Afrika Statistika

Afrika Statistika
Vol. 16 (3), 2021, pages 2911 - 2922.
DOI: http://dx.doi.org/10.16929/as/2021.2911.189

ISSN 2316-090X

A Stepwise Procedure for Toxicity Studies Based on Ratio of two means

Michael Jackson Adjabui [1,*] **John Ayuekanbey Awaab** [2] **and Jakperik Dioggban** [3]

[1] Department of Mathematics, C.K. Tedam University of Technology and Applied Sciences. Navrongo Ghana
[2] Department of Statistics, Bolgatanga Technical University, Bolgatanga, Ghana
[3] Department of Statistics, C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

Received on September 20, 2021; Accepted on October 27, 2021

Abstract. This paper proposes a stepwise confidence set procedure for identifying equivalence or safety of compounds in a toxicity study under heteroscedasticity of variances for a normally distributed data. The problem of statistical methodology for drug safety is the control of the familywise error rate (*FWER*). Hence, we construct a confidence set procedure for toxicological evaluation and incorporating the partitioning principle with a case of heteroscedascity of variances under normal assumption. Our simulation studies demonstrated that the power of the procedures for heterogeneity of variances increases with increasing in ratio of means.

Key words: equivalence; Fiellers confidence intervals; multiple comparisons; multiple Ratio; stepwise procedure.
AMS 2010 Mathematics Subject Classification Objects : 62G32; 62G30; 62N02.

[*] Michael Jackson Adjabui: madjabui@cktutas.edu.gh
John Ayuekanbey Awaab : awaabjohn@gmail.com
Jakperik Dioggban : jdioggban@cktutas.edu.gh

Résumé. (Abstract in French) Cet article propose une méthode de mise au point d'ensemble de confiance par étapes pour identifier l'équivalence ou l'innocuité des elements constitutifs dans une étude de toxicité sous hétéroscédasticité des variances, pour une données normales. Le problème de la méthodologie statistique pour l'innocuité des médicaments est le contrôle du taux d'erreur par groupe [familial] (*FWER*). Par conséquent, nous construisons une procédure d'ensemble de confiance pour l'évaluation toxicologique et en incorporant le principe de partitionnement avec un cas d'hétéroscédascité des variances dans une hypothèse normale. Nos études de simulation ont démontré que la puissance des procédures d'hétérogénéité des variances augmente avec l'augmentation du rapport des moyennes.

The authors.

Michael Jackson Adjabui, Ph.D., is a Senior Lecturer of Department of Mathematics at C. K. Tedam University of Technology and Applied Sciences. Navrongo Ghana.

John Ayuekanbey Awaab: MPhil in Statistics, is preparing a Ph.D, thesis under the supervision of the first author at CK Tedam University of Technology and Applied Sciences. Navrongo Ghana.

Jakperik Dioggban, Ph.D., is a Senior Lecturer of Department of Statistics at C. K. Tedam University of Technology and Applied Sciences. Navrongo Ghana.

1. Introduction

The central issue in toxicological evaluation is the proof of identity of equivalence between k experimental doses $(k \geq 2)$ and a standard dose or a placebo. Comparing k experimental treatments with a placebo involves multiple comparisons procedure, specifically multiple comparisons with control proposed by Dunnett (1955). This raises the problem of multiplicity adjustment, for which when care is not taken will inflate the *FWER* type I error . In a situation where the experimental group can be ordered a priori according to their treatment effect, then there is no need for multiplicity adjustment. For this reason, we employ stepwise method as in Bofinger (1987), Stefanson (1988), Hsu and Berger (1999). Statistical procedures for determination of equivalence of a drug agent have been proposed in literature; see, for example, Hauschke *et al.* (1999), Hsu and Berger (1999), Tao *el al.* (200), and Adjabui (2020) among others.

Stepwise confidence set procedure without multiplicity adjustment was proposed by Hsu and Berger (1999) in a dose-response and toxicity studies. Also, Hauschke *et al.* (1999) made equivalence/safety assessment for sample size determination based on ratio of two means for normally distributed data. Unfortunately,

their procedures necessitates homogeneity of variances across dose groups. However, variation of responses under different dose levels is usually different with a change of dose level because patients in different dose groups turn to response differently for some biological factors or toxicity effect at various dose levels. Hence, statistical procedures that assumed homogeneity of variances may lead to liberal or incorrect decision. For this reason, equal variance assumptions are seldom satisfy in practice. But the issue of heteroscedasticity of variances has been a long standing problem in multiple comparison procedures since the initial work of Welch (1938). This situation is far from being resolved. Therefore, we propose a stepwise confidence set procedure without multiplicity adjustment and incorporating the partitioning principle proposed by Finner and Strassburger (2002) for identifying equivalence/safety of a drug agent under the ratio of two means for a normally distributed data. In other words, we extend the procedure proposed by Hsu and Berger (1999) to a situation of unequal variances across dose groups by using the partitioning principle in a step-by-step fashion.

The outline of this article is as follows. In Section 2, the concept of the partitioning principle is discussed and the problem is formulated, the testing and the Fieller's confidence interval procedure for ratio of two means for normally distributed data is derived. A stepwise confidence set procedure is proposed and results discussed in Section 3. We conducted simulation studies to investigate the performance the *FWER* and the power of our procedure in Section 3.3. An illustrative example for clinical trial in assessing toxicity Section 3.2 . Inevitably our conclusion is in Section 4.

2. Preliminaries

2.1. Partitioning Principles

The validity of statistical procedure in clinical trials is the strong control of the *FWER*, especially in a case where the false discovery rate is not valid. For this reason, we employ the partitioning principle proposed by Stefanson (1988) and Finner and Strassburger (2002). To throw more light on the concept of the partitioning principle, we consider testing the following null hypotheses as an example of multiple testing problem in a clinical study.

$$H_{0i} : \lambda_i \leq \delta \text{ for } i = 1, \cdots, k.$$

where $\delta > 0$ is pre-specified clinical margin determine by medical expert. Using the partitioning principle, we have 2^k parameter subspaces and with $2^k - 1$ null hypotheses to be tested. The $H_{0i} : \lambda_i \leq \delta$ hypotheses is decomposed into $2^k - 1$ union of disjoint subspaces and each of these subsets is tested at level α. Only one of these contain the true parameter of interest. In this setting, multiplicity adjustment is needless. For example, letting $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ for $k = 3$, we partition it into eight disjoint parameter subsets :

$$\begin{aligned}
\Lambda_1 &= \{\lambda_1 \leq \delta \text{ and } \lambda_2 \leq \delta \text{ and } \lambda_3 \leq \delta\} \\
\Lambda_2 &= \{\lambda_1 \leq \delta \text{ and } \lambda_2 \leq \delta \text{ and } \lambda_3 > \delta\} \\
\Lambda_3 &= \{\lambda_1 \leq \delta \text{ and } \lambda_2 \leq \delta \text{ and } \lambda_3 < \delta\} \\
\Lambda_4 &= \{\lambda_1 > \delta \text{ and } \lambda_2 \leq \delta \text{ and } \lambda_3 \leq \delta\} \\
\Lambda_5 &= \{\lambda_1 \leq \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 > \delta\} \\
\Lambda_6 &= \{\lambda_1 > \delta \text{ and } \lambda_2 \leq \delta \text{ and} \lambda_3 > \delta\} \\
\Lambda_7 &= \{\lambda_1 > \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 \leq \delta\} \\
\Lambda_8 &= \{\lambda_1 > \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 > \delta\}.
\end{aligned}$$

and then testing each of these null hypotheses at the nominal level α as:

$$\begin{aligned}
H_{0\{123\}}&: \lambda_1 \leq \delta \text{ and } \lambda_2 \leq \delta \text{ and } \lambda_3 \leq \delta \\
H_{0\{12\}}&: \lambda_1 \leq \delta \text{ and} \lambda_2 \leq \delta \text{ and } \lambda_3 > \delta \\
H_{0\{13\}}&: \lambda_1 \leq \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 \leq \delta \\
H_{0\{23\}}&: \lambda_1 > \delta \text{ and } \lambda_2 \leq \delta \text{ and } \lambda_3 \leq \delta \\
H_{0\{1\}}&: \lambda_1 \leq \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 > \delta \\
H_{0\{2\}}&: \lambda_1 > \delta \text{ and } \lambda_2 \leq \delta \text{ and} \lambda_3 > \delta \\
H_{0\{3\}}&: \lambda_1 > \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 \leq \delta \\
H_{0\{\emptyset\}}&: \lambda_1 > \delta \text{ and } \lambda_2 > \delta \text{ and } \lambda_3 > \delta
\end{aligned}$$

These decompositions will guarantee strong control of *FWER* in strong sense. Hence, we extend the concept of partition principle to a confidence interval procedure.

2.2. Testing Procedure

Let X_i and X_0 denote mutually independent normally distributed variables of interest for new treatments and placebo effect, that is $X_i \sim N(\mu_i, \sigma_i^2)$ and $X_0 \sim N(\mu_0, \sigma_0^2)$ with $X_{0j} \sim N(\mu_o, \sigma_0^2)$, for $i = 1, 2, \ldots, k$ and $j = 1, 2, \cdots, n_0$ respectively, where μ_i and μ_o are population means for treatment and placebo groups with unknown but unequal variance, that is σ_o^2, and $\sigma_i^2 \neq \sigma_j^2$ for variances of placebo and for any two treatments groups i, j respectively. In this case, no pooled variance estimator is needed. For equivalence testing it is reasonable to assume that the signs of the corresponding population means μ_i and μ_o are both positive. Let the equivalence interval (δ_1, δ_2) denote the pre-specified equivalence range, so the corresponding test problem is formulated as follows:

$$H_{0i} : \frac{\mu_i}{\mu_o} \leq \delta_1 \text{ or } \frac{\mu_i}{\mu_o} \geq \delta_2 \;\; \textit{vs} \;\; H_{1i} : \delta_1 < \frac{\mu_i}{\mu_o} < \delta_2 \text{ for } i = 1, 2, \ldots, k, \tag{1}$$

where $\delta_1 < 0$ and $\delta_2 > 0$ are pre-specified quantities. In practice, δ_2 is usually chosen to be $-\delta_1$. Let $\lambda_i = \frac{\mu_i}{\mu_o}$, the ratio of means, then the hypothesis in Equation (1) can be rewritten as:

$$H_{i1} : \lambda_i \leq \delta_1 \text{ or } \lambda_i \geq \delta_2 \text{ versus } H_{1i} : \delta_1 < \lambda_i < \delta_2 \text{ for } i = 1, 2, \ldots, k. \tag{2}$$

The problem in Equation (2) is union-intersection test formulated by Berger (1982) because the global null hypothesis can be expressed as the intersection of the local null hypothesis. That is

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

If H_{0i} is rejected, then $1, 2, \ldots, i-1$ hypothesis is also rejected as well in a stepwise manner.

For all $i \in \{1, \cdots, n_i\}$, we set

$$\overline{X}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} X_i$$

and

$$\overline{X}_0 = \frac{1}{n_0} \sum_{j=1}^{n_j} X_{0j},$$

with $j = 1, \cdots, n_0$. The random variables

$$T_i = \frac{\bar{X}_i - \delta_i \bar{X}_0}{\frac{S_i^2}{ni} + \delta_i^2 \frac{S_0^2}{n_0}} \tag{3}$$

for $i = 1, 2, \ldots, k$ are the test statistics for the testing problem in Equation 2, where $\delta_i = \delta_1$ or δ_2 and T_i has t- distribution with

$$\nu_i = \frac{\left(\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}\right)^2}{\left(\frac{S_i^4}{n_i^2(n_i-1)}\right) + \left(\frac{S_0^4}{n_0^2(n_0-1)}\right)}$$

degrees of freedom. Equivalence/safety in Equation (2) is considered if $T_i > t_{i-\alpha,\nu_i}$ for $i = 1, 2, \ldots, k$ where $t_{i-\alpha,\nu_i}$, is $(1-\alpha)100\%$ percentile of the central t-distribution with ν_i degrees of freedom. The problem of this nature was first identified by Welch (1938) and subsequently by Satterthwaite (1946).

2.3. Fiellers confidence intervals

Dilba *et al.* (2006) constructed simultaneous confidence for multiple ratio by employing Fieller's (1954) generalized confidence intervals method. These confidence intervals can be estimated as follows: the generalized k-quadratic equations are $A_iX^2 + B_iX + C_i = 0.$ and the corresponding confidence limits are stated as:

$$\begin{aligned}\delta_{low,i} &= \frac{-B_i - \sqrt{(B_i)^2 - 4A_iC_i}}{2A_i}(i = 1, 2, \ldots, k)\\ \delta_{upp,i} &= \frac{-B_i + \sqrt{(B_i)^2 - 4A_iC_i}}{2A_i}(i = 1, 2, \ldots, k),\end{aligned}$$

where $\delta_{low,i}$ and $\delta_{upp,i}$ are the lower and the upper confidence limits respectively, with

$$\begin{aligned}A_i &= (\bar{X}_0)^2 - t^2_{k,k-\alpha(\nu_i)}\left(\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}\right)(i = 1, 2, \ldots, k)\\ B_i &= -2\bar{X}_i\bar{X}_0\\ C_i &= (\bar{X}_i)^2 - t^2_{k,1-\alpha(\nu_i)}\left(\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}\right)(i = 1, 2, \ldots, k)\end{aligned}$$

and the $(1 - \alpha)$ quantile $t^2_{k,k-\alpha(\nu_i)}$ of k-variate t-distribution with ν_i degrees of freedom. The above confidence intervals are valid if and only if $A_i > 0$. This would occur if μ_o is significantly greater than zero.

In order to generalize Hsu and Berger (1999) stepwise confidence method to the case of unknown unequal variances, we rewrite some results.

Definition 1. Let 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 Θ 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 $C(X) \subset \Theta^*$.

Equation (2) of our problem can be rewritten as:

$$H_{i0} : \lambda_i \in \Theta_i \ \ vs. \ \ H_{i1} : \lambda_i \in \Theta_i^c \ \text{ for } \ i = 1, ..., k,$$

where $\Theta_i^c = \{\delta_1 < \lambda_i < \delta_2\}$ and $\Theta_i = \{\lambda_i \leq \delta_1 \ \text{ or } \ \lambda_i \geq \delta_2\}$. From the above definition, the confidence set $C(X)$ is directed towards Θ_i^c. Notice that $\Theta_i^c \subseteq \Theta_i^*$. For a given sample point X. the confidence set $C_i(X)$ contains the alternative space $\Theta^* = (\delta_1 < \lambda_i < \delta_2)$ or the confidence set is contained in the alternative space $(\delta_1 \leq \lambda_i \ \text{ or } \ \lambda_i \geq \delta_2)$. For the ith dose level $(i = 1, 2, \ldots, k)$, let

$$D_i^-(X) = \min\left\{\frac{-B_i - \sqrt{(B_i)^2 - 4A_iC_i}}{2A_i}, 0\right\}$$

and

$$D_i^+(X) = \max\left\{\frac{-B_i + \sqrt{(B_i)^2 - 4A_iC_i}}{2A_i}, 0\right\}$$

Then

$$D_i(X) = \begin{cases} D_i^-(X).D_i^+(X), & \text{if } D_i^-(X) < 0 < D_i^+(X), \\ [0, D_i^+(X)), & \text{if } D_i^-(X) = 0 \\ (D_i^-(X), 0), & \text{if } D_i^+(X) = 0 \end{cases}$$

is a $100(1-\alpha)\%$ confidence interval for λ_i.

Moreover, let

$$C_i(X) = \begin{cases} D_i(X), & \text{if } D_i(X) \subset (\delta_1, \delta_2) \\ D_i(X) \cup (\delta_1, \delta_2), & \text{otherwise.} \end{cases}$$

Then $C_i(X)$ is a $100(1-\alpha)\%$ confidence interval for λ_i directed towards the range (δ_1, δ_2).

In equivalence studies, one would like to drive a method that does not declare the equivalent/safety of the new drug at higher dosages prior to the declaration of the equivalent/safety at lower dosages. This can be achieved by answering the question is $\delta_1 < \lambda_i. < \delta_2$ in stepwise fashion (see details in Hsu and Berger (1999); Tao *el al.* (200) continuing only while the answer is in the affirmative.

3. Results

3.1. The propose stepwise procedure

We establish practical equivalent via Hsu and Berger (1999) stepwise confidence set procedure as follows:

Let $D_i(X)$ be confidence intervals for λ_i for $i = 1, \cdots, k$, where k is the total number of treatments doses to be tested. In step one, we assess the adequacy of the

procedure by proving that A_i for $i = 1, \cdots, k$ is significantly greater than zero. Otherwise, the procedure is stopped, and we declare that the experiment is inadequate. In step two, we scan for toxicological equivalence/safety dose by scanning the lowest dose at $D_1(X)$ for the first equivalence drug if it exists and sequentially scan the subsequent doses for $i = 2, 3, \cdots, k$ without adjusting the α levels in each of the steps in ascending fashion searching for the first integer M $(1 \leq M \leq k)$, if it exists such that $D_M(X) \subset (\delta_1, \delta_2)$ and $D_{M+1}(X) \not\subset (\delta_1, \delta_2)$ (this scans the first non-equivalence or unsafe dose). In this set up, doses at $D_1(X), D_2(X), \cdots, D_M(X)$ are established as equivalence while doses at $D_{M+1}(X), D_{M+2}(X), \cdots, D_k(X)$ are non-equivalence. Notice that the confidence intervals at each step are computed without multiplicity adjustments. Admittedly, the values of δ_1 and δ_2 must be predetermined by clinical experts. Nevertheless, the magnitude of δ_1 and δ_2 may affect the choice of the first integer M.

To throw more light on the above procedure, let $M(1 \leq M \leq k)$ be the step at which the procedure is stopped . If $M = 1$, then the sensitivity of the experiment is inadequate, and the lower confidence bound for λ_k is given. If $1 < M < k$, then a confidence set for λ_M that contains (δ_1, δ_2) is given, and the confidence intervals $\lambda_i \in (\delta_1, \delta_2)$ for $i = 2, \cdots, M-1$ are given if $M > 2$. If $M = k$, then a confidence interval for all $\lambda_i . i = 1, \cdots, k$ which are entirely within the range (δ_1, δ_2) is given. Hence,we state and prove the following theorem based on the above procedure.

Theorem 1. *Let X represent a sample data point and let Θ be the parameter space for parameter vector $\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_k\}$. For any $i = 1, \cdots, k$. let $D_i(X)$ be any $100(1-\alpha)\%$ confidence interval for λ_i, also let $C_i(X)$ be confidence set directed towards $\delta_1 < \lambda_i < \delta_2$. Denote M the largest integer of i such that $D_M(X) \subset (\delta_1, \delta_2)$ if such an $i(1 \leq i \leq k)$ exits; otherwise let $M = 0$. Then for any $\Lambda \in \Theta$, for*

$$\Sigma = \Theta_1^c \bigcap D_2(X) \subset \Theta_2^c, \cdots, D_M(X) \subset \Theta_M^c \bigcap D_{M+1}(X) \subset \Theta_{M+1} \bigcap C_{M+1}(X),$$

we have

$$\mathbb{P}(D_1(X) \subset \Sigma) \geq 1 - \alpha.$$

Proof. Let step $M(1 \leq M \leq K)$ be the step at which the stepwise procedure stops. If $M = 1$, then $A_i \leq 0$. In this situation, the sensitivity of the experiment is inadequate and the lower confidence bound for λ_i is given. If for each $D_i(X) \subset (\delta_1, \delta_2)$, there is a $100(1-\alpha)\%$ confidence interval for λ_i for $M > 1$, then $C_i(X)$ is a $100(1-\alpha)$ confidence intervals for λ_i that is directed towards $\Theta^* = (\delta_1 < \lambda_i < \delta_2)$ for $i = 2, \cdots, k$. The rest of the proof follows Theorem 1 of Hsu and Berger (1999). □

Remark. Therefore, it can be inferred from Theorem 1 that, *FWER* is properly controlled in strong sense. That is to say, all declaration can be guaranteed to be correct with a probability higher than $100(1-\alpha)$. ⋄

3.2. An illustrative example

A 90-day routine rat study was conducted to evaluate toxicity of a crop protection compound. Test substance was added directly to the rodent diet and was thoroughly mixed to ensure homogeneous distribution. Three doses of the compound with a zero dose control. The sample sizes in the four groups were $n_0 = 18, n_1 = 20, n_2 = 19$ and $n_3 = 18$. The variable of interest was the kidney weight to the body weight ratio. This data set was published by Tamhane and Logan (2004)

	Dose			
	0	1	2	3
	6.593	7.062	7.006	9.569
	7.480	7.347	8.706	9.362
	6.930	7.733	7.257	10.911
	5.622	7.369	7.743	9.961
	6.789	8.173	7.026	9.497
	7.268	6.938	8.561	9.911
	6.647	6.988	7.674	8.544
	6.443	6.621	7.450	10.404
	6.713	7.508	8.188	10.421
	6.057	6.657	8.150	10065
	6.253	7.787	7.619	9.670
	7.045	6.537	8.723	8.194
	6.552	7.369	7.387	8.989
	5.668	6.623	6.798	7.347
	6.354	6.456	7.617	7.260
	6.511	6.507	8.071	9.017
	7.111	6.154	7.020	8.847
		6.909	7.063	
		7.252		
n_j	18	20	19	18
Mean	6.5606	6.9975	7.6778	9.2606
SD	0.5094	0.5755	0.5949	1.0052

Table 1. Kidney Wt/body Wt $\times 10^3$

Hence, setting $\delta_1 = -42$, and $\delta_2 = 42$, the following are established:

$D_1\ (X)$ =(-0.03, 38,55) $\subset (-42\ , 42)$. Equivalent has been demonstrated.

D_2(X)= (-0.03, 41.75) $\subset (-42\ , 42)$. Equivalent has been demonstrated.

D_3(X)= (-0.09, 23.19) $\subset (-42,\ 42)$. Equivalent has been demonstrated.

3.3. Simulation Studies

3.3.1. *FWER* study

Simulation studies were conducted to investigate the performance of the *FWER* using **R** software program. Without loss generality, we set $-\delta_{(1)} = \delta_{(2)} = 0.8$ and $\alpha = 0.025$. In this study, observations were generated with 10000 replications from a normal distribution based on the assumption of unknown and unequal variances. We set increasing in both the number of observations in the treatments (n_i) and that of placebo (n_0) as $(n_i, n_0) = (4, 5), \cdots, (28, 29)$. The heterogeneous case is simulated with $\mu_i = 24, \mu_0 = 30, \sigma_i = 4, \sigma_0 = 8$ and compare with that of the homogeneous $\mu_i = 24, \mu_0 = 30, \sigma_i = 8, \sigma_0 = 8$ when the assumption of heterogeneity is violated. In Table 2, the *FWER* is generally controlled at the nominal value $\alpha = 0.025$ in the situation of heteroscedasticity except when the treatment sample sizes $n_i = 16, 17$ and placebo 27. In the case of homogeneity the *FWER* was poorly controlled because the simulated values exceed the nominal value of $\alpha = 0.025$.

(n_i, n_o)	Heteroscedastic procedure	Homoscedastic procedure
4(5)	0.0162 (0.0207)	0.0354(0.0307)
6(7)	0.0198 (0.0232)	0.0351 (0.0295)
8(9)	0.0217 (0.0245)	0.0309 (0.0321)
10(11)	0.0210 (0.0246)	0.0301 (0.0286)
12(13)	0.0247 (0.0233)	0.0304 (0.0290)
14(15)	0.0219 (0.0230)	0.0313(0.0283)
16(17)	0.0255(0.0249)	0.0323 (0.0281)
18(19)	0.0255 (0.0223)	0.0295(0.0264)
20(21)	0.0237 (0.0247)	0.0273(0.0284)
22(23)	0.0239 (0.0237)	0.03021 (0.0294)
24(25)	0.0241 (0.0243)	0.0293 (0.0283)
26(27)	0.0224 (0.0252)	0.0308 (0.0271)
28(29)	0.0237 (0.0249)	0.0299 (0.0318)

Table 2. The simulation of *FWER* study

3.3.2. Power of estimation of the procedure

In multiple comparisons procedures, there are many definition of power concept proposed in literature. In this study, we use the power concept introduced by Ramsey for all-pairwise comparisons procedure, that is, probability rejecting incorrect null hypotheses. The simulations results in Table 3 revealed that, the power of our procedure increases with increasing values μ_i/μ_0. This result is consistent with Hasler *et al.* (2008).

$\frac{\mu_i}{\mu_o}$	**Power**
0.800	0.0250
0.825	0.0573
0.850	0.1158
0.875	0.2078
0.900	0.3328
0.925	0.4800
0.95	0.6299
0.975	0.7622
1.000	0.8632
1.025	0.9301
1.050	0.9680
1.075	0.9874
1.100	0.9956
1.125	0.9986
1.150	0.9996
1.175	0.9999
1.200	0.9999
1.225	0.9999
1.250	0.9999
1.275	1.0000
1.300	1.0000

Table 3. Shows power of the heteroscedastic procedure.

4. Conclusion

In practical toxicological study, it is risky and undesirable to mistakenly proclaim a non-equivalent dose (unsafe) to be equivalent (safe). Therefore, a good statistical procedure should strictly control *FWER* in a strong sense. For this reason, we propose stepwise confidence set-based procedure and incorporated the partitioning principle when the treatment variances are heterogeneous across dosages. The strong control of the *FWER* was validated by the partitioning principle. Our simulation results confirm our theoretical findings of *FWER* and the power of our procedure increases with increasing with the ratio μ_i/μ_0 values under the heterogeneous settings.

Acknowledgments. We thank the chief editor and anonymous reviewers for their useful comments which improved the quality of the paper.

References

Adjabui, M. J. (2020). A note on stepwise confidence procedure for toxicological evaluation under unknown and unequal variances. *Africa Journal of Applied Statistics* 7(2) 983-992.

Adjabui, M . J., Howard, N. K. and Akamba, M. (2019). Biostatistical assessment ot mutagenicity studies: A stepwise confidence procedure. *Journal of Probability and Statistics, Hindawi*

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

Bonfinger, E.B (1987). Stepdown procedure for comparison with control. *Australia Journal of Statistics*,29:348-368.

Dunnett C. W. (1955). Multiple comparisons procedures for comparison for severals treatments with control' *Journal of America Statistics Assoc.* 50, 1096-1121

Dilba g., Bretz, F., and Guard, V. (2006) Simultaneous Confidence sets and Confidence interval for Multiple ratio *Journal of Statistical Planning and inferences* 136(8) 2640-2658

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society*, Series B 16, 175 -185.

Finner, H., and Strassburger, K. (2002). The partitioning principle: a powerful tool in multiple decision theory. *Annals of Statistics*, 30, 1194 –1213.

Hasler M., Vonk, R., Hothorn A. L. (2008). Assessing non-inferiority of a new treatment in three-arm clinical trial in the presence of heteroscedasticity.*Satistics in Medicine*,27:490-503

Hauschke, D. Kieser, M. Diletti, E. and Burke M. (1999) Sample size determination for proving equivalence based on ratio of two means for normally distributed data. *Statistics Med.* 18, 93 - 105

Hsu, J. C. and Berger, R. (1999). Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *Journal of the American Statistical Association*, 94, 468 - 482.

Stefansson, G. (1988). On confidence sets in multiple comparison. *Journal Statistical Decision Theory*, 2:89-104.

Tamhane A. C. and Logan B. R. (2004). Finding the Maximum Safe Dose Level for Heteroscedastic Data. *Journal of Biopharnaceutical Statistics* 14, 843-856

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

Tao, J., Guo, J and Shi, N. (2002) Stepwise procedure under unknown variances for toxicological evaluation. *Biometric Journal* 44 2 149 -160.

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