

Controlling False Discoveries in Multidimensional Directional  
Decisions, with Applications to Gene Expression Data on  
Ordered Categories

Wenge Guo

Biostatistics Branch

National Institute of Environmental Health Sciences

Research Triangle Park, NC 27709, U.S.A.

wenge.guo@gmail.com

Sanat K. Sarkar

Department of Statistics, Temple University

Philadelphia, PA 19122, U.S.A.

sanat@temple.edu

Shyamal D. Peddada

Biostatistics Branch

National Institute of Environmental Health Sciences

Research Triangle Park, NC 27709, U.S.A.

peddada@niehs.nih.gov

June 19, 2008

**Abstract**

Microarray gene expression studies over ordered categories are routinely conducted to gain insights into biological functions of genes and the underlying biological processes. Some common experiments are time-course/dose-response experiments where a tissue or cell-line

is exposed for different doses and/or durations of time to a chemical. A goal of such studies is to identify gene expression patterns/profiles over the ordered categories. This problem can be formulated as a multiple testing problem where for each gene the null hypothesis of no difference between the successive mean gene expressions are tested and further directional decisions are made if it is rejected. Much of the existing multiple testing procedures are devised for controlling the usual false discovery rate (FDR) rather than the mixed directional FDR, the expected proportion of Type I and directional errors among all rejections. Benjamini and Yekutieli (2005) proved that an augmentation of the usual Benjamini-Hochberg (BH) procedure can control the mixed directional FDR while testing simple null hypotheses against two-sided alternatives in terms of one dimensional parameters. In this article, we consider the problem of controlling the mixed directional FDR involving multidimensional parameters. To deal with this problem, we develop a procedure extending that of Benjamini and Yekutieli based on the Bonferroni test for each gene. A proof is given for its mixed directional FDR control when the underlying test statistics are independent across the genes. The results of a simulation study evaluating its performance under independence as well as under dependence of the underlying test statistics across the genes relative to other relevant procedures are reported. Finally, the proposed methodology is applied to a time-course microarray data obtained by Lobenhofer et al. (2002). We identified several important cell-cycle genes, such as DNA replication/repair gene MCM4 and replication factor subunit C2, which were not identified by the previous analyses of the same data by Lobenhofer et al. (2002) and Peddada et al. (2003). Although some of our findings overlap with previous findings, we identify several other genes that compliment the results of Lobenhofer et al. (2002).

*Key words and phrases:* Benjamini-Hochberg procedure; Directional FDR; Dose-response; Microarray; Multiple testing; Ordered categories; Time-course

## 1 Introduction

In many applications researchers are interested in identifying trends in mean response over ordered categories in large scale experiments. With the advent of microarray technology such experiments are common in the literature where investigators are routinely conducting experiments to investigate changes in mean gene expressions over time or dose of a chemical or cancer

stage, etc. For example, Lobenhofer et al. (2002) studied the effect of  $17 - \beta$  estradiol on the gene expression of MCF-7 breast cancer cells as the cells progressed through various phases of cell division cycle. In another experiment, Tamoto et al. (2004) investigated the changes in gene expression with tumor progression in esophageal cancer and identified genes implicated in the early stages of esophageal squamous cell carcinoma. Recently, Bochkina and Richardson (2007) discussed the analysis of a time-course gene expression data where cells from the H2Kb muscle cell line of mouse were treated by insulin (0, 2 or 12 hours of exposure).

In studies such as those described above, identification of statistically significant genes that have similar mean expression profiles over ordered categories is often an important goal to researchers. By identifying such genes, the researchers may potentially discover co-regulated genes belonging to similar pathways and gain insights into biological functions and processes of groups of genes with similar patterns of expressions.

Peddada et al. (2003) introduced an order restricted inference based method for identifying significant genes and group them according to various patterns of inequalities. Implicitly in their methodology, two decisions are being taken for each gene. First, it is decided whether or not a gene is significant using a method exercising a control over gene specific Type I error rate. Then, a suitable inequality pattern is assigned for each selected significant gene based on the values of the underlying test statistics. The Type III or directional error that can potentially occur in addition to the usual Type I error due to assigning wrong inequality pattern to a selected significant gene has not been addressed in that paper. In this paper, we take care of both the Type I and directional errors. We do that by taking a multiple testing approach to the main problem where for each gene the mean expressions are successively compared across the ordered categories and a null hypothesis signifying no particular directional pattern is formed to test against the union of all possible directional patterns. We develop a method of simultaneously testing these null hypotheses and determining a directional pattern upon rejection of each of them controlling both the Type I and directional errors in an overall sense.

Although directional error has not been discussed extensively in the literature, it is perhaps a common error that occurs in applications. While testing a null hypothesis  $H_0 : \theta = 0$  against the two-sided alternative  $H_1 : \theta \neq 0$ , for some single parameter  $\theta$  of interest, researchers commonly

conclude either  $\theta > 0$  or  $\theta < 0$  upon rejection of  $H_0$  depending on the sign of the underlying test statistic. This problem is greatly exacerbated when multiple hypotheses are tested and the number of parameters describing directional patterns is larger, even moderately, than one, as is the case with time course or dose-response microarray data.

A traditional approach to dealing with directional as well as Type I errors from a multiple testing point of view is to apply a method that controls the so called mixed directional familywise error rate (mdFWER), which is the probability of one or more Type I or Type III errors, a variant of the classical familywise error rate (FWER) (Finner, 1994, 1999; Liu, 1996; Sarkar, Sen and Finner, 2004 and Shaffer, 1980). However, when the number of null hypotheses is large, as in the context of microarray experiments, the notion of mdFWER, just like the FWER, is too stringent, allowing little chance to make true directional as well as non-directional discoveries. The FDR (False Discovery Rate), due to Benjamini and Hochberg (1995), is a more powerful concept of overall Type I error rate than the FWER in the context of multiple testing and is now most commonly used in large scale scientific investigations, especially in microarray gene expression studies. A variant of it while controlling both Type I and directional errors would be more powerful than the mdFWER. Two such variants have been introduced in the literature (Benjamini, Hochberg and Kling, 1993), the pure directional FDR which is the expected proportion of Type III errors among rejected hypotheses and the mixed directional FDR (mdFDR) which is the expected proportion of Type I and III errors among rejected hypotheses. In this article, we will focus on procedures controlling the mdFDR.

Benjamini and Yekutieli (2005) gave a method with independent tests that controls the mdFDR when testing multiple simple hypotheses against two-sided alternatives. They proved that the original Benjamini and Hochberg [BH (1995)] procedure controlling the FDR at  $\alpha$  can be augmented to make directional decision upon rejecting a null hypothesis according to the value of the corresponding test statistic without causing the mdFDR to exceed  $\alpha$ , a result conjectured before by several authors (Benjamini and Hochberg, 2000; Shaffer, 2002 and William et al., 1999). Clearly, this method, referred to as the directional BH procedure, can be applied to analyze dose-response microarray data if there are only two ordered categories, but often this is not the case, as such data typically involve more than two ordered categories and the method

needs to be suitably extended to accommodate such multiple categories.

We extend the BH directional FDR procedure in this article to develop our proposed multiple testing method that allows us to make a decision on the directional pattern involving multiple parameters once a null hypothesis of no pattern is rejected and maintains a control over the mdFDR. The proposed methodology is then evaluated through a simulation study and applied to the time-course microarray data in Lobenhofer et al. (2002). Our analysis of Lobenhofer’s data resulted in the discovery of several cell-cycle genes that were not previously identified by Lobenhofer et al. (2002) and Peddada et al. (2003). Some of our findings complement the previous findings as detailed in Section 5. An important and unique feature our methodology is that it permits us to specify the time interval of up (or down) regulation of a gene during the 48 hour period of the cell-cycle. One of the usual objectives for conducting cell-cycle time course experiments is to determine the phase of peak expression for a cell-cycle gene and our methodology allows us to make such determinations.

## 2 Notations, Definitions and Problem Formulation

In this section, we present the multiple testing formulation of the problem of identifying expression patterns/trends over ordered categories simultaneously for all the genes, having introduced some notations and definitions related to multiple testing.

Let  $\mu_{ij}$  denote the mean response of the  $j^{th}$  variable (e.g. gene),  $j = 1, \dots, m$ , in the  $i^{th}$  ordered category,  $i = 1, \dots, p$ . A problem of biological interest is to group genes by the inequalities among the mean responses, known as directional patterns or order restrictions. Some common inequalities of interest are  $\mu_{1j} \leq \mu_{2j} \leq \dots \leq \mu_{pj}$  (monotone pattern),  $\mu_{1j} \leq \mu_{2j} \leq \dots \leq \mu_{ij} \geq \mu_{(i+1)j} \geq \dots \geq \mu_{pj}$ ,  $i = 2, \dots, p - 1$  (umbrella order with peak  $\mu_{ij}$ ). Let  $\delta_{ij} = \mu_{i+1j} - \mu_{ij}$ ,  $i = 1, \dots, p - 1$ ,  $j = 1, \dots, m$ . Then, the above inequalities of interest or any other inequalities can be restated in terms of the signs of the  $\delta_{ij}$ ’s. Let  $\boldsymbol{\delta}_j = (\delta_{1j}, \dots, \delta_{qj})'$ , where  $q = p - 1$ . Suppose we test

$$H_{0j} : \boldsymbol{\delta}_j = \mathbf{0} \text{ against } H_{1j} : \boldsymbol{\delta}_j \neq \mathbf{0}, \tag{1}$$

and decide for a rejected  $H_{0j}$  which component  $\delta_{ij}$ 's are non-zero before declaring their signs to be positive or negative depending on the values of the corresponding test statistics. The declared signs of the  $\delta_{ij}$ 's then determine a possible inequality or directional pattern. For instance, in the case of  $q = 4$ , suppose for a given gene  $j$ , the  $\boldsymbol{\delta}_j = (\delta_{1j}, \dots, \delta_{4j})$  is found significantly different from a null vector, with  $\delta_{1j}$  and  $\delta_{2j}$  declared to be positive and negative, respectively, and  $\delta_{3j}$  and  $\delta_{4j}$  zeros. Then, the corresponding directional pattern is  $\mu_{1j} > \mu_{2j} < \mu_{3j} = \mu_{4j} = \mu_{5j}$ . We can test  $H_{0j}$  against  $H_{1j}$  for all the genes applying a suitable multiple testing method.

Thus, given  $p$  ordered categories for each gene, the task of identifying directional patterns of the mean expressions over these categories for all the genes is being formulated as a multiple testing problem where  $H_{0j}$  is tested against  $H_{1j}$  simultaneously for all the genes and the signs of the  $\delta_{ij}$ 's are determined subsequent to the rejection of the corresponding  $H_{0j}$ .

For multiple testing of  $H_{0j}$  against  $H_{1j}$ ,  $j = 1, \dots, m$ , we need  $p$ -values that will provide a valid test for each of these individual testing problems and will allow us to make decisions on the individual  $\delta_{ij}$ 's once a  $H_{0j}$  is rejected without losing a control over the Type I error. For that, we consider for each  $j$  the  $p$ -value available for testing each component null hypothesis  $H_{0j}^i : \delta_{ij} = 0$  against the corresponding component alternative hypothesis  $H_{1j}^i : \delta_{ij} \neq 0$ , for  $i = 1, \dots, q$ , and apply a suitable combination method pooling these  $q$   $p$ -values by treating  $H_{0j}$  as an intersection of the subfamily of these  $q$  component null hypotheses, that is,  $H_{0j} = \bigcap_{i=1}^q H_{0j}^i$ , and  $H_{1j}$  as a union of the corresponding  $q$  alternative hypotheses, that is,  $H_{1j} = \bigcup_{i=1}^q H_{1j}^i$ . Before we discuss appropriate combination methods to be used, let us explain how to obtain these component  $p$ -values and state the underlying assumptions.

For every  $i = 1, \dots, q$  and  $j = 1, \dots, m$ , suppose we use the absolute value of a test statistic  $T_{ij}$  for testing  $H_{0j}^i$  against  $H_{1j}^i$ . Let  $T_{ij}$  be symmetric about 0 under  $H_{0j}^i$  and get stochastically larger or smaller as  $\delta_{ij}$  either increases or decreases from 0. In other words, with  $F_{ij}(t, \delta_{ij})$  denoting the cdf of  $T_{ij}$  at  $t$  under the parameter  $\delta_{ij}$ , we have  $F_{ij}(t, \delta_{ij}) \leq$  or  $\geq F_{ij}(t, 0)$  according as  $\delta_{ij} >$  or  $< 0$ , and  $F_{ij}(0, 0) = \frac{1}{2}$ . Under this setting, a right-tailed test based on the absolute value of  $T_{ij}$  will be optimal for testing  $H_{0j}^i$  against  $H_{1j}^i$ , with the corresponding two-sided  $p$ -value being defined as  $\tilde{P}_{ij} = 2 \min \{F_{ij}(T_{ij}, 0), 1 - F_{ij}(T_{ij}, 0)\}$ . By the distributional property of  $T_{ij}$ ,

it is easy to verify that under  $H_{0j}^i$ , the two-sided  $p$ -value  $\tilde{P}_{ij}$  satisfies

$$Pr\{\tilde{P}_{ij} \leq p\} \leq p, \quad \text{for any } p \in (0, 1). \quad (2)$$

Given  $p$ -values for testing  $H_{0j}^i$  against  $H_{1j}^i$ , for  $i = 1, \dots, q$ , a number of combination methods (or methods of pooling the  $p$ -values) are available in the literature for testing the intersection null hypothesis  $H_{0j} = \bigcap_{i=1}^q H_{0j}^i$  against the alternative  $H_{1j} = \bigcup_{i=1}^q H_{1j}^i$ . Among these, however, the Bonferroni and Simes methods are often used in multiple testing and allow one to make decisions on the individual  $\delta_{ij}$ 's. For a review of these methods, one may see Bernhard, Klein and Hommel (2004). Let  $\tilde{P}_{(1)j} \leq \dots \leq \tilde{P}_{(q)j}$  be the ordered versions of  $\tilde{P}_{ij}$ ,  $i = 1, \dots, q$ , for a fixed  $j = 1, \dots, m$ . Then, in the Bonferroni test, the pooled (or adjusted)  $p$ -value is given by  $P_j = q\tilde{P}_{(1)j}$ ; whereas, in the Simes test, it is given by  $P_j = \min_{1 \leq i \leq q} \{q\tilde{P}_{(i)j}/i\}$ . While the Bonferroni test does not require any dependence structure in the underlying  $p$ -values, the Simes test requires a certain type of positive dependence condition that is often satisfied in multiple testing applications [Sarkar and Chang (1997)]. Upon rejection of  $H_{0j}$  using the Bonferroni pooled  $p$ -value at a level  $\alpha$ , the  $i$ th component null hypothesis  $H_{0j}^i$  can be rejected if  $\tilde{P}_{ij} < \alpha/q$ . For the test based on the Simes pooled  $p$ -value,  $H_{0j}^i$  corresponding to every  $\tilde{P}_{ij} \leq \tilde{P}_{(R_j)j}$  is rejected, where  $R_j = \max \left\{ i : \tilde{P}_{(i)j} \leq \frac{i}{q}\alpha \right\}$ , if the maximum exists; otherwise,  $R_j = 0$ .

Now, suppose the pooled  $p$ -value  $P_j$ , based on either Bonferroni or Simes test, is available to us for every  $j = 1, \dots, m$ , to carry out a multiple testing procedure to test  $H_{0j}$  against  $H_{1j}$  simultaneously for all  $j = 1, \dots, m$ . We will use the multiple testing method of Benjamini and Hochberg (1995) (the BH method) that is designed to control the false discovery rate (FDR). The FDR, for any given multiple testing procedure, is the expected proportion of false rejections (Type I errors) among all rejections, an overall measure of Type I error rate commonly used in microarray studies. More formally, with  $V$  the number of false rejections and  $R$  the total number of rejections, it is defined as

$$FDR = E \left\{ \frac{V}{R \vee 1} \right\}, \quad (3)$$

where  $R \vee 1 = \max(R, 1)$ . This method with a control of the FDR at a given level  $\alpha$  is a stepup test that, given ordered  $p$ -values  $P_{(1)} \leq \dots \leq P_{(m)}$  with the corresponding null hypotheses  $H_{(1)}, \dots, H_{(m)}$ , finds  $k = \max\{1 \leq j \leq m : P_{(j)} \leq j\alpha/m\}$  and rejects those  $H_{(j)}$  for which  $P_{(j)} \leq P_{(k)}$ , provided this maximum exists, otherwise, accepts all the hypotheses.

When a  $H_{0j}$  is rejected using the BH method and further decisions are being made on the signs of the component  $\delta_{ij}$ 's in the corresponding  $\delta_j$ , a Type III or directional errors might occur due to wrong assignments of the signs. For instance, if there is a component  $\delta_{ij}$  in  $\delta_j = (\delta_{1j}, \dots, \delta_{qj})$  that is truly positive (or negative) but declared to be negative (or positive) while deciding on the signs of the  $\delta_{ij}$ 's upon rejection of  $H_{0j} : \delta_j = \mathbf{0}$ , a directional error occurs. So, we need to control such directional errors as well. A convenient and practical way of doing that would be to use an error rate combining both Type I and directional errors in the FDR framework and make sure that it is controlled. One such error rate is the mixed directional FDR (mdFDR), the sum of the FDR and the pure directional FDR (dFDR). The dFDR is defined as

$$dFDR = E \left\{ \frac{S}{R \vee 1} \right\}, \quad (4)$$

where  $S$  denotes the total number of (pure) directional errors. Thus, more formally, the mdFDR is defined as

$$mdFDR = FDR + dFDR = E \left\{ \frac{V + S}{R \vee 1} \right\}, \quad (5)$$

the expected proportion of Type I and III errors among all rejections.

In the next section, we will develop methods to control the mdFDR. This will extend the following directional BH procedure of Benjamini and Yekutieli (2005) from dimension one (i.e.,  $q = 1$ ) to a general dimension.

DEFINITION 1 (The level- $\alpha$  directional BH Procedure)

1. *Apply the BH method at level  $\alpha$  to test  $H_{0j} : \delta_{1j} = 0$  against  $H_{1j} : \delta_{1j} \neq 0$  simultaneously for  $j = 1, \dots, m$ , based on the two-sided  $p$ -values  $\tilde{P}_{1j}$ ,  $j = 1, \dots, m$ .*
2. *Let  $R$  denote the total number of null hypotheses rejected.*

3. For every  $j = 1, \dots, m$ , with  $\tilde{P}_{1j} \leq \frac{R}{m}\alpha$ , declare  $\delta_{1j} >$  or  $< 0$  according as  $T_{1j} > 0$  or  $< 0$ .

It controls the mdFDR at  $\alpha$  under independence of the underlying test statistics.

### 3 Multidimensional Directional FDR Controlling Procedures

We introduce in this section our proposed method of controlling the mdFDR while testing  $H_{0j} : \boldsymbol{\delta}_j = \mathbf{0}$  against  $H_{1j} : \boldsymbol{\delta}_j \neq \mathbf{0}$ , simultaneously for all  $j = 1, \dots, m$ , and making further decisions on the signs of the  $\delta_{ij}$ 's upon rejection of the corresponding  $H_{0j}$ . It is based on the Bonferroni pooled  $p$ -values.

#### PROCEDURE 1

1. Apply the BH method at level  $\alpha$  to test  $H_{0j}$  against  $H_{1j}$  simultaneously for  $j = 1, \dots, m$ , based on the Bonferroni pooled  $p$ -values  $P_j, j = 1, \dots, m$ .
2. Let  $R$  denote the total number of null hypotheses rejected.
3. For every  $i = 1, \dots, q$  and  $j = 1, \dots, m$  with  $\tilde{P}_{ij} \leq \frac{R}{qm}\alpha$ , if  $T_{ij} > 0$ , declare  $\delta_{ij} >$  or  $< 0$  according as  $T_{ij} > 0$  or  $< 0$ .

**Theorem 1** *With independent  $q$ -dimensional test statistics  $\mathbf{T}_j = (T_{1j}, \dots, T_{qj})$ ,  $j = 1, \dots, m$ , the mdFDR of Procedure 1 satisfies*

$$mdFDR \leq \alpha/2 \cdot \left(1 + \frac{m_0}{m}\right) \leq \alpha ; \quad (6)$$

*that is, the mdFDR is controlled at level  $\alpha$ .*

*Remark 1.* We offer a proof of Theorem 1 in the Appendix. Benjamini and Yekutieli (2005) gave an indirect proof of this theorem in the special case when  $q = 1$  using an approach that relates to the FDR-adjusted confidence intervals for selected parameters they developed in the same paper. However it is not apparent how one could adapt their proof to the present case involving multiple parameters. So, we provide a direct proof in a more general setting.

An alternative method based on the Simes pooled  $p$ -values is given in the following. As Simes test is known to be more powerful than the Bonferroni test [Simes (1986)], this would be more powerful than Procedure 1. Unfortunately, however, a proof of its mdFDR control, unlike that for Procedure 1, seems relatively more difficult, even under independence. So, we do not formally propose it in this article as a multidimensional directional FDR controlling procedure, though we will consider it along with Procedure 1 in our simulation studies in the next section.

PROCEDURE 2

1. *Apply the BH method at level  $\alpha$  to test  $H_{0j}$  against  $H_{1j}$  simultaneously for  $j = 1, \dots, m$ , based on the Simes pooled  $p$ -values  $P_j, j = 1, \dots, m$ .*
2. *Let  $R$  denote the total number of null hypotheses rejected.*
3. *For every  $j = 1, \dots, m$ , let  $\tilde{P}_{(1)j} \leq \dots \leq \tilde{P}_{(q)j}$  be the ordered values of  $\tilde{P}_{ij}, i = 1, \dots, q$ . Let  $R_j = \max\{i : \tilde{P}_{(i)j} \leq \frac{i}{q} \cdot \frac{R}{m} \alpha\}$ , if the maximum exists; otherwise  $R_j = 0$ . For every  $i$  and  $j$  with  $\tilde{P}_{ij} \leq \frac{R_j}{q} \cdot \frac{R}{m} \alpha$ , declare  $\delta_{ij} > 0$  or  $< 0$  according as  $T_{ij} > 0$  or  $< 0$ .*

## 4 A Simulation Study

A simulation study was performed to evaluate the performance of our proposed method, Procedure 1. Specifically, we wanted to investigate the following three questions:

- (i) How does it perform in terms of its control of the FDR, dFDR and mdFDR and also power under independence as well as under different dependence situations across the genes?
- (ii) How dose it perform in terms of the same operating characteristics under the independence across the genes when we benchmark it against Procedure 2 (based on the Simes pooled  $p$ -values) and the procedure that makes no adjustment to the gene specific  $p$ -values, that is, simply uses  $\tilde{P}_{(1)j}$  as the pooled  $p$ -value?
- (iii) How does the performance of Procedure 1 under the independence across the genes change as the the dimension  $q$  increases?

We generated  $q + 1$  independently distributed  $m$ -dimensional random normal vectors  $\mathbf{Z}_1, \dots, \mathbf{Z}_{q+1}$ , where the components  $Z_{ij}, j = 1, \dots, m$ , in each  $\mathbf{Z}_i$  are dependent with  $Z_{ij} \sim N(\mu_{ij}, 1)$

and have a common correlation  $\rho$ . Let  $\delta_{ij} = (\mu_{i+1,j} - \mu_{ij})/\sqrt{2}$ ,  $i = 1, \dots, q; j = 1, \dots, m$ . Out of the  $m$  parameter vectors  $\boldsymbol{\delta}_j = (\delta_{1j}, \dots, \delta_{qj})$ ,  $j = 1, \dots, m$ ,  $m_0$  were set to a null vector each, and all the  $\delta_{ij}$ 's in 50%, 25% and 25% of the remaining  $m - m_0$   $\boldsymbol{\delta}_j$ 's were selected randomly from the intervals  $(-0.75, 0.75)$ ,  $(-4.25, -2.75)$  and  $(2.75, 4.25)$  respectively. For each  $i = 1, \dots, q$ , and  $j = 1, \dots, m$ , the statistic  $T_{ij} = (Z_{i+1,j} - Z_{ij})/\sqrt{2}$  for testing  $H_{0j}^i : \delta_{ij} = 0$  vs.  $H_{1j}^i : \delta_{ij} \neq 0$  and the corresponding two-sided  $p$ -value  $\tilde{P}_{ij} = 2\{1 - \Phi(|T_{ij}|)\}$  were then computed, where  $\Phi(\cdot)$  is the standard normal cdf. The pooled  $p$ -values were calculated according to the Bonferroni, Simes and 'no-adjustment' tests for each  $j = 1, \dots, m$ , and Procedures 1, 2 and the so called no-adjustment procedure were applied to their respective lists of pooled  $p$ -values for testing the  $m$  null hypotheses described in (1). For each of these procedures, the number of true null hypotheses that are rejected (Type I errors), the number of  $\boldsymbol{\delta}_j$ 's corresponding to the false null hypotheses the signs of whose components do not completely match with those assigned by the procedure (directional errors) and the sum of these two numbers (Type I and directional errors) were noted. Finally, the following three proportions among the total number of rejected null hypotheses were calculated, the proportion of Type I errors (the observed value of  $V/R \vee 1$ ), the proportion of directional errors (the observed value of  $S/R \vee 1$ ) and the proportion of Type I and directional errors (the observed value of  $(V + S)/R \vee 1$ ). These steps were repeated 2,000 times and the simulated values of the FDR, dFDR and mdFDR were obtained by respectively averaging out the 2,000 values of the above three proportions.

Figure 1 presents the simulated FDR, dFDR and mdFDR and Figure 2 presents the average power (simulated proportion of false null hypotheses that are correctly rejected and assigned signs) of Procedure 1 plotted against the number of false null hypotheses for  $m = 240$ ,  $q = 5$ ,  $\alpha = 0.05$  and  $\rho = 0$  (independence), 0.2, 0.5 and 0.8. As we can see from Figure 1, the FDR and mdFDR both decrease, that is, Procedure 1 becomes more conservative, as the dependence across the genes increases, while the dFDR basically remains the same, at a level below 0.01. As seen from Figure 2, the change in the power performance of Procedure 1 as the dependence across the genes increases is minimal; it decreases slightly with increasing dependence. Overall, the effect of the dependence across the genes on the performance of the proposed multidimensional directional BH procedure is relatively weak.

Figure 3 presents an answer to question (ii). As we can see from this figure, Procedures 1 and 2 behave quite similarly, at least when the dependence across the genes is not of concern, in terms of controlling the FDR, dFDR and mdFDR and the power, though Procedure 2 is slightly more liberal as expected. Also, as expected, if no adjustment is made to the gene specific  $p$ -values, we lose the control of the FDR and mdFDR, with the maximum reaching 0.15. It seems surprising that, even without any adjustment to the gene specific  $p$ -values, the dFDR always remains low, though it becomes larger compared to that for Procedures 1 and 2 as the number of false nulls increases.

Figures 4 and 5 provide an answer to question (iii). It is interesting to note that the performance of Procedure 1 in terms of controlling the FDR, dFDR and mdFDR is unaffected by the dimension  $q$  when the dependence across the genes is not present. The power, of course, increases with increasing dimension.

## 5 An Application to Time-Course Gene Expression Data

Lobenhofer et al. (2002) investigated the effect of estrogen on the expression of cell-cycle genes as MCF-7 breast cancer cells go through the cell division cycle. A normal cell division cycle consists of four major phases, namely, the G1 (or Gap 1), S (Synthesis), G2 (or Gap 2) and M (Mitosis) phase. Genes involved in the cell cycle (known as *cell-cycle genes*) are expected to attain peak gene expression during the phase in which they have a specific biological function in the cell cycle.

According to Lobenhofer et al. (2002), most estradiol treated MCF-7 cells are expected to go through S, G2 and M phases in 12 to 36 hours after treatment and complete the cycle in 48 hours. Genes involved in cell growth and related activities are expected to have maximum expression (or minimum expression if they are anti-growth) during 1 or 4 hours and then monotonically decrease (or increase) in expression as cells go through the remaining phases. On the other hand, genes involved in DNA synthesis, repair and mitosis would have maximum (or minimum) expression during 12 to 36 hours. Thus, such genes may have an *Umbrella* (or *Inverted umbrella*) shaped pattern with a peak or trough during 12 to 36 hours time period. However, according

to Lobenhofer et al. (2002), the cells may be asynchronous as they complete the cell division cycle at 48 hours after exposure. For this reason, the expression of some of the cell-cycle genes may not return to their baseline values at 48 hours but may attain a plateau.

Before exposing the MCF-7 breast cancer cells to estrogen, Lobenhofer et al. (2002) first synchronized all the cells to G1 phase by depriving the cells of serum for 24 hours. Synchronization of cells to the same phase at the beginning of the experiment is important for obtaining reliable gene expression data. They then harvested estradiol treated cells after 1, 4, 12, 24, 36 or 48 hours of treatment. Gene expressions using cDNA microarray chips were obtained at each time point. Each cDNA microarray chip consisted of 1900 gene probes. With 8 replicates at each time point, there were a total of 48 microarray chips across the 6 time points.

Motivated by the above observations, in this section we apply the proposed methodology to identify some cell-cycle genes by considering four ordered categories of time points, namely, 1 hour ( $T1$ ), 4 hours ( $T2$ ), mid group (i.e., the union of 12, 24, 36 hours) ( $T3$ ) and 48 hours ( $T4$ ) after treatment. Thus the sample sizes in the four groups are 8, 8, 24 and 8 respectively. Since the major cell division related activity takes place during the 12 to 36 hours time interval, we combined those time periods together to contrast that period from initial cell growth period (1, 4 hours) and the end of mitosis (48 hours).

Suppose  $\theta_{T1,j}$ ,  $\theta_{T2,j}$ ,  $\theta_{T3,j}$  and  $\theta_{T4,j}$  denote the mean gene expression of gene  $j$ ,  $j = 1, 2, \dots, 1900$ , during time periods  $T1$ ,  $T2$ ,  $T3$  and  $T4$ , respectively. Using notations from the previous section, we let  $\delta_{1j} = \theta_{T1,j} - \theta_{T2,j}$ ,  $\delta_{2j} = \theta_{T2,j} - \theta_{T3,j}$  and  $\delta_{3j} = \theta_{T3,j} - \theta_{T4,j}$ .

Let  $T_{ij}$  denote the test statistic associated with the parameter  $\delta_{ij}$ ,  $i = 1, 2, 3$  and  $j = 1, 2, \dots, 1900$ . In this application  $T_{ij}$  is the usual two-sample  $t$ -test statistic and since the underlying data are not necessarily normally distributed, nonparametric bootstrap methodology based on 10,000 bootstrap samples is used for computing the  $p$ -values  $\tilde{P}_{ij}$  associated with hypotheses on  $\delta_{ij}$ . We then calculate the Bonferroni pooled  $p$ -value  $P_j$  for each gene  $j$  after computing  $\tilde{P}_{ij}$ ,  $i = 1, 2, 3$ .

By applying our proposed method, Procedure 1 to the list of the pooled  $p$ -values  $P_j$ 's, we identified 86 differentially expressed genes at level  $\alpha = 0.05$  of which 19 had an umbrella shape response, 3 inverted umbrella, 32 increased in expression from  $T1$  to  $T3$  and then plateaued

(i.e. for some gene  $j$ ,  $\theta_{T1,j} \leq \theta_{T2,j} \leq \theta_{T3,j} = \theta_{T4,j}$ , with at least one strict inequality). An opposite response was seen with 22 genes which had decreased expression from  $T1$  to  $T3$  and then plateaued (i.e. for some gene  $j$ ,  $\theta_{T1,j} \geq \theta_{T2,j} \geq \theta_{T3,j} = \theta_{T4,j}$ , with at least one strict inequality). We also discovered 10 genes that had a flat expression until  $T3$  and then a decrease in response from  $T3$  to  $T4$ .

Comparing our results with those of Lobenhofer et al. (2002) and Peddada et al. (2003), we found that of the 86 genes we identified, 39 were also identified in at least one of the two previous papers. This included 8 of 13 DNA replication/repair genes identified by Lobenhofer et al. (2002). Among the 5 that were not identified by our procedure, we note that except for MCM7, which may be significant at  $\alpha = 0.10$ , all others had large  $p$ -values that were not significant even at  $\alpha = 0.20$ . Interestingly, in addition to MCM3 that was identified by both Lobenhofer et al. (2002) and Peddada et al. (2003), we identified a well known cell-cycle gene MCM4 (<http://www.cyclebase.org>).

An important step in DNA synthesis during the S phase is the binding of complex proteins to DNA for recruiting other proteins necessary for DNA synthesis. One such complex protein is the replication factor C. Lobenhofer et al. (2002) identified one subunit of this protein, known as replication factor C3. Later the order-restricted inference based methodology of Peddada et al. (2003) identified two additional subunits of this protein, namely, replication factors C4 and C5. Interestingly, the newly introduced methodology identified subunits C2, C3 and C5 as significant genes, thus reinforcing the earlier findings and adding one more subunit to the previous list of replication factor C. Furthermore, based on the proposed methodology it is possible to conclude that the subunits C2, C3 and C5 have peak expression during the 12, 24 or 36 hours time period where the DNA synthesis and replication takes place.

Furthermore, similar to the order-restricted inference procedure of Peddada et al. (2003), the proposed methodology identified the cyclin-dependent kinase inhibitor 1 A (p21 and Cip 1) as repressed during 12 to 36 hours. This gene was not identified by Lobenhofer et al. (2002).

A complete list of all genes identified by this procedure is provided in the *Supplementary Materials* accompanying this article.

## 6 Concluding Remarks

In microarray gene expression studies, researchers are often not only interested in identifying differentially expressed genes under different biological conditions, but are also interested in detecting trends in mean response over ordered categories. For instance, in the simple case of two categories (normal versus tumor tissue), researchers are not only interested in identifying significant genes across these two categories, but they are also interested in further identifying the down and up-regulated genes. As the number of ordered categories increases, the trends or directional patterns become complex and the number of directional patterns increases. Except for the usual Type I errors, this also potentially results in a relatively high frequency of directional errors. Hence, it is important to develop statistical methods of identifying trends in mean response over ordered categories while maintaining a control over both the Type I and directional errors.

The approach proposed in this article provides such a methodology. Differently from existing statistical methods (Peddada et al., 2003; Lin et al., 2007), we have formulated the problem of identifying trends in mean response over ordered categories as a multiple testing problem involving successive comparisons and further directional decisions on the multidimensional parameter of each gene. To deal with this problem, we have first suggested a general multidimensional BH-type directional procedure using the Bonferroni test for controlling the mixed directional FDR (mdFDR), an overall measure of both Type I and directional errors within the framework of the FDR, and theoretically proved that the proposed procedure controls the mdFDR at a pre-specified level when the underlying test statistics are independent across the genes. We evaluated the performance of the introduced procedure in the case of dependence through a simulation study. Finally, the whole proposed methodology has been applied to analyze a time-course microarray data and some interesting results have been obtained.

Although our focus was on identifying individual gene expression profile or trend over the ordered categories in some common microarray experiments such as time-course or dose-response experiments, the proposed methodology can be also applied in National Toxicology Program (NTP) studies, where researchers are interested in assessing whether a test agent is carcinogenic

to some certain tumor types and then identifying its dose-response profile for each tumor type.

The proposed methodology in this article provides an interesting starting point towards addressing the complex yet important problem of controlling both Type I and directional errors in multiple testing involving multidimensional parameters. A proof of the mdFDR control of Procedure 2, which as noted in the paper should be more powerful than Procedure 1, would be an important result among possible future work. The mdFDR controlling property of the proposed directional BH procedure has been established under the assumption that the underlying test statistics are independent across the genes. When gene expressions are obtained by drawing samples from same subjects over time, such an assumption need not be valid. In such cases, not only we have dependence between gene expressions at a given time point but there may be temporal dependence within a gene. It will be interesting to theoretically investigate the performance of the proposed directional BH procedure under such complex dependence structures. In addition, it will also be interesting to develop more powerful adaptive BH directional FDR procedure by exploiting knowledge of the proportion of true null hypotheses.

## 7 Acknowledgements

The research of the first and third authors is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences[Z01 ES101744-04] and the research of the second author is supported by a grant from U.S. National Science Foundation (DMS-0603868). The authors thank Grace Kissling and David Dunson for several useful comments that led to improved presentation of the manuscript.

## References

- [1] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289-300.
- [2] BENJAMINI, Y. and HOCHBERG, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral*

*Statistics* **25** 60-83.

- [3] BENJAMINI, Y., HOCHBERG, Y. and KLING, Y. (1993). False discovery rate control in pairwise comparisons. Working paper 93-2, Tel Aviv University, Dept. of Statistics and Operation Research.
- [4] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165-1188.
- [5] BENJAMINI, Y. and YEKUTIELI, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J Amer. Statist. Assoc.* **100** 71-93.
- [6] BERNHARD, G., KLEIN, M. and HOMMEL, G. (2004) Global and multiple test procedures using ordered p-values – A review. *Statist. Pap.*, **45** 1-14.
- [7] BOCHKINA, N. and RICHARDSON, S. (2007). Tail Posterior Probability for Inference in Pairwise and Multiclass Gene Expression Data. *Biometrics* **63** 1117-1125.
- [8] FINNER, H. (1994). Testing multiple hypotheses: general theory, specific problems, and relationships to other multiple decision procedures. Habilitationsschrift, Fachbereich IV Mathematik, Univ. Trier.
- [9] FINNER, H. (1999). Stepwise multiple test procedures and control of directional errors. *Ann. Statist.* **27** 274-289.
- [10] LIN, D., SHKEDY, Z., YEKUTIELI, D., BURZYKOWSKI, T., GÖHLMANN, H., BONDT, A., PERERA, T., GEERTS, T. and BIJNENS, L. (2007). Testing for trends in dose-response microarray experiments: a comparison of several testing procedures, multiplicity and resampling-based inference. *Statist. App. Gen. Mol. Bio.* **6**(1): Article 26.
- [11] LIU, W. (1996). Control of directional errors with step-up multiple tests. *Statist. Probab. Lett.* **31** 239-242.

- [12] LOBENHOFER, E., BENNETT, L., CABLE, P., LI, L., BUSHEL, P. and AFSHARI, C. (2002). Regulation of DNA replication fork genes by 17 beta-estradiol. *Molecular Endocrinology* **16** 1215–1229.
- [13] PEDDADA, S., LOBENHOFER, E., LI, L., AFSHARI, C., WEINBERG, C. and UMBACH, D. (2003). Gene selection and clustering for time-course and dose response microarray experiments using order-restricted inference. *Bioinformatics* **19** 834-841.
- [14] SARKAR, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30** 239-257.
- [15] SARKAR, S. K. and CHANG, C.-K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.* **92** 1601-1608.
- [16] SARKAR, S. K., SEN, P. K. and FINNER, H. (2004). On two results in multiple testing. In *Recent Developments in Multiple Comparisons*. IMS Lectures Notes-Monograph Series, 47, Y. Benjamini, F. Bretz and S. Sarkar, eds. 89-99, Institute of Mathematical Statistics, Beachwood.
- [17] SHAFFER, J. P. (1980). Control of directional errors with stagewise multiple test procedures. *Ann. Statist.* **8** 1342-1347.
- [18] SHAFFER, J. P. (2002). Multiplicity, directional (type III) errors, and the null hypothesis. *Psychological Methods* **7** 356-369.
- [19] SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751-754.
- [20] TAMOTO, E., TADA, M., MURAKAWA, K., TAKADA, M., SHINDO, G., TERAMOTO, K., MATSUNAGA, A., KOMURO, K., KANAI, M., KAWAKAMI, A., FUJIWARA, Y., KOBAYASHI, N., SHIRATA, K., NISHIMURA, N., OKUSHIBA, S., KONDO, S., HAMADA, J., YOSHIKI, T., MORIUCHI, T. and KATO, H. (2004). Gene-expression profile changes correlated with tumor progression and lymph node metastasis in esophageal cancer. *Clinical Cancer Research* **10** 3629-3638.

- [21] WILLIAMS, V. S., JONES, V. and TUKEY, J. W. (1999). Control errors in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics* **24** 42-69.

## APPENDIX

*Proof of Theorem 1:* We begin by calculating the pure directional FDR (dFDR),  $E \left\{ \frac{S}{R \vee 1} \right\}$ . Let  $I_1 = \{1 \leq j \leq m : \delta_j \neq 0\}$  be the set of indices of false null hypotheses.

$$\begin{aligned}
dFDR &= E \left\{ \frac{S}{R \vee 1} \right\} = E \left\{ \frac{E(S | R)}{R \vee 1} \right\} \\
&= E \left[ \frac{\sum_{j \in I_1} Pr \left\{ \bigcup_{i=1}^q \left( \tilde{P}_{ij} \leq \frac{R}{qm} \alpha, T_{ij} \delta_{ij} < 0 \right) \mid R \right\}}{R \vee 1} \right] \\
&= \sum_{r=1}^m \sum_{j \in I_1} \frac{1}{r} Pr \left\{ \bigcup_{i=1}^q \left( \tilde{P}_{ij} \leq \frac{r}{qm} \alpha, T_{ij} \delta_{ij} < 0, R = r \right) \right\} \\
&\leq \sum_{r=1}^m \sum_{j \in I_1} \sum_{i=1}^q \frac{1}{r} Pr \left\{ \tilde{P}_{ij} \leq \frac{r}{qm} \alpha, T_{ij} \delta_{ij} < 0, R = r \right\}. \tag{7}
\end{aligned}$$

The inequality follows from the Bonferroni inequality.

For any given  $i$  and  $j$ , without loss of generality, we assume  $\delta_{ij} > 0$ . Noting that under given parameter  $\delta_{ij}$ ,

$$\begin{aligned}
&Pr \left\{ \tilde{P}_{ij} \leq \frac{r}{qm} \alpha, T_{ij} \delta_{ij} < 0, R = r \right\} \\
&= Pr \left\{ \tilde{P}_{ij} \leq \frac{r}{qm} \alpha, T_{ij} < 0, R = r \right\} \\
&\leq Pr \left\{ F_{ij}(T_{ij}, 0) \leq \frac{r}{2qm} \alpha, R = r \right\} \\
&= Pr \left\{ T_{ij} \leq F_{ij}^{-1} \left( \frac{r}{2qm} \alpha, 0 \right), R = r \right\}, \tag{8}
\end{aligned}$$

where  $F_{ij}^{-1}(\cdot, 0)$  is the inverse function of  $F_{ij}(\cdot, 0)$ . The inequality in the above calculations follows from the definition of  $\tilde{P}_{ij}$  and the assumption  $F_{ij}(0, 0) = \frac{1}{2}$ .

When  $\mathbf{T}_j = (T_{1j}, \dots, T_{qj}), j = 1, \dots, m$ , are independent of each other, the last probability in (8) can be simplified to

$$\begin{aligned}
& Pr \left\{ T_{ij} \leq F_{ij}^{-1} \left( \frac{r}{2qm} \alpha, 0 \right) \right\} \cdot Pr \left( R^{(-j)} = r - 1 \right) \\
&= F_{ij} \left( F_{ij}^{-1} \left( \frac{r}{2qm} \alpha, 0 \right), \delta_{ij} \right) \cdot Pr \left( R^{(-j)} = r - 1 \right) \\
&\leq F_{ij} \left( F_{ij}^{-1} \left( \frac{r}{2qm} \alpha, 0 \right), 0 \right) \cdot Pr \left( R^{(-j)} = r - 1 \right) \\
&= \frac{r}{2qm} \alpha \cdot Pr \left( R^{(-j)} = r - 1 \right), \tag{9}
\end{aligned}$$

where  $R^{(-j)}$  denotes the number of rejections in the stepup procedure with critical constants  $\alpha_j = \frac{j+1}{m} \alpha, j = 1, \dots, m - 1$  based on  $\{P_1, \dots, P_m\} \setminus \{P_j\}$ . The above inequality follows from the assumption that  $F_{ij}(\cdot, \delta_{ij})$  is stochastically larger than  $U(0, 1)$  when  $\delta_{ij} > 0$ .

Using (8) and (9) in (7), we have

$$\begin{aligned}
dFDR &\leq \sum_{r=1}^m \sum_{j \in I_1} \sum_{i=1}^q \frac{\alpha}{2qm} Pr \left( R^{(-j)} = r - 1 \right) \\
&= \frac{m_1}{2m} \alpha. \tag{10}
\end{aligned}$$

Noting that the pooled  $p$ -values  $P_j, j = 1, \dots, m$ , satisfy the condition (2), then for independent  $p$ -values  $P_j$ 's, the usual FDR of the  $q$ -dimensional directional BH procedure satisfies the following inequality,

$$FDR \leq \frac{m_0}{m} \alpha; \tag{11}$$

see Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001) or Sarkar (2002). Combining (10) and (11), we have

$$\begin{aligned}
mdFDR &= FDR + dFDR \\
&\leq \frac{m_0}{m} \alpha + \frac{m_1}{2m} \alpha = \alpha/2 \cdot \left( 1 + \frac{m_0}{m} \right), \tag{12}
\end{aligned}$$

and hence the proof is complete. ■

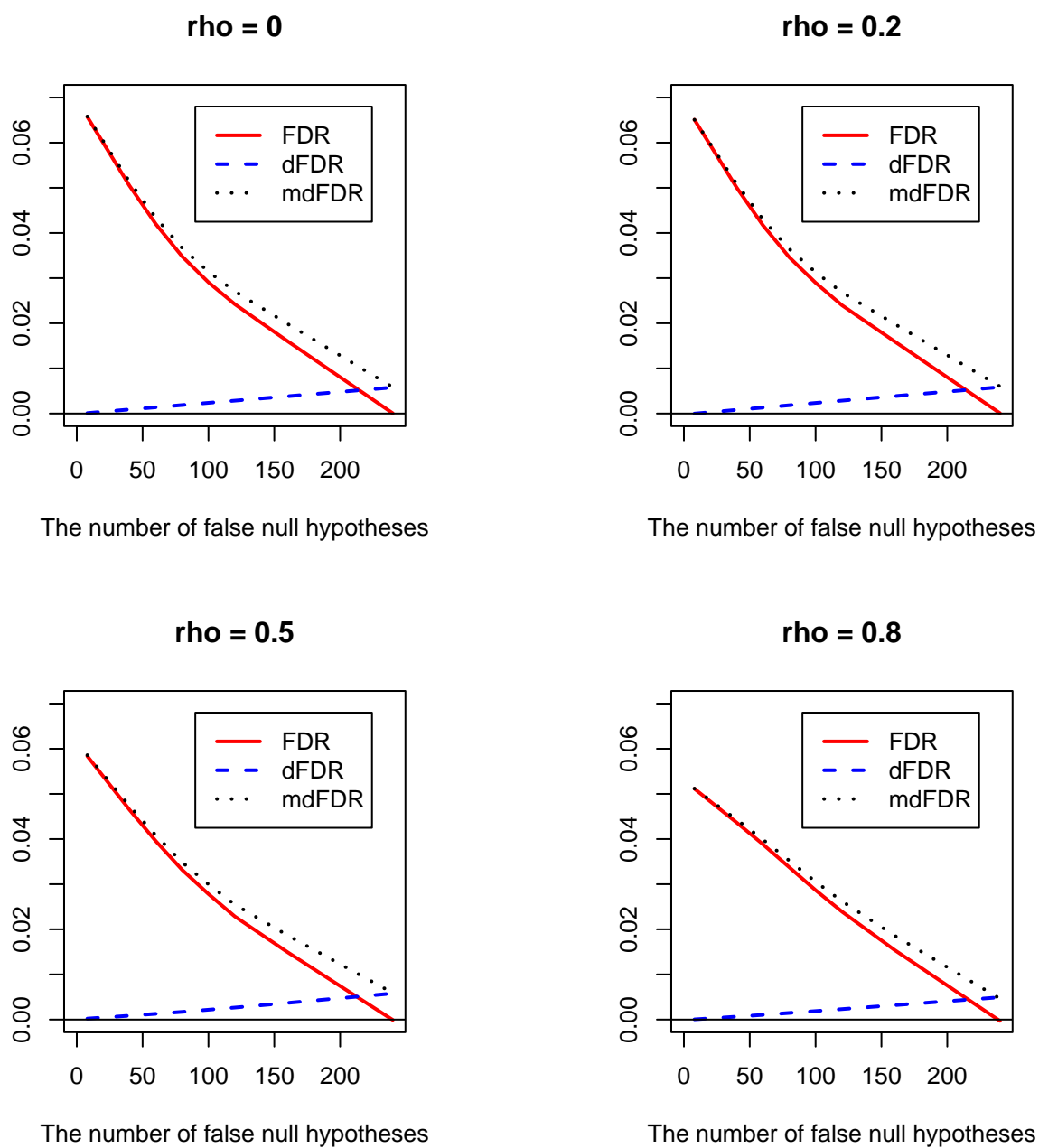


Figure 1: Performance of Procedure 1 in terms of its control of the FDR, dFDR and mdFDR for  $m = 240, q = 5, \alpha = 0.05$ , and  $\rho = 0, 0.2, 0.5$  and  $0.8$ .

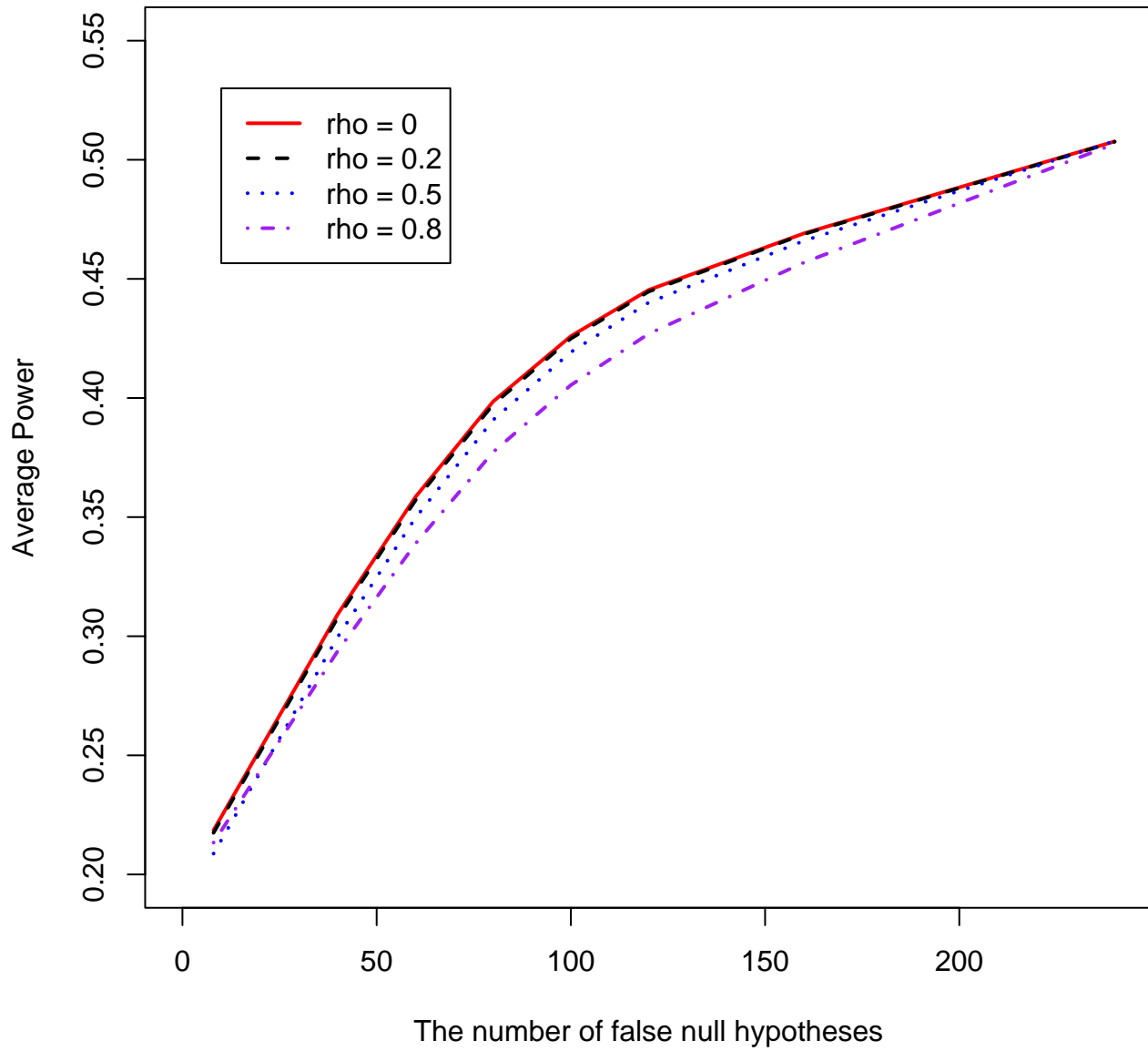


Figure 2: Power of Procedure 1 for  $m = 240, q = 5, \alpha = 0.05$  and  $\rho = 0, 0.2, 0.5$  and  $0.8$ .

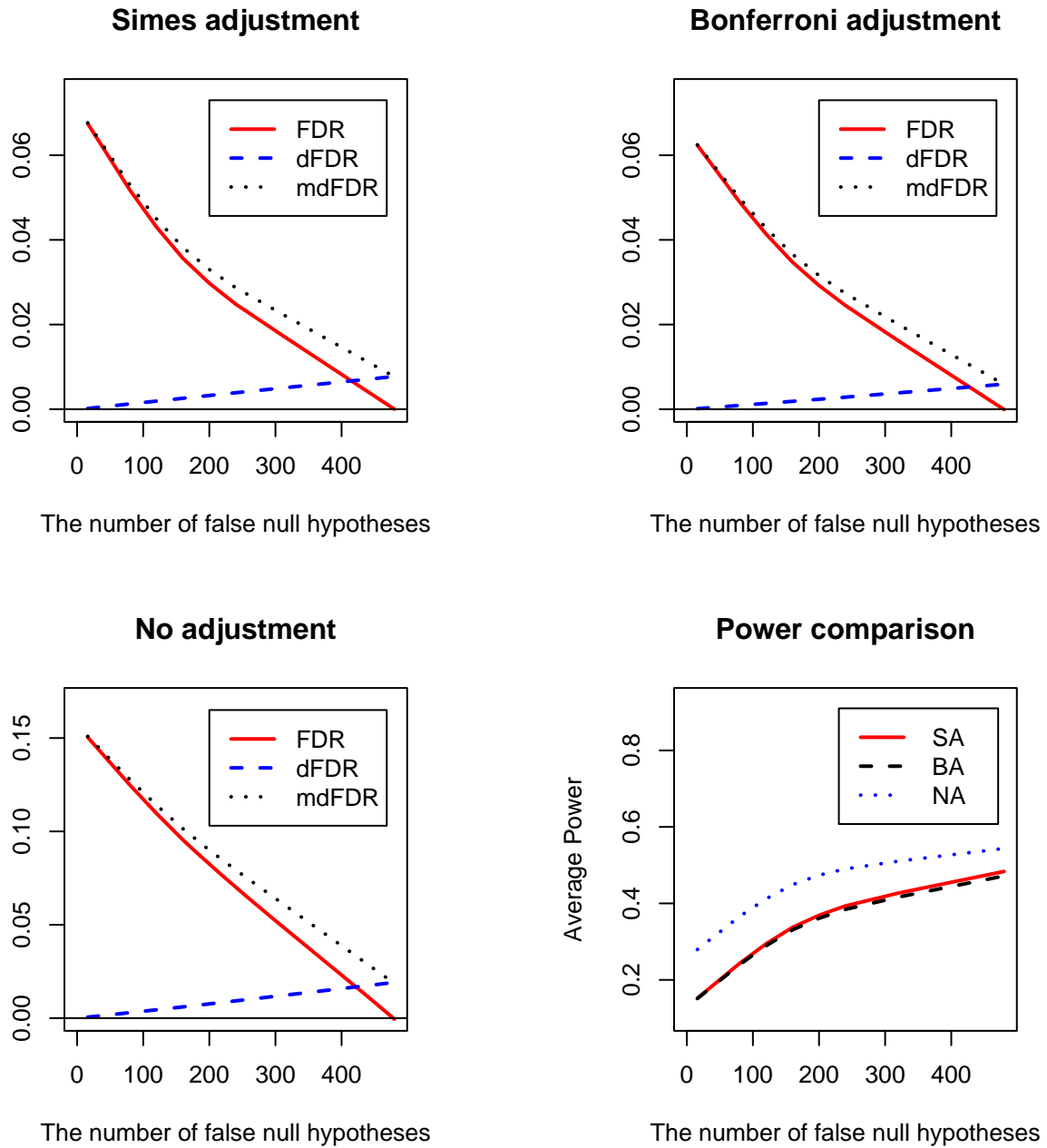


Figure 3: A numerical comparison of Procedures 1 and 2 and the ‘no-adjustment’ procedure in terms of the control of the FDR, dFDR and mdFDR and also power for  $m = 480, q = 3, \rho = 0$ , and  $\alpha = 0.05$ .

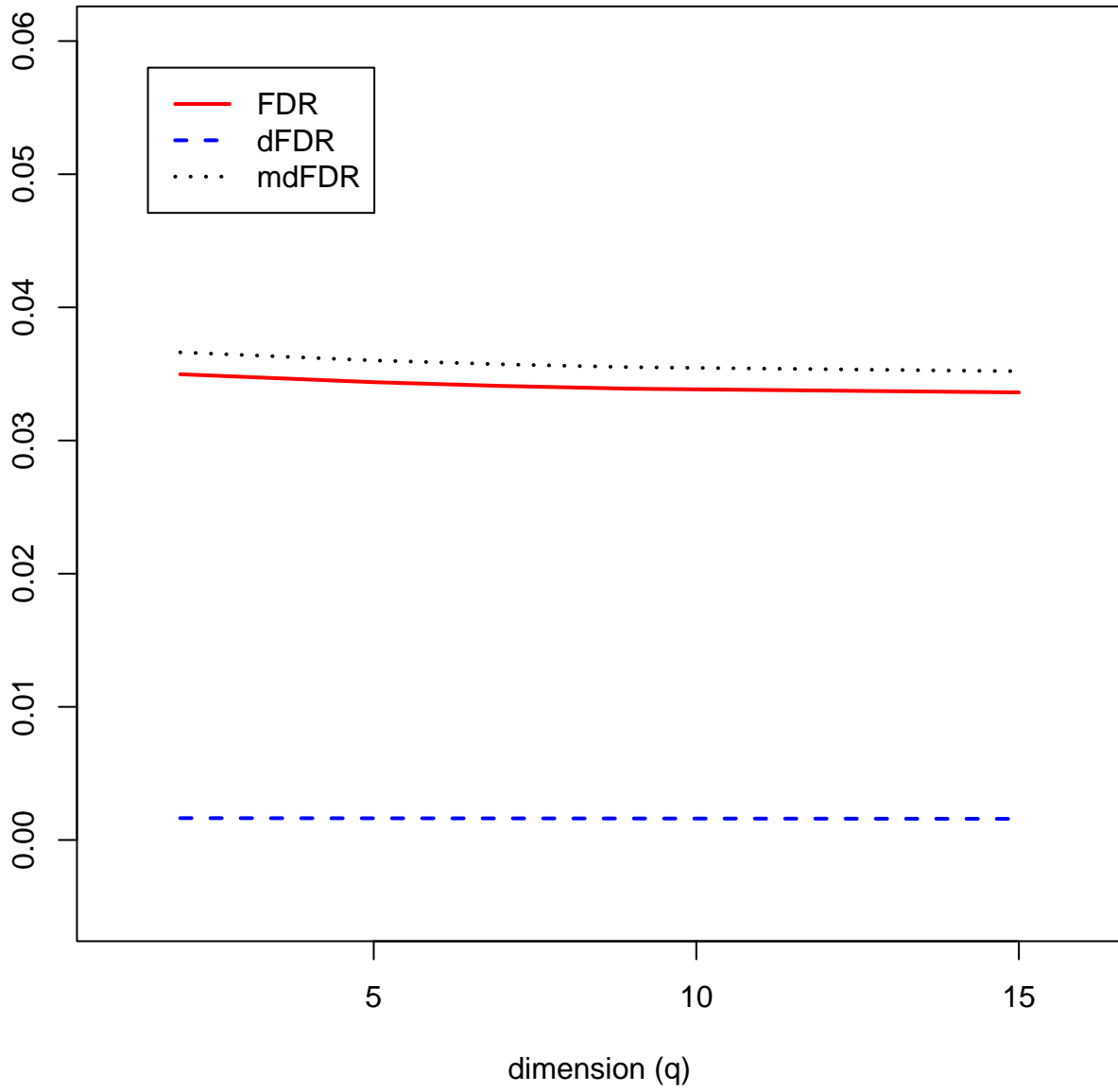


Figure 4: Performance of Procedure 1 with respect to the dimension  $q$  in terms of its control of the FDR, dFDR and mdFDR for  $m = 600, m_0 = 420, \alpha = 0.05$  and  $\rho = 0$ .

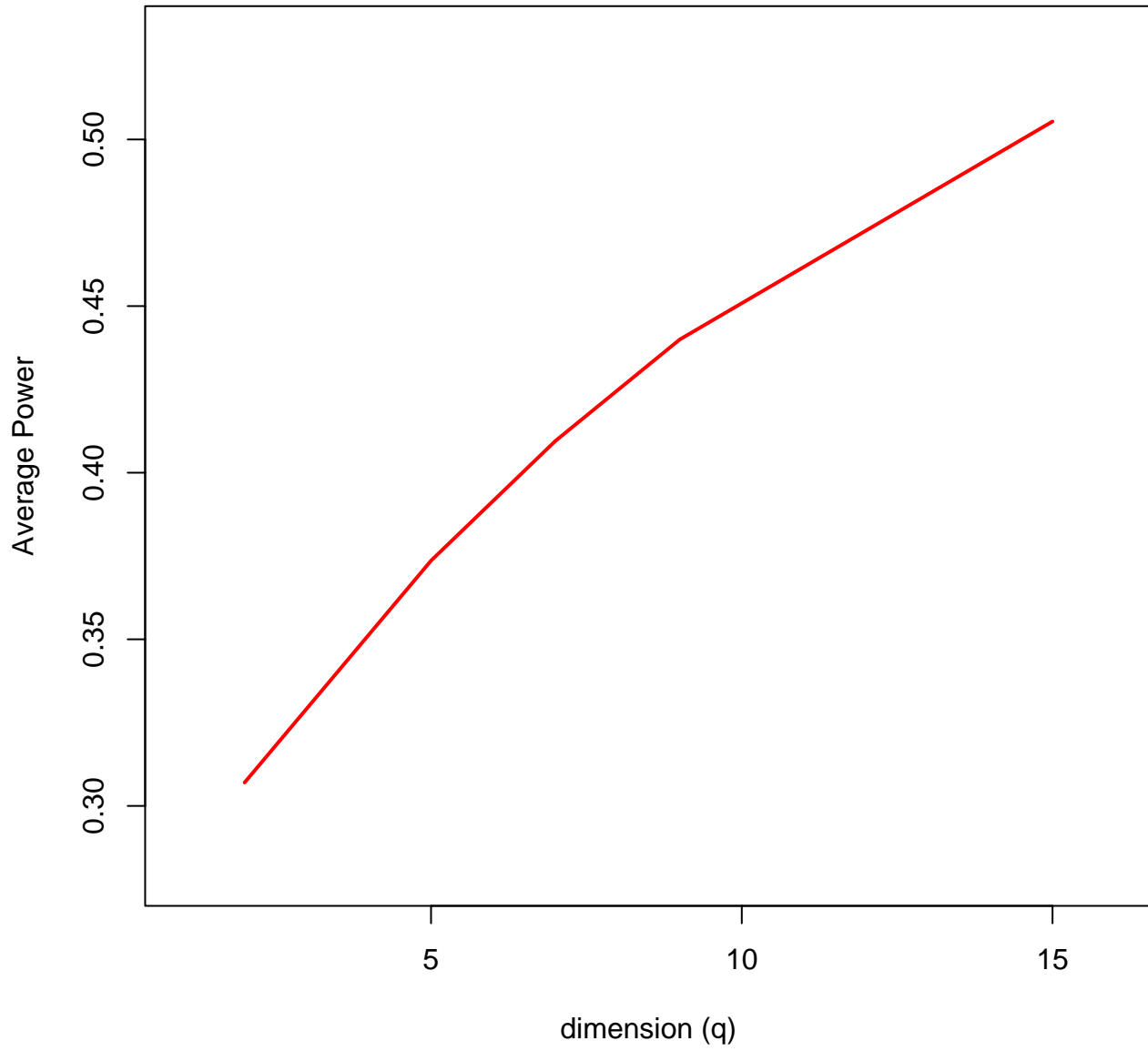


Figure 5: Power performance of Procedure 1 with respect to the dimension  $q$  for  $m = 600, m_0 = 420, \alpha = 0.05$  and  $\rho = 0$ .