AJAS / SPAS

African Journal of Applied Statistics
Vol. 6 (1), 2019, pages 635 – 647.
DOI: http://dx.doi.org//10.16929/ajas/2019.635.235

ISSN 2316-0861

Identifying maximum safe dose (MSD) in a (k+2)-arm trial under unequal variances

Michael Jackson Adjabui [1,*], **Emmanuel Dodzi Kpeglo** [2] **and Christian Akrong Hesse** [2]

[1] Department of Mathematics, University for Development Studies, Navrongo Ghana.
[2] Department of Mathematical Sciences, Methodist University, Dansoman-Accra Ghana.

Received on February 12, 2019, Accepted on May 20, 2019. Online on June 16, 2019.

Abstract. In this paper, we address the problem of identification of maximum safe dose, where the maximum safe is defined as the highest dose that is non-inferior to the reference treatment (negative control) dose such that any other lower dose is also non-inferior to the reference dose. Statistical methodology for non-inferiority trial for several multiple treatments in establishing efficacy of new treatment readily available for multiplicity adjustment. We extend this methodology to identification of maximum safe dose in a toxicological experiment by incorporating the confidence set-based stepwise procedure using Fieller's confidence interval method under heteroscedasticity without multiplicity adjustment. Our simulation studies indicate that the familywise error rate (FWER) is properly controlled in a strong sense, and the power of our procedure depicts that power increases with increasing in sample size and the ratio of mean difference. Results from our studies support the theoretical findings.

Key words: Confidence set; Fieller's confidence intervals; Multiple treatments; Non-inferiority trials.
AMS 2010 Mathematics Subject Classification Objects : 62-02; 62H15; 62J15.

Presented by Dr. Richard Minkah, University of Ghana, Legon, Accra, Ghana
Member of the Editorial Board.

---

*Corresponding author Michael Jackson Adjabui: madjabui@uds.edu.gh
Emmanuel Dodzi Kpeglo :ekpeglo@mucg.edu.gh
Christian Akrong Hesse : akrongh@yahoo.com

Résumé. (Abstract in French). Nous traitons le problème de la détermination de la dose maximale de securité dans le domaine médical. La methode statistique pour les essais de no-inforieté pour des traitements multiples permettent de mettre en place un nouveu traitement rapidement disponible avec des ajustements mutilples. Nous étendons cette méthode en incorporant des etapes utilisant des intervalles de confiance de type Fieller avec hétérocedasticity sans répétition des ajustements. Nos résultats théoriques sont simulés et les résultats de la simulation démontre des gains de notre methodology en terme d'erreurs par famille (FWER), de puissance des tests et de rapport des moyennes

1. Introduction

The main purpose of non-inferiority clinical three-arm trial is to demonstrate both the assay sensitivity of trial and non-inferiority of the new treatment compared with the reference treatment by some pre-specified non-inferiority margin. This is necessary in assessing whether the new treatment is clinically not worse than the reference treatment in so far as the new treatment maintain sizable fraction of the effectiveness of the reference treatment. The reference treatment is replaced with the new treatment if it can be proven that the benefits of the new treatment outweigh the loss in efficacy of the reference treatment at an acceptable margin. There has been a lot of research on three-arm non-inferiority trials: in assessing the efficacy of a new treatment [see Pigeot *et al.* (2003), Kwong *et al.* (2014), Huang *et al.* (2014), Huang *et al.* (2015)] and as proof of safety in toxicological experiments [see Hauschke *et al.* (2005), Hasler *et al.* (2008), Hasler (2012)]. For example, Hauschke *et al.* (2005) considered mutagenicity data set for micronucleus assay published by Adler and Kliesch (1990).They evaluated proof of safety by using the concept of maximum safe dose by formulating a problem based on Bartholomew test and incorporated a biological meaningful threshold value as a fraction of the difference between a positive and negative control. An inherent issue about proof of safety is the control of the consumer risk. That is probability of erroneously concluding on safety at a pre-specified margin. In other words, the control of the FWER is critical in such investigation in order to maintain α level.

Statistical procedures have been developed to analyze data collected from non-inferiority trials but they focused on situations with only one experimental treatment [see Pigeot *et al.* (2003), Hasler *et al.* (2008)]. However, non-inferiority clinical trials may involve different combination of several new drugs or different doses of a new drug. In this case, appropriate statistical methods involving several experimental treatments (multiple treatments) is required [see Hasler (2012), Kwong *et al.* (2014), Huang *et al.* (2014), Huang *et al.* (2015)]. For example Hasler (2012) extended the theory of three-arm trial with one experimental treatment to $(k+2)$-arm trials for safety endpoint by considering cases with k experimental treatments. This necessitates multiple comparisons and multiple testing procedures. For this reasons, they employed single step procedure and adjusted for multiplicity effect in order to maintain the FWER at a designated level α. On the other hand, no multiplicity adjustment is necessary if the experimental

groups (new treatments) can be ordered a priori according to their treatment effect. It is known in literature that stepwise procedure is more powerful than single step procedure. Stepwise procedure for multiple comparisons settings were proposed by Bonfiger (1987), Stefansson *et al.* (1988), Cao *et al.* (2015), Adjabui *et al.* (2016): where the individual inferences are made in stepwise fashion if the sequence of individual inferences is in a specific order . Hsu and Berger (1999) proposed stepwise confidence set-based procedure in a dose-response study for identification of minimum effective dose without multiplicity adjustment. In their procedure, unknown and equal variances across dose group under normality were assumed and they also incorporated the partitioning principle proposed by Finner and Strassburger (2002) in their investigations. However, equal variance assumption is seldom satisfied in practice. Tao *et al.* (2002) extended Hsu and Berger (1999) procedure for a situation of unknown and unequal variances across dose groups by employing two-stage sampling procedure proposed by Stein (1945) and incorporating the partitioning principle proposed by Finner and Strassburger (2002) for identifying minimum effective dose. But the issue of unknown and unequal variances or heteroscedasticity has been a long standing problem in multiple comparison procedures since the pioneering work of Welch (1938). The situation is far from being settled. Therefore, we propose a stepwise confidence set procedure without multiplicity adjustment and incorporating the partitioning principle for identifying maximum safe dose within the framework of non-inferiority trials under the ratio of mean difference for a normally distributed data. In other words, we extend the procedure proposed by Hsu and Berger (1999) to a situation of unequal variances by using the partition principle proposed by Finner and Strassburger (2002).

The outline of the article is as follows. In Section 2 the test statistics and the confidence intervals for the problem are formulated and derived. Algorithm for our proposed procedure is discussed in Section 3. Simulation studies is conducted to assess the performance of the power and the FWER of our stepwise procedure in Section 4. In Section 5 we employ our proposed procedure to examine a real data set. Conclusion is presented in Section 6.

2. Statistical background

Let X_{ij} denote observations for $i = 0, \cdots k+1$ and $j = 1, 2, \cdots n_i$, where $i = 1, 2, \cdots, k$ are the experimental treatment groups, $i = 0$ for the negative control (placebo group (P)) and $i = k+1$ for the positive control (reference group (R)) respectively. Assume that $X'_{ij}s$ are mutually independent and follows normal distribution with means μ_i and unknown variances σ_i^2, in other words $X_{ij} \sim N(\mu_i, \sigma_i^2)$ for $i = 0, 1, \cdots, k+1$. The sample variances and the sample means are denoted as S_i^2 and $\bar{X}_i$ respectively.

The test problem is formulated as:

$$H_{0i} : \mu_i - \mu_{k+1} \geq \delta \text{ versus } H_{1i} : \mu_i - \mu_{k+1} < \delta \quad \text{for} \quad i = 1, 2, \cdots k. \tag{1}$$

When a placebo group is added to the set up problem in Equation (1) for some ethical reasons, the safety threshold and non-inferiority margin δ can be expressed

as a fraction of difference between reference and placebo drug by $\delta = f(\mu_{k+1} - \mu_0)$, this is a as result of Pigeot *et al.* (2003). The testing problem in Equation (1) can be written as:

$H_{0i} : \mu_i - \mu_{k+1} \geq f(\mu_{k+1} - \mu_0)$ versus $H_{1i} : \mu_i - \mu_{k+1} < f(\mu_{k+1} - \mu_0)$ for $i = 1, 2, \cdots k$
or
$H_{0i} : \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} \geq 1 + f$ versus $H_{0i} : \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} < 1 + f.$

Letting $\theta = 1 + f$, this can be rewritten as:

$$H_{0i} : \lambda_i \geq \theta \text{ versus } H_{1i} : \lambda_i < \theta \tag{2}$$

where λ_i is the ratio of difference in means denoted as :

$$\lambda_i = \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} \text{ for } i = 1, 2, \cdots k, \tag{3}$$

where $\theta \in (0, 1)$ being the pre-specified maximum fraction of the effect of reference group relative to the placebo, which is in effect of treatment groups relative to placebo group. This requirement has to be preserved in order to demonstrate non-inferiority. The specification of θ depends on combination of clinical relevance and statistical judgment. Equation (3) is valid if and only if $\mu_{k+1} - \mu_0 > 0$, this is necessary for assessment of assay sensitivity of the trial and must be established in the first step in our stepwise procedure. The corresponding estimators for Equation (2) is

$$\hat{\lambda}_i = \frac{\bar{X}_i - \bar{X}_0}{\bar{X}_{k+1} - \bar{X}_0} \quad (i = 1, 2, \cdots, k).$$

The test statistics

$$T_i = \frac{\bar{X}_i - \theta\bar{X}_{k+1} - (1-\theta)\bar{X}_0}{\sqrt{\frac{S_i^2}{n_i} + \frac{\theta^2 S_{k+1}^2}{n_{k+1}} + \frac{(1-\theta)^2 S_0^2}{n_0}}} \quad \text{for} \quad i = 1, 2, \cdots k \tag{4}$$

are $t-$distributed with degrees of freedom given by

$$\hat{\nu}_i = \frac{\left(\frac{S_i^2}{n_i} + \frac{\theta^2 S_{k+1}^2}{n_{k+1}} + \frac{(1-\theta)^2 S_0^2}{n_0}\right)^2}{\frac{S_i^4}{n_i^2(n_i-1)} + \frac{\theta^4 S_{k+1}^4}{n_{k+1}^2(n_{k+1}-1)} + \frac{(1-\theta)^4 S_0^4}{n_0^2(n_0-1)}} \quad \text{for} \quad i = 1, 2, \cdots, k \tag{5}$$

The confidence intervals for λ_i are derived from the random variable in Formula (4) using Fieller's (1954) method. The upper confidence limits are obtained as:

$$\theta_{i,1-\alpha} = \left(\frac{-B_i + \sqrt{(B_i)^2 - 4A_iC_i}}{2A_i}\right) \quad i = 1, 2, \cdots k$$

where

$$A_i = \left(\bar{X}_{K+1} - \bar{X}_0\right)^2 - t_{k,(1-\alpha)\nu_i}\left(\frac{S_{k+1}^2}{n_{k+1}} + \frac{S_0^2}{n_0}\right)$$

$$B_i = -2\left(\bar{X}_i - \bar{X}_0\right)\left(\bar{X}_{k+1} - \bar{X}_0\right) - t_{k,(1-\alpha)\nu_i}\frac{S_0^2}{n_0}$$

and

$$C_i = \left(\bar{X}_i - \bar{X}_0\right)^2 - t_{k,(1-\alpha)\nu_i}\left(\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}\right).$$

3. The proposed procedure

3.1. Algorithm : a stepwise procedure

We employ stepwise confidence set procedure proposed by Hsu and Berger (1999) for the case of unequal variances across dose groups for identifying maximum safe dose that is non-inferior to the vehicle control. In this sense all doses lower than the maximum safe dose are also non-inferior to the vehicle control. This is because in Hsu and Berger (1999) procedure, it is desirable not to declare safety for higher dose of the experimental drug prior to the declaration of safety at lower dose . Let $\theta_{i,1-\alpha}$ be a $100(1-\alpha)\%$ upper one-sided individual confidence bound for the ratio $\lambda_i, i = 1, 2, \cdots k$. The stepwise procedure takes the following form:

Step 0:

Is $A_i > 0$ at $\min\{\nu_i : i = 1, 2, \cdots, k\}$? If yes continue to step 1, otherwise stop and declare that the assay sensitivity of the experiment is inadequate.

Step 1:

If $\theta_{1,1-\alpha} \not\subset (-\infty, \theta)$, then conclude that $\lambda_1 \in \theta_{1,1-\alpha}$ and stop. Otherwise conclude that $\lambda_1 \in (-\infty, \theta)$ and continue.

Step 2:

If $\theta_{2,1-\alpha} \not\subset (-\infty, \theta)$, then conclude that $\lambda_2 \in \theta_{2,1-\alpha}$ and stop. Otherwise conclude that $\lambda_2 \in (-\infty, \theta)$ and continue.

$\vdots$

Step m:

If $\theta_{m,1-\alpha} \not\subset (-\infty, \theta)$, then conclude that $\lambda_m \in \theta_{m,1-\alpha}$ and stop. Otherwise conclude that $\lambda_m \in (-\infty, \theta)$ and continue.

$\vdots$

Step k:

If $\theta_{k,1-\alpha} \not\subset (-\infty, \theta)$, then conclude that $\lambda_k \in \theta_{k,1-\alpha}$ and stop. Otherwise conclude that $\lambda_k \in (-\infty, \theta)$ and continue.

step k+1:

Conclude that $\max_{i=1,2,\cdots k} \lambda_i < \max_{i=1,2,\cdots k}\left(\frac{B_i + \sqrt{(B_i)^2 - 4A_iC_i}}{2A_i}\right)$.

To understand the algorithm 3.1, we start from $i = 1$, screen lowest dose and then sequentially screen up toward $i = k$ in a step-by-step manner searching for the first integer M $(1 \leq M \leq k)$ if it exists, such that $\lambda_M < \theta$ and $\lambda_{M+1} \geq \theta$ (where the first statistical insignificant for safety effect occurred to screen the first unsafe dose). In other words, the subsequent comparison in a more higher dose levels for steps $M+1, M+2, \cdots k$ are insignificant and needless for screening.

3.2. Validity of the stepwise procedure

Definition 1. Suppose that the data X have a distribution determined by a parameter $\Lambda = \{\lambda_1, \lambda_2 \cdots \lambda_k\} \in \Theta$. A confidence set $C(X)$ for Λ is said to be directed towards a subset of the parameter space $\Theta^* \subset \Theta$, if for every sample point X, either $\Theta^* \subset C(X)$ or $C(X) \subset \Theta^*$.

To validate our procedure, let $\lambda_1, \lambda_2, \cdots \lambda_k$ be parameter of interest in Θ, (the parameter space). Before one can construct $100(1-\alpha)\%$ simultaneous confidence set, it is required that, the individual confidence intervals should have $100(1-\alpha)\%$ confidence level. For $i = 1, 2, \cdots k$, let $\Theta_i^c = (-\infty, \theta)$, and $\Theta_i = [\theta, \infty)$. Let also, $\Theta_k^c = \Theta$, $\Theta_k = \emptyset$, and $\Theta_i^* = \Theta_1^c \cap \Theta_2^c \cap \cdots \cap \Theta_{i-1}^c \cap \Theta_i$ for $i = 2, 3, \cdots, k-1, k$ with $\Theta_1^* = \Theta_1$. Then, $\Theta_1^*, \Theta_2^*, \cdots \Theta_k^*$ constitute a partition in the entire parameter space Θ. Note that, $\Theta = \Theta_1^* \cup \Theta_2^* \cup \cdots \cup \Theta_k^*$ constitute disjoint subsets, this is premised on the fact that the true parameter of interest lies in one and only one of these disjoint subsets. Each of these subset Θ_i^* is tested at a local level α. Hence the simultaneous confidence sets for $\frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0}$ are directed toward $\Theta_i^c = \{\lambda_i < \theta\}$ for $i = 1, 2, \cdots k$ Hsu and Berger (1999). This construction leads to multiple comparison procedure which guarantee the control of FWER in the strong sense. In handling the problem in Equation (1), we construct simultaneous confidence set-based procedure using intersection-union principle formulated by Berger (1982). The overall null hypothesis can be expressed as the union of the local null hypotheses H_0 against the intersection of the alternative hypotheses H_1. Thus

$$H_0 = \bigcup_{i=1}^{k} H_{0i} \quad \text{against} \quad H_1 = \bigcap_{i=1}^{k} H_{1i}$$

The rationale behind the intersection-union principle is that, if the overall null hypothesis $H_0 : \theta \in \bigcup_{i=1}^{k} \Theta_i$ can be rejected then each of the individual null hypotheses $H_{0i} : \theta \in \Theta_i$ can also be rejected. Hence Berger (1982) proved the following theorem.

Theorem 1. *If Θ_i is a level α test of H_{0i}, for $i = 1, 2, \cdots, k$, then the intersection-union test with rejection region Θ^c is a level α test of $H_0 = \bigcup_{i=1}^{k} \lambda_i \in \Theta_i$ against $H_1 = \bigcap_{i=1}^{k} \lambda_i \in \Theta_i^c$.*

Remark 1. The discernible feature about Theorem 1 is that if each of the individual test is performed at level α, then the overall test is also performed at level α. In this manner, the test does remit multiplicity adjustment for performing the multiple test at level α.Therefore, the procedure in this article is an extended intersection-union principle.

Proposition 1. *Suppose that $\theta_1, \theta_2 \cdots \theta_k$ are the $100(1-\alpha)\%$ confidence bounds for $\lambda_1, \lambda_2 \cdots \lambda_k$ respectively, with confidence level $1-\alpha$. Then for all $\lambda_1, \lambda_2 \cdots \lambda_k \in \Theta$. we have*

$$P\left(\lambda_1 < \theta_1, \lambda_2 < \theta_2 \cdots \lambda_{M-1} < \theta_{M-1}, \lambda_M < \theta_M\right) \geq 1 - \alpha.$$

The proof of Proposition 1 is similar to the proof of Theorem 3.1 of Tao *et al.* (2002). See appendix A.

Table 1. Simulated FWER study for our procedure. setting $n_R = 10, n_P = 20, \sigma_R = 15, \sigma_P = 5, \sigma_{E_i} = 10 \;\; i = 1, 2$

$n_{E_1}(n_{E_2})$	Balanced design	Unbalanced design
4 (5)	0.0216 (0.0223)	0.0256 (0.0240)
6 (7)	0.0236 (0.0235	0.0244 (0.0236)
8 (9)	0.0246 (0.0251)	0.0241 (0.0248)
10 (11)	0.0242 (0.0246)	0.0247 (0.0241)
12 (13)	0.0240 (0.0249)	0.0244 (0.0242)
14 (15)	0.0247 (0.0241)	0.0249 (0.0249)
16 (17)	0.0250 (0.0249)	0.0247 (0.0258)
18 (19)	0.0244 (0.0249)	0.0255 (0.0255)
20 (21)	0.0239 (0.0245)	0.0252 (0.0255)
22 (23)	0.0247 (0.0245)	0.0255 (0.0255)
24 (25)	0.0243 (0.0239)	0.0259 (0.0260)
26 (27)	0.0251 (0.0238)	0.0259 (0.0269)
28 (29)	0.0251 (0.0246)	0.0269 (0.0255)

Remark 2. The proposition guarantee a meaningful protection against incorrect decision because the overall coverage probability is at least $(1-\alpha)100\%$. Therefore probability of at least one type I error rate should be less than or equal to pre-specified level α. Hence the FWER is properly controlled, this is validated by the partition principle.

4. Simulation studies

We considered $k = 2$ experimental treatments for the simulation studies. Without lost of generality, we set $\alpha = 0.025$ and $\theta = 0.8$. We adapted Hasler *et al.* (2008) mean configurations $\mu_P = 16.5, \mu_R = 36.7, \mu_{E_i} = 32.66$ for $i = 1, 2$ for our simulation studies. The data set for the simulation was generated from normal distribution under the cases of heteroscedasticity with 100,000 replications. Computations were done using R software codes. In these studies, we assess the performance of our procedure based on FWER and the power evaluation.

4.1. FWER study

The assessment of the FWER was carried out for cases of balanced and unbalanced designs by using Welch method in our construction of confidence interval, because its approximates the degrees of freedom in order to ensure the control of type I error. Result from Table 1 shows that, our procedure performed well for balanced and unbalanced designs at the nominal level of $\alpha = 0.025$. However, in the case of unbalanced design, the procedure failed to control the FWER for sample sizes 4, and at least 16.

4.2. Power calculation

In most cases, confidence interval procedures for analyzing clinical trials are insufficient. Therefore, power estimation is imperative for a well-design clinical study. There are many definitions of power in multiple comparison procedures, but in our study, we will define power in the case of maximum safe dose. The maximum safe dose i is established when $\lambda_M < \theta$ and $\lambda_{M+1} \geq \theta$ where M is the first integer at which the stepwise procedure stops. Notice that M is a random variable because it depends on the sample under consideration. That is

$$P(\widehat{MSD} = i) = P\left(\bigcap_{j=1}^{i} \{T_j > t_{1-\alpha,v_i}\} \cap \{T_{i+1} \leq t_{1-\alpha,v_i}\}\right)$$

However, in our setting, power is defined as the probability of incorrectly rejecting the null hypothesis. Hence Equation (6) can be rewritten as:

$$P(\text{Reject } H_{jo} \text{ for } j = 1, 2, \cdots i) = P\left(\bigcap_{j=1}^{i} \{T_j > t_{1-\alpha,v_i}\}\right) \tag{6}$$

where the

$$\nu_i = \frac{\left(\frac{\sigma_i^2}{n_i} + \frac{\theta^2\sigma_{k+1}^2}{n_{k+1}} + \frac{(1-\theta)^2\sigma_0^2}{n_0}\right)^2}{\frac{\sigma_i^4}{n_i^2(n_i-1)} + \frac{\theta^4\sigma_{k+1}^4}{n_{k+1}^2(n_{k+1}-1)} + \frac{(1-\theta)^4\sigma_0^4}{n_0^2(n_0-1)}} \quad \text{for} \quad i = 1, 2, \cdots, k. \tag{7}$$

Notice that, the sample variances S_i^2, S_{k+1}^2 and S_0^2 in Equation (5) are approximations from $\sigma_i^2, \sigma_{k+1}^2$ and σ_0^2 in Equation (7) based on original ideas of Welch (1938) and Satterthwaite (1946) for a situation of unequal and unknown variances. Therefore Equation (6)can be calculated from a k variate non-central t-distribution with ν_i degrees of freedom. Thus non-centrality parameters for $i = 1, 2, \cdots, k$ are:

$$\Theta_i = \frac{\mu_i - \theta\mu_{k+1} - (1-\theta)\mu_0}{\sqrt{\left\{\frac{\sigma_i}{n_i} + \frac{\theta^2\sigma_{k+1}}{n_{k+1}} + \frac{(1-\theta)^2\sigma_0}{n_0}\right\}}}.$$

Hence exact power calculation is impossible in this settings: this implies the power calculation must be approximated. The result of our power estimation are tabulated in Table 2. In this table, it can be observed that in general, power increases with increasing in both the ratio of mean differences, λ and the sample sizes. This is consistent with earlier findings Hasler *et al.* (2008).

5. Practical Application

We apply our stepwise confidence set procedure to a data set published by Adler and Kliesch (1990) for evaluation of mutagenic experiment. The data was used for micronucleus assay on hydroquinone using a positive control of 25mg/kg cyclphosphamide. The goal is to show that whether or not the investigational substance is

Table 2. Power Estimation of the confidence intervals for $\sigma_R = 13.2, \sigma_P = 7.5, \sigma_{E_i} = 10.4 \;\; i = 1, 2$

Ratio(λ_i)	$n_{E_i=1,2}$	n_R	n_P	Power
0.70	20	20	20	0.0054
0.70	30	30	30	0.0035
0.70	40	40	40	0.0024
0.70	60	60	60	0.0013
0.80	20	20	20	0.0250
0.80	30	30	30	0.02500
0.80	40	40	40	0.0250
0.80	60	60	60	0.0250
0.90	20	20	20	0.0560
0.90	30	30	30	0.1094
0.90	40	40	40	0.1329
0.90	60	60	60	0.1793
1.00	20	20	20	0.2160
1.00	30	30	30	0.3090
1.00	40	40	40	0.3953
1.00	60	60	60	0.5493
1.10	20	20	20	0.4262
1.10	30	30	30	0.5917
1.10	40	40	40	0.7196
1.10	60	60	60	0.8781
1.20	20	20	20	0.6570
1.20	30	30	30	0.8320
1.20	40	40	40	0.9234
1.20	60	60	60	0.9864
1.30	20	20	20	0.8404
1.30	30	30	30	0.9547
1.30	40	40	40	0.9885
1.30	60	60	60	0.9994

able to induce chromosome damage or interact with the mitotic spindle apparatus. Consequently, the result of male mice at 21h sampling time is given in Table 3. Counts of micronuclei in polychromatic erythrocytes after 24 h are taken as a measure for the potency to induce chromosome damage. A scatter plot in Figure 4 for the data set in Table 2 in Gamalo *et al.* (2013) depicts heterogeneity of variances among the reference treatment and the placebo groups. For our analysis in Table 4, we set and without loss of generality $\alpha = 0.05$ and $\theta = 0.5$, where θ is the safety threshold. The following results were obtained.

$\hat{\theta_1} = 0.24 < \theta = 0.5$ we reject H_{01}
$\hat{\theta_2} = 0.35 < \theta = 0.5$ we reject H_{02}
$\hat{\theta_3} = 0.74 \not< \theta = 0.5$ we do not reject H_{03}
$\hat{\theta_4} = 1.04 \not< \theta = 0.5$ we do not reject H_{04}
the procedure then stops at step 3, implies it is needless to step it down further.

Table 3. Number of micronuclei per animal and 2000 scored cells for the negative control, four doses of hydroquinone and positive control cyclophosphamide

Treatment group	Number of micronuclie/per 2000 cells
Negative control	1, 2, 2, 2, 3, 5
30mg/kg	2, 4, 4, 4, 5
50mg/kg	4, 6, 6, 7, 8
75mg/kg	9, 12, 13, 13, 18, 18
100mg/kg	13, 20, 22, 22, 23
Positive control	15, 20, 32, 33

Table 4. 95% upper bound for λ_i, for $i = 1, 2, 3, 4$

Treatment groups	Comparisons	Upper bound
30mg/kg	λ_1	0.24
50mg/kg	λ_2	0.35
75mg/kg	λ_3	0.75
100mg/kg	λ_4	1.04

In this analysis, our stepwise procedure concluded that doses 75mg/kg and 100mg/kg are unsafe at level α, while safety is concluded on the two lower doses 30mg/kg and 50mg/kg. Since $\hat{\theta}_3 = 0.74 \not< \theta = 0.5$ is the last step at which the procedure stop, that is $M = 3$ 50mg/kg is recommended as the maximum safe dose (MSD), which is the highest dose that is non-inferior to the reference drug such that any lower dose is also non-inferior at level α. Notice that 30mg/kg is also non-inferior to the reference drug but lower.

6. Conclusions

A typical three-arm clinical trial involves positive control, negative control and exactly one experimental treatment. But certain therapeutic situations necessitate more than one experimental treatment. For example, cases that involve different combinations of several new drugs or different doses of a new drug require multiple experimental treatments. This calls for multiple comparisons procedure with its concomitant multiplicity effect. Adjustment of multiplicity for the maintenance of the FWER is unneeded if the new treatment can be ordered a priori according to their treatment effect. In this article, we extended the three-arm trial to a case of $(k + 2)$-arm trial (where $k \geq 2$) for identification of maximum safe dose. Where the maximum safe dose in this context is defined as the highest dose that is non-inferior to the the reference dose such that any other lower dose is also non-inferior to the reference dose.

We employed the partitioning method in constructing confidence set-based procedure under unknown unequal variances across dose groups. The procedure properly controlled the FWER in the sense at the nominal level α. This was validated by the partitioning principle. Results indicated that the power of the procedure increases with increasing in sample size and the ratio of mean differences.

Appendix A: Proof of Proposition 3.1

Let M be the step at which the proposed procedure stop. Where M is a random variable that assume the values $(0 \leq M \leq k)$. There are two cases to consider in the proof of Proposition 3.1.

Case 1. Let $M = 0$, this means $A_i \leq 0$ at min$\{\nu_i : i = 1, 2, \cdots, k\}$. This implies the sensitivity of the experiment is not adequate to proceed further and the procedure is terminated.

Case 2. Let $1 \leq M \leq k$ and for j = $1, \cdots k$, let
(i)$C_j(X) = \{\lambda_{k-j+1} < \theta_{k-j+1}\}$; and
(ii) $\Theta_1 = \{\lambda_k \geq \theta\}$ and $\Theta_j = \bigcap_{l=1}^{j-1} \{\lambda_{k-l+1} < \theta\} \cap \{\lambda_{k-j+1} \geq \theta\}$

for $j = 2, \cdots, k$. Then, the parameter space Θ is partitioned by $\Theta_j, j = 1, \cdots k+1$. Moreover,

$$\bigcup_{j=1}^{k} (C_j(X) \cap \Theta_j)$$

provides a $100(1-\alpha)$ confidence set for $\lambda_1 \cdots \lambda_k$ because if $(\lambda_1 \cdots \lambda_k) \in \Theta$ then

$$P_{\lambda_1, \cdots, \lambda_k} \left\{ (\lambda_1, \cdots, \lambda_k) \in \bigcup_{j=1}^{k} (C_j(X) \cap \Theta_j \right\} = P_{\lambda_1, \cdots, \lambda_k} \{(\lambda_1 \cdots \lambda_K) \in C_j(X)\} \geq 1 - \alpha.$$

In this setup, the unionized confidence set can be decomposed as follows:

$$\begin{aligned}
\bigcup_{j=1}^{k}(C_j(X)\cap\Theta_j) = & \quad \{\bigcup_{j=1}^{M-1}(C_j(X)\cap\Theta_j)\}\bigcup\{\bigcup_{j=M}^{k}(C_j(X)\cap\Theta_j)\} \\
= & \quad \bigcup_{j=M}^{k}(C_j(X)\cap\Theta_j) \\
& \quad \subset (C_M(X)\cap\Theta_M)\cup\left(\bigcap_{j=1}^{M}\{\lambda_{k-j+1}<\theta\}\right) \\
= & \quad \left(\bigcap_{j=1}^{M-1}\{\lambda_{k-j+1}<\theta\}\cap\{\lambda_{k-M+1}\geq\theta\}\cap C_M(X)\right) \\
& \quad \bigcup\left(\bigcap_{j=1}^{M}\{\lambda_{k-j+1}<\theta\}\right) \\
= & \quad \left(\bigcap_{j=1}^{M-1}\{\lambda_{k-j+1}<\theta\}\cap\{\lambda_{k-M+1}\geq\theta\}\cap C_M(X)\right) \\
& \quad \bigcup\left(\bigcap_{j=1}^{M}\{\lambda_{k-j+1}<\theta\}\cap C_M(X)\right) \\
= & \quad \bigcap_{j=1}^{M-1}\{\lambda_{k-j+1}<\theta\}\cap C_M(X)
\end{aligned}$$

Finally, we have

$$\begin{aligned}
& P_{\lambda_1,\cdots\lambda_k}\left((\lambda_1,\cdots,\lambda_k)\in\bigcap_{j=1}^{M-1}\{\lambda_{k-j+1}<\theta\}\cap C_M(X)\right) \\
& = P_{\lambda_1,\cdots,\lambda_k}\left\{(\lambda_1,\cdots,\lambda_k)\in\bigcup_{j=1}^{k}(C_j(X)\cap\Theta_j\right\}\geq 1-\alpha
\end{aligned}$$

This completes the proof of Proposition 1. □

References

Adjabui M. J., Howard, N., and Luguterah, A. (2016). Nonparametric stepwise procedure for identification of minimum effective dose(MED) *International Journal of Statistics and System* 11(1); 77-88.

Adler, I. D., and Kliesch, U. (1990). Comparison of single and multiple treatments in the mouse bone marrow micronucleus assay for hydroquinone and cyclophosphamide *mutation Research*, 234, 115-123.

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometric*,24, 295-300.

Bonfiger, E. (1987). Stepdown procedure for comparison with control *Aust. J Statistics* 29:348-364.

Cao, L., ,Tao, J., Shi N. Liu W. (2015). A stepwise confidence interval procedure under unknown variance based on an asymmetric loss function for toxicological evaluation *Aust. N. Z. J. Stat.* 57(1), 73-98.

Finner, H.,Strassburger, K. (2002). The partitioning principle : a powerful tool in multiple decision theory. *Annals of statistics,* pp.1194-1213.

Gamalo, M. A., Muthukumanara, S.,Ghosh, P. and Tiwari R. C., (2013). Agenerilized P-value approach for assessing noninferiority in three-arm trial. *Statistical Method in Med Res* 22(3):261-277.

Hasler M., Vonk R. and Hothorn L. A. (2008). Assessing non-inferiority of a new treatment in a three-arm clinical trial including placebo. *Statistics in Medicine* 27: 490-503.

Hasler M. (2012). Multiple comparisons to both negative and positive control *Pharmaceutical Statistics* 11:74-81.

Hauschke, D., Slacl-Erben,R., Hansen., S.,Kaufmann R.(2005). Biotatistical assessment of mutagenicity studies by including the positive control *Biometrical Journal*47 1, 82-87.

Hsu , J . C , and Berger , R . L . (1999). (Stepwise confidence interval without multiplicity adjustment for dose response and toxicity studies. *Journal of America Statistics Association* 94: 468 -482

Huang L. C., Wen M, J.,Cheung S. H. (2014). Non-inferiority studies with multiple new treatment and Heterogenenous Variances. *Journal Biophamaceutical Statistics* 25:958-971.

Huang L. C., Wen M, J.,Cheung S. H (2015). Non-inferiority studies with multiple reference treatment *Statistical methods in medical Research* 26:2095-1307.

Kwong, S K., Cheung S. H., Hayter A. J. Wen M. J,. (2014). Extension of three-arm non-inferiority studies to trials with multiple new treatments *Statistics in medicine* 13:2833-2843.

Pigeot, I. Shaffer, J,. Rohmel J., and Hauschke D. (2003). Assessing non-inferiority of new treatment in a three-arm clinical trial including a placebo *Statist Med,* Vol.22, pp. 883-889.

Satterthwaite, F. E. (1946) An approximate distribution of estimates variances component *Biometrics* 2: 110:114.

Stefansson, G., Kim, W. and Hsu, J. C.(1988). *On confidence set in multiple comparisons.* In statistics decision theory and Related topic, IV(2) eds, S. S. Gupta and J. O.Berger pp 89-104. New York: Springer Verlag

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variances *Ann. Math. Statist.* 16, 243-258.

Tao, J., Shi, N., Guo, J. Gao W., (2002). Stepwise Procedure for identification of minimum effective dose with Unknowm variances *Statistics and Probability Letters* 57: 121-131.

Welch, B. L. (1938). The significance of difference between two means when the population variances unequal *Biometrika* 29: 350-362