When $L \geq 2$, the maximum value of the elements of $I_i^d$ is less than the minimum value of $I_{i+1}^d$ for any integer $i$ such that $1 \leq i \leq L - 1$.

Then Theorem 2.4.1 of Robertson et al. (1988) gives, for $L = 2, \ldots, k - 1$,

$$P(L,k;\lambda) = \sum_{\{I_1^d, I_2^d, \ldots, I_L^d\}} P\left(L, L; \Lambda(I_1^d), \Lambda(I_2^d), \ldots, \Lambda(I_L^d)\right) \cdot \prod_{i=1}^L P(1, \#(I_s^d);\lambda(I_s^d)),$$

(5)

where $\#(I_s^d)$ denotes the number of elements of $I_s^d$, $\Lambda(I_s^d) = \sum_{i \in I_s^d} \lambda_i$, $\lambda(I_s^d) = (\lambda_i, \lambda_{i+1}, \ldots, \lambda_j)$ for $I_s^d = \{i, i+1, \ldots, j\}$, and $\sum_{\{I_1^d, I_2^d, \ldots, I_L^d\}}$ denotes the sum over all partitions of $\{1, 2, \ldots, k\}$ satisfying (P1). $\#(I_s^d)$ of (5) is less than or equal to $k - 1$. Furthermore we get

$$P(1,k;\lambda) = 1 - \sum_{L=2}^k P(L,k;\lambda)$$

and

$$P(1,1;\lambda_i) = 1 \quad (1 \leq i \leq k)$$

(6)

$$P(1,2;\lambda_i,\lambda_j) = P(2,2;\lambda_i,\lambda_j) = \frac{1}{2} \quad (1 \leq i < j \leq k).$$

(7)
Since $P(L, k; \lambda)$ depends on $L$ and $k$ for
\begin{equation}
\lambda_1 = \cdots = \lambda_k = 1/k,
\end{equation}
we write simply $P(L, k)$ instead of $P(L, k; \lambda)$. Barlow et al. (1972) gives the following recurrence formula.
\begin{align*}
P(1, k) &= \frac{1}{k}, \\
P(L, k) &= \frac{1}{k} \left\{ (k - 1)P(L, k - 1) + P(L - 1, k - 1) \right\} \quad (2 \leq L \leq k - 1), \\
P(k, k) &= \frac{1}{k!}.
\end{align*}
For given $\alpha$ such that $0 < \alpha < 1$, we give the following equation of $t$.
\begin{equation}
k \sum_{L=2}^{L} P(L, k; \lambda) P(\chi^2_{L-1} \geq t) = \alpha.
\end{equation}
We denote a solution of this equation by $\tilde{c}_k(\lambda; \alpha)$. Hence, from (3), we can reject $H_0$ when the value of $\chi^2_k$ is larger than $\tilde{c}_k(\lambda; \alpha)$.

### 2.3 All Pairwise Comparisons

We consider test procedures for all pairwise comparisons of
\begin{align*}
\{ \text{the null hypothesis } H_{(i, i')} : p_i = p_{i'} \text{ vs. the alternative } H_{(i, i')}^{A*} : p_i \neq p_{i'} \mid 1 \leq i < i' \leq k \}\).
\end{align*}
Let us put
\begin{equation}
T_{ii'} = \frac{\hat{\nu}_{i'} - \hat{\nu}_i}{\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}} \quad (i < i').
\end{equation}
Then, Shiraishi (2011b) gives for $t > 0$,
\begin{equation}
A(t) \leq \lim_{n \to \infty} P_0 \left( \max_{1 \leq i < i' \leq k} |T_{ii'}| \leq t \right) \leq A_\lambda(t),
\end{equation}
where $P_0(\cdot)$ denotes the probability measure under $H_0$,
\begin{align*}
A(t) &= k \int_{-\infty}^{\infty} \left\{ \Phi(x) - \Phi(x - \sqrt{2} \cdot t) \right\}^{k-1} d\Phi(x), \\
A_\lambda(t) &= \int_{-\infty}^{\infty} \sum_{j=1}^{k} \prod_{i=1}^{k} \left\{ \Phi \left( \sqrt{\frac{\lambda_i}{\lambda_j}} \cdot x \right) - \Phi \left( \sqrt{\frac{\lambda_i}{\lambda_j}} \cdot x - \sqrt{\frac{\lambda_i + \lambda_j}{\lambda_j}} \cdot t \right) \right\} d\Phi(x)
\end{align*}
and $\Phi(x)$ denotes the standard normal distribution function. The inequality of the left hand side of (10) is derived by using the main theorem of Hayter (1984). When (8) is satisfied, both equalities of (10) hold. Suppose that $a(k; \alpha)$ is a solution of $t$ satisfying the equation $A(t) = 1 - \alpha$. Then, from the inequality of the left hand side of (10), Shiraishi (2011b) proposed the Tukey-Kramer type procedure given by the following (I).
(I) Tukey-Kramer type procedure

Asymptotic simultaneous tests of level \( \alpha \) for the null hypotheses \( \{ H_{(i,i')} : p_i = p_{i'} \mid 1 \leq i < i' \leq k \} \) consist in rejecting \( H_{(i,i')} \) for \( 1 \leq i < i' \leq k \) such that \( |T_{i,i'}| > a(k; \alpha) \).

The values of \( A_\lambda(a(k; \alpha)) \) for \( \alpha = 0.05, 0.01 \) and \( k = 3(1)10 \) were provided in Table I of Shiraishi (2007). From Table I of Shiraishi (2007), it can be seen that the value of \( A_\lambda(a(k; \alpha)) \) is nearly equal to \( 1 - \alpha \) when the following condition (C2) is satisfied.

\[
(C2) \quad 1 < \max\{\lambda_i \mid i = 1, \ldots, k\}/\min\{\lambda_i \mid i = 1, \ldots, k\} \leq 2.
\]

Since \( 1 < \max\{n_i \mid i = 1, \ldots, k\}/\min\{n_i \mid i = 1, \ldots, k\} \leq 2 \) implies (C2), most data satisfy (C2). Hence, the conservativeness for the procedure of (I) is a little.

2.4 All Pairwise Comparisons Under Simple Ordered Restrictions

Under the ordered restrictions (1), we consider test procedures for all pairwise comparisons of

\[\{ \text{the null hypothesis } H_{(i,i')} : p_i = p_{i'} \text{ vs. the alternative } H_{(i,i')}^A : p_i < p_{i'} \mid 1 \leq i < i' \leq k \}.\]

We add the condition (C3) of same sample sizes.

\[n_1 = n_2 = \cdots = n_k.\]

We put

\[
D_1(t) = P \left( \max_{1 \leq i < i' \leq k} \frac{Z_{i'} - Z_i}{\sqrt{2}} \leq t \right),
\]

where \( Z_i \sim N(0,1) \) and \( Z_1, \ldots, Z_k \) are independent. Shiraishi (2014b) gives

\[
\lim_{n \to \infty} P_0 \left( \max_{1 \leq i < i' \leq k} T_{i,i'} \leq t \right) = D_1(t).
\]

From Hayter and Liu (1996), we can get the value of \( D_1(t) \) by using the recurrence relation

\[
h_1(t,y) = \Phi(\sqrt{2} \cdot t + y),
\]

\[
h_r(t,y) = \int_{-\infty}^{y} h_{r-1}(t,y) \varphi(y)dy + h_{r-1}(t,y)\{ \Phi(\sqrt{2} \cdot t + y) - \Phi(y) \},
\]

\[
D_1(t) = \int_{-\infty}^{\infty} h_{k-1}(t,y) \varphi(y)dy,
\]

where \( \varphi(y) \) denotes a standard normal density function. We denote the solution of \( D_1(t) = 1 - \alpha \) by \( d_1(k; \alpha) \), i.e., \( D_1(d_1(k; \alpha)) = 1 - \alpha \). Then, from (12), Shiraishi (2014b) proposed the following procedure similar to the normal theory of Hayter (1990).

(II) Hayter-type procedure

Asymptotic simultaneous tests of level \( \alpha \) for \{ the null hypothesis \( H_{(i,i')} \) vs. the alternative \( H_{(i,i')}^A \mid 1 \leq i < i' \leq k \} \) consist in rejecting \( H_{(i,i')} \) for \( (i,i') \) such that \( T_{i,i'} > d_1(k; \alpha) \) and \( 1 \leq i < i' \leq k \).
The values of $d_1(k; \alpha)$ for $\alpha = 0.05, 0.01$ and $k = 3(1)10$ were provided in Table 3 of Shiraishi (2014b).

3. Closed Testing Procedures

Next, we will discuss closed testing procedures.

3.1 All Pairwise Comparisons Under Simple Ordered Restrictions

Assume that the ordered restrictions (1) is satisfied. Then we put

$$H_1 = \{H_{(i,i')} \mid (i,i') \in U\} \text{ and } U = \{(i,i') \mid 1 \leq i < i' \leq k\}.$$ 

The closure of $H_1$ is given by

$$\overline{H_1} = \left\{ \bigwedge_{v \in V} H_v \mid \emptyset \subsetneq V \subset U \right\},$$

where $\bigwedge$ denotes the conjunction symbol (Refer to Enderton (2001)). Then, we get

$$\bigwedge_{v \in V} H_v: \text{ for any } (i,i') \in V, \quad p_i = p_{i'}. \quad (14)$$

Let $I_1, \ldots, I_J$ be disjoint sets satisfying the following property (C4).

(C4) There exist integers $\ell_1, \ldots, \ell_J \geq 2$ and integers $0 \leq s_1 < \cdots < s_J < k$ such that

$$I_j = \{s_j + 1, s_j + 2, \ldots, s_j + \ell_j\} \quad (j = 1, \ldots, J),$$

$$s_j + \ell_j \leq s_{j+1} \quad (j = 1, \ldots, J-1) \text{ and } s_J \leq k.$$ 

We define the null hypothesis $H(I_1, \ldots, I_J)$ by

$$H(I_1, \ldots, I_J): \text{ for any } j \text{ such that } 1 \leq j \leq J \text{ and for any } i, i' \in I_j, \quad p_i = p_{i'} \text{ holds.} \quad (15)$$

The elements of $I_j$ are consecutive integers and $\#(I_j) \geq 2$. From (14) and (15), for any nonempty $V \subset U$, there exist an integer $J$ and some subsets $I_1, \ldots, I_J \subset \{1, \ldots, k\}$ satisfying (C4) such that

$$\bigwedge_{v \in V} H_v = H(I_1, \ldots, I_J). \quad (16)$$

Furthermore $H(I_1, \ldots, I_J)$ is expressed as

$$H(I_1, \ldots, I_J): p_{s_j+1} = p_{s_j+2} = \cdots = p_{s_j+\ell_j} \quad (j = 1, \ldots, J). \quad (17)$$

For $j = 1, \ldots, J$, we define $(\hat{\nu}_{s_j+1}(I_j), \ldots, \hat{\nu}_{s_j+\ell_j}(I_j))$ by $(u_{s_j+1}, \ldots, u_{s_j+\ell_j})$ which minimize $\sum_{i \in I_j} \lambda_{ni}(u_i - \hat{\nu}_i)^2$ under simple ordered restrictions $u_{s_j+1} \leq u_{s_j+2} \leq \cdots \leq u_{s_j+\ell_j}$, i.e.,

$$\sum_{i \in I_j} \lambda_{ni}(\hat{\nu}_i(I_j) - \hat{\nu}_i)^2 = \min_{u_{s_j+1} \leq \cdots \leq u_{s_j+\ell_j}} \sum_{i \in I_j} \lambda_{ni}(u_i - \hat{\nu}_i)^2,$$

where $I_j$, $s_j$ and $\ell_j$ are defined in (C4). For $H(I_1, \ldots, I_J)$ of (16), we set

$$M = M(I_1, \ldots, I_J) = \sum_{j=1}^{J} \ell_j, \quad \ell_j = \#(I_j).$$

Jpn J Biomet Vol. 37, No. 2, 2016
For $\ell = \ell_1, \ldots, \ell_J$, we define $\alpha(M, \ell)$ by

$$\alpha(M, \ell) = 1 - (1 - \alpha)^{\ell/M}.$$  

Then, we propose a stepwise procedure.

**(III) Proposed stepwise procedure**

Let us put

$$\tilde{\chi}_2^2(I_j) = \sum_{i \in I_j} n_i \left( \frac{n_i}{n(I_j)} - \frac{n_i}{n(I_j)} \right)^2 (j = 1, \ldots, J),$$

where $n(I_j) = \sum_{i \in I_j} n_i$.  

(a) $J \geq 2$

Whenever $\tilde{c}_{\ell_j}(\lambda(I_j); \alpha(M, \ell_j)) < \tilde{\chi}_2^2(I_j)$ holds for an integer $j$ such that $1 \leq j \leq J$, we reject the hypothesis $\bigwedge_{w \in V} H_w$, where $\lambda(I_j) = (\lambda_{s_j+1}, \lambda_{s_j+2}, \ldots, \lambda_{s_j+\ell_j})$.

(b) $J = 1$ ($M = \ell_1$)

Whenever $\tilde{c}_{\ell_1}(\lambda(I_1); \alpha) < \tilde{\chi}_1^2(I_1)$, we reject the hypothesis $\bigwedge_{w \in V} H_w$.

By using the methods of (a) and (b), when $\bigwedge_{w \in V} H_w$ is rejected for any $V$ such that $(i, i') \in V \subset U$, the null hypothesis $H(i, i')$ is rejected as a multiple comparison test.

**Theorem 1.** Under (C1), the level for test procedure (III) is asymptotically $\alpha$ as a multiple comparison test.

The proof of Theorem 1 is stated in Appendix A.

When the condition (C3) is satisfied, $\tilde{c}_{\ell}(\lambda(I); \alpha(M, \ell))$ does not depend on $\lambda(I)$ for $(\ell, I) = (\ell_1, I_1), \ldots, (\ell_J, I_J)$. Hence, under (C3), we rewrite $\tilde{c}_{\ell}(\lambda(I); \alpha(M, \ell))$ as $\tilde{c}_{\ell}(\alpha(M, \ell))$ shortly.

For $\alpha = 0.05, 0.01$, we give the values of $\tilde{c}_{\ell}(\alpha(M, \ell))$ in Table 1.

We limited attention to $2 \leq M \leq 10$. The values of $\tilde{c}_{\ell}(\alpha(M, \ell))$ for $\ell = M - 1$ are not used.

From (16), we find

$$\mathcal{H}_1 = \left\{ H(I_1, \ldots, I_J) | \right. \text{There exists } J \text{ such that } \bigcup_{j=1}^{J} I_j \subset \{1, \ldots, k\}, \right.$$

$I_j$ satisfies (C4), $\#(I_j) \geq 2 (1 \leq j \leq J)$, and $I_j \cap I_j' = \emptyset (1 \leq j < j' \leq J)$ for $J \geq 2$ \}.

For $(i, i') \in U$, we put

$$\mathcal{H}_{(i, i')} = \left\{ H(I_1, \ldots, I_J) \in \mathcal{H}_1 | \right. \text{There exists } j \text{ such that } 1 \leq j \leq J \text{ and } \{i, i'\} \subset I_j \}.$$
Table 1. Critical values $\tilde{\ell}(\alpha(M, \ell))$ for the powerful stepwise procedure

| $M \setminus \ell$ | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 10                  | 5.376 | 6.022 | 6.302 | 6.445 | 6.521 | 6.558 | 6.572 | *   | 6.560 |
| 9                   | 5.194 | 5.825 | 6.095 | 6.231 | 6.301 | 6.334 | *   | 6.339 |
| 8                   | 4.991 | 5.606 | 5.865 | 5.993 | 6.056 | *   | 6.088 |
| 7                   | 4.762 | 5.359 | 5.606 | 5.724 | *   | 5.800 |
| 6                   | 4.499 | 5.075 | 5.307 | *   | 5.460 |
| 5                   | 4.192 | 4.740 | *   | 5.049 |
| 4                   | 3.820 | *   | 4.528 |
| 3                   | *   | 3.820 |
| 2                   | 2.706 |

| $M \setminus \ell$ | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 10                  | 8.277 | 9.118 | 9.532 | 9.779 | 9.939 | 10.048 | 10.124 | *   | 10.216 |
| 9                   | 8.086 | 8.916 | 9.322 | 9.562 | 9.717 | 9.822 | *   | 9.945 |
| 8                   | 7.873 | 8.690 | 9.087 | 9.320 | 9.470 | *   | 9.638 |
| 7                   | 7.632 | 8.434 | 8.821 | 9.046 | *   | 9.284 |
| 6                   | 7.355 | 8.139 | 8.514 | *   | 8.865 |
| 5                   | 7.028 | 7.792 | *   | 8.356 |
| 4                   | 6.630 | *   | 7.709 |
| 3                   | *   | 6.823 |
| 2                   | 5.412 |

Then we get

$$\mathcal{H}_1 = \bigcup_{(i,i') \in \mathcal{U}} \mathcal{H}_{(i,i')}, \text{ and } H_{(i,i')}, H_0 \in \mathcal{H}_{(i,i')}.$$  

Furthermore, for $1 \leq i_1 \leq i_2 < i'_2 \leq i'_1 \leq k$,

$$\mathcal{H}_{(i_1,i'_1)} \subset \mathcal{H}_{(i_2,i'_2)} \quad (18)$$

holds.

We consider the example of $k = 4$. By using the multi-step procedure (III) as a multiple test procedure, $H(I_1, \ldots, I_J)$’s tested to reject specified null hypothesis $H_{(i,i')}$ are given in Tables 2, 3. The specified null hypotheses are as follows.

- $H(\{1,2,3,4\})$: $p_1 = p_2 = p_3 = p_4$; $J = 1$, $s_1 = 0$, $\ell_1 = 4$
- $H(\{1,2\},\{3,4\})$: $p_1 = p_2$, $p_3 = p_4$; $J = 2$, $s_1 = 0$, $\ell_1 = 2$, $s_2 = 2$, $\ell_2 = 2$
- $H(\{1,2,3\})$: $p_1 = p_2 = p_3$; $J = 1$, $s_1 = 0$, $\ell_1 = 3$
- $H(\{1,2\}) = H(1,2)$: $p_1 = p_2$; $J = 1$, $s_1 = 0$, $\ell_1 = 2$.

In Table 2(a), the elements of $\mathcal{H}_{(1,2)}$ are stated. From Table 2(a), to reject $H_{(1,2)}$ as a multiple test, the rejection of the four null hypotheses is required. Therefore if the following
becomes a closed test.

(IV) Shiraishi’s stepwise procedure

The condition (C3) is assumed. Let us put
\[ T(I_j) = \max_{s_j + 1 \leq i' \leq s_j + \ell_j} T_{i'} \quad (j = 1, \ldots, J). \]

In the procedure (III), replace \( \tilde{c}_{i_j} (\lambda(I_j); \alpha(M, \ell_j)) < \tilde{\chi}^2_{i_j}(I_j) \) and \( \tilde{c}_{\ell_j} (\lambda(I_1); \alpha) < \tilde{\chi}^2_{\ell}(I_1) \) with \( d_1(\ell_j; \alpha(M, \ell_j)) < T(I_j) \) and \( d_1(M; \alpha) < T(I_1) \), respectively. Then, this procedure also becomes a closed test.

For \( k = 4 \), to execute the closed test procedures of level 0.05, from Table 3 of Shiraishi (2014b), we use the following relations (19)–(21).

\[
\begin{align*}
\lim_{n \to \infty} P_0(T(\{1,2,3,4\}) &\geq 2.569) = \lim_{n \to \infty} P_0 \left( \max_{1 \leq i' \leq 4} |T_{i'}| \geq 2.569 \right) = 0.05, \\
\lim_{n \to \infty} P_0(T(\{1,2\}, \{3,4\}) &\geq 2.236) = \lim_{n \to \infty} P_0(\max \{|T_{12}|, |T_{34}|\} \geq 2.236) = 0.05.
\end{align*}
\]
Furthermore we put (V) A stepwise procedure based on asymptotic $\chi^2$-statistics

We introduce the stepwise procedure based on asymptotic $\chi^2$-statistics. Let us put

\begin{equation}
\lim_{n \to \infty} P_0(T\{\{1,2,3\}\}) \geq 2.344 = \lim_{n \to \infty} P_0(|T_{12}| \geq 1.960) = 0.05. \tag{21}
\end{equation}

**Table 3.** $H(I_1,\ldots,I_j) \in \mathcal{H}_{(2,3)}$, $\mathcal{H}_{(2,4)}$ or $\mathcal{H}_{(3,4)}$ for $k = 4$

| (a) $H(I_1,\ldots,I_j) \in \mathcal{H}_{(2,3)}$ |  
|-----|-----|-----|
| $M$ | $H(I_1,\ldots,I_j)$ |  
| $4$ | $H(\{1,2,3,4\})$ |  
| $3$ | $H(\{1,2,3\}), H(\{2,3,4\})$ |  
| $2$ | $H(\{2,3\})$ |  

| (b) $H(I_1,\ldots,I_j) \in \mathcal{H}_{(2,4)}$ |  
|-----|-----|-----|
| $M$ | $H(I_1,\ldots,I_j)$ |  
| $4$ | $H(\{1,2,3,4\})$ |  
| $3$ | $H(\{2,3,4\})$ |  

| (c) $H(I_1,\ldots,I_j) \in \mathcal{H}_{(3,4)}$ |  
|-----|-----|-----|
| $M$ | $H(I_1,\ldots,I_j)$ |  
| $4$ | $H(\{1,2,3,4\}), H(\{1,2\},\{3,4\})$ |  
| $3$ | $H(\{2,3,4\})$ |  
| $2$ | $H(\{3,4\})$ |  

(V) A stepwise procedure based on asymptotic $\chi^2$-statistics

We put

\[
S(I_j) = \sum_{i \in I_j} n_i \left( \bar{\nu}_i - \sum_{j' \in I_j} \left( \frac{n_{j'}}{n(I_j)} \right) \hat{\nu}_{j'} \right)^2.
\]

In procedures (III), replace \(\widetilde{c}_{i,\ell} (\lambda(I_j); \alpha(M,\ell_j)) < \tilde{\chi}_{\nu}^2(I_j)\) and \(\bar{c}_{i,\ell} (\lambda(I_1); \alpha) < \tilde{\chi}_{\nu}^2(I_1)\) with \(\chi^2_{i,\ell-1}(\alpha(M,\ell_j)) < S(I_j)\) and \(\chi^2_{\nu-1}(\alpha) < S(I_1)\), respectively. Then, the new procedure also becomes a closed test.

3.2 Comparisons with a Control Under Simple Ordered Restrictions

We assume that the ordered restrictions (1) is satisfied. The first treatment is regarded as a control with which the remaining $k - 1$ treatments are to be compared. In one comparison, we test the null hypothesis $H_i: p_1 = p_i$ vs. the alternative hypothesis $H^A_i: p_1 < p_i$. We put

\[
\mathcal{H}_2 = \{H_i \mid 2 \leq i \leq k\}. \tag{22}
\]

Furthermore we put \(\mathcal{I}_\ell = \{i \mid 1 \leq i \leq \ell\} (\ell = 2,\ldots,k)\). For \(\ell = 2,\ldots,k\), we define \((\hat{\nu}^*_1(\mathcal{I}_\ell),\ldots,\hat{\nu}^*_\ell(\mathcal{I}_\ell))\) by \((u_1,\ldots,u_\ell)\) which minimize \(\sum_{i=1}^\ell \lambda_{n_i} (u_i - \hat{\nu}_i)^2\) under simple ordered restrictions \(u_1 \leq u_2 \leq \cdots \leq u_\ell\), i.e.,

\[
\sum_{i=1}^\ell \lambda_{n_i} (\hat{\nu}^*_i(\mathcal{I}_\ell) - \hat{\nu}_i)^2 = \min_{u_1 \leq \cdots \leq u_\ell} \sum_{i=1}^\ell \lambda_{n_i} (u_i - \hat{\nu}_i)^2.
\]

Let us put

Jpn J Biomet Vol. 37, No. 2, 2016
\[ \bar{\chi}^2(I_\ell) = \sum_{i=1}^{\ell} n_i \left( \hat{\nu}_i^*(I_\ell) - \frac{\sum_{i=1}^{\ell} n_i}{n(I_\ell)} \hat{\nu}_i \right)^2 \quad (\ell = 2, \ldots, k), \] (23)

where \( n(I_\ell) = \sum_{i=1}^{\ell} n_i \). We define \( \tilde{\nu}_1^*(I_\ell), \ldots, \tilde{\nu}_\ell^*(I_\ell) \) by

\[ \sum_{i=1}^{\ell} \lambda_i (\hat{\nu}_i^*(I_\ell) - Y_i)^2 = \min_{u_1 \leq \cdots \leq u_l} \sum_{i=1}^{\ell} \lambda_i (u_i - Y_i)^2, \]

where \( Y_i \) is defined in (2). \( P(L, \ell; \lambda(I_\ell)) \) becomes the probability that \( \tilde{\nu}_1^*(I_\ell), \ldots, \tilde{\nu}_\ell^*(I_\ell) \) takes exactly \( L \) distinct values, where \( \lambda(I_\ell) = (\lambda_1, \ldots, \lambda_\ell) \). From the discussion similar to (2), we get, under (C1),

\[ \lim_{n \to \infty} P_{0} (\bar{\chi}^2(I_\ell) \geq t) = \sum_{L=2}^{\ell} P(L, \ell; \lambda(I_\ell))P(\chi^2_{L-1} \geq t) \quad (t > 0). \] (24)

Then, we propose a stepwise procedure.

(VI) **Proposed stepwise procedure**

Whenever \( \tilde{c}_\ell (\lambda(I_\ell); \alpha) < \bar{\chi}^2(I_\ell) \) holds for any integer \( \ell \) such that \( i \leq \ell \leq k \), we reject the null hypothesis \( H_i \).

**Theorem 2.** Under (C1), the level for test procedure (VI) is asymptotically \( \alpha \) as a multiple comparison test.

The proof of Theorem 2 is stated in Appendix B.

When the condition (C3) is satisfied, \( \tilde{c}_\ell (\lambda(I); \alpha) \) does not depend on \( \lambda(I) \) for \( (\ell, I) = (\ell_1, I_1), \ldots, (\ell_J, I_J) \). Hence, under (C3), we rewrite \( \tilde{c}_\ell (\lambda(I); \alpha) \) as \( \tilde{c}_\ell (\alpha) \) shortly. Table 4 shows the critical values for \( \alpha = 0.05 \) or \( 0.01 \).

Last we introduce the procedure similar to the normal theory procedure of Williams (1972). We assume the following condition

\( \text{(C5)} \quad n_2 = \cdots = n_k. \)

Under the condition (C5), for \( \ell \) such that \( 2 \leq \ell \leq k \), we define \( T_\ell \) and \( \hat{\mu}_\ell \) by

**Table 4.** Critical values \( \tilde{c}_\ell (\alpha) \) for the proposed stepwise procedure

| \( \ell \) | \( 2 \) | \( 3 \) | \( 4 \) | \( 5 \) | \( 6 \) | \( 7 \) | \( 8 \) | \( 9 \) | \( 10 \) |
|---|---|---|---|---|---|---|---|---|---|
| \( \alpha = 0.05 \) | \( \tilde{c}_\ell (\alpha) \) | 2.706 | 3.820 | 4.528 | 5.049 | 5.460 | 5.800 | 6.088 | 6.339 | 6.560 |
| \( \alpha = 0.01 \) | \( \tilde{c}_\ell (\alpha) \) | 5.412 | 6.823 | 7.709 | 8.356 | 8.865 | 9.284 | 9.638 | 9.945 | 10.216 |
\[ T_\ell = \frac{\hat{\mu}_\ell - \hat{\nu}_1}{\sqrt{\frac{1}{\pi_1} + \frac{1}{\pi_1}}} \quad \text{and} \quad \hat{\mu}_\ell = \max_{2 \leq s \leq \ell} \frac{\sum_{i=s}^{\ell} \hat{\nu}_i}{\ell - s + 1}, \]

respectively.

Assume that \( W_1, Z_2, \ldots, Z_k \) are independent and that \( W_1 \sim N(0, \lambda_2/\lambda_1) \) and \( Z_i \sim N(0, 1) \).

We set
\[ D_2(t \mid \ell, \lambda_2/\lambda_1) = P\left( \frac{\hat{\mu}_\ell^* - W_1}{\sqrt{1 + \lambda_2/\lambda_1}} \leq t \right), \quad (25) \]
where \( \hat{\mu}_\ell^* = \max_{2 \leq s \leq \ell} \left\{ \frac{\sum_{i=s}^{\ell} Z_i}{(\ell - s + 1)} \right\} \).

Then under \( H_0 \),
\[ \lim_{n \to \infty} P_0(T_\ell \leq t) = D_2(t \mid \ell, \lambda_2/\lambda_1) \]
holds.

We denote the solution of \( D_2(t) = 1 - \alpha \) by \( d_2(\ell, \lambda_2/\lambda_1; \alpha) \), i.e., \( D_2(d_2(\ell, \lambda_2/\lambda_1; \alpha)) = 1 - \alpha \).

When the condition (C3) is satisfied, for \( \alpha = 0.05, 0.025, 0.01 \), the values of \( d_2(\ell, 1; \alpha) \) are stated in Tables 1 and 2 of Williams (1971). Shiraishi and Sugiura (2015) give the algorithm based on sinc method to calculate \( d_2(\ell, 1; \alpha) \). The sinc method is described in Lund and Bowers (1992) and Stenger (1993). The algorithm is more efficient than that of Williams (1971).

(VII) The Williams-type procedure

Whenever \( d_2(\ell, \lambda_2/\lambda_1; \alpha) < T_\ell \) holds for any integer \( \ell \) such that \( i \leq \ell \leq k \), we reject the null hypothesis \( H_i \).

4. Simulation Studies

We give results of several simulations to evaluate the performance of the procedures, each based on 100,000 Monte Carlo replicates. We deal two types, i.e. all pairwise comparisons in sections 2 and 3.1, and comparisons with a control in section 3.2. The former type has (I) Tukey-Kramer type procedure, denoted by [T1], (II) Hayter-type procedure, [H2], (III) proposed stepwise procedure, [P3], (IV) Shiraishi’s stepwise procedure, [S4], and (V) a stepwise procedure based on asymptotic \( \chi^2 \)-statistics, [C5]. The later type has (VI) proposed stepwise procedure, denoted by [P6], and (VII) the Williams-type procedure, [W7].

First, we investigate whether each procedure protects the significant level on the null hypothesis, i.e. familywise error rates (FWER). In section 2, we give two estimators, \( \hat{p}_i = X_i/n_i \) or \( (X_i + 0.5)/(n_i + 1) \). We get values of FWER on the following situations based on each estimator:
\[ k = 4; \quad n = 20, 50, 100, 200; \quad p = 0.1(0.1)0.9; \quad \alpha = 0.05, 0.01. \]

Table 5 shows the results of them for [P3].

The former estimator: \( \hat{p}_i = X_i/n_i \) shows liberal performance especially on \( n = 20 \) and \( p = 0.1, 0.9 \) and the later estimator: \( \hat{p}_i = (X_i + 0.5)/(n_i + 1) \) shows conservative one especially on the same settings. The other procedures show similar performance. Then, we use the later estimator.
in the following simulation.

Next, we investigate the all-pairs power, which is shown on Ramsey (1978), on alternative hypotheses. Alternative hypotheses are the following one:

\[
H^{A_1}: p_1 = 0.2, p_2 = 0.4, p_3 = 0.6, p_4 = 0.8, \\
H^{A_2}: p_1 = 0.4, p_2 = 0.4, p_3 = 0.6, p_4 = 0.6e.
\]

Table 6 shows the results of all-pairs power for all pairwise comparisons.
For Tables 6(a) and (b), the order of the power is the following:

\[ [P3] = [S4] > [C5] > [H2] > [T1]. \]

But, for Tables 6(c) and (d), the order of the power is the following:

\[ [P3] \simeq [S4] > [H2] > [C5] > [T1]. \]

Now we check the rate of reject of a pair. Table 7 is the result of \( n = 50 \) and \( \alpha = 0.05 \).

The performance of a pair is similar to one of all-pairs power. In detail, \([P3]\) is slight better than \([S4]\) for \( H^{A_1} \), and \([C5]\) is better than \([H2]\) for a near pair in \( H^{A_2} \) but worse for a far pair.
Table 6. Values of all-pairs power for [T1]–[C5]

(a) \( \alpha = 0.05 \) and \( H^{A1} \)

| n \( \) procedure | [T1]   | [H2]   | [P3]   | [S4]   | [C5]   |
|---------------------|--------|--------|--------|--------|--------|
| 20                  | 0.0000 | 0.0000 | 0.0043 | 0.0043 | 0.0008 |
| 50                  | 0.0030 | 0.0119 | 0.2029 | 0.2029 | 0.0776 |
| 100                 | 0.1927 | 0.3465 | 0.7547 | 0.7547 | 0.5952 |
| 200                 | 0.8664 | 0.9163 | 0.9865 | 0.9865 | 0.9675 |

(b) \( \alpha = 0.01 \) and \( H^{A1} \)

| n \( \) procedure | [T1]   | [H2]   | [P3]   | [S4]   | [C5]   |
|---------------------|--------|--------|--------|--------|--------|
| 20                  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 50                  | 0.0000 | 0.0001 | 0.0104 | 0.0104 | 0.0025 |
| 100                 | 0.0207 | 0.0624 | 0.3401 | 0.3401 | 0.1821 |
| 200                 | 0.6299 | 0.7340 | 0.9168 | 0.9168 | 0.8659 |

(c) \( \alpha = 0.05 \) and \( H^{A2} \)

| n \( \) procedure | [T1]   | [H2]   | [P3]   | [S4]   | [C5]   |
|---------------------|--------|--------|--------|--------|--------|
| 20                  | 0.0024 | 0.1289 | 0.1618 | 0.1692 | 0.1195 |
| 50                  | 0.0546 | 0.3805 | 0.4709 | 0.4591 | 0.3644 |
| 100                 | 0.2461 | 0.6968 | 0.8325 | 0.8140 | 0.7402 |
| 200                 | 0.8004 | 0.9534 | 0.9895 | 0.9868 | 0.9742 |

(d) \( \alpha = 0.01 \) and \( H^{A2} \)

| n \( \) procedure | [T1]   | [H2]   | [P3]   | [S4]   | [C5]   |
|---------------------|--------|--------|--------|--------|--------|
| 20                  | 0.0005 | 0.0406 | 0.0447 | 0.0478 | 0.0349 |
| 50                  | 0.0080 | 0.1801 | 0.2378 | 0.2169 | 0.1719 |
| 100                 | 0.0842 | 0.4748 | 0.5994 | 0.5512 | 0.4998 |
| 200                 | 0.5446 | 0.8588 | 0.9430 | 0.9341 | 0.9091 |

Table 7. Rate of reject of a pair for [T1]–[C5] with \( n = 50, \alpha = 0.05 \)

(a) \( H^{A1} \)

| Pair \( \) procedure | [T1]   | [H2]   | [P3]   | [S4]   | [C5]   |
|----------------------|--------|--------|--------|--------|--------|
| (1,2)                | 0.3459 | 0.4379 | 0.6640 | 0.6624 | 0.5281 |
| (1,3)                | 0.9519 | 0.9735 | 0.9924 | 0.9858 | 0.9725 |
| (1,4)                | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| (2,3)                | 0.3073 | 0.3805 | 0.6211 | 0.6188 | 0.5396 |
| (2,4)                | 0.9536 | 0.9737 | 0.9931 | 0.9854 | 0.9728 |
| (3,4)                | 0.3434 | 0.4340 | 0.6625 | 0.6605 | 0.5262 |

(b) \( H^{A2} \)

| Pair \( \) procedure | [T1]   | [H2]   | [P3]   | [S4]   | [C5]   |
|----------------------|--------|--------|--------|--------|--------|
| (1,2)                | 0.0110 | 0.0095 | 0.0266 | 0.0251 | 0.0228 |
| (1,3)                | 0.3075 | 0.5411 | 0.6388 | 0.5862 | 0.4857 |
| (1,4)                | 0.3073 | 0.7257 | 0.8112 | 0.7257 | 0.6480 |
| (2,3)                | 0.3073 | 0.3805 | 0.4709 | 0.4591 | 0.3644 |
| (2,4)                | 0.3051 | 0.5414 | 0.6387 | 0.5856 | 0.4851 |
| (3,4)                | 0.0106 | 0.0099 | 0.0266 | 0.0251 | 0.0227 |
The proposed procedure [P3] shows the best performance among all procedures.

Table 8 shows the results of comparisons with a control.

In this situation, we notice that [P6] and [W7] have similar performance with [P3] and [S4]. Then the proposed procedure [P6] also shows the best performance.

We also simulate for other type of alternative hypotheses, i.e. \( p_1 = p_2 = p_3 = 0.4, p_4 = 0.6 \) and \( p_1 = 0.4, p_2 = p_3 = p_4 = 0.6 \), those results are similar to one of \( H^{A2} \) for all procedures [T1]–[W7].

5. Application to dose-finding trial data

In this section we illustrate the application of the proposed procedures. We use the notation [T1]–[W7] of procedures appeared in section 4. Macdonald et al. (2008) is the study of placebo-controlled dose-finding trial of clazosentan for angiographic vasospasm occurring after subarachnoid hemorrhage. The primary efficacy end point was moderate or severe vasospasm. If clazosentan has the effect for this disease, those rates are decreased with dose. Then, we correct this data to mild vasospasm and apply the proposed procedures. Table 9 show the result of this study for PPS (per-protocol set) with original sample size \( n \).

| Table 8. Values of all-pairs power for [P6] and [W7] |
|---|---|---|
| \( n \) \| Procedure \( [P6] \) \| Procedure \( [W7] \) |
| 20 \| 0.3661 \| 0.3641 |
| 50 \| 0.7095 \| 0.7095 |
| 100 \| 0.9304 \| 0.9304 |
| 200 \| 0.9973 \| 0.9973 |

| \( \alpha = 0.05 \) and \( H^{A1} \) |
|---|---|---|
| \( \alpha = 0.01 \) and \( H^{A1} \) |

| \( n \) \| Procedure \( [P6] \) \| Procedure \( [W7] \) |
| --- | --- | --- |
| 20 \| 0.1473 \| 0.1439 |
| 50 \| 0.4373 \| 0.4368 |
| 100 \| 0.7824 \| 0.7824 |
| 200 \| 0.9815 \| 0.9815 |

| \( \alpha = 0.05 \) and \( H^{A2} \) |
|---|---|---|
| \( \alpha = 0.01 \) and \( H^{A2} \) |

| \( n \) \| Procedure \( [P6] \) \| Procedure \( [W7] \) |
| --- | --- | --- |
| 20 \| 0.1008 \| 0.0054 |
| 50 \| 0.3803 \| 0.2784 |
| 100 \| 0.7588 \| 0.6288 |
| 200 \| 0.9820 \| 0.9468 |

Since this data is not same value of PPS, we should calculate critical values for this situation.

Table 10 is the critical values for the procedure [P6] for \( \alpha = 0.05, 0.01 \).

We get that the values of \( \hat{\chi}^2(I_{I_c}) \) in equation (23) are 9.295, 14.878 and 32.978. Then, we reject placebo to dose 15mg/h with 32.978 > 4.585(7.774), reject placebo to dose 5mg/h.
for \( [P3] \) consists testing the set of null hypothesis in Tables 2 and 3. If we want to get the result \( \alpha \) for the procedure \([P3]\) for \( \alpha \) hypotheses for each \( H \), then we reject \( H \) as above, then we reject \( H \). We could calculate those values on any sample size situation in this way. The procedure \([P3]\) consists testing the set of null hypothesis in Tables 2(a), 2(c) and 3. For Table 2(b), we get the values of \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) and \( H(\{1,2,3\}) \) as above, then we reject \( H(\{1,2\}) \). For Table 2(c), we get \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) as above, then we reject \( H(\{1,4\}) \). For Table 3(a), we get the values of \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) and \( H(\{1,2,3\}) \) as 32.978, 14.878 and 9.295, respectively, and find 32.978 > 4.585, 9.295 > 3.820 or 5.192 > 3.820, 14.878 > 3.834, and 9.295 > 2.706, then we reject \( H(\{1,2\}) \). In the same manner, we check the set of null hypothesis in Tables 2(b), 2(c) and 3. For Table 2(b), we get the values of \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) and \( H(\{1,2,3\}) \) as above, then we reject \( H(\{1,3\}) \). For Table 2(c), we get \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) as above, then we reject \( H(\{1,4\}) \). For Table 3(a), we get the values of \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) and \( H(\{1,2,3\}) \) as 32.978, 14.878, 8.797 and 0.341, respectively, and find 32.978 > 4.585, 14.878 > 3.834, 8.797 > 3.843 and 0.341 < 2.706, then we retain \( H(\{2,3\}) \). For Table 3(b), we get the values of \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \) and \( H(\{1,2,3\}) \) as above, then we reject \( H(\{2,4\}) \). For Table 3(c), we get the values of \( \bar{\chi}^2 \) for \( H(\{1,2,3,4\}) \), \( H(\{1,2\}, \{3,4\}) \), \( H(\{2,3,4\}) \) and \( H(\{3,4\}) \) as above without the last value: 5.192, then we reject \( H(\{3,4\}) \) with 5.192 > 2.706. We obtain the result of rejecting for all pairs without the pair of dose 1mg/h and dose 5mg/h.

We do not apply the procedures \([H2], [S4] and [W7]\) for this data, because those procedures need same sample sizes.

We check the rest procedures \([T1], [C5]\). For the procedure \([T1]\), we get \( T_{12} = 3.049, T_{13} = 3.616, T_{14} = 5.675, T_{23} = 0.584, T_{24} = 2.835 \) and \( T_{34} = 2.279 \) in equation (9). The critical value

| \( \alpha \) | placebo to 1mg/h | placebo to 5mg/h | placebo to 10mg/h |
|---|---|---|---|
| 0.05 | 2.706 | 3.834 | 4.585 |
| 0.01 | 5.412 | 6.839 | 7.774 |

Table 10. Critical values for \([P6]\) on Macdonald et al. (2008)

| Hypotheses | critical values (\( \alpha = 0.05 \)) |
|---|---|
| \( H(\{1,2,3,4\}) \) | 4.585 |
| \( H(\{1,2\}, \{3,4\}) \) | 3.820, 3.820 |
| \( H(\{1,2,3\}) \) | 3.834 |
| \( H(\{1,2\}) \) | 2.706 |
| \( H(\{2,3,4\}) \) | 3.843 |
| \( H(\{2,3\}) \) | 2.706 |
| \( H(\{3,4\}) \) | 2.706 |

Table 11. Critical values for \([P3]\) on Macdonald et al. (2008)

with 14.878 > 3.834(6.839), and reject placebo to dose 1mg/h with 9.295 > 2.706(5.412). All hypotheses for each \( \alpha \) are rejected by the procedure \([P6]\) and all doses are different from placebo.

Otherwise, we demonstrate applying the procedure \([P3]\). Table 11 gives the critical values for the procedure \([P3]\) for \( \alpha = 0.05 \).

Jpn J Biomet Vol. 37, No. 2, 2016
of \([T1]\) is 2.569, then only \(H(\{2,3\})\) and \(H(\{3,4\})\) are retained.

For the procedure \([C5]\), we get same estimators as \([P3]\), because \(\hat{\nu}_i = \hat{\nu}_i\) for all \(i\). Critical values for \(H(\{1,2,3,4\})\), \(H(\{1,2\}, \{3,4\})\), \(H(\{1,2,3\})\) and \(H(\{1,2\})\) are 7.815, 5.002, 5.991 and 3.841, respectively. In the same manner of \([P3]\), we get the same result as \([P3]\).

6. Discussion

We considered \(k\) sample models with Bernoulli responses. For multiple comparisons of all proportion parameters, Shiraishi (2012) proposed single-step multiple comparison procedures based on the upper 100\(\alpha\)% points of the \(F\)-distribution. Next closed testing procedures were derived based on the proposed single-step multiple comparison tests. These proposed procedures are exactly conservative. Furthermore the asymptotic theory for the multiple comparisons was discussed. Especially as multi-step multiple comparison tests, sequentially rejective procedures could be constructed under unequal sample sizes in the asymptotic theory. Shiraishi (2011b) discussed multiple comparison tests for the differences among proportions. The simultaneous confidence intervals for all the pairwise differences among the proportions are expressed in Hochberg and Tamhane (1987). We may propose the Tukey-Kramer type multiple comparison tests similar to the simultaneous confidence intervals. However the degree of the conservativeness for the multiple comparison tests depends on unknown parameters. Therefore multiple comparison tests based on arcsine transformation were proposed in Shiraishi (2011b). It was shown that the degree of the conservativeness for the proposed tests was controlled by the sizes of the samples.

Shiraishi (2014b) discussed multiple comparison procedures based on maximum of two-sample statistics for the differences under the simple ordered restrictions (1). We could propose single step procedures based on arcsine transformation similar to Hayter (1990) of normal distribution theory. Shiraishi (2014b) proposed a closed testing procedure which is superior to the Hayter type test procedure. He showed that the closed testing procedure is superior to the closed testing procedure proposed by Shiraishi (2011b) under (1). The assumption of equality of sizes is needed in these procedures. In the present paper, we show that the closed testing procedure based on \(\bar{\chi}^2\)-tests is superior to the procedures of Shiraishi (2014b) in many cases.

For the multiple comparisons with a control, we proposed closed testing procedures based on \(\bar{\chi}^2\)-tests which is superior to the Williams type tests.

Suppose the reverse order restrictions:

\[
p_1 \geq p_2 \geq \cdots \geq p_k.
\]

(26)

Then we put \(Y_{ij} = 1 - X_{ij} (j = 1, \ldots, n_i; \ i = 1, \ldots, k)\). \((Y_{i1}, \ldots, Y_{in_i})\) is a random sample of size \(n_i\) from the \(i\)-th Bernoulli population with unknown success probability \(q_i = 1 - p_i \ (i = 1, \ldots, k)\).

(26) is equivalent to the simple ordered restrictions of \(q_i\)’s:
By replacing $X_i$ with $Y_i = \sum_{j=1}^{n_i} Y_{ij}$ in all statistics of sections 2 and 3, we can discuss the multiple comparison procedures under the restrictions (26).

The condition (C3) of same sample sizes was needed in the normal theory of all pairwise comparisons of Hayter (1990) and Shiraishi (2014a). (C3) was also needed in the Bernoulli theory of all pairwise comparisons of Shiraishi (2014b). The condition (C5) of same sample sizes was needed in the normal theory of Williams (1972) for comparisons with a control. In the present paper, we find that the proposed procedures based on $\chi^2$-statistics are applicable for the models with unequal sample sizes.

Acknowledgements

The authors would like to thank the referees for some valuable comments. The authors are grateful to Mr. Takanori Tanase of Taiho Pharmaceutical Co., LTD. for his helpful comment to example data. This research was supported in part by Nanzan University Pache Research Subsidy I-A-2 for 2016 academic year.

REFERENCES

Agresti, A. and Coull, B. A. (1996). Order-restricted tests for stratified comparisons of binomial proportions. *Biometrics* **52**, 1103–1111.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, Wiley, London.

Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika* **46**, 36–48.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics* **25**, 573–578.

Enderton, H. B. (2001). *A Mathematical Introduction to Logic: Second Edition*, Academic Press.

Hayter, A. J. (1984). A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *Annals of Statistics* **12**, 61–75.

Hayter, A. J. (1990). A one-sided Studentized range test for testing against a simple ordered alternative. *Journal of the American Statistical Association* **85**, 778–785.

Hayter, A. J. and Liu, W. (1996). On the exact calculation of the one-sided studentized range test. *Computational Statistics and Data Analysis* **22**, 17–25.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*, John Wiley, New York.
Lund, J. and Bowers, K. L. (1992). *Sinc Methods for Quadrature and Differential Equations*, siam.

Macdonald, R. L. et al. (2008). Clazosentan to Overcome Neurological Ischemia and Infarction Occurring After Subarachnoid Hemorrhage (CONSCIOUS-1) Randomized, Double-Blind, Placebo-Controlled Phase 2 Dose-Finding Trial. *Stroke* 39, 3015–3021.

Ramsey, P. H. (1978). Power differences between pairwise multiple comparisons. *Journal of the American Statistical Association* 73, 479–485.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

Shi, N. Z. (1991). A test of homogeneity of odds ratios against order restrictions. *Journal of the American Statistical Association* 86, 154–158.

Shiraishi, T. (2007). The upper bound for the distribution of Tukey-Kramer’s statistic. *Bulletin of the Computational Statistics of Japan* 19, 77-87.

Shiraishi, T. (2011a). Closed Testing Procedures for Pairwise Comparisons in Multi-Sample Models. *Japanese Journal of Biometrics* 32, 33–47 (in Japanese).

Shiraishi, T. (2011b). Multiple tests based on arcsin transformation in multi-sample models with Bernoulli responses. *Japanese Journal of Applied Statistics* 40, 1–17 (in Japanese).

Shiraishi, T. (2012). Multiple Comparisons for All Parameters in Multi-Sample Models with Bernoulli and Poisson Responses. *Journal of the Japan Statistical Society* 42, 55–90 (in Japanese).

Shiraishi, T. (2014a). Closed Testing Procedures in Multi-Sample Models under a Simple Ordered Restriction. *Journal of the Japan Statistical Society* 43, 215–245 (in Japanese).

Shiraishi, T. (2014b). Multiple Comparison Procedures for a Simple Ordered Restriction in Multi-Sample Models with Bernoulli Responses. *Japanese Journal of Applied Statistics* 43, 1–22 (in Japanese).

Shiraishi, T. and Sugiura, H. (2015). The Upper $100\alpha^*$th Percentiles of the Distributions Used in Multiple Comparison Procedures Under a Simple Order Restriction. *Journal of the Japan Statistical Society* 44, 271–314 (in Japanese).

Stenger, F. (1993). *Numerical Methods Based on Sinc and Analytic Function*, Springer-Verlag.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics* 27, 103–117.

Williams, D. A. (1972). The comparison of several dose levels are compared with a zero dose control. *Biometrics* 28, 519–531.
Appendix

A. The proof of Theorem 1

It is trivial to verify that the level of the test procedure of (b) is $\alpha$. We show that the level of the test procedure of (a) is $\alpha$.

Since $\bar{\chi}^2_{\ell_j}(I_j), \ldots, \bar{\chi}^2_{\ell_J}(I_J)$ are independent, we have

$$\lim_{n \to \infty} P_0(\bar{\chi}^2_{\ell_j}(I_j) \leq \bar{c}_{\ell_j}(\lambda(I_j); \alpha(M, \ell_j))) = \prod_{j=1}^J \left( \lim_{n \to \infty} P_0(\bar{\chi}^2_{\ell_j}(I_j) \leq \bar{c}_{\ell_j}(\lambda(I_j); \alpha(M, \ell_j))) \right)$$

$$= \prod_{j=1}^J (1 - \alpha(M, \ell_j))$$

$$= \prod_{j=1}^J \{(1 - \alpha)^{\ell_j/M}\}$$

$$= 1 - \alpha. \quad (A.1)$$

From (A.1), we have

$$\lim_{n \to \infty} P_0(\text{for some } j, \, \bar{\chi}^2_{\ell_j}(I_j) > \bar{c}_{\ell_j}(\lambda(I_j); \alpha(M, \ell_j)))$$

$$= 1 - \lim_{n \to \infty} P_0(\bar{\chi}^2_{\ell_j}(I_j) \leq \bar{c}_{\ell_j}(\lambda(I_j); \alpha(M, \ell_j)), \, j = 1, \ldots, J)$$

$$= \alpha.$$

Hence, the level of the test procedure of (a) is asymptotically $\alpha$.

B. The proof of Theorem 2

We put

$$\Theta_0 = \{p \mid \text{there exists integer } i \text{ such that } p_1 = p_i \text{ and } 2 \leq i \leq k\},$$

where $p = (p_1, \ldots, p_k)$. We take any $p \in \Theta_0$ and we put $i_0 = \max\{i \mid p_i = p_1\}$. We define $D_\ell$ by

$$D_\ell = \{\bar{c}_\ell(\lambda(I_\ell); \alpha) < \bar{\chi}^2_{\ell}(I_\ell)\}.$$  

For integer $i$ satisfying $2 \leq i \leq i_0$, the event that the procedure (VI) rejects the null hypothesis $H_i$ is given by $\bigcap_{\ell=1}^k D_\ell$. The family-wise error rate of the procedure (VI) is equal to

$$P_p \left( \bigcup_{i=2}^{i_0} \left\{ \bigcap_{\ell=1}^k D_\ell \right\} \right) \leq P_p(D_{i_0}) = P_0(D_{i_0}) \leq \alpha.$$

Hence the assertion of the theorem is proved.

Jpn J Biomet Vol. 37, No. 2, 2016

Williams, M. L. (1990). HIV seroprevalence among male IVDUs in Houston, Texas. American Journal of Public Health 80, 1507–1509.