Data Science in Kyoto Women's University 1 (2026) 79 − 91

A Study on Group Sequential Procedure for Contrast in Several Populations

Reina Hara*, Rin Tomioka* and Tomohiro Nakamura*

Key Words and Phrases: repeated confidence boundaries; several treatments with a control; hypotheses of no difference among several populations; probability of Type II error

1. Introduction

The group sequential procedure proposed by Pocock (1977) is frequently employed to test for differences in effect between two treatments (drugs) in clinical trials involving long-term therapy. The group sequential procedure is known as a testing method that reduces the average sample size, as conclusions at an early stage. Furthermore, procedure for determining the repeated confidence boundaries for testing based on a pre-specified significance level include those by O'Brien and Fleming (1979), DeMets and Ware (1980), Lan and DeMets (1983), Whitehead (1983) and Fleming Harrington and O'Brien (1984). This method involves repeated testing, which introduces the problem of multiple comparisons leading to an inflated significance level. However, this issue can be resolved by setting repeated confidence boundaries as proposed by Armitage (1969). Typically, group sequential procedure is used to test the difference in effect between two treatments.

In this study, we extend the group sequential procedure for control comparisons in multi-sample problems based on the approach of Dunnett (1955) for multiple comparisons. Based on the simultaneous hypotheses at each stage, we define test statistics and adopt the probability that at least one true null hypothesis is incorrectly rejected among multiple comparisons with the type I familywise error rate. Repeated confidence boundaries are constructed so that the cumulative nominal significance levels across stages equal a prespecified overall significance level. However, it is not practically feasible to determine the critical values requires evaluating multiple integration that depend on the number of stages K. Therefore, we derive an integral formula at each stage based on the nominal significance level at each stage. In addition, a new decision rule is introduced in this study. To construct the repeated confidence boundaries, the Pocock's method, the O'Brien and Fleming's method, and the Lan and DeMets' method are employed. Since this study considers group sequential testing for simultaneous hypotheses among several groups, Ramsey's all-pairs power is adopted as the power of the test in the

* Faculty of Data Science, Kyoto Women's University, Japan
Corresponding Author: Tomohiro Nakamura E-mail: nakamurat@kyoto-wu.ac.jp

multiple comparison. Finally, simulation studies are conducted to compare the repeated confidence boundaries and required sample size by Pocock, O'Brien and Fleming, and Lan and DeMets methods in the settings of several treatments with stages.

2. Group sequential procedure for contrast in several populations

Group sequential procedure is a method that sequentially tests the difference in effect between two treatments based on the group observations obtained at each stage. Hereafter, the point at which tests are performed based on the obtained group observations is referred to as a stage, and the maximum number of tests K is specified in advance. Here, let the number of treatments be L ($L \geq 3$), and let the response to each treatment be represented by the random variables $X_1, \cdots, X_L$. At this point, let $X_1, \cdots, X_L$ follow normal distributions $N(\mu_1, \sigma^2), \cdots, N(\mu_L, \sigma^2)$ respectively, where $\mu_1, \cdots, \mu_L$ denote the population means and σ^2 represents the known population variance. We will write these distributions as

$$X_l \sim N(\mu_l, \sigma^2)(l = 1, \cdots, L). \tag{2.1}$$

Table 2.1 Group observations at each stage

Stage	Group observations			
1	$x_{1,1}$	$x_{1,2}$	$\cdots$	x_{1,n_1}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$x_{L,1}$	$x_{L,2}$	$\cdots$	x_{L,n_1}
$\vdots$		$\vdots$		
k	$x_{1,n_{k-1}+1}$	$x_{1,n_{k-1}+2}$	$\cdots$	x_{1,n_k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$x_{L,n_{k-1}+1}$	$x_{L,n_{k-1}+2}$	$\cdots$	x_{L,n_k}
$\vdots$		$\vdots$		
K	$x_{1,n_{K-1}+1}$	$x_{n_{1,K-1}+2}$	$\cdots$	x_{1,n_K}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$x_{L,n_{K-1}+1}$	$x_{L,n_{K-1}+2}$	$\cdots$	x_{L,n_K}

Let g denote the number of samples extracted at each stage. Then, for each treatment, the cumulative number of sample n_k obtained by the stage $k(k = 1, \cdots, K)$ is given by $n_k = g \times k$.

Table 2.1 shows the group observations obtained from stage 1 to stage K. In this study, treatment 1 serves as the control, and comparative analyses are conducted with treatment $l(l = 2, \cdots, L)$. Specifically, the following simultaneous hypotheses will be tested sequentially at each stage.

$$\begin{cases} \text{Null hypotheses } H_{12}^{(0)} : \mu_2 = \mu_1, \ \cdots, \ H_{1l}^{(0)} : \mu_l = \mu_1, \ \cdots, \ H_{1L}^{(0)} : \mu_L = \mu_1, \\ \text{Alternative hypotheses } H_{12}^{(1)} : \mu_2 \neq \mu_1, \ \cdots, \ H_{1l}^{(1)} : \mu_l \neq \mu_1, \ \cdots, \ H_{1L}^{(1)} : \mu_L \neq \mu_1. \end{cases} \tag{2.2}$$

The sample mean $\bar{X}_{l,k}$ of treatment $l(l=1,\cdots,L)$ at stage $k(k=1,\cdots,K)$ is given by

$$\bar{X}_{l,k} = \sum_{j=n_{k-1}+1}^{n_k} \frac{X_{l,j}}{g} \sim N\left(\mu_l, \frac{\sigma^2}{g}\right),\ l=1,\cdots,L,\ k=1,\cdots,K \tag{2.3}$$

and let

$$Z_{l,k} = \frac{\sqrt{g}\bar{X}_{l,k}}{\sigma} \sim N\left(\frac{\sqrt{g}\mu_l}{\sigma}, 1\right),\ l=1,\cdots,L,\ k=1,\cdots,K. \tag{2.4}$$

Then

$$Z_{l,k} - Z_{1,k} \sim N\left(\frac{\sqrt{g}(\mu_l-\mu_1)}{\sigma}, 2\right),\ l=2,\cdots,L,\ k=1,\cdots,K, \tag{2.5}$$

$$U_{l,k} = \sum_{i=1}^{k} Z_{l,i} \sim N\left(\frac{k\sqrt{g}\mu_l}{\sigma}, k\right),\ l=1,\cdots,L,\ k=1,\cdots,K. \tag{2.6}$$

The test statistic at stage $k(k=1,\cdots,K)$ can be expressed as follows :

$$S_{1l,k} = U_{l,k} - U_{1,k} = \sum_{i=1}^{k}(Z_{l,i} - Z_{1,i}) \sim N\left(\frac{k\sqrt{g}(\mu_l-\mu_1)}{\sigma}, 2k\right),\ l=2,\cdots,L,\ k=1,\cdots,K. \tag{2.7}$$

2.1. Determination of repeated confidence boundaries

Conventionally developed group sequential procedure is characterized by their approach to setting repeated confidence boundaries, their foundation lies in the theory of Armitage (1969), which treated the problem as a stochastic process.

In this study, to further address simultaneous hypotheses at each stage, we adopt the type I familywise error rate, the probability that at least one correct null hypothesis is erroneously rejected in multiple comparison methods. Under the assumption that all $(L-1)$ null hypotheses $H_{1l}^{(0)} : \mu_l = \mu_1 (l=2,\cdots,L)$ hold, the test statistics $\{S_{12,1},\cdots,S_{1L,1}\}$ at stage 1, the test statistics $\{S_{12,2},\cdots,S_{1L,2}\}$ at stage 2, and the test statistics $\{S_{12,K},\cdots,S_{1L,K}\}$ at stage K correspond to the repeated confidence boundaries $b_1, b_2, \cdots, b_K$ determined as follows :

$$\begin{aligned}
\alpha_1^* &= \Pr(|S_{12,1}| > b_1 \text{ or } \cdots \text{ or } |S_{1L,1}| > b_1),\\
\alpha_2^* &= \Pr(|S_{12,1}| \le b_1,\cdots,|S_{1L,1}| \le b_1, \{|S_{12,2}| > b_2 \text{ or } \cdots \text{ or } |S_{1L,2}| > b_2\}),\\
&\vdots\\
\alpha_k^* &= \Pr(|S_{12,1}| \le b_1,\cdots,|S_{1L,1}| \le b_1,\cdots,|S_{12,k-1}| \le b_{k-1},\cdots,|S_{1L,k-1}| \le b_{k-1},\\
&\qquad \{|S_{12,k}| > b_k \text{ or } \cdots \text{ or } |S_{1L,k}| > b_k\}),\\
&\vdots\\
\alpha_K^* &= \Pr(|S_{12,1}| \le b_1,\cdots,|S_{1L,1}| \le b_1,\cdots,|S_{12,K-1}| \le b_{K-1},\cdots,|S_{1L,K-1}| \le b_{K-1},\\
&\qquad \{|S_{12,K}| > b_K \text{ or } \cdots \text{ or } |S_{1L,K}| > b_K\}).
\end{aligned} \tag{2.8}$$

Under the conditions that this relationship holds, for a specified significance level α, the critical values $b_1,\cdots,b_K$ are sequentially determined such that

$$\alpha = \sum_{i=1}^{K} \alpha_i^*. \tag{2.9}$$

To determine the repeated confidence boundaries $b_k (k = 1, \cdots, K)$ involve the joint distribution of $S_{1l,k} (l = 2, \cdots, L, k = 1, \cdots, K)$ in equation (2.8), we need to derive the probability density function of the $K(L-1)$ -variable normal distribution of $S_{1l,k} (l = 2, \cdots, L, k = 1, \cdots, K)$. However, multiple integrations depend on the number of treatments L and the number of stages K are impractical. In this study, we attempt to derive repeated integral formula to compute b_k at each stage.

Here, under the assumption that all $(L-1)$ null hypotheses $H_{1l}^{(0)} : \mu_l = \mu_1 (l = 2, \cdots, L)$ hold, we have $\mu_1 = \cdots = \mu_L$. Furthermore, when we compute the repeated confidence boundaries, we may set $\mu_1 = \cdots = \mu_L = 0$ without loss of generality. Therefore, (2.4) and (2.6) are respectively as follows :

$$Z_{l,k} = \frac{\sqrt{g}\bar{X}_{l,k}}{\sigma} \sim N\Big(\frac{\sqrt{g}\mu_l}{\sigma}, 1\Big) \equiv N(0,1),\ l = 1, \cdots, L,\ k = 1, \cdots, K, \tag{2.10}$$

$$U_{l,k} = \sum_{i=1}^{k} Z_{l,i} \sim N\Big(\frac{k\sqrt{g}\mu_l}{\sigma}, k\Big) \equiv N(0,k),\ l = 1, \cdots, L,\ k = 1, \cdots, K. \tag{2.11}$$

Stage 1

From (2.8),

$$\alpha_1^* = \Pr(|S_{12,1}| > b_1 \text{ or } \cdots \text{ or } |S_{1L,1}| > b_1).$$

This expression can be transformed as follows :

$$\begin{aligned} 1 - \alpha_1^* &= \Pr(|S_{12,1}| \le b_1, \cdots, |S_{1L,1}| \le b_1) = \Pr(|Z_{2,1} - Z_{1,1}| \le b_1, \cdots, |Z_{L,1} - Z_{1,1}| \le b_1) \\ &= \int_{-\infty}^{\infty} \Pr(|Z_{2,1} - z_{1,1}| \le b_1, \cdots, |Z_{L,1} - z_{1,1}| \le b_1 | Z_{1,1} = z_{1,1}) g(z_{1,1}) dz_{1,1} \\ &= \int_{-\infty}^{\infty} \Big(\int_{-b_1 + \tilde{z}_1}^{b_1 + \tilde{z}_1} f(z_1) dz_1 \Big)^{L-1} \times g(\tilde{z}_1) d\tilde{z}_1, \end{aligned} \tag{2.12}$$

where $g(z_{l,1})(l = 1, \cdots, L)$ are the probability density functions of $Z_{l,1} \sim N(0,1)(l = 1, \cdots, L)$ respectively and where $z_{l,1} \equiv z_1 \equiv \tilde{z}_1$.

Stage $k(k = 2, \cdots, K)$

From (2.8),

$$\begin{aligned} \alpha_k^* = \Pr(&|S_{12,1}| \le b_1, \cdots, |S_{1L,1}| \le b_1, \cdots, |S_{12,k-1}| \le b_{k-1}, \cdots, |S_{1L,k-1}| \le b_{k-1}, \\ &\{|S_{12,k}| > b_k \text{ or } \cdots \text{ or } |S_{1L,k}| > b_k\}). \end{aligned}$$

This expression can be transformed as follows :

$$
\begin{aligned}
&1-(\alpha_1^*+\cdots+\alpha_k^*)\\
&=\Pr(|S_{12,1}|\le b_1,\cdots,|S_{1L,1}|\le b_1,\cdots,|S_{12,k-1}|\le b_{k-1},\cdots,|S_{1L,k-1}|\le b_{k-1},|S_{12,k}|\le b_k,\cdots,|S_{1L,k}|\le b_k\})\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Pr(|Z_{2,1}-z_{1,1}|\le b_1,\cdots,|Z_{L,1}-z_{1,1}|\le b_1,\cdots,\\
&\quad|Z_{2,1}+\cdots+Z_{2,k-1}-(z_{1,1}+\cdots+z_{1,k-1})|\le b_{k-1},\cdots,|Z_{L,1}+\cdots+Z_{L,k-1}-(z_{1,1}+\cdots+z_{1,k-1})|\le b_{k-1},\\
&\quad|Z_{2,1}+\cdots+Z_{2,k}-(z_{1,1}+\cdots+z_{1,k})|\le b_k,\cdots,|Z_{L,1}+\cdots+Z_{L,k}-(z_{1,1}+\cdots+z_{1,k})|\le b_k\\
&\quad|Z_{1,1}=z_{1,1},\cdots,Z_{1,k}=z_{1,k})g(z_{1,1})\cdots g(z_{1,k-1})g(z_{1,k})dz_{1,1}\cdots dz_{1,k-1}dz_{1,k}\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\int_{-b_k+z_{1,1}+\cdots+z_{1,k}}^{b_k+z_{1,1}+\cdots+z_{1,k}}\int_{-b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}^{b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}f(u_{2,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{2,k}-u_{2,k-1})^2}{2}\right\}du_{2,k-1}dz_{2,k}\right)\\
&\quad\times\cdots\times\left(\int_{-b_k+z_{1,1}+\cdots+z_{1,k}}^{b_k+z_{1,1}+\cdots+z_{1,k}}\int_{-b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}^{b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}f(u_{L,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{L,k}-u_{L,k-1})^2}{2}\right\}du_{L,k-1}dz_{L,k}\right)\\
&\quad\times g(z_{1,1})\cdots g(z_{1,k-1})g(z_{1,k})dz_{1,1}\cdots dz_{1,k-1}dz_{1,k}\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-b_k+\tilde{u}_{k-1}+\tilde{z}_k}^{b_k+\tilde{u}_{k-1}+\tilde{z}_k}\int_{-b_{k-1}+\tilde{u}_{k-1}}^{b_{k-1}+\tilde{u}_{k-1}}f(u_{k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_k-u_{k-1})^2}{2}\right\}du_{k-1}dz_k\right)^{L-1}\\
&\quad\times f(\tilde{u}_{k-1})g(\tilde{z}_k)d\tilde{u}_{k-1}d\tilde{z}_k,
\end{aligned}
\tag{2.13}
$$

where $g(z_{l,k})$, $f(u_{l,k-1})$ are the probability density functions of $Z_{l,k}\sim N(0,1)$, $U_{l,k-1}=Z_{l,1}+\cdots+Z_{l,k-1}\sim N(0,k-1)(l=1,\cdots,L,k=1,\cdots,K)$ respectively and $u_{l,k-1}\equiv u_{k-1}\equiv\tilde{u}_{k-1}(l=1,\cdots,L-1,k=2,\cdots,K)$, $z_{l,k}\equiv z_k\equiv\tilde{z}_k(l=1,\cdots,L,k=1,\cdots,K)$.

In our group sequential procedure, given the repeated confidence boundaries $b_1,\cdots,b_K$ at each stage and the real values $s_{12,1},\cdots,s_{1L,1},\cdots,s_{12,K},\cdots,s_{1L,K}$ of the test statistics $S_{12,1}$, $\cdots,S_{1L,1},\cdots,S_{12,K},\cdots,S_{1L,K}$ are given, the decision rule is as follows :

• **Stage** $k(k=1,\cdots,K-1)$

(a) If $|s_{1l,k}|\le b_k$ holds for all $l(l=2,\cdots,L)$, then refrain from making a decision and continue to the next stage.

(b) If $|s_{1l,k}|>b_k$ for at least one among $l(l=2,\cdots,L)$, terminate the test based on the following decision.

- If $|s_{1l,k}|\le b_k(l=2,\cdots,L)$, then accept the corresponding null hypothesis $H_{1l}^{(0)}$.
- If $|s_{1l,k}|>b_k(l=2,\cdots,L)$, then reject the corresponding null hypothesis $H_{1l}^{(0)}$.

• **Stage** K

(c) If $|s_{1l,K}|\le b_K$ holds for all l $(l=2,\cdots,L)$, then accept the corresponding null hypothesis $H_{1l}^{(0)}$ and terminate the test.

(d) If $|s_{1l,K}|>b_K$ for at least one among $l(l=2,\cdots,L)$, terminate the test based on the following decision.

- If $|s_{1l,K}|\le b_K(l=2,\cdots,L)$, then accept the corresponding null hypothesis $H_{1l}^{(0)}$.
- If $|s_{1l,K}|>b_K(l=2,\cdots,L)$, then reject the corresponding null hypothesis $H_{1l}^{(0)}$.

2.2. α spending function and repeated confidence boundaries

Here, we describe three representative methods for determining the repeated confidence boundary.

(1) Pocock's method (1977)

This method uses a common value *Z* at all stages. The method for determining *Z* involves applying the Armitage's method to the repeated confidence boundaries

$$b_k = Z \cdot \sqrt{k}, \quad k = 1, \cdots, K \tag{2.14}$$

obtaining α_k^*.The value of *Z* is determined by iteratively varying it until it satisfies $\alpha = \sum_{i=1}^{k} \alpha_k^*$. Specifically, it is a method that repeatedly determines the confidence boundaries by fixing *Z* at each stage and matching a pre-specified significance level *α* with a maximum of *K* tests. It is more likely to be terminated early, however if the test is not terminated early, the final analysis will be conservative.

(2) O'Brien and Fleming's method (1979)

This method uses $U\sqrt{K/k}$. The value of *U* is determined using

$$b_k = U\sqrt{\frac{K}{k}}\sqrt{k} = U \cdot \sqrt{K}, \quad k = 1, \cdots, K \tag{2.15}$$

and is determined via Monte Carlo simulation.

(3) Lan and DeMets' method (1983)

This method does not require pre-setting the number of interim analyses and can be configured for either equidistant or non-equidistant intervals, allowing the repeated confidence boundaries to be determined at any point. This method assumes the following *α* spending function, allocates the significance level across each stage, and applies the Armitage's method based on this allocation to determine the repeated confidence boundaries. The Lan and DeMets' method primarily consists of three approaches. These are characterized by their allocation of the significance level *α* and are referred to as the Normal method, Log method, and *αt* method, respectively.

(a) Normal method

In this method, assuming Φ is the standard normal distribution and $Z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ percentile point of the standard normal distribution, *α* spending function α_k is defined as follows :

$$\alpha_k = 2\left\{1 - \Phi\left(Z_{\frac{\alpha}{2}}\sqrt{\frac{K}{k}}\right)\right\}, \quad k = 1, \cdots, K. \tag{2.16}$$

This method has the characteristic of assigning a low probability at an early stage, with that probability increasing as the number of stages increases.

(b) Log method

In the log method, α spending function α_k is defined as the following logarithmic function.

$$\alpha_k = \alpha \log\left\{1 + (e-1)\frac{k}{K}\right\}, \quad k = 1, \cdots, K. \tag{2.17}$$

This method has the characteristic of assigning a large probability at an early stage, with that probability decreasing as the number of stages increases.

(c) *αt* method

In the αt method, let $t = k/K$ and define α spending function α_k as the following linear function.

$$\alpha_k = \alpha \times \frac{k}{K} = \alpha t, \quad k = 1, \cdots, K. \tag{2.18}$$

This method assumes the probability α_k at each stage is constant and seeks to determine b_k.

From the above methods, determine the significance level α_k for each stage. Then, from $\alpha_k^* = \alpha_k - \alpha_{k-1}$, determine the α_k^* for each stage. This α_k^* satisfies the following relationship.

$$\alpha = \sum_{k=1}^{K} \alpha_k^*. \tag{2.19}$$

2.3. Power and required sample size

The power of the test of group sequential procedure for the difference between two population means is the probability exceeds the repeated confidence boundaries under the alternative hypothesis. Since this study employs a group sequential procedure for simultaneous hypotheses involving three or more population means, we adopt the all-pairs power defined by Ramsey (1978) as the probability of detecting all pairs with differences between population means for multiple comparisons.

Now,

$$\frac{\sqrt{g}\mu_l}{\sigma} = \Delta_l (l = 1, \cdots, L), \tag{2.20}$$

then (2.4) and (2.6) become respectively as follows.

$$Z_{l,k} = \frac{\sqrt{g}\bar{X}_{l,k}}{\sigma} \sim N\left(\frac{\sqrt{g}\mu_l}{\sigma}, 1\right) = N(\Delta_l, 1),\ U_{l,k} = \sum_{i=1}^{k} Z_{l,i} \sim N\left(\frac{k\sqrt{g}\mu_l}{\sigma}, k\right) = N(k\Delta_l, k),\ l = 1, \cdots, L.$$

(2.21)

The probability $\pi_k^*(k=1,\cdots,K)$ of accepting all $(L-1)$ alternative hypotheses $H_{1l}^{(1)}$: $\mu_l \neq \mu_1 (l=2,\cdots,L)$ that reperesent differences between population means at each stage is as follows :

$$\begin{aligned}
\pi_1^* &= \Pr(|S_{12,1}| > b_1, \cdots, |S_{1L,1}| > b_1), \\
\pi_2^* &= \Pr(|S_{12,1}| \leq b_1, \cdots, |S_{1L,1}| \leq b_1, |S_{12,2}| > b_2, \cdots, |S_{1L,2}| > b_2), \\
&\vdots \\
\pi_k^* &= \Pr(|S_{12,1}| \leq b_1, \cdots, |S_{1L,1}| \leq b_1, \cdots, |S_{12,k-1}| \leq b_{k-1}, \cdots, |S_{1L,k-1}| \leq b_{k-1}, \\
&\qquad |S_{12,k}| > b_k, \cdots, |S_{1L,k}| > b_k), \\
&\vdots \\
\pi_K^* &= \Pr(|S_{12,1}| \leq b_1, \cdots, |S_{1L,1}| \leq b_1, \cdots, |S_{12,K-1}| \leq b_{K-1}, \cdots, |S_{1L,K-1}| \leq b_{K-1}, \\
&\qquad |S_{12,K}| > b_K, \cdots, |S_{1L,K}| > b_K).
\end{aligned} \tag{2.22}$$

For the specified Type II error probability β,

$$1-\beta = \pi_1^* + \cdots + \pi_K^*. \tag{2.23}$$

The value of $\Delta_l = \sqrt{g}\mu_l/\sigma (l=1,\cdots,L)$ is sequentially adjusted until this equation is satisfied. Here, σ is known, and we sequentially modify the value of g for $\mu_1,\cdots,\mu_L$ under the alternative hypotheses. $\pi_k^*(k=1,\cdots,K)$ is obtained via the following integrations.

Stage 1

From (2.22),

$$\begin{aligned}
\pi_1^* &= \Pr(|S_{12,1}| > b_1, \cdots, |S_{1L,1}| > b_1) = \Pr(|Z_{2,1} - Z_{1,1}| > b_1, \cdots, |Z_{L,1} - Z_{1,1}| > b_1) \\
&= \int_{-\infty}^{\infty} \Pr(|Z_{2,1} - z_{1,1}| > b_1, \cdots, |Z_{L,1} - z_{1,1}| > b_1 | Z_{1,1} = z_{1,1}) g(z_{1,1}) dz_{1,1} \\
&= \int_{-\infty}^{\infty} \prod_{l=2}^{L} \left(\int_{-\infty}^{-b_1+z_{1,1}} f(z_{l,1}) dz_{l,1} + \int_{b_1+z_{1,1}}^{\infty} f(z_{l,1}) dz_{l,1} \right) g(z_{1,1}) dz_{1,1},
\end{aligned} \tag{2.24}$$

where $g(z_{1,1}), f(z_{l,1}) (l = 2,\cdots,L)$ are both probability density functions of $Z_{l,1} \sim N(\Delta_l, 1)$ $(l=1,\cdots,L)$ respectively.

Stage $k(k=2,\cdots,K)$

From (2.22),

$$
\begin{aligned}
&\pi_k^* \\
&= \Pr(|S_{12,1}| \le b_1, \cdots, |S_{1L,1}| \le b_1, \cdots, |S_{12,k-1}| \le b_{k-1}, \cdots, |S_{1L,k-1}| \le b_{k-1}, |S_{12,k}| > b_k, \cdots, |S_{1L,k}| > b_k) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \Pr(|Z_{2,1}-z_{1,1}| \le b_1, \cdots, |Z_{L,1}-z_{1,1}| \le b_1, \cdots, \\
&\quad |Z_{2,1}+\cdots+Z_{2,k-1}-(z_{1,1}+\cdots+z_{1,k-1})| \le b_{k-1}, \cdots, |Z_{L,1}+\cdots+Z_{L,k-1}-(z_{1,1}+\cdots+z_{1,k-1})| \le b_{k-1}, \\
&\quad |Z_{2,1}+\cdots+Z_{2,k}-(z_{1,1}+\cdots+z_{1,k})| > b_k, \cdots, |Z_{L,1}+\cdots+Z_{L,k}-(z_{1,1}+\cdots+z_{1,k})| > b_k \\
&\quad |Z_{1,1}=z_{1,1}, \cdots, Z_{1,k}=z_{1,k}) g(z_{1,1})\cdots g(z_{1,k-1}) g(z_{1,k}) dz_{1,1}\cdots dz_{1,k-1}dz_{1,k} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Bigg(\int_{-\infty}^{-b_k+z_{1,1}+\cdots+z_{1,k}}\int_{-b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}^{b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}} \\
&\qquad f(u_{2,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{2,k}-u_{2,k-1}-\Delta_2)^2}{2}\right\}du_{2,k-1}dz_{2,k} \\
&\quad + \int_{b_k+z_{1,1}+\cdots+z_{1,k}}^{\infty}\int_{-b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}^{b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}} f(u_{2,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{2,k}-u_{2,k-1}-\Delta_2)^2}{2}\right\}du_{2,k-1}dz_{2,k}\Bigg)\times\cdots \\
&\quad \times\Bigg(\int_{-\infty}^{-b_k+z_{1,1}+\cdots+z_{1,k}}\int_{-b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}^{b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}} f(u_{L,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{L,k}-u_{L,k-1}-\Delta_L)^2}{2}\right\}du_{L,k-1}dz_{L,k} \\
&\quad + \int_{b_k+z_{1,1}+\cdots+z_{1,k}}^{\infty}\int_{-b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}}^{b_{k-1}+z_{1,1}+\cdots+z_{1,k-1}} f(u_{L,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{L,k}-u_{L,k-1}-\Delta_L)^2}{2}\right\}du_{L,k-1}dz_{L,k}\Bigg) \\
&\quad \times g(z_{1,1})\cdots g(z_{1,k})dz_{1,1}\cdots dz_{1,k} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{l=2}^{L}\Bigg(\int_{-\infty}^{-b_k+u_{1,k-1}+z_{1,k}}\int_{-b_{k-1}+u_{1,k-1}}^{b_{k-1}+u_{1,k-1}} f(u_{l,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{l,k}-u_{l,k-1}-\Delta_l)^2}{2}\right\}du_{l,k-1}dz_{l,k} \\
&\quad + \int_{b_k+u_{1,k-1}+z_{1,k}}^{\infty}\int_{-b_{k-1}+u_{1,k-1}}^{b_{k-1}+u_{1,k-1}} f(u_{l,k-1})\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z_{l,k}-u_{l,k-1}-\Delta_l)^2}{2}\right\}du_{l,k-1}dz_{l,k}\Bigg) \\
&\quad \times f(u_{1,k-1})g(z_{1,k})du_{1,k-1}dz_{1,k}, \qquad (2.25)
\end{aligned}
$$

where $g(z_{1,k})$, $f(u_{l,k-1})$ are the probability density functions of $Z_{1,k} \sim N(\Delta_1, 1)(k=1,\cdots,K)$ respectively and $U_{l,k-1} = Z_{l,1}+\cdots+Z_{l,k-1} \sim N((k-1)\Delta_l, k-1)(l=1,\cdots,L, k=2,\cdots,K)$, respectively.

Since $\sqrt{g}\mu_l/\sigma = \Delta_l(l=1,\cdots,L)$ by (2.20),

$$g = \frac{\sigma^2\Delta_l^2}{\mu_l^2},\ l=1,\cdots,L. \qquad (2.26)$$

The maximum sample size per treatment is given by $n_{\max} = gK$.

3. Simulations

Table 3.1 shows the repeated confidence boundaries of Pocock's method, O'Brien and Fleming's method, and Lan and DeMets' method when α =0.05 and K =3 in the existing group sequential procedure for difference between two population means with L =2.

Table 3.1 Comparison of repeated confidence boundaries (L =2, K = 3)

	b_1	b_2	b_3
Pocock′s method	2.298	3.250	3.980
O′Brien and Fleming′s method	3.471	3.471	3.471
Normal method	3.395	3.403	3.492
αt method	2.394	3.244	3.843
Log method	2.279	3.245	4.020

The O'Brien and Fleming's method maintains the same value throughout all stages and exhibits characteristics nearly identical to the Normal method. The other three methods show an increase in the repeated confidence boundaries as the stages progress. Next, Table 3.2 and 3.3 show the repeated confidence boundaries in our study for Pocock's method, O'Brien and Fleming's method and Lan and DeMets' method when α =0.05 and (L, K) = (3, 3), (4, 4) respectively.

Table 3.2 Comparison of repeated confidence boundaries (L =3, K = 3)

	b_1	b_2	b_3
Pocock′s method	3.386	4.788	5.864
O′Brien and Fleming′s method	5.498	5.498	5.498
Normal method	5.058	5.269	5.554
αt method	3.708	5.069	5.637
Log method	3.555	5.075	5.634

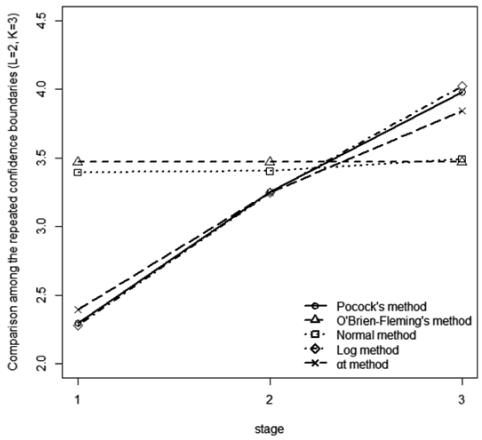

Fig.3.1 Comparison among the repeated confidence boundaries (L =2, K = 3)

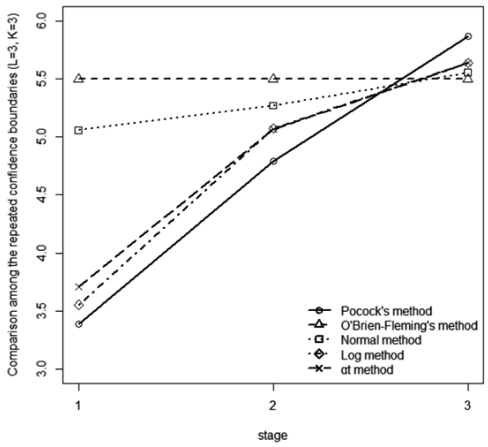

Fig.3.2 Comparison among the repeated confidence boundaries (L =3, K = 3)

Table 3.3 Comparison of repeated confidence boundaries (L =4, K = 4)

	b_1	b_2	b_3	b_4
Pocock′s method	3.559	5.033	6.164	7.118
O′Brien and Fleming′s method	6.782	6.782	6.782	6.782
Normal method	5.905	6.198	6.528	6.863
αt method	4.020	5.540	6.238	7.043
Log method	3.852	5.470	6.146	7.139

Similar to the conventional group sequential procedure for the difference between two population means, the O'Brien and Fleming's method maintains the same value throughout all stages, exhibiting characteristics nearly identical to the Normal method. The other three methods show increasing repeated confidence baundaries as the stage progresses. Furthermore, as the number of L and K increases, the number of tests performed also increases, making the group sequential test more conservative. Consequently, the values of the repeated confidence boundaries also increase as the number of L and K increases.

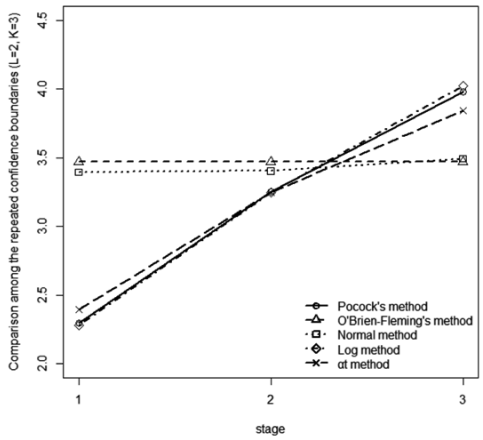

Fig.3.3 Comparison among the repeated confidence boundaries (L =4, K = 4)

Next, Table 3.4, 3.5 show the required sample sizes g at each stage of Pocock's method, O'Brien and Fleming's method and Lan and DeMets' method for the configurations of population means (μ_1, μ_2, μ_3) under K =3, L =3 and $(\mu_1, \mu_2, \mu_3, \mu_4)$ under K =4, L =4 with α =0.05, σ^2 =1, and power 1−β =0.8 respectively.

Table 3.4 Comparison of required sample size g at each stage(L =3, K =3, $1-\beta$ =0.8)

(μ_1, μ_2, μ_3)	$(1, 1.5, 2)$	$(1, 2, 3)$	$(1, 3, 5)$
Pocock′s method	84	21	6
O′Brien and Fleming′s method	179	45	12
Normal method	157	40	10
αt method	96	24	6
Log method	91	23	6

Table 3.5 Comparison of required sample size g at each stage(L =4, K =4, $1-\beta$ =0.8)

$(\mu_1, \mu_2, \mu_3, \mu_4)$	$(1, 1.5, 2, 2.5)$	$(1, 2, 3, 4)$	$(1, 3, 5, 7)$
Pocock′s method	91	23	6
O′Brien and Fleming′s method	255	64	16
Normal method	202	51	13
αt method	109	28	7
Log method	102	26	7

The required sample size g for O'Brien and Fleming's method and Normal method, which repeatedly use large repeated confidence boundaries from the stage 1, is the largest. Furthermore, as the number of L and K increases, the number of tests performed increases, making the tests more conservative. Consequently, the required sample size g also increases as the number of L and K increases.

4. Further discussions

This study broadens conventional group sequential procedure typically formulated for one or two sample settings to situations involving three or more groups. In addition, drawing on Tukey (1953) in multiple comparison theory, we explore a sequential framework that enables stepwise evaluation of all pairwise differences among several population means. Also we simulated to calculate the repeated confidence boundaries and the required sample size among Pocock (1977), O'Brien and Fleming (1979) and Lan and DeMets (1983) in the settings of several treatments with stages.

Nonetheless, several issues arise. We adopt Ramsey (1978) of power of the test for multiple comparisons, known as all-pairs power. In practice, however, simultaneously identifying every true difference among population means is often difficult. For this reason, an alternative criterion the any-pair power, derived by replacing the intersection with a union in the definition also warrants consideration. We will obtain an analogous integral expression for this quantity and

examine how the two notions of power differ in future.

Typically, in group sequential procedures, acceptance of null hypotheses is deferred until the final stage. In contrast, our setup halts the procedure immediately once any null hypothesis is rejected, leaving the remaining hypotheses unresolved when termination occurs. Moving forward, we intend to study a rule that would allow unrejected hypotheses to proceed to subsequent stages. Under such a scheme, statistics corresponding to rejected hypotheses would no longer be examined later, reducing the number of active comparisons as the test progresses. Consequently, the repeated confidence boundaries must be recomputed at each stage. This modification has the potential to yield a group sequential procedure with improved statistical efficiency. If future, we develop a new procedure to correct the above points and we would like to apply this procedure to adaptive design.

References

[1] Armitage, P., McPherson, C. K. and Rowe, B. C. (1969). Repeated Significance Tests on Accumulating data. J. R. Statist. Soc. A, 132, 235-244.

[2] DeMets, D. L. and Ware, J. H. (1980). Group Sequential Methods for Clinical Trials with a One-sided Hypothesis. Biometrics, 67, 3, 651-660.

[3] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 50, 1096-1121.

[4] Fleming, T. R.,Harrington, D. P. and O'Brian, P. C. (1984). Designs for Group Sequential Tests. Controlled Clinical Trials , 5, 348-361.

[5] Lan, K. K. G. and DeMets, D. L. (1983). Discrete Sequential Boundaries for Clinical Trials. Biometrika, 70, 659-663.

[6] O'Brien, P. C. and Fleming, T. R. (1979). A Multiple Testing Procedure for Clinical Trials. Biometrics, 35, 549-556.

[7] Pocock, S. J. (1977). Group Sequential Methods in the Design and Analysis of Clinical Trials. Biometrika, 64, 191-199.

[8] Ramsey, P. H. (1978). Differences Between Pairwise Multiple Comparisons. Journal of the American Statistical Association, 73, 479-485.

[9] Whitehead, J. (1983). The Design and Analysis of Sequential Clinical Trials. Chichester, Ellis Horwood.

Received: 19 February 2026