Hindawi
Journal of Probability and Statistics
Volume 2019, Article ID 3249097, 8 pages
https://doi.org/10.1155/2019/3249097

Research Article

Biostatistical Assessment of Mutagenicity Studies: A Stepwise Confidence Procedure

Michael J. Adjabui,[1] **Nathaniel K. Howard**,[2] **and Mathias Akamba**[3]

[1]*Department of Mathematics, University for Development Studies, Navrongo, Ghana*
[2]*Department of Statistics, University of Cape Coast, Cape Coast, Ghana*
[3]*Department of Statistics, University for Development Studies, Navrongo, Ghana*

Correspondence should be addressed to Michael J. Adjabui; madjabui@uds.edu.gh

Received 13 February 2019; Revised 26 March 2019; Accepted 28 March 2019; Published 14 April 2019

Academic Editor: Anna Karczewska

The paper addresses the issue of identifying the maximum safe dose in the context of noninferiority trials where several doses of toxicological compounds exist. Statistical methodology for identifying the maximum safe dose is available for three-arm noninferiority designs with only one experimental drug treatment. Extension of this methodology for several experimental groups exists but with multiplicity adjustment. However, if the experimental or the treatment groups can be ordered a priori according to their treatment effect, then multiplicity adjustment is unneeded. Assuming homogeneity of variances across dose group in normality settings, we employed the generalized Fieller's confidence interval method in a multiple comparison stepwise procedure by incorporating the partitioning principle in order to control the familywise error rate (FWER). Simulation results revealed that the procedure properly controlled the FWER in strong sense. Also, the power of our procedure increases with increasing sample size and the ratio of mean differences. We illustrate our procedure with mutagenicity dataset from a clinical study.

1. Introduction

Assessing an investigational substance for mutagenic activity is one of the vital concerns of genetic toxicologists. This is because it is unacceptable to declare a substance as nonmutagenic when in actual fact it is mutagenic. Hence, the objective of mutagenicity assay in regulatory toxicology is the decision on mutagenicity or nonmutagenicity of an investigational substance (Hothorn et al., [1]). Therefore, it is important to adopt reliable biostatistical procedure to properly control (FWER) in a strong sense. However, a deep-seated problem of a statistical procedure is the possibility of a false decision. A typical experimental design used in this assay for genotoxicity assessment in one-way model in $k + 2$ groups is as follows:

$$\{Negative\ control, treatment_1, \cdots, treatment_k, positive\ control\}. \quad (1)$$

In this setup, we have two objectives to achieve. Firstly, we need to assess the sensitivity of the experiment in order to ensure the validity of the study by comparing the the positive control to negative control. Secondly, we simultaneously compare each of the k treatments with the negative control. Statistical decision in this settings involves multiple comparison and stepwise procedures: that is, individual inferences are made in stepwise manner if the sequence of individual inferences is in a specific order, as used in Stefensson et al. [2], Cao et al. [3], Chen [4], and Adjabui et al. [5]. Some simultaneous inferences remit multiplicity adjustments by invoking the partition principle proposed by Finner and Strassburger [6]: where the parameter space is partitioned into many disjoint subsets and only one of these nonempty disjoint subsets contains the true parameter of interest, so that the FWER will be properly controlled. In literature, mutagenicity dataset has been assessed according to the proof of safety by utilizing the concept of the maximum safe dose (Hothorn and Hauschke [7], by numerous authors, among them Hauschke and Hothorn [8], Hauschke et al. [9], Hothorn and Bretz [10]).

As a result, this article discusses statistical aspects in terms of design and analysis using stepwise confidence

set-based procedure for identification of maximum safe dose: that is, the highest experiment dose with no biological relevant increase in safety effect in comparison with negative control (Hothorn amd Hauschke [9]). We organize the article as follows. In Section 2, we provide both the testing and confidence notations, which are essential for the construction of our proposed stepwise confidence procedure. We proposed stepwise confidence interval procedure for identifying maximum safe dose for a normally distributed data with equal variances across dose group in Section 3. In Section 4, we carried out simulation studies to investigate the performance of our stepwise confidence interval procedure in terms of FWER and power estimation. We apply our proposed procedure to analyze real dataset as an example in Section 5. We end with conclusion of our study in Section 6.

2. Preliminaries

2.1. Testing Procedure. Let a random sample $X_{i1}, X_{i2}, \cdots, X_{in_i}$ be the observations from ith group ($i = 0, 1, \cdots k+1$). Consider a one-way model as follows:

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

where X_{ij} represent the genetic response for the jth experimental unit, $j = 1, 2, \cdots n_i$ in the $(1, 2, \ldots, k)th$ treatment group, where $i = 0$ denote the negative control group and $i = k+1$ denote a positive control group, respectively. Suppose that the random sample variables X_{ij} are mutually independent and follow a normal distribution with means μ_i, μ_{k+1}, and μ_0 with their respective sample sizes n_i, n_{k+1}, and n_0 which are not necessarily equal. The random error has $N(0, \sigma^2)$, where σ^2 is unknown constant variance. Without loss of generality, assume larger values of μ_i imply better safety of the ith treatment group.

The test problem is formulated as

$$\begin{aligned} &H_{0i} : \mu_i - \mu_{k+1} \geq \delta \\ &\text{versus } H_{1i} : \mu_i - \mu_{k+1} < \delta \\ &\qquad\qquad\qquad \text{for } i = 1, 2, \cdots k, \end{aligned} \tag{3}$$

where δ is a relevant safety threshold. Practitioners, that is, genetic toxicologists, are often reluctant to define δ as an absolute value. However, Hauschke et al. [12] express the δ value as a fraction of difference between negative and positive control groups by $\delta = (\theta - 1)(\mu_{k+1} - \mu_0)$, for $\theta \in (0, 1)$. For some ethical reasons, a negative control group can be included in trial in (3). Therefore, the testing problem can be written as

$$\begin{aligned} &H_0 : \gamma_i \geq \theta \\ &\text{versus } H_1 : \gamma_i < \theta, \end{aligned} \tag{4}$$

where γ_i is the ratio of difference in means denoted as

$$\gamma_i = \frac{\mu_i - \mu_0}{\mu_{k+1} - \mu_0} \quad \text{for } i = 1, 2, \cdots k. \tag{5}$$

Equation (4) is valid if and only if $\mu_{k+1} - \mu_0 > 0$; this is inescapable condition and must be determined in the first step in our stepwise procedure in order to assess the sensitivity of the trial. We can rearrange and express (3) as

$$\begin{aligned} &H_{0i} : \mu_i - \theta\mu_{k+1} - (1-\theta)\,\mu_0 \geq 0 \\ &\text{versus } H_{1i} : \mu_i - \theta\mu_{k+1} - (1-\theta)\,\mu_0 < 0. \end{aligned} \tag{6}$$

Let the sample mean estimates be

$$\begin{aligned} \overline{X_i} &= \frac{1}{n_i}\sum_{i=1}^{n_i} X_i, \quad i = 1, 2, \cdots k \\ \overline{X}_{k+1} &= \frac{1}{n_{k+1}}\sum_{j=1}^{n_{k+1}} X_{k+1,j}, \\ \overline{X_0} &= \frac{1}{n_0}\sum_{j=1}^{n_0} X_{0,j}. \end{aligned} \tag{7}$$

The unknown and common variance σ^2 can be estimated as

$$\widehat{\sigma^2} = \frac{(n_i - 1)\,S_i + (n_{k+1} - 1)\,S_{k+1} + (n_0 - 1)\,S_0}{n_i + n_{k+1} + n_0 - 3} \quad \text{for } i = 1, 2, \cdots k, \tag{8}$$

where $\widehat{\sigma}^2$ is the pooled estimator of the variance σ^2 and S_i^2, S_{k+1}^2, and S_0^2 denote the sample variances for the experiment and positive and negative groups, respectively. Then, the random variables

$$T_i = \frac{\overline{X_i} - \theta\overline{X}_{k+1} - (1-\theta)\,\overline{X_0}}{\widehat{\sigma}\left(1/n_i + \theta^2/n_{k+1} + \left(1-\theta^2\right)/n_0\right)} \tag{9}$$

for $i = 1, 2, \cdots k$ are the test statistics for the testing problem in (3), which has t distribution with $\nu = n_i + n_{k+1} + n_0 - 3$ degrees of freedom. Pigeot et al. [13] have proved that one can claim safety if

$$T_i > t_{1-\alpha,\nu} \quad \text{for } i = 1, 2, \cdots k, \tag{10}$$

where $t_{1-\alpha,\nu}$ is $(1-\alpha)$-percentile of the central t- distribution with ν d.f. There are two approaches in solving the problem in (2), namely, the p-value approach and the confidence interval approach. It is noted in literature that the confidence interval approach is preferred to p-value approach. Therefore, in this study, we will construct a confidence set-based approach for γ_i for $i = 1, 2, \cdots k$ that remits multiplicity adjustment. The concept of maximum safe dose (MSD) for the proof of safety was defined by Hothorn and Hauschke [7] as

$$MSD = \max\left\{i : \gamma_i < \theta, \;\; i = 1, 2, \cdots k\right\} \tag{11}$$

which means that H_0 is rejected if $T_i > t_{1-\alpha,\nu}$ ($\gamma_i < \theta$) at a given level α. Then, safety can be concluded for treatments i ($i = 1, 2, \cdots k$).

In solving the testing problem in (3), we construct simultaneous confidence sets using intersection-union principle formulated by Berger [14]: the global null hypothesis can be expressed as the union of the subsets $\{H_{0i}\}$ of the null

hypotheses, H_0 against the intersection of the alternatives hypotheses H_1, that is,

$$H_0 = \bigcup_{i=1}^{k} H_{0i} \quad \text{against } H_1 = \bigcap_{i=1}^{k} H_{1i}. \tag{12}$$

If H_{0i} is rejected, then $j = 1, 2, \cdots i-1$ are all rejected too in a stepwise fashion. In this case, no multiplicity adjustment is needed. Notice that these hypotheses are a priori ordered according to their importance and one's interest and beliefs but they assume no order restrictions.

2.2. Fieller's Confidence Interval. We employed the generalized Fieller's theorem [15] to construct confidence interval for γ_i for $i = 1, 2, \cdots k$. We need to solve k quadratic equations and then adapt the following notation from Hasler et al. [11]:

$$\begin{aligned} Z_i &= \overline{X}_i - \overline{X}_0 \\ Z_{k+1} &= \overline{X}_{k+1} - \overline{X}_0 \\ Y_i &= \frac{t^2_{1-\alpha,\nu}}{n_i}, \\ Y_{k+1} &= \frac{t^2_{1-\alpha,\nu}}{n_{k+1}}, \\ Y_0 &= \frac{t^2_{1-\alpha,\nu}}{n_0}, \end{aligned} \tag{13}$$

thus yielding the upper confidence bounds as

$$\theta_{i,1-\alpha} = \left(-\infty, \frac{Z_i Z_{k+1} - Y_0 + \sqrt{\left(Z_i Z_{k+1} - Y_0\right)^2 - \left(Z^2_{k+1} - Y_{k+1} - Y_0\right)\left(Z^2_i - Y_i - Y_0\right)}}{Z^2_{k+1} - Y_{k+1} - Y_0} \right). \tag{14}$$

The above confidence interval is only valid as long as $Z^2_{k+1} > Y_{k+1} - Y_0$ by Fieller's theorem [15]. The upper confidence limits for one-sided $100(1-\alpha)\%$ confidence interval are

$$\theta_i = \frac{Z_i Z_{k+1} - Y_0 + \sqrt{\left(Z_i Z_{k+1} - Y_0\right)^2 - \left(Z^2_{k+1} - Y_{k+1} - Y_0\right)\left(Z^2_i - Y_i - Y_0\right)}}{Z^2_{k+1} - Y_{k+1} - Y_0} \quad \text{for } i = 1, \cdots k \tag{15}$$

for the parameters γ_i.

3. The Proposed Procedure

3.1. Stepwise Confidence Interval for Identifying Maximum Safe Dose Based on Ratio of Mean Differences. We identify maximum safe dose via Hsu-Berger [16] stepwise confidence set procedure: In the first step, we establish the assay sensitivity of the procedure by proving that $Z^2_{k+1} > Y_{k+1} - Y_0$. If not, the procedure stops, indicating that the sensitivity of experiment is inadequate. We estimate the upper confidence limits in the second step as

$$\widehat{\theta_i} = \frac{Z_i Z_{k+1} - Y_0 + \sqrt{\left(Z_i Z_{k+1} - Y_0\right)^2 - \left(Z^2_{k+1} - Y_{k+1} - Y_0\right)\left(Z^2_i - Y_i - Y_0\right)}}{Z^2_{k+1} - Y_{k+1} - Y_0} \quad \text{for } i = 1, \cdots k, \tag{16}$$

where k is the total number of treatment doses to be tested. In step three, we start screening the drug by screening the lowest dose (that is at $i = 1$) for the first safety drug and sequentially screen the subsequent doses for $i = 2, 3, \cdots k$ without adjusting the α levels in each of the steps in ascending manner searching for the first integer M, if it exists $\{1 \le M \le k\}$ such that $\theta_M < \theta$ and $\theta_{M+1} \ge \theta$ (this screens the first unsafe dose that is inferior to the reference dose). In this set up, dose level at step M is estimated as *MSD*: the highest estimated safe dose that is noninferior to the reference doses, such that it and all lower doses at steps $1, 2, \cdots M-1$ are also noninferior.

Once dose at step M is estimated as $\widehat{MSD}$, then the upper confidence bound for doses at $M+2, M+3, \cdots k$ steps is unneeded and should not be computed. A discernible property of this procedure is theoretically more powerful than Bonferroni-Holm step-down procedure (Holm [17]).

This is because the α value in our procedure is inexhaustible and hence in each step the entire α is used without multiplicity adjustment while in Bonferroni-Holm step-down procedure the α is exhaustible: that is, $\alpha/k, \alpha/(k-1), \cdots, \alpha/2, \alpha$ is exhausted and hence conservative. This may lead to liberal decision especially when k is large. The conservativeness of Bonferroni-Holm step-down procedure is overcome by the partition principle employed in our procedure.

3.2. Validity of the Stepwise Procedure. To construct and validate $100(1-\alpha)\%$ simultaneous confidence sets in the above procedure in estimating MSD, the individual confidence intervals should have $100(1-\alpha)\%$ confidence level. For a given parameter space Θ, we set $\Theta_i^c = (-\infty, \theta)$ as the rejection region and the alternative $\Theta_i = [\theta, \infty)$ as the acceptance. We can construct simultaneous confidence set for the parameter vector $\Gamma = \{\gamma_1, \gamma_2, \cdots \gamma_k\}$ by employing the partitioning principle (Bretz et al. [18]). In identifying the MSD, the parameter space Θ can be decomposed into nonempty disjoint subset as follows:

$$\begin{aligned}
\Theta_1^* &= \Theta_1 \\
\Theta_2^* &= \Theta_1^c \cap \Theta_2 \\
&\vdots \\
\Theta_i^* &= \Theta_1^c \cap \cdots \cap \Theta_{i-1}^c \cap \Theta_i \\
&\vdots \\
&\vdots \\
\Theta_k^* &= \Theta_1^c \cap \Theta_2^c \cap \cdots \cap \Theta_{k-1}^c \cap \Theta_k.
\end{aligned} \tag{17}$$

Therefore, $\Theta_1^*, \Theta_2^*, \cdots \Theta_k^*$ partition the entire parameter space Θ. That is, $\Theta = \Theta_1^* \cup \Theta_2^* \cup \cdots \cup \Theta_k^*$. Each of these subsets Θ_i^* is tested at a local level α with the conviction that the true parameter of interest can be found in one and only one of the nonempty disjoint subsets. This construction leads to multiple comparison procedure which guarantees the control of family-wise error in the strong sense. Hence, (12) can be rewritten as

$$\begin{aligned}
H_0 &= \bigcup_{i=1}^{k} \gamma_i \in \Theta_i \\
\text{against } H_1 &= \bigcap_{i=1}^{k} \gamma_i \in \Theta_i^c.
\end{aligned} \tag{18}$$

Theorem 1. *Suppose that $\theta_1, \theta_2 \cdots \theta_k$ are the $100(1-\alpha)\%$ confidence bounds for $\gamma_1, \gamma_2 \cdots \gamma_k$, respectively, with confidence level $1-\alpha$. Then, for all $\gamma_1, \gamma_2 \cdots \gamma_k \in \Theta$, we have*

$$\begin{aligned}
&P\left(\gamma_1 < \theta_1, \gamma_2 < \theta_2 \cdots \gamma_{M-1} < \theta_{M-1}, \gamma_M < \theta_M\right) \\
&\quad \geq 1 - \alpha.
\end{aligned} \tag{19}$$

The proof of Theorem 1 is a direct application of Theorem 1 of Hsu and Berger [16].

Proof.

Case 1. Let M=1 be the step at which the procedure stops. In such a situation, the assay sensitivity of the experiment cannot be assessed

Case 2. $2 \leq M \leq k$: For j = $1, \cdots k$, let

(i) $C_j(X) = \{\gamma_{k-j+1} < \theta_{k-j+1}\}$

(ii) $\Theta_1 = \{\gamma_k \geq \theta\}$ *and* $\Theta_j = \bigcap_{l=1}^{j-1}\{\gamma_{k-l+1} < \theta\} \cap \{\gamma_{k-j+1} \geq \theta\}$

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

$$\bigcup_{j=1}^{k} \left(C_j(X) \cap \Theta_j\right) \tag{20}$$

provides a $100(1-\alpha)$ confidence set for $\Gamma = \{\gamma_1 \cdots \gamma_k\}$ because if $\Gamma \in \Theta$ then

$$\begin{aligned}
&P_\Gamma\left\{\Gamma \in \bigcup_{j=1}^{k}\left(C_j(X) \cap \Theta_j\right\}\right. = P_\Gamma\left\{\Gamma \in C_j(X)\right\} \\
&\quad \geq 1 - \alpha.
\end{aligned} \tag{21}$$

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

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

Table 1: Simulated FWER, given $\alpha = 0.05, n_R = 20, n_P = 20$, and, $\theta = 0.8$.

$n_{E_1}(n_{E_2})$	HOMO	HETRO
5 (6)	0.0252 (0.0250)	0.0299 (0.0305)
7 (8)	0.0249 (0.02480	0.0184 (0.0177)
9 (10)	0.0251 (0.0251)	0.0109 (0.0160)
11 (12)	0.0249 (0.0247)	0.0157 (0.0153)
13 (14)	0.0244 (0.0249)	0.0149 (0.0114)
15 (16)	0.0247 (0.0248)	0.0141 (0.0136)
17 (18)	0.0249 (0.0250)	0.0129 (0.0128)
19 (20)	0.0249 (0.0248)	0.0124 (0.0119)
21 (22)	0.0249 (0.0250)	0.0117 (0.0115)
23 (24)	0.0205 (0.0247)	0.0110 (0.0109)
25 (26)	0.0250 (0.0249)	0.0106 (0.0106)
27 (28)	0.0251 (0.0245)	0.0100 (0.0009)
29 (30)	0.0250 (0.0248)	0.0096 (0.0093)

$$\cup\left(\bigcap_{j=1}^{M}\left\{\gamma_{k-j+1}<\theta\right\}\cap C_M(X)\right)$$
$$=\bigcap_{j=1}^{M-1}\left\{\gamma_{k-j+1}<\theta\right\}\cap C_M(X). \tag{22}$$

Finally, we have

$$P_\Gamma\left(\Gamma\in\bigcap_{j=1}^{M-1}\left\{\gamma_{k-j+1}<\theta\right\}\cap C_M(X)\right)$$
$$=P_\Gamma\left\{\Gamma\in\bigcup_{j=1}^{k}\left(C_j(X)\cap\Theta_j\right\}\right\}\geq 1-\alpha. \tag{23}$$

□

Remark 2. The resulting proof of Theorem 1 warrants the control of FWER at level 1-α in a strong sense.

For this reason, we state and prove the following proposition.

Proposition 3. *The stepwise simultaneous inferences procedure for ratio of difference in means strongly controls the FWER at level α.*

Proof. Let I be any unknown subset of $\{1, 2, \cdots k\}$. Suppose that $I = \emptyset$, then no FWER will ever exist. Thus, assume that $I \neq \emptyset$ and $I = \{i_1, i_2, \cdots i_m\}$, where $1 \leq i_1 < i_2 < \cdots < i_m \leq k$. Without loss of generality, let

$$P\left(\text{Reject one of } H_{0i}, i\in I \mid H_{0i}, i\in I \text{ is true}\right)=1$$
$$-P\left(\text{do not reject all } H_{0i}, i\in I \mid H_{0i}, i\right.$$
$$\left.\in I \text{ is true}\right)\leq 1-P\left(\text{do not reject } H_{0i_m}\mid H_{0i}, i\right.$$
$$\left.\in I \text{ is true}\right)\text{the procedure then stops at step } i_m$$
$$=1-P\left(\left\{\theta_{i_m}\notin(-\infty,\theta)\mid H_{0i}, i\in I \text{ is true}\right\}\right)\leq 1$$
$$-P\left(\gamma_1<\theta_1,\gamma_2<\theta_2\cdots\gamma_{M-1}<\theta_{M-1},\gamma_M<\theta_M\right)$$
$$\geq 1-\alpha\leq 1-(1-\alpha)\left(\text{By Theorem 1}\right)=\alpha. \tag{24}$$

□

Remark 4. Proposition 3 guarantees that FWER is properly controlled at prespecified nominal level α. This is a critical requirement by Food and Drug Administration (FDA) for statistical procedures in dose-findings.

To confirm these theoretical results, the following simulation studies were carried out at Section 4.

4. Simulation Studies

4.1. FWER. We conducted simulation studies to investigate the performance of the (FWER). Without loss of generality, we set $\theta = 0.8$, $\alpha = 0.025$. In this study, observations were generated with 1million replications from a normal distribution based on the assumption of equal variance across dose groups. This is indicated in Table 1 as HOMO. We also explored the effect of violation of this assumption as a way of comparing the two situations and this is indicated in Table 1 as HETRO. We used Hasler et al. [11] means configuration $\mu_P = 16.5, \mu_R = 16.5, \mu_{E_1} = 32.66, \mu_{E_2} = 32.66$. For HOMO=($\sigma_p = \sigma_R = \sigma_{E_i} = 5$ for $i = 1, 2$) and the HETRO= ($\sigma_p = 5, \sigma_R = 12, \sigma_{E_i} = 9$ for $i = 1, 2$). In the simulation study, we considered only $k = 2$ experimental treatment. Results from Table 1 indicated that the FWER is properly controlled at a nominal value $\alpha = 0.025$ in the case of equal variances but that of unequal variances is seriously conservative because simulated values are far below or above 0.025, the nominal level, and hence, poorly controlled the FWER.

4.2. Power Estimation. Power estimation is imperative for a well-design clinical study. There are many definitions of power in multiple comparisons procedures, but in this study,

Table 2: Power Estimation of the confidence intervals for $\sigma_R = 10, \sigma_P = 10, \sigma_{E_i} = 10\ i = 1, 2$.

Ratio(γ_i)	$n_{E_i=1,2}$	$\epsilon = 0.25$	$\epsilon = 0.5$	$\epsilon = 1$
0.85	5	0.0623	0.0402	0.0319
0.85	20	0.1161	0.0574	0.00385
0.85	30	0.1409	0.0644	0.0410
0.85	40	0.1606	0.700	0.00429
0.90	5	0.1332	0.0623	0.0402
0.90	20	0.3336	0.1410	0.0573
0.90	30	0.4234	0.1409	0.0645
0.90	40	0.4903	0.1606	0.0700
0.95	5	0.2460	0.0928	0.0503
0.95	20	0.6312	0.2082	0.0828
0.95	30	0.7550	0.2627	0.1085
0.95	40	0.8273	0.3056	0.1086
1.00	5	0.3964	0.1332	0.0623
1.00	20	0.8643	0.3336	0.1161
1.00	30	0.9422	0.4230	0.1409
1.00	40	0.9720	0.4903	0.1606
1.05	5	0.5641	0.11842	0.0764
1.05	20	0.9689	0.4830	0.1578
1.05	30	0.9930	0.5982	0.1961
1.05	40	0.9980	0.6771	0.2266
1.10	5	0.7300	0.2460	0.0928
1.10	20	0.9957	0.6312	0.2082
1.10	30	0.9996	0,7550	0.2630
1.10	40	0.9999	0.8273	0.3057
1.15	5	0.8437	0.3124	0.1113
1.15	20	0.9965	0.7635	0.2672
1.15	30	0.9996	0.9232	0.3335
1.15	40	0.9999	0.9232	0.3951
1.20	5	0.9242	0.3963	0.1332
1.20	20	0.9999	0.8643	0.3335
1.20	30	0.9999	0.9422	0.4235
1.20	40	1.0000	0.9730	0.4903

we will define power in the case of maximum safe dose. The maximum safe dose i is established when $\theta_j < \theta$ and $\theta_{j+1} \geq \theta$ for $j = 1, 2, \cdots i$. That is,

$$P\left(\widehat{MSD} = i\right) = P\left(\bigcap_{j=1}^{i}\left\{T_j > t_{1-\alpha,\nu_i}\right\} \cap \left\{T_{i+1} \leq t_{1-\alpha,\nu_i}\right\}\right). \tag{25}$$

Hence, in this setting, power is defined as the probability of rejecting the incorrect null hypotheses. This power concept is directly related to all-pairs power definition introduced by Ramsay [19]. Therefore, (25) expression can be rewritten as

$$P\left(\text{Reject } H_{jo} \text{ for } j = 1, 2, \cdots i\right) = P\left(\bigcap_{j=1}^{i}\left\{T_j > t_{1-\alpha,\nu_i}\right\}\right). \tag{26}$$

Therefore, (26) can be calculated from a k variate noncentral t-distribution with ν_i degree of freedom and noncentrality parameters for $i = 1, 2, \cdots, k$:

$$\Theta_i = \frac{\mu_i - \theta\mu_{k+1} - (1-\theta)\,\mu_0}{\sigma\sqrt{\left\{1/n_i + \theta^2/n_{k+1} + (1-\theta)^2/n_0\right\}}}. \tag{27}$$

It is possible to express common variance σ as a fraction of difference $\mu_{k+1} - \mu_0$, that is, $\sigma = \epsilon(\mu_{k+1} - \mu_0)$, $\epsilon > 0$. Hence, the following representation of noncentrality parameter based on the ratio of mean differences is stated as

$$\Theta_i = \frac{\gamma_i - \theta}{\epsilon\sqrt{\left\{1/n_i + \theta^2/n_{k+1} + (1-\theta)^2/n_0\right\}}}. \tag{28}$$

From (28), it is clear that the expected values of power are a function of γ_i, the ratio of mean differences, and the sample sizes. From Table 2, it can be seen that power increases with increasing γ_i and sample size but decreases with increasing

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

Experimental group	Mean	Standard deviation	Sample size
Vehicle control	2.57	1.27	7
Hydro30	3.80	1.10	5
Hydro50	6.30	1.48	5
Hydro75	14.0	3.97	5
Hyro100	20.0	4.06	5
Positive control	25	8.91	4

Table 4: Summary of the test for micronucleus assay data from Hasler et al. [11].

Treatment groups	Unadjusted p-values	Upper bound
30 mg/kg	0.0088	0.24
50 mg/kg	0.0182	0.35
75 mg/kg	0.0288	0.74
100 mg/kg	0.09639	1.04

values of ϵ. This is consistent with the results of Pigeot et al. [13].

5. Example

To illustrate our procedure, we used raw data published by Adler and Kliesch [20] for a micronucleus assay by applying 30mg/kg, 50mg/kg, 75mg/kg, and 100mg/kg doses of hydroquinone (Hydro) with positive control 25mg/kg cyclophosphamide. Their primary interest is to demonstrate whether the underlying substance is able to induce chromosome damage or interact with spindle apparatus. The male mice studies results of 24h sampling time are given in Table 3. and summary of the test for micronucleus assay data from Hasler et al. [11] is given in Table 4.

In evaluation of the mutagenicity data from Table 3 and setting $\alpha = 0.05$ and $\theta = 0.5$, where θ is the safety threshold, the following results were obtained:

$$\begin{aligned} \widehat{\theta_1} = 0.24 < \theta = 0.5 \quad &\text{we reject } H_{01} \\ \widehat{\theta_2} = 0.35 < \theta = 0.5 \quad &\text{we reject } H_{02} \\ \widehat{\theta_3} = 0.74 \not< \theta = 0.5 \quad &\text{we do not reject } H_{03}; \end{aligned} \tag{29}$$

the procedure then stop at step 3, which implies that it is needless to step it further down.

From this analysis, the doses 30mg/kg and 50mg/kg are declared safe while doses 75mg/kg and 100mg/kg are unsafe at level α. Since $\widehat{\theta_3} = 0.74 \not< \theta = 0.5$, 50mg/kg is recommended as the maximum safe dose, which the highest dose that is noninferior to the reference drug at level α. Note that 30mg/kg is also noninferior to the reference drug but lower.

6. Conclusion

In this paper, we have proposed a stepdown confidence set approach for identification of maximum safe dose within the framework of noninferiority clinical trials. The classical three-arm trial for noninferiority investigations involves only one experimental treatment but in clinical trials some therapeutic situations necessitate comparisons with several experimental compounds. Therefore, the proposed $(k + 2)$ – *arm* trial is an extended three-arm noninferiority trial with only one treatment to multiple treatments without multiplicity adjustment. Our simulations results revealed strong control of the familywise type I error rate when we assumed equal variances across dose groups for a normally distributed dataset. This was validated by the partitioning principle.

Data Availability

I used data from literature for illustrative purposes.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] L. A. Hothorn, M. Hayash, and D. Siedel, "Dose-response relationship in mutagenicity assay including an appropriate positive control group; a multiple test approach," *Enviromental and ecologic statistics*, vol. 7, pp. 27–42, 2000.

[2] G. Stefansson, W. C. Kim, and J. C. Hsu, "On confidence set in multiple comparisons," in *Statistics Decision Theory and Related Topic, IV*, S. Gupta and J. O. Berger, Eds., vol. 2, pp. 89–104, Springer, New York, NY, USA, 1988.

[3] L. Cao, J. Tao, N.-Z. Shi, and W. Liu, "A stepwise confidence interval procedure under unknown variance based on an asymetric loss fuction for toxicological evaluation," *Australian & New Zealand Journal of Statistics*, vol. 57, no. 1, pp. 73–98, 2015.

[4] J. T. Chen, "A nonparametric coherent confidence procedure," *Communications in Statistics—Theory and Methods*, vol. 45, no. 11, pp. 3397–3409, 2016.

[5] M. Adjabui, N. Howard, and A. Luguterah, "Nonparametric stepwise procedure foe identification of maximum safe

dose(MSD)," *Asian Research Journal of Mathematics*, vol. 6, no. 3, pp. 1–12, 2017.

[6] H. Finner and K. Strassburger, "The partitioning principle: a powerful tool in multiple decision theory," *The Annals of Statistics*, vol. 30, no. 4, pp. 1194–1213, 2002.

[7] L. A. Hothorn and D. Hauschke, "Identification of maximum safe dose: a multiple test approach," *Journal of Biophamarceutical Statistics*, vol. 10, pp. 15–30, 2000.

[8] D. Hauschke and L. A. Hothorn, "Two-stage testing of safety: A statistical view," *ATLA Alternatives to Laboratory Animals*, vol. 31, no. 1, pp. 77–80, 2003.

[9] D. Hauschke, R. Slacik-Erben, S. Hensen, and R. Kaufmann, "Biastatistical assessment of mutagenicity studiesby including the positive control," *Biometrical Journal*, vol. 47, no. 1, pp. 82–87, 2005.

[10] L. A. Hothorn and F. Bretz, "Dose-response and threshold in mutagencity studies: a statistical testing approach," *ATL*, vol. 31, no. 1, pp. 97–103, 2003.

[11] M. Hasler, R. Vonk, and L. A. Hothorn, "Assessing non-inferiority of a new treatment in a three-arm clinical trial in the presence of heteroscedasticity," *Statistics in Medicine*, vol. 27, no. 4, pp. 490–503, 2008.

[12] D. Hauschke, T. Hothorn, and J. Schäfer, "The role of control group in mutagenicity studies:maching biological and statistical relevance," *Alternatives to Laboratory Animals*, vol. 31, no. 1, pp. 65–75, 2003.

[13] I. Pigeot, J. Schäfer, J. Röhmel, and D. Hauschke, "Assessing non-inferiority of new treatment in a three-arm clinical trial including a placebo," *Statistics in Medicine*, vol. 22, no. 6, pp. 883–899, 2003.

[14] R. L. Berger, "Multiparameter hypothesis testing and acceptance sampling," *Technometrics. A Journal of Statistics for the Physical, Chemical and Engineering Sciences*, vol. 24, no. 4, pp. 295–300, 1982.

[15] E. C. Fieller, "Some problems in interval estimation," *Journal of Royal Statistical Society Series B*, vol. 16, pp. 175–185, 1954.

[16] J. C. Hsu and R. L. Berger, "Stepwise confidence interval without multiplicity adjustment for dose response and toxicity studies," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 468–482, 1999.

[17] S. Holm, "A simple sequential rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[18] F. Bretz, L. A. Hothorn, and J. C. Hsu, "Identifying effective and/or safe doses by stepwise confidence interval for ratio," *Statistics in Medicine*, vol. 22, no. 6, pp. 847–858, 2003.

[19] P. H. Ramsey, "Power differences between pairwise multiple comparisons," *Journal of the American Statistical Association*, vol. 73, no. 363, pp. 479–485, 1978.

[20] I.-D. Adler and U. Kliesch, "Comparison of single and multiple treatments in the mouse bone marrow micronucleus assay for hydroquinone and cyclophosphamide," *Mutation Research/Environmental Mutagenesis and Related Subjects*, vol. 234, no. 3-4, pp. 115–123, 1990.