Finding Minimum Effective Dose With Multiple Testing Procedures

Flora F Opoku Asantewaa[1] and Lucy Kerns *[2]

[1]Department of Mathematics and Statistics, Youngstown State University, 1 University Plaza, Youngstown, OH 44555

Abstract

The dose-response studies of a drug is important in clinical trials, pharmacology and toxicology research. In dose-response clinical studies, humans or animals are exposed to the drug which has some effect depending on the dose level, and efficacy and/or toxicity are often the primary endpoints to be evaluated to determine the effective and safe ranges of the drug. Typical clinical dose-response studies consist of the comparison of several doses of a drug against a control, and one of the primary goals is to identify the minimum effective dose (MED) and the maximum safe dose (MSD).

Multiple comparison procedures have been one of the major techniques to analyze data from dose-response studies. We will discuss some multiple comparison techniques, which include multiple test procedures (closed testing procedures and testing procedures based on partition principle), and simultaneous confidence intervals under the monotonicity assumption. Monotonicity helps to avoid logical inconsistencies and as such is essential in multiple testing. Examples are given to illustrate and compare these different approaches using summary data from Wöhr, M., Borta, A., and Schwarting, R. (2005).

Keywords: Dose-response studies, minimum effective dose, multiple comparison procedures, closed principle, partition principle.

1 Introduction

Dose-response studies are critical procedures under clinical trials since it borders around the safety of consumers. There is a rising need to find the safest and effective dose which will be ideal to consumers. The effective dose range is found within the range of the minimum effective dose (MED) and the maximum tolerated dose (MTD).

The MED of any drug can be defined as the minimum dose for which a clinically significant response is observed and the mean response at that dose is significantly better than the mean response of the control dose. The MTD is relatively the highest possible but tolerable dose level in reference to a prespecified toxicity threshold. This study focusses on finding the MED.

*Corresponding author: luxjing@gmail.com

Under dose-response experiments, several doses are given to separate groups and in most cases compared against a standard group that usually receives a zero dose or placebo. The main purpose is to find the lowest dose level which produces a desirable effect over the control dose if a dose response effect exists. Testing the groups against the placebo or standard group gives rise to multiple testing which is a major challenge.

In clinical trials, multiplicity or multiple testing is virtually present in every procedure, ranging from considering several treatments, different dose groups, subgroups, the use of different statistical models,etc.

Multiple testing occurs in a lot of fields but is found more commonly in biology related fields where it is common to have more than 20 simultaneous tests whiles studying a phenomena. One obvious and easy approach may be to test each hypothesis from a set of hypothesis separately using a significance level. This may seem like a good option, except it allows for errors since the chances of observing at least one significant result is highly inflated only due to chance. As the number of independent tests increases for the same significant level, the chances of obtaining a significant result and consequently making a false decision increases.

A need to adjust for such errors has been a controversial debate in biostatistics. Different methods have been suggested to help in adjusting for multiplicity. Methods for dealing with multiple testing usuallly requires adjusting the significance level α in some way, such that the probability of observing at least one significant result due to chance remains under the desired significance level.

The errors that arise due to multiple testing are of concern under dose-response studies and hence needs to be adjusted for to address safety issues. Different researchers have proposed various methods to test hypotheses in a stepwise manner to adjust for errors and also help in determining the minimum effective dose.

This current study explores three different stepwise procedures that seek to adjust for errors and subsequently determine the minimum effective dose under a monotonicity assumption, that is, $\mu_1 \leq \ldots \leq \mu_n$, where μ_i is the mean response of dose responses and indicates a better average outcome when large. A monotone procedure tends to reject a hypothesis whenever it rejects another hypothesis with a bigger p-value. A rejection of a hypothesis H_j implies the rejection of H_i if $\mu_i < \mu_j$. The Bonferroni-Holm procedure and Hochberg procedures are both closed testing methods that are performed in different stepwise directions. The approach using the Simultaneous Confidence intervals employs the partition principle.

The Bonferroni-Holm procedure is an extension of the Bonferroni test which only computes a new pairwise alpha that keeps the familywise alpha value at 0.05 to use in testing. The Holm procedure is a sequential stepdown procedure which is based on the Bonferroni test. This method orders all p-values from minimum to maximum with corresponding hypothesis and then the procedure scans forward and stops as soon as the first non-significant p-value is obtained. The procedure then declares all remaining p-values not significant and the dose level associated with the very last significant p-value the MED.

The Bonferronni test by itself lacks power due to the assumption that all null hypothesis are true which is not usually likely the case. The use of the Holm procedure with the Bonferroni tests gives a stepwise procedure that helps to correct for multiple testing errors and also compensate for the loss in power that arises with the use of only the Bonferroni procedure.

The Hochberg procedure works in a similar manner to the Bonferroni-Holm procedure. It is also an improvement on the Bonferroni test. This procedure simply works in the opposite direction to the Bonferroni-Holm procedure. This is a step-up procedure which scans forward through the set of ordered p-values from maximum to minimum and stops as soon as the first significant p-value is obtained. For the Hochberg procedure, the dose level associated with the very first significant p-value is the MED.

The simultaneous confidence interval approach employs the use of testing confidence intervals against prespecified clinically significant thresholds. Confidence intervals are known to be clinically better than the use of p-values or yes/no decisions. Two sided confidence intervals may be constructed but for the purposes of this project, the lower confidence bound is computed and used in testing simultaneously.

The procedure starts with computing the lower confidence bound for the difference between the mean response of the highest dose level and the that of the control. This lower bound is compared with the specified clinically significant threshold and if it is found to be bigger than this threshold, then the dose level is claimed significantly better than the control dose level and the procedure continues to the next highest dose. The procedure continues until the last dose level or until a bound fails to clear its threshold. If a lower bound fails to clear its threshold, the dose level tested before this lower bound becomes the MED as defined by the simultaneous confidence approachand otherwise declares all dosages being tested significantly better than the control dose if the least dose level clears its threshold.

Summary data from a study by Wöhr, M., Borta, A., and Schwarting, R. (2005) will be used in the proposed procedures.

2 Methodology

2.1 Multiple Test Procedures

Multiple testing refers to testing more than one hypothesis at a time. Dose-response studies typically employ the use of multiple comparison procedures due to comparisons that are made between different dosage levels and a placebo or zero dosage level to identify the effective and safe ranges of a drug in clinical trials. Generally, multiple test procedures yield multiple type I errors in analysis due to the multiple hypotheses that needs to be tested.

The elevation of false-positive (type I) error rate is not acceptable and more particularly in clinical trials where safety of participants is at stake. A high incidence of these false positives can cause researchers to claim significance by including a high number of inefficient treatments by chance only [1].

In dealing with single hypothesis testing, type I error rate is controlled at a designated α level. The probability that at least some type I error are committed increases greatly and may exceed the given level α when testing multiple hypotheses with each hypothesis having a specified type I error probability under the assumption that all null hypothesis are simultaneously true [2] [3].

The ordered nature of doses that are compared with the zero dose indicates the use of stepwise multiple test procedures. Two procedures that are known to make use of such stepwise approaches are the Closed test procedures and the Partitioning test procedures.

2.2 Closed Testing Procedures

The closed testing procedure which was introduced by Marcus, Peritz and Gabriel (1976) is identified to be one of the most basic and efficient ways used to controll the multiple type I errors that occur with multiple testing.

The closed testing procedures are designed to test a set of hypothesis that is closed under intersection. A set is closed under intersection if for a set C containing a collection of subsets from a set E, the intersection of any two of the subsets in set C belongs to the set C whenever the two sets are also in set C.This testing procedure is characterised by two essential features, namely the closure principle and the familywise error rate.

1. **Closure Principle**
   Let each of $\{H_i\}_{i=1}^n$ and $\{H_j\}_{j=1}^n$ be a family of hypothesis and let $p_1, \ldots, p_n$ be p-values associated with each hypothesis, then we can define the intersection null

$$H_{ij} = H_i \cap H_j = \{p_i, p_j \backsim U([0,1])\}$$

   The closure of this family can be generalized as (Marcus et'al; 1976)

$$H_I = \bigcap_{i \in I} H_i \quad \text{for all } I \subset \{1, 2, \ldots, n\}$$

   The closure principle tests each hypothesis H_I and if it is not rejected, then the closure principle accepts all H_i since the representation $H_I = \bigcap_{i \in I} H_i$ shows that all H_i for $i < I$ are implied by H_I. This leads to a step-down test procedure for the closed family where H_n is first tested.

   For the family of hypothesis,$\{H_i\}_{i=1}^n$, where H_i is the hypothesis of homogeneity of the first $i+1$ dose groups, it can be observed under the monotonicity of dose response that the family is closed under intersection since $H_i \cap H_j = H_{max\{i,j\}}$ which also belongs to the family for every i, j between 1 and n.

   Consider a dose response study which compares two doses of a drug with a placebo with associated hypotheses given by $H_1 : \mu_1 \leq \mu_0$ and $H_2 : \mu_2 \leq \mu_0$ and p-values p_1 and p_2

respectively for testing the hypotheses. A closed test is constructed by considering all possible intersections of all hypotheses associated with the doses to form a closed family of hypotheses and uses the natural order therein to ensure that the family of hypotheses is closed under intersection. The family contains the following three intersection hypotheses:

$$H_1, H_2 \text{ and } H_{12} = H_1 \cap H_2$$

Each hypothesis in the closed family is tested using a suitable local α-level test and the procedure rejects a hypothesis if all intersection hypotheses containing this hypothesis are rejected by the associated local tests. In effect, both H_1 and H_{12} would have to be rejected to conclude that $\mu_1 > \mu_0$ since both hypotheses contain H_1 and the same goes for concluding that $\mu_2 > \mu_0$, both H_2 and H_{12} will be rejected due to the closure of intersections.[13]

Under the dose comparison hypothesis tests, if H_n is not rejected, then any other hypothesis below H_n in the heirachy is accepted and no dose is considered effective. If H_n is rejected, then H_{n-1} is tested next and the step-wise sequence is continued until the last rejected hypothesis.

2. Familywise Error Rate

The process of multiple testing yields a higher probability of having false-positives.

$$\begin{gathered}\mathbb{P}(\text{making an error}) = \alpha \\ \mathbb{P}(\text{not making an error}) = (1-\alpha) \\ \mathbb{P}(\text{not making an error in n tests}) = (1-\alpha)^n \\ \mathbb{P}(\text{making an error in n tests}) = 1-(1-\alpha)^n\end{gathered}$$

It is obvious that as the number of comparisons, n increases, this probability gets progressively higher. If each hypothesis is tested independently, then each comparison may have some level of error rate known as the Per Comparison Error Rate (PCER). The PCER is the expected proportion of Type I errors among the n hypotheses being tested. Controlling only PCER at level α is not adequate and hence we need consider an error rate that takes into all n hypotheses as a joint family.

The concept of family in regards to multiple testing refers to any collection of hypotheses for which some joint measure of errors need to be taken into account.

Generally, step testing procedures are known to control the familywise error rate. If any step testing procedure satisfies the conditions of closed procedure then that procedure controls the family-wise error rate strongly [5].

The Familywise Error Rate (FWER) is the probability of making one or more false discoveries, or type I errors, among all the hypotheses when performing multiple hypotheses tests. It can also be referred to as the combined type I error rate.

Many procedures have been developed to control the FWER. FWER control is more desirable when the number of tests is small, so that a good number of rejections can be made, and all

can be trusted to be true findings. Some of the well known procedures include the Bonferroni/ Bonferroni - Holm procedures and the Hochberg procedures.

2.2.1 Bonferroni/ Bonferroni - Holm Procedures (1979)

The Bonferroni correction approach to correcting FWER is the most widely used approach.This approach simply computes a new pairwise alpha to keep the familywise alpha value at 0.05 (or the specified alpha value). Let the Familywise Error(FWE) be defined as

$$\alpha_{FWE} \leq 1 - (1 - \alpha)^n$$

Then the formula to compute a new pairwise alpha is as follows

$$\alpha_B = \frac{\alpha_{FWE}}{n}$$

where α_B is the new alpha based on the Bonferroni test that is now used in each comparison, α_{FWE} is as defined and n is the number of comparisons to be performed.
The Bonferroni test is most commonly used due to its flexibility and simple computation. It however lacks power because of the assumption that all null hypothesis are true which is unlikely to be the case.

To account for the loss of power in the classical Bonferroni test, several alternatives that is based on the Bonferroni approach have been developed. One of such is the Holm procedure. This test has much greater power than the single-step Bonferroni procedure because it begins at the same significance level as the Bonferroni procedure and tests the other hypotheses at successively higher levels but still maintains the flexibility that allows for use with any set of statistical tests.

The Holm's procedure is a sequential step-down process that scans forward and stops as soon as a p-value fails to clear its threshold.

Let $H_1,\ldots,H_n$ be a family of hypothesis and their corresponding p-values after conducting tests with significance level α/n be $P_1,\ldots,\ P_n$. The steps involved in this procedure are as outlined below.

- $P_1,\ldots,\ P_n$ are arranged in order from smallest to largest as $P_{(1)},\ldots,\ P_{(n)}$ with corresponding hypothesis $H_{(1)},\ldots,H_{(n)}$ and each p-value is compared with a significance level, $\alpha/n + 1 - k$, where k is the smallest index being tested.
- The first smallest p-value is compared with α/n. If this p-value is found to be greater than or equal to α/n, the procedure stops and no p-values are significant. The process continues otherwise.
- If the first p-value is significant, the second smallest p-value is next tested. It is compared to $\alpha/(n-1)$ and if it is found to be greater then the procedure is stopped and no further p-values are found to be significant. The process continues otherwise.
- If for all n, no p-value is found to be greater than its corresponding significance level,then all of the null hypotheses is rejected.

The power gain obtained by using a sequentially rejective Bonferroni test (Bonferroni-Holm) instead of a classical or traditional Bonferroni test depends very much upon the alternative. It is small if all the hypotheses are 'almost true', but it may be considerable if a number of hypotheses are 'completely wrong'[6].

2.2.2 Hochberg Procedure

The Bonferroni procedure in itself is conservative in that the Bonferroni inequality does not take into account the correlations amongst test statistics and hence lacks power if several highly correlated tests are undertaken.

Simes (1986) proposed an adjustment to the Bonferroni inequality that rejects H_o if $P_{(n)} \leq k\alpha/n$ for at least one $k(1 \leq k \leq n)$. This procedure does not always produce a level α test, however, Simes proved that a level α test can be achieved provided test statistics used are independent. In such a case, the Simes procedure is strictly more powerful than other Bonferroni adjustments. Applying the closure principle to the Simes procedure yields the Hochberg procedure.

The Hochberg's (1988) [7] procedure is also an improvement to the classical Bonferronni test designed to control FWER. It offers an adjustment which creates a more powerful test than the Holm procedure.Hochberg's method is thought of as a step-up version of the Bonferroni test. Hochberg's method is more powerful than Holm's method, but the test statistics need to be independent as indicated by Simes procedure or have a distribution with multivariate total positivity of order two or a scale mixture thereof for its validity [9].

Hochberg's procedure can be seen as a reversed Holm procedure, since it uses the same critical values, but in a reversed testing sequence. Let H_1,...,H_n be a family of hypothesis and their corresponding p-values after conducting tests with significance level $k\alpha/n$ be P_1,..., P_n. The p-values are ordered with corresponding hypothesis and compared to $\alpha/n + 1 - k$, where k is the largest index being tested. The test proceeds as follows

- The hypothesis $H_{(n)}$ associated with the largest p-value, $P_{(n)}$ is first tested. If $P_{(n)} \leq \alpha$, the procedure stops and all hypotheses $H_{(1)}$,...,$H_{(n)}$ are rejected. $H_{(n)}$ is retained if not found to be significant and the procedure continues testing $H_{(n-1)}$ at the smaller significance level $\alpha/2$.
- If $P_{(n-1)} \leq \alpha/2$, the procedure stops and all hypotheses $H_{(1)}, ..., H_{(n-1)}$ are rejected.
- The steps are repeated for all n until either the first rejection occurs or all null hypotheses $H_{(1)}$,...,$H_{(n)}$ are retained

2.3 Partitioning Testing Principle

Stefansson, Kim and Hsu (1988) introduced the partitioning principle as another method of testing multiple hypothesis by setting up disjoint hypothesis from a family of hypothesis. It was further

advanced by Finner and Strassburger (2002) [10]. Partitioning principles are used to construct procedures that are more powerful than the procedures derived using the closed testing procedures. The mutually exclusive hypotheses that are created by partitioning can each be tested at level α without compromising the FWER control since they are disjoint. Consider testing the hypothesis

$$H_{0i} : \theta_i \leq 0, \qquad i = 1, \ldots, k$$

A partitioning test to test this hypothesis follows the following steps.

- Let $I \subseteq \{1, \ldots, k\}$ and $I \neq \varnothing$, form mutually exclusive partitions such that $H^*_{0I} : \theta_i \leq 0$ for all $i \in I$ and $\theta_j > 0$ for $j \notin I$. There are 2^k parameter subspaces and $2^k - 1$ hypothesis to be tested.
- Each hypothesis H^*_{0I} is tested at level α.
- For each i we can infer that $\theta_i > 0$ if and only if we reject all H^*_{0I} with $i \in I$ because H_{0i} is the union of H^*_{0I} with $i \in I$.

To illustrate the partitioning process above, consider the following example involving two doses and a placebo in a clinical trial. The first step involves the partitioning of the union of the hypothesis

$$H_1 : \mu_1 \leq \mu_0, \quad H_2 : \mu_2 \leq \mu_0$$

into three mutually exclusive hypotheses since $2^2 - 1 = 3$. The partitioned hypothesis is as follows:

$$\begin{array}{lll} H^*_1 : \mu_1 \leq \mu_0 & \text{and} & \mu_2 \leq \mu_0, \\ H^*_2 : \mu_1 \leq \mu_0 & \text{and} & \mu_2 > \mu_0, \\ H^*_3 : \mu_1 > \mu_0 & \text{and} & \mu_2 \leq \mu_0 \end{array}$$

The three hypotheses are disjoint and hence testing each at α does not compromise the FWER. The final decision rule is constructed by considering all possible outcomes for the three mutually exclusive hypotheses. We can consider the following examples based on the outcomes

- If H^*_1 is rejected, we conclude that $\mu_1 > \mu_0$ or $\mu_2 > \mu_0$.
- We conclude that $\mu_1 > \mu_0$ if H^*_1 and H^*_2 are rejected.
- Similarly, a rejection of H^*_1 and H^*_3 will imply that $\mu_2 > \mu_0$.
- If H^*_1, H^*_2 and H^*_3 are all rejected, then the conclusion is that $\mu_1 > \mu_0$ and $\mu_2 > \mu_0$.

The test appears to be conceptually similar to the closed test procedure. The partition test, however, does not deal with the hypothesis in a closed family but rather with mutually exclusive hypotheses.

2.4 Simultaneous Confidence Intervals

Confidence intervals are clinically more appropriate than P-values or yes/no decisions and hence simultaneos confidence intervals are also proposed for several designs and aims in clinical trials. In this instance, 'simultaneous' implies multiplicity adjusted.

The choice of a one-sided or two-sided confidence intervals depends on the design of the study and the formulation of hypotheses. Two sided intervals for example are deemed appropriate and recommended when information on both the minimum and maximum likely is of interest per the design of further studies.

The lower confidence bound for the difference between the mean response of any nonzero-dose level and that of the control is of interest because its size may be useful in assessing the actual treatment effect between the largest dose and the control.

Let

$$\hat{\mu_i} = \overline{Y_i} = \sum_{a=1}^{n_i} Y_{ia}/n_i$$

Where $\bar{Y_i}$ follows from a one - way model defined by

$$Y_{ia} = \mu_i + \varepsilon_{ia}, \quad i = 1, \ldots, k, \quad a = 1, \ldots, n$$

Where Y_{ia} is the ath observation of the ith treatment and error term is iid normal with mean 0 and variance σ^2 unknown.

Also let the minimum effective dose be defined as

$$\text{MED} = \min \{i : \mu_i > \mu_1 + \delta\}$$

where MED is the minimum dose such that the mean response at that dose is clinically significantly better than the mean response of the negative controls and $\delta > 0$ defines a clinically significant difference.

The MED can be obtained using the simultaneous confidence interval approach in a stepwise manner. The stepwise confidence set is constructed under a one-way model and it takes the following form: [11]

Step 1.

If $\bar{Y_k} - \bar{Y_1} - t_{\alpha,\nu}\hat{\sigma}\sqrt{1/n_k + 1/n_1} \geq \delta$, assert $\mu_k > \mu_1 + \delta$ and go to step 2, otherwise claim that there is no nonzero-dose level which is significantly better than the control and $\mu_k - \mu_1 > \bar{Y_k} - \bar{Y_1} - t_{\alpha,\nu}\hat{\sigma}\sqrt{1/n_k + 1/n_1}$ and stop.

Step 2.

If $\bar{Y_{k-1}} - \bar{Y_1} - t_{\alpha,\nu}\hat{\sigma}\sqrt{1/n_{k-1} + 1/n_1} \geq \delta$, assert $\mu_{k-1} > \mu_1 + \delta$ and go to step 3, else assert $\mu_{k-1} - \mu_1 > \bar{Y_{k-1}} - \bar{Y_1} - t_{\alpha,\nu}\hat{\sigma}\sqrt{1/n_{k-1} + 1/n_1}$ and stop.

$\vdots$

Step k - 1.

If $\bar{Y_2} - \bar{Y_1} - t_{\alpha,\nu}\hat{\sigma}\sqrt{1/n_2 + 1/n_1} \geq \delta$, assert $\mu_2 > \mu_1 + \delta$ and go to step k, else assert $\mu_2 - \mu_1 > \bar{Y_2} - \bar{Y_1} - t_{\alpha,\nu}\hat{\sigma}\sqrt{1/n_2 + 1/n_1}$ and stop.

Step k.

Claim every dose level is significantly better than the control and stop.

To help in understanding the stepwise process and how it works, let step $M(1 \leq M \leq k)$ be the step at which the stepwise method stops. Then the stepwise method declares doses $k-M+2, \ldots, k$ to be efficacious if $M > 1$.

If $M < k$, then the stepwise method fails to declare doses $2, \ldots, k-M+1$ to be efficacious and gives a lower bound (which is less than δ) for $\mu_{k-M+1} - \mu_1$.

If $M = k$, then the stepwise method gives a lower bound on how every dose is. This lower bound is greater than δ.

3 Analysis and Findings

The three methods discussed above were employed in finding the minimum effective dose (MinEd) in this section. To illustrate these procedures, consider the data below obtained from Table 1 of Wöhr et al. (2005).

Table 1: Sample Dose Response Data

Dosage	Sample Size	Mean Response	SEM response
0.0 mA	7	8.89	3.96
0.2 mA	7	5.36	1.87
0.5 mA	7	32.01	6.29
0.8 mA	7	42.75	4.93
1.1 mA	5	48.06	3.55

The data comprises of five dosage levels including a zero-dose, the sample size for each dosage, its mean response and the standard error of the mean response. Four dose levels are compared with the zero-dose control.

A test for difference in means was first performed between each of the dose level and zero-dose and the following p-values associated with each dose were obtained as shown in table 2.

Table 2: P-Value Table

Dosage	P-Values
0.2 mA	2.195573e-01
0.5 mA	5.580436e-03
0.8 mA	1.655207e-04
1.1 mA	1.263558e-05

The same p-values are employed in both the Bonferroni-Holm and Hochberg methods in just a differrent stepwise approaches as follow.

3.1 Applying Bonferroni-Holm Procedure

The p-values obtained were reordered from minimum to maximum along with the hypothesis and dosage levels that correspond to the order of p-values. The first p-value that was tested was $1.26e-05$ which is associated with dosage level 1.1 mA. A critical value of 0.0125 was obtained for this step and since that is bigger than the p-value being tested,the test continues. This implies that the p-value for dosage level 1.1 mA is significant.

The hypothesis for the 0.8 mA dosage was next tested since it had the next smallest p-value of approximately $1.655e-04$ and it also produced a significant critical value of 0.0167 with the test. The stepwise procedure continued with testing the rest of the hypothesis in the ordered sequence. The 0.5 mA and 0.2 mA were next tested in that order and they produced critical values of 0.025 and 0.05 respectively.

In the testing sequence, the p-value $(2.196e-01)$ associated with dosage level 0.2 mA was the first found to be bigger than its corresponding Bonferroni-Holm critical value of 0.05. Although that is the last dose to be tested, it is also where the Bonferroni- Holm procedure stops since it had the first non significant p-value.

The combined results from the test from R is summarized in the table below showing the order in which test was conducted and the first nonsignificant critical value obtained labeled with stop in the interpretation column.

The test failed to reject the hypothesis that the 0.2 mA dosage was significantly different from the zero dose and hence this implies that the minimum efffective dose detremined by the Bonferroni-Holm test is the 0.5 mA dosage level.

Table 3: Summary of Bonferroni-Holm Test

Dosage	P-Values	critical values	Interpretation
1.1 mA	1.263558e-05	0.125	Continue
0.8 mA	1.655207e-04	0.0167	Continue
0.5 mA	5.580436e-03	0.025	Continue
0.2 mA	2.195573e-01	0.05	Stop

3.2 Applying the Hochberg Procedure

The p-values were reordered again from maximum to minimum in the order corresponding to the dosage levels as 0.2 mA, 0.5 mA, 0.8 mA and 1.1 mA to use in the Hochberg process. This test obtains its first true statement at the first significant p-value. All subsequent p-values together with this first significant p-value are rejected and the Hochberg Procedure identifies the dosage level associated with this first significant p-value to be the minimum effective dosage (MED).

The table below summarises the testing results for the Hochberg procedure. The first p-value found to be significantly smaller than its corresponding critical value was 0.0055804 which is associated with the dose level of 0.5 mA. According to the Hochberg procedure, this also indicates that the minimum effective dose is dose 0.5 mA.

Table 4: Summary of Hochberg Test

Dosage	P-Values	critical values	Interpretation
0.2 mA	2.195573e-01	0.05	Continue
0.5 mA	5.580436e-03	0.025	Stop
0.8 mA	1.655207e-04	0.0167	Stop
1.1 mA	1.263558e-05	0.125	Stop

3.3 The Simultaneous Confidence Interval

In this approach, $\delta = 10$, $\delta = 11$ and $\delta = 12$ were chosen as three different δ values to be used. The mean for each dosage level starting with the highest 1.1 mA down to the least 0.2 mA was tested with the mean for the placebo or zero dose for each of the three δ values at 95% lower confidence bounds.

Table 5 below summarises the sequence of testing for the simultaneous Confidence interval procedure, where μ_5 corresponds to dosage level 1.1 mA and μ_2 corresponds to the 0.2 mA dose in that order.

When $\delta = 10$,dosage levels 1.1 mA and 0.8 mA were both found to be effectively better than the zero dose and hence the test continued. At the 0.5 mA dose level, the test failed to clear its threshold and therefore by the Simultaneous Confidence interval procedure, the MinED will be that

Table 5: Simultaneous Confidence Interval Test

$\mu_i - \mu_1$	$\delta = 10$	$\delta = 11$	$\delta = 12$
$\mu_5 - \mu_1$	Continue	Continue	Continue
$\mu_4 - \mu_1$	Continue	Continue	Continue
$\mu_3 - \mu_1$	Stop	Stop	Stop
$\mu_2 - \mu_1$	Stop	Stop	Stop

of 0.8 mA since that was the dosage level whose mean was tested just before the 0.5 mA dose level which failed to clear its threshhold. The same results were observed for $\delta = 11$ and $\delta = 12$ as seen above.

4 Conclusions

Methods to help in making correct decisons when one has to use multiple testing has been studied and used widely as already discussed. In dose response studies, it is essentially critical to avoid any form of errors that may arise due to multiple testing.

This research considered both closed testing methods and partition principle methods that have been proven to reduce if not eliminate errors arising from multiple testing under the monotonicity assumption. The methods considered included the Bonferroni-Holm procedure, Hochberg procedure and the Simultaneous Confidence Interval approach.

The Bonferroni-Holm and the Hochberg procedures produced much about the same result since they are actually direct opposites with regards to the testing procedure. They both produced results that indicated that the minimum effective dose (MED) was ideally the 0.5mA.

The Simultaneous confidence produced a slightly different results from the other two. Despite the use of three different thresholds, they all determined the 0.8mA dosage level to be the MinED.

References

[1] Hothorn, L. A. (2007). How to deal with multiple treatment or dose groups in randomized clinical trials? Fundamental & Clinical Pharmacology, 21(2), 1472-8206. http://dx.doi.org/10.1111/j.1472-8206.2007.00469.x.

[2] Banken, L., Burger, H. U., Kristiansen, S., & Pasquier, M. The Closed Test Procedure.

[3] Shaffer, J. P. (1995). Multiple hypothesis testing. Annual Review of Psychology, 46, 561. Retrieved from http://search.proquest.com/docview/205849207?accountid=29141

[4] Tamhane, A. C., Hochberg, Y., & Dunnett, C. W.. (1996). Multiple Test Procedures for Dose Finding. Biometrics, 52(1), 21-37. http://doi.org/10.2307/2533141

[5] Jia, G. (2005) Use of Simultaneous Inference Under Order Restriction, Stepdown Testing Procedure and Stage-wise Sequential Optimal Design in Clinical Dose Study. Doctoral Dissertation, University of Pittsburgh.

[6] Holm, S. (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 6, 65-70

[7] Hochberg Y.(1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika;75:800-802

[8] Benjamini, Y. and Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B. 57: 289-300.

[9] Sarkar, S. (1998). Some probability inequalities for ordered MTP2 random variables: A proof of the Simes conjecture. Ann. Statist. 26, 494-504.

[10] Finner, H., and Strassburger, K. (2002). The partitioning principle: a powerful tool in multiple decision theory. Ann. Statist., 30, 1194-1213

[11] Peng, J., Lee, C.C., Davis, K. A., Wang, W. (2008). Stepwise Confidence Intervals for Monotone Dose-Response Studies. Biometrics 64,877-885.doi: 10.1111/j.1541-0420.2007.00958.x.

[12] Wöhr, M., Borta, A., and Schwarting, R. (2005). Overt behavior and ultrasonic vocalization in a fear conditioning paradigm: A dose-response study in the rat. Neurobiology ofLearning and Memory 84, 228-240.

[13] Dmitrienko, A., Bretz, F., Westfall, P. H., Troendle, J., Wiens, B.L., Tamhane, A. C., Hsu, J.(2010). Multiple Testing Methodology. Taylor and Francis Group,LLC.