

Pairwise comparisons in the analysis of carcinogenicity data*

Mohammad A. Rahman^{1#}, Ram C. Tiwari²

¹Division of Biometrics-6, Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, USA; [#]Corresponding Author: mohammad.rahman@fda.hhs.gov, HAtiarsemail@gmail.com

²Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, USA

Received 25 June 2012; revised 23 July 2012; accepted 6 August 2012

ABSTRACT

Analysis of carcinogenicity data generally involves a trend test across all dose groups and a pairwise comparison of the high dose group with the control. The most commonly used test for a positive trend is the Cochran-Armitage test. This test is asymptotically normal. For the pairwise comparison of the high dose group with the control group, we propose two modifications: the first modification is to apply the test on the data from high dose and control groups after dropping the data from the low and the medium dose groups; the second modification is to adjust the test conditional on data from all dose groups. We compare the power performance of these two modifications for the pairwise comparisons.

Keywords: Carcinogenicity Study; Trend Test; Pairwise Test; Exact Test

1. INTRODUCTION

The standard design for a long term carcinogenicity study of a new drug development in clinical research includes three treatment groups of increasing doses of the study drug (low, medium, and high) and one untreated control. The group sizes are about 50 animals per group. The statistical analyses include a trend test for positive dose response relationship in tumor incidence rates across all dose groups and pairwise comparisons of treated groups with the control group by organ/tumor combination.

The most common test for positive trend is the Cochran-Armitage [CA] test, see e.g. Cochran [1] and Armitage [2]. There are several extensions of the CA test see e.g. Tarone [3,4], Hoel and Yanagawa [5], and Tamura and Young [6] among others. Since difference in mortalities among treatment groups is a concern, there are vari-

ous mortality adjusted tests suggested by different authors see e.g. Peto *et al.* [7], Bailer and Portier [8]. Both of these mortality adjusted tests can be approached from CA test. The CA test is asymptotically normal. An exact test was proposed by Mehta *et al.* [9]. For the pairwise comparison of a treated group e.g. high dose group with the control group, both the asymptotic CA test and the exact test can be modified in two different ways. The first way is to drop the data from the low and medium dose groups and apply the trend tests to the remaining data from the high dose group and the control group. The second way is to modify the tests for pairwise comparison of the high dose group and the control group conditional on the data from all dose groups. We shall refer to these tests as unconditional pairwise test and conditional pairwise tests, respectively. The purpose of this work is to compare the power performance of these two modifications of pairwise tests.

It may be noted that a significant trend test may not necessarily indicate one of the pairwise tests to be statistically significant (see **Table 1**) and also a non-significant trend test may not necessarily indicate no pairwise test to be significant (see **Table 2**).

These tables show that it is important to check the pairwise tests after significant or non-significant trend test.

The rest of the paper is organized as follows. In Section 2, we review the CA and exact trend tests and present the modifications for the pairwise comparisons. In

Table 1. Asymptotic and exact *p*-values showing significant Trend with non-significant pairwise comparisons at $\alpha = 0.05$.

Group size	50	50	50	50	<i>p</i> -value	
Dose	0	1	2	3	Trend	Pair HC
#TBA	0	0	0	2	0.029 (Asymp) 0.062 (Exact)	0.078 (Asymp) 0.248 (Exact)

Group size	50	50	50	50	<i>p</i> -value	
Dose	0	1	2	3	Trend	Pair HC
#TBA	0	0	0	3	0.010 (Asymp) 0.015 (Exact)	0.040 (Asymp) 0.121 (Exact)

*Disclaimer: This article reflects the views of the authors and should not be construed to represent FDA's views or policies.

Table 2. Asymptotic and exact p -values showing non-significant trend with significant pairwise comparison at $\alpha = 0.05$.

Group size	50	50	50	50	p -Value	
Dose	0	1	2	3	Trend	Pair LC
#TBA	0	3	2	1	0.366 (Asymp) 0.440 (Exact)	0.038 (Asymp) 0.117 (Exact)

Group size	50	50	50	50	p -Value	
Dose	0	1	2	3	Trend	Pair HC
#TBA	0	5	2	1	0.326 (Asymp) 0.3846 (Exact)	0.011 (Asymp) 0.028 (Exact)

Section 3, we illustrate the application of the two modified pairwise tests on a dataset, and carry out a simulation study in Section 4 to evaluate their power performances. In Section 5, we make some concluding remarks.

2. THEORETICAL DEVELOPMENT

Consider a carcinogenicity study with $r + 1$ dose groups consisting of one control and r treated groups. Let n_i be the number of animals assigned to the i th treatment group, x_i be the number of tumor bearing animals observed in the i th treatment group, and d_i be the dose level for the i th treatment group, with $d_0 = 0$ for control group. Assume that x_i has a Binomial distribution as $x_i \sim \text{Bin}(n_i, p_i)$, where p_i is the probability of developing tumor by an animal in the i th dose group. The value of p_i is generally modeled as $p_i = H(a + bd_i)$ with $H(x) = \exp(x) / (1 + \exp(x))$, the logistic distribution. The value $p_0 = H(a)$ with $d_i = 0$ corresponds to the control group. Here, a is a nuisance parameter and b is the parameter of interest.

2.1. Test for Positive Trend

The positive trend is tested by the hypothesis $H_0 : b = 0$ versus the alternative hypothesis $H_1 : b > 0$, or equivalently by testing $H_0 : p_0 = p_1 = \dots = p_r = p^*$, vs. $H_1 : p_i \leq p_{i+1}$ for all i with strict inequality for at least one i . The value of p^* is the overall probability of developing tumor by an animal or overall proportion of tumor bearing animals in the population under null hypothesis. It can be easily shown that the pair $(n_+, \sum_{i=0}^r x_i d_i)$ is jointly sufficient, where $n_+ = \sum_{i=0}^r n_i$ is the total number of subjects on test in $(r + 1)$ groups. The CA test is based on the sufficient statistic.

The CA test for testing the null hypothesis that there is no trend, $H_0 : b = 0$, versus the alternative hypothesis, $H_1 : b > 0$, is given by

$$T_{\text{trend}} = \frac{\sum_{i=0}^r x_i (d_i - \bar{d})}{\sqrt{p^* (1 - p^*) \sum_{i=0}^r n_i (d_i - \bar{d})^2}} H_{2,0}^2$$

$$N(0, 1), \text{ where } \bar{d} = \frac{1}{n_+} \sum_{i=0}^r n_i d_i \text{ and } n_+ = \sum_{i=0}^r n_i.$$

In applications, we replace p^* by $\hat{p}^* = \frac{x_+}{n_+}$,

with $x_+ = \sum_{i=0}^r x_i$. The CA test is an asymptotically normal.

The exact test, as derived by Mehta *et al.*, is as follows: Let $\Gamma_{x_+} = \{x_- : x_- = (x_0, \dots, x_r), \sum_{i=0}^r x_i = x_+\}$ be the sample space which is the collection of all permutations of (x_0, \dots, x_r) such that $\sum_{i=0}^r x_i = x_+$, the observed total number of tumor bearing animals. Define the critical region for trend test:

$$C_{x_+}(t_{\text{trend}}) = \{x_- \in \Gamma_{x_+} : x_0 d_0 + x_1 d_1 + \dots + x_r d_r \geq t_{\text{obs,trend}}\},$$

where $t_{\text{obs,trend}} = x_0 d_0 + \dots + x_r d_r$ is the realization of the statistic $T_{\text{trend}} = \sum_{i=0}^r X_i d_i$, based on the observed data (x_0, \dots, x_r) . Using the hyper geometric distribution, the probability of each realization of

$$x_- = (x_0, \dots, x_r) \in \Gamma_{x_+} \text{ is } \frac{\prod_{i=0}^r \binom{n_i}{x_i}}{\binom{n_+}{x_+}}.$$

This probability is also known as the table probability, signifying the probability of each table in the all possible permutation of the observed number of tumor bearing animals. The exact p -value for testing H_0 (right hand tail) is then

$$P(T_{\text{trend}} > t_{\text{obs,trend}} | x_+, H_0) = \sum_{x_- \in C_{x_+}(t_{\text{trend}})} \frac{\prod_{i=0}^r \binom{n_i}{x_i}}{\binom{n_+}{x_+}}$$

2.2. Pairwise Comparisons

Since the highest dose for a regulatory carcinogenicity study is selected mostly based on the maximum tolerated dose (MTD) criterion, the pairwise comparisons between the high dose group with the control group has special regulatory interest. In this paper we present some results related to pairwise comparisons of high dose group with the control group. The results, however, can be used for the pairwise comparison of any treatment group with control. If we were interested in testing simultaneous multiple contrasts, such as Williams type contrast, the approach described in Hothorn *et al.* [10] can be used. These methods are based on the quantiles of multivariate normal distribution taking care of the correlation into account as the package MVTNORM.

For pairwise comparison of the highest dose Group r

with control with the null hypothesis $H_0^* : p_r - p_0 = 0$, and the alternative hypothesis $H_1^* : p_r - p_0 > 0$, we describe the following two approaches. First note that $\hat{p}_r - \hat{p}_0 = \hat{b}(d_r - d_0)$, and

$$\begin{aligned} & V(\hat{p}_r - \hat{p}_0) \\ &= V(\hat{b})(d_r - d_0)^2 \\ &=_{H_0} \frac{p^*(1-p^*)}{\sum_{i=0}^r n_i (d_i - \bar{d})^2} (d_r - d_0)^2. \end{aligned}$$

In our first approach, we delete the data from all dose groups except data from dose groups 0 and r , estimate overall proportion of tumor bearing animals as

$$\hat{p}_{0,r}^* = \frac{\sum_{i=0,r} x_i}{\sum_{i=0,r} n_i} \text{ and define the test statistic as}$$

$$\begin{aligned} T_{\text{pairwise1}} &= \frac{\hat{p}_r - \hat{p}_0}{\sqrt{V_{H_0}(\hat{p}_r - \hat{p}_0)}} \\ &= \frac{\hat{p}_r - \hat{p}_0}{\sqrt{\frac{\hat{p}_{0,r}^*(1-\hat{p}_{0,r}^*)}{\sum_{i=0,r} n_i (d_i - \bar{d})^2} (d_r - \bar{d})^2}} \end{aligned} \tag{1}$$

As mentioned, in the derivation of the above test, the variance of $\hat{p}_r - \hat{p}_0$ is estimated based on the data from Group r and control only. We will refer to this test as the

$$\begin{aligned} ARE(T_{\text{pairwise2}}, T_{\text{pairwise1}}) &= \frac{Var_{H_0^*}(T_{\text{pairwise1}})}{Var_{H_0^*}(T_{\text{pairwise2}})} = \frac{n_0(d_0 - \bar{d})^2 + n_1(d_1 - \bar{d})^2 + \dots + n_{r-1}(d_{r-1} - \bar{d})^2 + n_r(d_r - \bar{d})^2}{n_0(d_0 - \bar{d})^2 + n_r(d_r - \bar{d})^2} \\ &= 1 + \frac{n_1(d_1 - \bar{d})^2 + \dots + n_{r-1}(d_{r-1} - \bar{d})^2}{n_0(d_0 - \bar{d})^2 + n_3(d_3 - \bar{d})^2} > 1 \end{aligned}$$

showing that $T_{\text{pairwise2}}$ is asymptotically more efficient than $T_{\text{pairwise1}}$.

Hothorn and Bretz [11] proposed (asymptotic) tests for positive trend based on single and multiple contrasts under the assumption of equally spaced dose-levels. For single contrast, test is defined as

$$T_{HB} = \frac{\sum_{i=0}^r \hat{p}_i c_i}{\sqrt{p_{0,r}^*(1-p_{0,r}^*) \sum_{i=0}^r \frac{c_i^2}{n_i}}}, \text{ where } \sum_{i=0}^r c_i = 0.$$

For pairwise comparison of Group r and control ($r = 0$) with $c_0 = -1$ and $c_r = 1$, this test statistic is

unconditional test. However, under the null hypothesis of no dose effect, a better estimate of variance of $\hat{p}_r - \hat{p}_0$ can be obtained from the complete data *i.e.*

$$V_{H_0}(\hat{p}_r - \hat{p}_0) = \frac{p^*(1-p^*)}{\sum_{i=0}^r n_i (d_i - \bar{d})^2} (d_r - d_0)^2,$$

where now p^* is estimated as $\hat{p}_{\text{all}}^* = \frac{\sum_{i=0}^r x_i}{\sum_{i=0}^r n_i}$ based on data

from all dose groups. Using this estimate in the denominator, our second approach is to define the test statistic as

$$T_{\text{pairwise2}} = \frac{\hat{p}_r - \hat{p}_0}{\sqrt{\frac{\hat{p}_{\text{all}}^*(1-\hat{p}_{\text{all}}^*)}{\sum_{i=0}^r n_i (d_i - \bar{d})^2} (d_r - d_0)^2}} \tag{2}$$

We will refer to this test as the conditional test.

It should be noted that under the linearity assumption of p_i with d_i (the denominator of) the above test is same as the Cochran-Armitage trend test.

2.3. Asymptotic Relative Efficiency of the Conditional and the Unconditional Pairwise Tests

The asymptotic relative efficiency (ARE) of $T_{\text{pairwise2}}$ and $T_{\text{pairwise1}}$ is

$$T_{\text{pairwiseHB}} = \frac{\hat{p}_r - \hat{p}_0}{\sqrt{p_{0,r}^*(1-p_{0,r}^*) \left(\frac{1}{n_r} + \frac{1}{n_0} \right)}} \tag{3}$$

The $p_{0,r}^*$ is estimated by $\hat{p}_{0,r}^*$, as defined earlier. If the group sizes are equal (*i.e.* if $n_0 = n_1 = n_2 = n_3$) then it can be shown that the statistics $T_{\text{pairwise1}}$ and $T_{\text{pairwiseHB}}$ are identical.

2.4. Exact Pairwise Test

We now consider the derivation of the exact pairwise tests. Following Mehta *et al.*, the exact p -value for unconditional exact test $T_{\text{exact,pairwise1}}$ based on the data from the Group r and Control, for testing H_0^* , is cal-

culated as

$$p\left(T_{\text{exact,pairwise1}} > t_{\text{obs,pairwise1}} \mid x_r + x_0, H_0^*\right) = \sum_{x \in C_{x_j+x_j} (t_{\text{pairwise1}})} \frac{\binom{n_r}{x_r} \binom{n_0}{x_0}}{\binom{n_r+n_0}{x_r+x_0}}$$

The exact test for our second approach is as follows. As before, let $\Gamma_{x_+} = \left\{ \underline{x} : \underline{x} = (x_0, \dots, x_r), \sum_{i=0} x_i = x_+ \right\}$, and define the critical region

$$C_{x_+} (t_{\text{exact,pairwise2}}) = \left\{ \underline{x} \in \Gamma_{x_+} : x_0 d_0 + x_r d_r \geq t_{\text{obs,pairwise2}} \right\},$$

where $t_{\text{obs,pairwise2}} = x_0 d_0 + x_r d_r$ is the realization of $T_{\text{Exact,Pairwise2}} = x_0 d_0 + x_r d_r$ based on the observed data (x_0, x_r) . The exact p -value for testing H_0^* is calculated as

$$p\left(T_{\text{exact,pairwise2}} \geq t_{\text{obs,pairwise2}} \mid x_+, H_0^*\right) = \sum_{x \in C_{x_+} (t_{\text{pairwise2}})} \frac{\prod_{l=0}^r \binom{n_l}{x_l}}{\binom{n_+}{x_+}}$$

We will refer to this test as the conditional exact pairwise test.

Proceeding along the lines of Mehta *et al.*, the power of pairwise test is calculated as follows. Let $t_\alpha(x_+)$ be such that $P(T_{\text{exact,pairwise2}} > t_\alpha(x_+) \mid x_+, H_0^*) \leq \alpha$. Then the power of the pairwise test conditional on x_+ is

$$\begin{aligned} \beta(x_+) &= p\left(T_{\text{exact,pairwise2}} > t_{\text{obs,pairwise2}} \mid x_+, H_1^*\right) \\ &= \sum_{x \in \Gamma_{x_+} (t_\alpha(x_+))} \frac{\prod_{l=0}^r \binom{n_l}{x_l} p_l^{x_l} (1-p_l)^{n_l-x_l}}{\sum_{x \in \Gamma_{x_+}} \prod_{l=0}^r \binom{n_l}{x_l} p_l^{x_l} (1-p_l)^{n_l-x_l}} \\ &= \sum_{x \in \Gamma_{x_+} (t_\alpha(x_+))} \frac{\prod_{l=0}^r \binom{n_l}{x_l} p_l^{x_l} (1-p_l)^{n_l-x_l}}{p(x_+)} \\ &= p\left(X_0 + \dots + X_r = x_+ \mid H_1^*\right), \end{aligned}$$

where $p(x_+) = \sum_{x \in \Gamma_{x_+}} \prod_{l=0}^r \binom{n_l}{x_l} p_l^{x_l} (1-p_l)^{n_l-x_l}$.

The above power can be evaluated for exact test using the hyper geometric distribution and appropriate critical regions under conditional and unconditional situations.

We compare the relative power of three asymptotic pairwise tests, as well as that of the two exact tests $T_{\text{Exact,pairwise1}}$ and $T_{\text{Exact,pairwise2}}$. For evaluation of their power functions, we performed simulations and calculated the percentage of times the null hypothesis was rejected when the alternative hypothesis was true. The SAS proc Stratify and SAS proc Multtest [12] are very convenient

for the calculation of these exact probabilities.

3. EXAMPLE

Consider a carcinogenicity study with four treatment groups namely, control, low, medium, and high dose groups each with 50 animals, and dose scores 0, 1, 2, and 3, respectively. Suppose we observe a total of 10 animals developed a certain tumor type with 0, 2, 3 and 5 tumor bearing animals in control, low, medium, and high dose groups, respectively. We would like to perform a pairwise comparison of the high dose group with the control. The null hypothesis $H_0 : p_3 - p_0 = 0$ against alternative $H_1 : p_3 - p_0 > 0$. The results using the normal approximation test are $T_{\text{pairwise1}} = 2.294$, $T_{\text{pairwise2}} = 2.418$, and $T_{\text{pairwiseHB}} = 2.294$ with corresponding p -values as 0.0109, 0.0078, and 0.0109, respectively.

For exact test we have $t_{\text{obs,pairwise1}} = t_{\text{obs,pairwise2}} = x_0 d_0 + x_3 d_3 = 15$. **Table 3**, given below, shows all possible values of $T_{\text{pairwise1}}$ along with their table probabilities and the right tail probabilities for pairwise comparison of high dose with control calculated from data after dropping low and medium dose groups using SAS proc Stratify.

Table 4 given below shows all possible values of $T_{\text{pairwise2}}$ along with their table probabilities and the right tail probabilities for pairwise comparison of high dose with control calculated from all data using the scores 0, 0, 0 and 3 in SAS proc Multtest.

The results from **Tables 3** and **4** show that both the table- and right-tail probabilities for the two pairwise exact

Table 3. Pair comparison of control with high dose group after deleting the low and medium dose groups.

$T_{\text{pairwise1}}$	Table	Rt. Hand
	Probability	Tail
0	0.02814	1.00000
3	0.15295	0.97186
6	0.31891	0.81891
9	0.31891	0.50000
12	0.15295	0.18109
15	0.02814	0.02814 (observed $p < 0.05$)

Table 4. Pairwise comparison of high dose with control using all data.

$T_{\text{pairwise2}}$	Table	Rt. Hand
	Probability	Tail
0	0.05209	1.00000
3	0.18473	0.94791
6	0.28685	0.76318
9	0.25676	0.47633
12	0.14666	0.21957
15	0.05583	0.07291 (observed $p > 0.05$)
18	0.01434	0.01708
21	0.00245	0.00274
24	0.00027	0.00028
27	0.00002	0.00002
30	0.00000	0.00000

tests may go in either direction. For example for the observed number of 0, 2, 3 and 5 tumor bearing animals, we have $t_{obs, pairwise1} = t_{obs, pairwise2} = x_0d_0 + x_3d_3 = 15$, and the p -value after deleting the low and medium dose groups is $p_{pairwise1} = 0.0281$, and that using data from all dose groups is $p_{pairwise2} = 0.0729$ i.e. the p -value after deleting the low and medium dose groups is smaller than the p -value using data from all dose groups. On the other hand if the observes number of tumor bearing animals were 2, 2, 3, and 3, then $t_{obs, pairwise1} = t_{obs, pairwise2} = x_0d_0 + x_3d_3 = 9$. The p -value after deleting the low and medium dose groups would be $p_{pairwise1} = 0.5$, and that using data from all dose groups would be $p_{pairwise2} = 0.4763$. In this case the p -value for pairwise exact test after deleting the low and medium dose groups would be larger than the p -value for the pairwise exact test using the data from all dose groups.

4. SIMULATION STUDY OF POWER CALCULATION

Consider a carcinogenicity study with four treatment groups namely, control, low, medium, and high dose groups each with 50 animals, and dose scores 0, 1, 2, and 3, respectively. The power was calculated for different choices background incidence rate in the control group (p_0). The incidence rate for the high dose group (p_3) was then chosen by a certain increment (δ) over p_0 . The incidence rate for the low dose group (p_1) and that for medium dose group (p_2) were calculated using a logistic model as follows:

$$\text{If } a = \log(p_0/(1-p_0))$$

and

$$b = \frac{\log(p_3/(1-p_3)) - \log(p_0/(1-p_0))}{d_3},$$

then

$$p_1 = \frac{e^{(a+bd_1)}}{1 + e^{(a+bd_1)}} \text{ and } p_2 = \frac{e^{(a+bd_2)}}{1 + e^{(a+bd_2)}},$$

with $d_i = i$ and $i = 0, 1, 2, \text{ and } 3$. The values of the power were calculated by finding the percentages of times the null hypothesis was rejected when the alternative was true in a simulation with 1000 loops. **Table 5** shows the calculated power using the asymptotic normal approximation and **Figure 1** gives the graphical representation of the results.

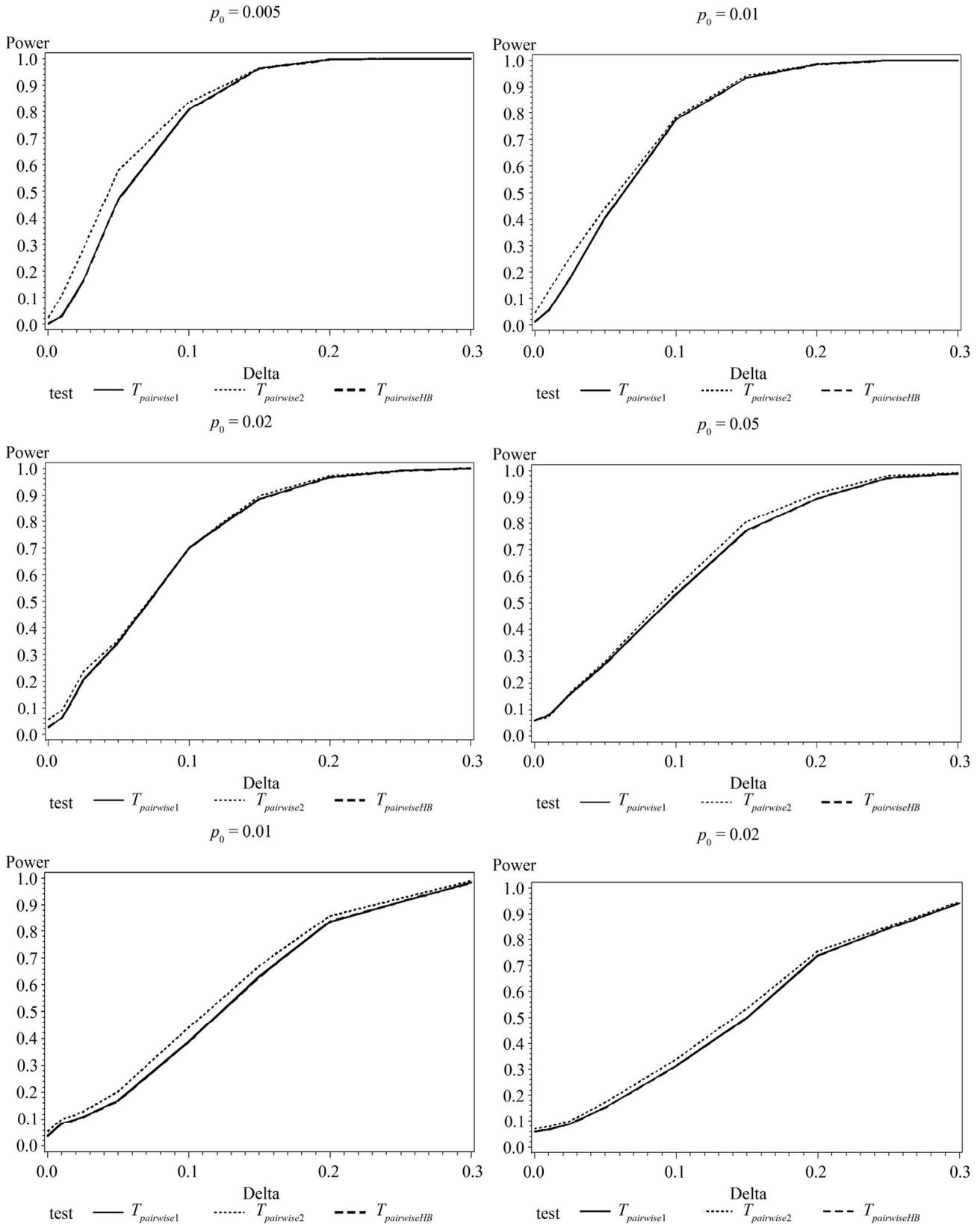
Table 6 shows the calculated power using the exact test and **Figure 2** gives the graphical representation of the results.

The simulation results show that asymptotic normal test $T_{pairwise2}$ is always a more powerful compared to $T_{pairwise1}$ or $T_{pairwiseHB}$. The two tests $T_{pairwise1}$ and $T_{pairwiseHB}$ have similar power (as sample sizes are taken to be same). The

pairwise exact test using data from all dose groups has more power compared to test based on data deleting the two middle dose groups for small values of p_0 and δ .

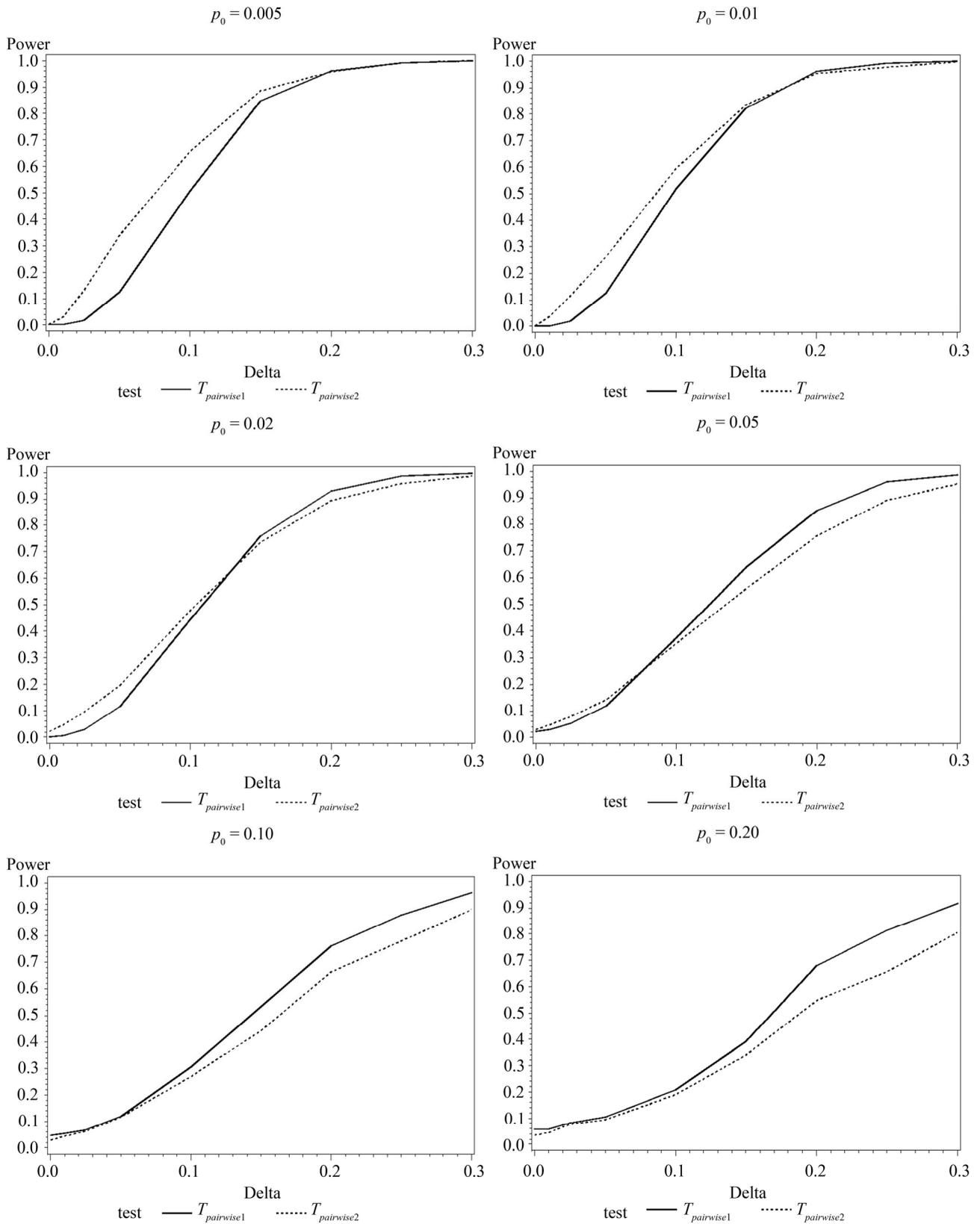
Table 5. Power calculated using the normal approximation for unconditional, conditional, and Hothorn Bretz tests.

Tumor rates in control group	Delta = $p_3 - p_0$	Power of unconditional test	Power of conditional test	Power of HB test
0.005	0	0.0010	0.0220	0.0010
	0.01	0.0310	0.1110	0.0310
	0.025	0.1640	0.2820	0.1640
	0.05	0.4710	0.5790	0.4710
	0.1	0.8070	0.8370	0.8070
	0.15	0.9630	0.9640	0.9630
	0.2	0.9950	0.9960	0.9950
	0.25	1.0000	0.9990	1.0000
	0.3	1.0000	1.0000	1.0000
	0	0.0110	0.0440	0.0110
	0.01	0.0560	0.1310	0.0560
	0.025	0.1790	0.2580	0.1790
	0.05	0.4070	0.4440	0.4070
	0.1	0.7740	0.7820	0.7740
	0.15	0.9340	0.9420	0.9340
0.01	0.2	0.9840	0.9860	0.9840
	0.25	0.9990	1.0000	0.9990
	0.3	1.0000	1.0000	1.0000
	0	0.0260	0.0520	0.0260
	0.01	0.0620	0.0930	0.0620
	0.025	0.2050	0.2350	0.2050
	0.05	0.3460	0.3560	0.3460
	0.1	0.7000	0.7000	0.7000
	0.15	0.8850	0.8960	0.8850
	0.2	0.9650	0.9720	0.9650
	0.25	0.9900	0.9920	0.9900
	0.3	1.0000	1.0000	1.0000
	0	0.0580	0.0580	0.0580
	0.01	0.0790	0.0740	0.0790
	0.025	0.1580	0.1630	0.1580
0.05	0.2710	0.2800	0.2710	
0.05	0.1	0.5330	0.5560	0.5330
	0.15	0.7710	0.8050	0.7710
	0.2	0.8940	0.9120	0.8940
	0.25	0.9710	0.9810	0.9710
	0.3	0.9890	0.9910	0.9890
	0	0.0370	0.0530	0.0370
	0.01	0.0830	0.0990	0.0830
	0.025	0.1080	0.1270	0.1080
	0.05	0.1690	0.2020	0.1690
	0.1	0.3870	0.4410	0.3870
	0.15	0.6300	0.6700	0.6300
	0.2	0.8370	0.8570	0.8370
	0.25	0.9110	0.9250	0.9110
	0.3	0.9820	0.9870	0.9820
	0	0.0600	0.0690	0.0600
0.01	0.0670	0.0770	0.0670	
0.025	0.0890	0.1010	0.0890	
0.05	0.1530	0.1720	0.1530	
0.2	0.1	0.3120	0.3390	0.3120
	0.15	0.4990	0.5340	0.4990
	0.2	0.7380	0.7570	0.7380
	0.25	0.8440	0.8530	0.8440
	0.3	0.9420	0.9470	0.9420



Where -----: power of unconditional test and: power of conditional test.

Figure 1. Graphical representation of power vs. delta for given p_0 using normal approximation for pairwise comparison of control and high dose group.



where -----: Power of unconditional test and: Power of conditional test.

Figure 2. Graphical representation of power vs. delta for given p_0 using exact test for pairwise comparison of control and high dose group.

Table 6. Power calculated using the exact test for unconditional and conditional tests.

Tumor rates in control group	Delta = $p_3 - p_0$	Power of unconditional test	Power of conditional test
0.005	0	0.0000	0.0013
	0.01	0.0007	0.0300
	0.025	0.0158	0.1292
	0.05	0.1249	0.3387
	0.1	0.5075	0.6573
	0.15	0.8482	0.8871
	0.2	0.9610	0.9570
	0.25	0.9920	0.9910
0.01	0.3	0.9997	0.9989
	0	0.0000	0.0000
	0.01	0.0025	0.0383
	0.025	0.0185	0.1148
	0.05	0.1239	0.2627
	0.1	0.5165	0.5944
	0.15	0.8212	0.8332
	0.2	0.9610	0.9530
0.02	0.25	0.9910	0.9750
	0.3	0.9988	0.9957
	0	0.0020	0.0214
	0.01	0.0073	0.0467
	0.025	0.0294	0.0959
	0.05	0.1179	0.1978
	0.1	0.4446	0.4795
	0.15	0.7592	0.7363
0.05	0.2	0.9292	0.8944
	0.25	0.9860	0.9580
	0.3	0.9972	0.9859
	0	0.0221	0.0291
	0.01	0.0308	0.0487
	0.025	0.0519	0.0789
	0.05	0.1199	0.1439
	0.1	0.3746	0.3556
0.1	0.15	0.6414	0.5614
	0.2	0.8511	0.7572
	0.25	0.9600	0.8901
	0.3	0.9859	0.9518
	0	0.0473	0.0308
	0.01	0.0542	0.0457
	0.025	0.0679	0.0629
	0.05	0.1209	0.1169
0.2	0.1	0.3050	0.2683
	0.15	0.5315	0.4426
	0.2	0.7622	0.6633
	0.25	0.8791	0.7812
	0.3	0.9628	0.8985
	0	0.0584	0.0349
	0.01	0.0590	0.0443
	0.025	0.0809	0.0749
0.2	0.05	0.1039	0.0929
	0.1	0.2068	0.1898
	0.15	0.3896	0.3397
	0.2	0.6803	0.5475
	0.25	0.8132	0.6573
	0.3	0.9170	0.8063

When p_0 and/or δ become (s) large the pairwise exact test that deletes the data from two middle dose groups showed better power.

5. CONCLUSIONS

In this paper, we discussed the topic of pairwise comparison of the high dose group with control in a typical carcinogenicity study. We proposed two tests procedure, one based on data only from the two dose groups to be compared and one based on data from all dose groups. We elaborated both exact and normal approximation version of our proposed tests.

Through a simulation, we compared the power performances of these tests. For the comparison of high dose group with control group in a typical four dose group carcinogenicity study, the simulation results showed that the power of the asymptotic normal test using data from all dose groups is asymptotically more efficient and hence is always more powerful than that of the test using data from high and control groups only. For exact test, neither of the two tests showed uniformly better power than the other. The pairwise exact test using data from all dose groups showed more power than that of the test based on data deleting the two middle dose groups for tumor types with low background rate and/or drug with small carcinogenic effect, while the pairwise exact test using data from all dose groups showed less power than that of the test based on data deleting the two middle dose groups for tumor types with high background rate and/or drug with large carcinogenic effect. However, since a test that drops part of the data is asymptotically less efficient, we recommend that for the pairwise comparison one uses tests that use the data from all dose groups.

6. ACKNOWLEDGEMENTS

The authors are deeply indebted to Drs. Stella G. Machado and Karl K. Lin, Division of Biometrics-6, US Food and Drug Administration, for their helpful advices and comments to improve and complete this work.

REFERENCES

- [1] Cochran, W.G. (1954) Some methods of strengthening the common χ^2 tests. *Biometrics*, **10**, 417-451. [doi:10.2307/3001616](https://doi.org/10.2307/3001616)
- [2] Armitage, P. (1955) Tests for linear trend in proportions and frequencies. *Biometrics*, **11**, 375-386. [doi:10.2307/3001775](https://doi.org/10.2307/3001775)
- [3] Tarone, R.E. (1975) Test for trend in life table analysis. *Biometrika*, **62**, 679-690. [doi:10.1093/biomet/62.3.679](https://doi.org/10.1093/biomet/62.3.679)
- [4] Tarone, R.E. (1982) The use of historical control information in testing for a trend in Poisson means. *Biometrics*, **38**, 457-462. [doi:10.2307/2530459](https://doi.org/10.2307/2530459)
- [5] Hoel, D.G. and Yanagawa, T. (1986) Incorporating historical controls in testing for a trend in proportions. *Journal of the American Statistical Association*, **81**, 1095-1099. [doi:10.1080/01621459.1986.10478379](https://doi.org/10.1080/01621459.1986.10478379)
- [6] Tamura, R.N. and Young, S.S. (1986) The incorporation

- of historical control information in tests and proportions: Simulation study of Tarone's Procedure. *Biometrics*, **42**, 343-349. [doi:10.2307/2531054](https://doi.org/10.2307/2531054)
- [7] Peto, R., Pike, M.C., Day, N.E., Gray, R.G.K, Lee, P.N. Parish, S., Peto, J., Richards, S. and Wahrendorf, J. (1980) Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Suppl. 2, Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal*, IARC, Lyon, 311-426.
- [8] Bailer, A.J. and Portier, C.J. (1988) Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*, **44**, 417-431. [doi:10.2307/2531856](https://doi.org/10.2307/2531856)
- [9] Mehta, C.R. and Patel, N.R. and Senchaudhuri, P. (1998) Exact power and sample-size computations for the Cochran-Armitage trend test. *Biometrics*, **54**, 1615-1621. [doi:10.2307/2533685](https://doi.org/10.2307/2533685)
- [10] Hothorn, L.A., Sill, M. and Schaarschmidt, F. (2010) Evaluation of incidence rates in pre-clinical studies using a Williams-type procedure. *The International Journal of Biostatistics*, **6**, 1557-4679. [doi:10.2202/1557-4679.1180](https://doi.org/10.2202/1557-4679.1180)
- [11] Hothorn, L.A. and Bretz, F. (2000) Evaluation of animal carcinogenicity studies: Cochran-armitage trend test vs. multiple contrast tests. *Biometrical Journal*, **42**, 553-567. [doi:10.1002/1521-4036\(200009\)42:5<553::AID-BIMJ553>3.0.CO;2-R](https://doi.org/10.1002/1521-4036(200009)42:5<553::AID-BIMJ553>3.0.CO;2-R)
- [12] SAS Institute Inc., (2009) Multtest procedure. SAS user's guide 9.2. 2nd Edition, SAS Institute Inc., Cary, 4176-4242.