AJAS / SPAS

African Journal of Applied Statistics
Vol. 7 (2), 2020, pages 983 - 992.
DOI: http://dx.doi.org/10.16929/ajas/2020.983.251

ISSN 2316-0861

A note on Stepwise Confidence Procedure for Toxicological Evaluation Under Unknown and Unequal variances

Michael Jackson Adjabui (*)

Department of Mathematics, University for Development Studies, Navrongo, Ghana.

Received on March 26, 2020. Accepted on December 1, 2020

Abstract. In this article, we propose stepwise confidence set procedure for toxicity evaluation based on ratio of means difference under unknown and unequal variance across dose groups. The procedure is based on construction of confidence interval at various dose levels without multiplicity adjustment so that the familywise type one error (FWER) will be well-controlled. The results from our simulation study indicate that the FWER was properly control in the case with balanced design but failed in some cases of sample sizes for situations of unbalanced design. Also, the power of the procedure increases with increasing with mean of ratio differences and the sample sizes.

Résumé (Abstract in French) Nous proposons une méthode fonctionnant par étapes pour la détermination d'intervalles de confiance, en se basant sur le ratio de différences des moyennes, pour des variances égales ou non. Les intervalles de confiance sont construit à différents seuils, évitant les ajustements multiples, de sorte que les erreurs de type I soient bien controllées (FWER). Des études de simulations sont menées pour supporter les résultats.

Key words: confidence set; family-wise error rate; multiplicity adjustment; ratio of mean differences; unknown and unequal variances.
AMS 2010 Mathematics Subject Classification Objects : 62F03; 62F25; 62H15.

Michael Jackson Adjabui: madjabui@uds.edu.gh

The author.

Michael Jackson Adjabui, Ph.D., is a senior lecturer at the Department of Mathematics, University for Development Studies, Navrongo, Ghana.

1. Introduction

Practical toxicological equivalent that focus on safety assessment of compounds such as pharmaceuticals, food additives and environmental contaminants is unending area of research for academicians, government health institutions and pharmaceutical companies. Statistical methodology evaluating these toxicological compounds involve multiple comparisons procedure. That is comparison of a new compound (treatments) to a control. A reliable statistical procedure that will control the FWER is desirable. This is because, it is completely unacceptable to wrongly declare non-equivalent/unsafe drug to be equivalent/safe. Various authors have proposed some statistical procedures in evaluating toxicological problems. For instance, Hsu and Berger (1999) proposed stepwise confidence interval without multiplicity adjustment for toxicity studies under a situation of unknown but equal variance across dose groups. However homogeneity of variances among different dose levels are seldom satisfied in practice. Tao *et al.* (2002) extended Hsu and Berger (1999) procedure for a situation of heteroscedasticity across dose groups by employing two-stage sampling procedure proposed by Stein (1945) for toxicological studies. But the problem of unknown and unequal variances has been a long standing matter in multiple comparison procedures since the pioneering work of Welch (1938). This is because, for some biological factors or toxicity effect, variation of response under different dose levels is usually different with the change of dosages because patients in different groups tend to response differently at various dose levels. Tao *et al.* (2010) extended the stepwise confidence set for toxicological evaluation based on asymmetric loss function. However, their procedure demands assumption of known variances, which is often not realistic in practice for some therapeutic situation. Cao *et al.* (2015) extended the work of Tao *et al.* (2010) to a case of unknown variances by employing stepwise confidence interval method for asymmetric loss function.

Previous evaluations of toxicological compounds used difference in means of location parameters for statistical analysis. However, in some cases, means of difference in ratio of location parameters are more desirable than mean of difference in location parameter. This is because ratio of location parameter has the advantage of being dimensionless. That is dimensionless of measurement of endpoint, as compare with difference Dilba *et al.* (2004). For example, the efficacy in in-vivo tumor inhibition experiments may be defined as ratio of tumor volumes of treatments and control. The paper is organized as follows. In Section 2, the problem is formulated and its corresponding test statistics are derived. Confidence intervals based on Fierller's generalized method for multiple ratios are constructed in Section 3. For the purpose of application, illustrative example for the formulated

problem is demonstrated in Section 4. In Section 5, simulation study was carried out to assess the performance of the FWER and the power of our procedure. The conclusion of the article is in Section 6.

2. Testing Procedure

Let $X_{i1}, X_{i2}, \cdots, X_{in_i}$ be safety response of a sample data points for the ith group $(i = 0, 1, \cdots, k+1)$ of a toxicological compound under study. Consider a one-way model as follows.

$$X_{ij} = \mu_i + \epsilon_{ij} \quad i = 0, 1, 2, \cdots, k+1, \quad j = 1, 2, \cdots, n_i, \tag{1}$$

where X_{ij} represent the safety response for the jth experimental unit, $j = 1, 2, \cdots, n_i$ in the $(1, 2, \cdots, k)th$ treatment group, where $i = 0$ denote the negative control group and $i = k+1$ a positive control group respectively. The justification for the inclusion of a positive control is that if the study fail to detect any significant difference between the positive and zero-dose control groups, which are known to be different, then any lack of observed significant difference between a dose group and zero-dose control group may be the result of failed experimentation rather than closeness of their mean responses Hsu and Berger (1999). Assume that $X'_{ij}s$ are mutually independent and follows normal distribution with means μ_i and unknown variances σ_i^2, that is $X_{ij} \sim N(\mu_i, \sigma_i^2)$ 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. This is formulated in terms of hypothesis as:

$$H_{0i} : \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} \leq \delta^{(L)} \text{ or } \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} \geq \delta^{(U)} \text{ versus } H_{1i} : \delta^{(L)} < \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} < \delta^{(U)}, \tag{2}$$

for $i = 1, 2, \cdots, k$, where $\delta^{(L)} < 0$ and $\delta^{(U)} > 0$ are some pre-specified quantities. In practice, $\delta^{(L)}$ could be chosen to be $-\delta^{(U)}$ as a relevant safety threshold quantities. Letting ζ_i for $i = 1, 2 \cdots, k$ be the ratio of difference in means. We have

$$\zeta_1 = \frac{\mu_1 - \mu_0}{\mu_{k+1} - \mu_0}, \zeta_2 = \frac{\mu_2 - \mu_0}{\mu_{k+1} - \mu_0}, \cdots, \zeta_k = \frac{\mu_k - \mu_0}{\mu_{k+1} - \mu_0} \tag{3}$$

and the sample mean estimates as:

$$\bar{X}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} X_i, i = 1, 2, \cdots k, \qquad \bar{X}_{k+1} = \frac{1}{n_{k+1}} \sum_{k=1}^{n_{K+1}} X_{k+1,k} \quad \text{and} \quad \bar{X}_0 = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{0,j}$$

The test statistics

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

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

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

where $r = \delta^{(L)}$ or $\delta^{(U)}$, which are obtained by plugging in the estimators $\frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0}$ for $i = 1, 2, \cdots, k$ into equation (5) to approximate the degrees of freedom. Notice that, the degrees of freedom in this setting are according to Welch (1938) and Satterthwaite (1946), which are estimated and therefore not exact. This is because they rely on unknown group variances. When the assumption of increasing values of the endpoints represent better treatment effect is of interest, then equivalence/safety can be concluded if H_{0i} is rejected. That is:

$$T_i > t_{k,1-\alpha(\nu_i)} \quad \text{for} \quad i = 1, 2, \cdots, k$$

,
with its corresponding $(1-\alpha)$-quantiles $t_{k,1-\alpha(\nu_i)}$ of central $k-$variate t-distribution with degrees of freedom $\nu_i (i = 1, \cdots, k)$.

Therefore in this studies, our main objective is to establish practical equivalent/safety for ζ_i at level α. In other words, we will construct a $100(1-\alpha)\%$ confidence interval for all the individuals ζ_i for $i = 1, 2, \cdots, k$ and to ascertain whether resultant confidence interval lies entirely in range $(\delta^{(L)}, \delta^{(U)})$ without multiplicity adjustment.

3. Construction of confidence intervals for ratio of mean difference

We employed the generalized Fieller's theorem Fieller (1954) to construct confidence interval for ζ_i for $i = 1, 2, \cdots, k$. We do this by solving k quadratic equations and then adapt the following notations from Hasler *et al.* (2008) and Adjabui *et al.* (2019a) to obtain:

$$Z_i = \bar{X}_i - \bar{X}_0 \quad Z_{k+1} = \bar{X}_{k+1} - \bar{X}_0$$

$$Y_i = \frac{t_{1-\alpha,\nu_i}^2 s_i^2}{n_i}, \quad Y_{k+1} = \frac{t_{1-\alpha,\nu_{k+1}}^2 s_{k+1}^2}{n_{k+1}}, \quad Y_0 = \frac{t_{1-\alpha,\nu_0}^2 s_0^2}{n_0}.$$

Thus yielding the lower and upper confidence bounds respectively as:

$$\theta_{L,1-\alpha} = \frac{Z_i Z_{k+1} - Y_0 - \sqrt{(Z_i Z_{k+1} - Y_0)^2 - (Z_{k+1}^2 - Y_{k+1} - Y_0)(Z_i^2 - Y_i - Y_0)}}{Z_{k+1}^2 - Y_{k+1} - Y_0} \tag{6}$$

$$\theta_{U,1-\alpha} = \frac{Z_i Z_{k+1} - Y_0 + \sqrt{(Z_i Z_{k+1} - Y_0)^2 - (Z_{k+1}^2 - Y_{k+1} - Y_0)(Z_i^2 - Y_i - Y_0)}}{Z_{k+1}^2 - Y_{k+1} - Y_0} \tag{7}$$

The above confidence limits for one-sided $100(1-\alpha)\%$ confidence interval are only valid as long as $Z_{k+1}^2 > Y_{k+1} - Y_0$ by Fieller's theorem Fieller (1954).

Definition 1. Suppose that the data X have a distribution determined by a parameter $Z = \{\zeta_1, \zeta_2, \cdots, \zeta_k\} \in \Theta$. A confidence set $C(X)$ for Z 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 $C(X) \subset \Theta^*$.

For $(i = 1, \cdots, k)$, let

$$D_i^-(X) = \min\left\{\frac{Z_iZ_{k+1} - Y_0 - \sqrt{(Z_iZ_{k+1} - Y_0)^2 - (Z_{k+1}^2 - Y_{k+1} - Y_0)(Z_i^2 - Y_i - Y_0)}}{Z_{k+1}^2 - Y_{k+1} - Y_0}, 0\right\}$$

and and

$$D_i^+(X) = \max\left\{\frac{Z_iZ_{k+1} - Y_0 + \sqrt{(Z_iZ_{k+1} - Y_0)^2 - (Z_{k+1}^2 - Y_{k+1} - Y_0)(Z_i^2 - Y_i - Y_0)}}{Z_{k+1}^2 - Y_{k+1} - Y_0}, 0\right\}$$

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 intervals for ζ_i.
Furthermore, let:

$$C_i(X) = \begin{cases} D_i(X) & \text{if } D_i(X) \subset (\delta^{(L)}, \delta^{(U)}) \\ D_i(X) \cup (\delta^{(L)}, \delta^{(U)}) & \text{otherwise.} \end{cases} \tag{8}$$

Then $C_i(X)$ is a $100(1-\alpha)$ confidence intervals for ζ_i directed towards $\Theta^* = (\delta^{(L)} < \zeta_i < \delta^{(U)})$ for $i = 1, 2 \cdots, k$.

In assessing safety in toxicological compounds, it is desirable to devise a procedure that does not claim safety of a drug agent at a higher doses prior to the proclamation of safety at a lower doses Hsu and Berger (1999). We can establish this by answering the question "Is $\delta^{(L)} < \zeta_i < \delta^{(U)}$ in stepwise manner, continuing only if the answer is in the affirmative. see for example Hsu and Berger (1999), Tao *et al.* (2010) and Cao *et al.* (2015). Now the stepwise confidence set procedure for ratio of mean difference is as follows:

Step 1:
If $Z_{k+1}^2 - Y_{k+1} - Y_0 > 0$, then assert that $\mu_{k+1} - \mu_0$ is significantly greater than zero and go to step 2; else assert that $\mu_{k+1} - \mu_0 > Z_{i,1-\alpha}$ for $i(i = 1, \cdots, k)$ and stop .

Step 2:
If $D_1 \subset (\delta^{(L)}, \delta^{(U)})$, then assert $\zeta_1 \in (\delta^{(L)}, \delta^{(U)})$ and go to step 3; else assert $\zeta_1 \in C_1$ and stop .

$\vdots$

Step m:
If $D_{m-1} \subset (\delta^{(L)}, \delta^{(U)})$, then assert $\zeta_{m-1} \in (\delta^{(L)}, \delta^{(U)})$ and go to step m+1; else assert $\zeta_{m-1} \in C_{m-1}$ and stop.

Step k:
If $D_{k-1} \subset (\delta^{(L)}, \delta^{(U)})$, then assert $\zeta_{k-1} \in (\delta^{(L)}, \delta^{(U)})$ and go to step k+1; else assert $\zeta_{k-1} \in C_{k-1}$ and stop.

step k+1:
Finally, we conclude that the drug is safe and stop.

Theorem 1. *Suppose that step $M(0 \leq M \leq k)$ is the step at which the confidence interval procedure stops, and for $i = 1, 2, \cdots, k$, we define $C_i(X)$ as in Equation (8). Then in this case,*

$$P_Z(\textit{at least one incorrect declaration up to step } M) \leq \alpha$$

The proof of this Theorem is similar to the one given by Tao *et al.* (2002) We will remark that Theorem 1 guarantees that, the FWER is properly controlled at pre-specified nominal level α

4. An illustratve example: Toxicity study (Bovine growth hormone)

To illustrate the procedure discussed in this article, we used bovine growth hormone for toxicity assessment. Writing for Food and Drug Administration (FDA), Juskevic and Guyer (1990) reported on a number of experiments that did not indicate bovine growth hormones are harmful if present in milk consumed by humans. A subset of this data was considered by Hsu and Berger (1999). Another group of data indicating liver weights of male rats treated in 90 days in rbGH (recombinant bovine growth hormone) is considered. The treatment included a negative control diet or placebo treatment (labeled level 0), a positive control treatment which was rbGH by injection (labeled 5) and four different doses of bovine growth hormone given orally (labeled 1- 4). Liver weights of male rats treated for 90days are given in the Table.

In an experiment of this nature, when comparing the positive and negative controls, significant directional difference is the prefer inference, in the direction that the positive control accelerates weight gain. The positive control is included in the experiment in order to validate that the measurement process is capable of detecting known difference. Thus a declaration of significant directional difference suffices; a confidence interval for difference is not needed. On the other hand,

when comparing the four levels of orally fed by the bovine growth hormone with the negative control, the desire inference is *practical equivalence*: weight gain in rats given any growth hormone are close to weights gains of rats given the placebo. Such a declaration can be made if the confidence intervals for the weight gain difference between rats given rats growth hormone and rats given the the negative control turnout to be tight around 0 Hsu (1996).

Table 1. Liver weight (g) of male rats (Juskevic and Guyer (1990))

Level	Method	Dose (mg/kg)	Sample Size	Mean Weight	Std.dev. Weight
0	oral	0	30	8.637	0.88
1	oral	0.1	30	8.302	0.57
2	oral	0.5	30	8.754	1.0
3	oral	5.0	30	8.446	0.90
4	oral	50.0	30	8.297	0.84
5	injection	1.0	30	11.146	1.43

For illustration purpose, we set $\alpha = 0.05, -\delta^{(L)} = \delta^{(U)} = 1$, $Z_1 = -0.33, Z_2 = 0.117, Z_3 = -0.191, Z_4 = -0.34, Z_{k+1} = 2.509$ and $Y_1 l = 0.037, Y_1 u = 0.034, Y_2 l = 0.111, Y_2 u = 0.103, Y_3 l = 0.089, Y_3 u = 0.084, Y_4 l = 0.673, Y_4 u = 0.078, Y_0 = 0.075, Y_{k+1} = 0.197$

$$Z_{k+1}^2 - Y_{k+1} - Y_0 = 6.0231 > 0$$

$$D_1(X) = (-0.3010, 0.0011)$$

$$D_2(X) = (-0.2381, 0.1156)$$

$$D_3(X) = (-0.0262, 0.0778)$$

$$D_4(X) = (-0.3242, 0.0163)$$

Since $Z_{k+1}^2 - Y_{k+1} - Y_0 > 0$, we can claim that the experiment is sufficiently sensitive to distinguish between positive and negative control. Since all the $D_i(X)$ for $i = 1, 2, 3$ and 4 lies entirely in the range $(\delta^{(L)}, \delta^{(U)})$. practical equivalence has been established and the result is consistent with FDA conclusion that "the use of rbGH in diary cattle present no health risk to consumer" Juskevic and Guyer (1990)

5. Simulation Studies

We conduct simulation studies to investigate the performance of the FWER : if the procedure incorrectly declare a non-equivalent dose to be equivalent, and the power : if the procedure correctly declare an equivalent dose to be non-equivalent of our

procedure under the assumption of unknown but unequal variances across dose group. The simulation studies for evaluation of FWER and power of this procedure is similar to the results obtained by Adjabui *et al.* (2019b). Hence, we adopt the results in the Tables below, which indicate that the FWER is properly controlled and the power increases with increasing with mean of ratio difference and the sample sizes.

5.1. FWER

Table 2. Simulated FWER study for our procedure. setting $-\delta^{(L)} = \delta^{(U)} = 0.8$

$n_{E_1}(n_{E_2})$	Balanced design	Unbalanced design
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)

5.2. Power calculation

Table 3. 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
0.80	20	20	20	0.0250
0.80	30	30	30	0.0250
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

6. Conclusion

Toxicological substances' assessments are vital to pharmaceutical industries because the danger they pose to the environment and patients. Regulators, for example, FDA deemed it unacceptable to declare a dose as safe/equivalent when in actual fact it is unsafe/non-eqivalent, for which reasons reliable statistical procedures are required for proper control of the FWER. Statistical methodologies for toxicity studies were previously designed to assess mean difference of various dosages which may not be reliable for some cases. Therefore, we have proposed ratio based approach for toxicity assessment by constructing stepwise confidence procedure for multiple ratio setting. Our simulation results indicate the procedure is useful in controlling the FWER for a balanced design than unbalanced design and also the larger the ratio difference and sample size the larger the power.

References

Adjabui J. M., Howard K. N., Akamba M. (2019a) Biostatistical assessment of mutaganicity studies; A stepwise confidence procedure *Journal of probability and statistics, Hindawi* 2019: 1-8

Adjabui J. M., Kpeglo D. E., Hesse A. C. (2019b) Identyfying maximum safe dose (MSD) in a $(K+2)$-arm trial under unequal variance *Africa journal of Applied Statistics* 6(1) :635-647

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.

Dilba, G., Bretz, F., Guiard, V., Hothon, LA. (2004). Simultaneous cnfidence interval for ratios with application to the comparisons of several treatment with control *Methods of Information in Medicine,*43(5) : 465-469

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

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

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.

Hsu, J. C. *Multiple Comparisons:Theory and Method* Chapmann and Hall London

Juskevic, J. C. and Guyer, C. G. (1990) Bovine growth hormone:Human food safety evaluation. *Science* 249, 875-884

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

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. , Tang, M.,Shi N. (2010) Stepwise confidence interval procedure based on an asymmetric loss function with application to toxicological evaluation *Aust. N. Z stat.* 52(1), 93-107

Tao, J., Guo, J., Shi, N. (2002). Stepwise Procedure under unknown variances for toxicological evaluation *Biometrical 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