

Spring 2015

Multiple testing procedures for complex structured hypotheses and directional decisions

Anjana Grandhi

New Jersey Institute of Technology

Follow this and additional works at: <https://digitalcommons.njit.edu/dissertations>



Part of the [Mathematics Commons](#)

Recommended Citation

Grandhi, Anjana, "Multiple testing procedures for complex structured hypotheses and directional decisions" (2015). *Dissertations*. 119.

<https://digitalcommons.njit.edu/dissertations/119>

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Digital Commons @ NJIT. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

MULTIPLE TESTING PROCEDURES FOR COMPLEX STRUCTURED HYPOTHESES AND DIRECTIONAL DECISIONS

by
Anjana Grandhi

Several multiple testing procedures are developed based on the inherent structure of the tested hypotheses and specific needs of data analysis. Incorporating the inherent structure of the hypotheses results in development of more powerful and situation-specific multiple testing procedures than existing ones. The focus of this dissertation is on developing multiple testing procedures that utilize the information on this structure of the hypotheses and aims at answering research questions while controlling appropriate error rates.

In the first part of the thesis, a mixed directional false discovery rate (mdFDR) controlling procedure is developed in the context of uterine fibroid gene expression data (Davis et al., 2013). The main question of interest that arises in this research is to discover genes associated with various stages of tumor progression, such as tumor onset, growth and development of tumors and large size tumors. To answer such questions, a three-step testing strategy is introduced and a general procedure is proposed that can be used with any mixed directional familywise error rate (mdFWER) controlling procedure for each gene, while controlling the mdFDR as the overall error rate. The procedure is proved to control mdFDR when the underlying test statistics are independent across the genes. A specific methodology, based on the Dunnett procedure, is developed and applied to the uterine fibroid gene expression data of Davis et al. (2013). Several important genes and pathways are identified that play important role in fibroid formation and growth.

In the second part, the problem of simultaneously testing many two-sided hypotheses is considered when rejections of null hypotheses are accompanied by

claims on the direction of the alternative. The fundamental goal is to construct methods that control the mdFWER, which is the probability of making a Type I or Type III (directional) error. In particular, attention is focused on cases where the hypotheses are ordered as H_1, \dots, H_n , so that H_{i+1} is tested only if H_1, \dots, H_i have all been previously rejected. This research proves that the conventional fixed sequence procedure, which tests each hypothesis at level α , when augmented with directional decisions, can control mdFWER under independence and positive regression dependence of the test statistics. Another more conservative directional procedure is also developed that strongly controls mdFWER under arbitrary dependence of test statistics.

Finally, in the third part, multiple testing procedures are developed for making real-time decisions while testing a sequence of a-priori ordered hypotheses. In large scale multiple testing problems in applications such as stream data, statistical process control, etc., the underlying process is regularly monitored and it is desired to control False Discovery Rate (FDR) while making real time decisions about the process being out of control or not. The existing stepwise FDR controlling procedures, such as the Benjamini-Hochberg procedure, are not applicable here because of the implicit assumption that all the p -values are available for applying the testing procedure. In this part of the thesis, powerful Fallback-type procedures are developed under various dependencies for controlling FDR that award the critical constants on rejection of a hypothesis. These procedures overcome the drawback of the conventional FDR controlling procedures by making real-time decisions based on partial information available when a hypothesis is tested and allowing testing of each a-priori ordered hypothesis. Simulation studies demonstrate the effectiveness of these procedures in terms of FDR control and average power.

**MULTIPLE TESTING PROCEDURES FOR COMPLEX
STRUCTURED HYPOTHESES AND DIRECTIONAL DECISIONS**

by
Anjana Grandhi

**A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
and Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences**

**Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark**

May 2015

Copyright © 2015 by Anjana Grandhi

ALL RIGHTS RESERVED

APPROVAL PAGE

**MULTIPLE TESTING PROCEDURES FOR COMPLEX
STRUCTURED HYPOTHESES AND DIRECTIONAL DECISIONS**

Anjana Grandhi

Dr. Wenge Guo, Dissertation Advisor Date
Associate Professor, New Jersey Institute of Technology, Department of
Mathematical Sciences

Dr. Ji Meng Loh, Dissertation Co-Advisor Date
Associate Professor, New Jersey Institute of Technology, Department of
Mathematical Sciences

Dr. Sunil Dhar, Committee Member Date
Professor, New Jersey Institute of Technology, Department of Mathematical
Sciences

Dr. Sundar Subramanian, Committee Member Date
Associate Professor, New Jersey Institute of Technology, Department of
Mathematical Sciences

Dr. Zhi Wei, Committee Member Date
Associate Professor, New Jersey Institute of Technology, Department of Computer
Science

BIOGRAPHICAL SKETCH

Author: Anjana Grandhi
Degree: Doctor of Philosophy
Date: May 2015

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
New Jersey Institute of Technology, Newark, NJ, 2015
- Master of Science in Statistics,
University of New Orleans, New Orleans, LA, 2011
- Master of Science in Statistics,
University of Calcutta, Kolkata, India, 2008
- Bachelor of Science in Statistics,
University of Calcutta, Kolkata, India, 2006

Major: Probability and Applied Statistics

Presentations and Publications:

A. Grandhi, W. Guo and S. D. Peddada. A multiple pairwise comparisons procedure reveals differential gene expression patterns and pathways in uterine fibroids by their size. Submitted.

A. Grandhi, W. Guo and J. P. Romano. Control of directional errors in fixed sequence multiple testing. *Submitted*.

A. Grandhi. FDR controlling procedures for testing a-priori ordered hypotheses. *Presentation*, Eastern North American Region International Biometric Society, Miami, 2015.

A. Grandhi. Testing of hierarchically structured families of hypotheses with multi-dimensional directional decisions. *Presentation*, Joint Statistical Meeting, Boston, 2014.

A. Grandhi. Multiple testing procedures for hierarchically structured families of hypotheses. *Poster Presentation*, Frontiers of Applied and Computational Mathematics, NJIT, 2014.

The best thing about being a statistician is that you get to play in everybody's backyard. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. Today, [software] is at least as important as the 'hardware' of tubes, transistors, wires, tapes and the like. Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise.

John Wilder Tukey

ACKNOWLEDGMENT

I am grateful to my co-advisors, Dr. Wenge Guo and Dr. Ji Meng Loh, who guided me in the right direction not only as academic advisors but also as true well-wishers and friends. It would not have been possible to complete this dissertation without their help, guidance and constant encouragement.

I am grateful to Dr. Shyamal D. Peddada, Dr. Joseph P. Romano and Dr. Gavin Lynch for their guidance and extremely important insights throughout my research.

I would like to extend my gratitude to the other members of my dissertation committee, Dr. Sunil Dhar, Dr. Sundarraman Subramanian and Dr. Zhi Wei for their encouragement and support throughout my PhD.

I take this opportunity to mention the names of my parents, Mr. G. K. V. Buchi Raju and Mrs. Sita Grandhi and my sister, Deepika Grandhi in an effort to convey my gratitude to them for their continuous guidance and sacrifice, which made me achieve this goal.

Last, but certainly not the least, I would like to thank my husband, Shoubhik Mondal, for his support, care and guidance throughout my study.

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION	1
1.1 Introduction	1
1.2 Basic Concepts of Multiple Hypothesis Testing	3
1.2.1 Error Measures	4
1.2.2 Strong and Weak Control	6
1.2.3 Definition of Power	6
1.3 Assumptions of MTPs	7
1.3.1 Assumptions on p -values	8
1.4 Procedures based on Ordered p -values	8
1.4.1 FWER Controlling Procedures	9
1.4.2 FDR Controlling Procedures	10
1.4.3 Directional Errors Controlling Procedures	10
1.5 Procedures based on Structure of Hypotheses	11
1.5.1 FWER Controlling Procedures	12
1.5.2 FDR Controlling Procedures	13
1.6 Motivation and Thesis Outline	16
2 CONTROL OF MDFDR IN TESTING HIERARCHICALLY STRUCTURED FAMILIES OF HYPOTHESES AND DIRECTIONAL DECISIONS . . .	19
2.1 Introduction	19
2.2 Background, Notations and Problem Formulation	22
2.3 A General Mixed Directional FDR Controlling Procedure	26
2.3.1 mdFDR Control under Independence	27
2.4 Methodology	30
2.4.1 Choices of Methods for P_{screen}^j	30
2.4.2 Choices of mdFWER Controlling Methods	32

TABLE OF CONTENTS
(Continued)

Chapter	Page
2.5 Discussion and Conclusion	34
3 METHODOLOGY FOR UTERINE FIBROID DATA	36
3.1 Introduction	36
3.2 Uterine Fibroid Gene Expression Data	37
3.3 Methodology for FGS Gene Expression Data	38
3.3.1 Step 1 - Identifying Differentially Expressed Genes	39
3.3.2 Step 2 - Identifying Significant Pairwise Comparisons	41
3.3.3 Step 3 - Directional Decisions	42
3.4 Simulation Study	42
3.4.1 Study Design	43
3.4.2 Results of The Simulation Study	44
3.5 Results of Data Analysis	46
3.6 Discussion and Conclusion	54
4 CONTROL OF DIRECTIONAL ERRORS IN FIXED SEQUENCE MULTIPLE TESTING	55
4.1 Introduction	55
4.2 Preliminaries	57
4.3 The mdFWER Control Under Arbitrary Dependence	58
4.4 The mdFWER Control Under Independence	60
4.5 Extension to Positive Dependence	63
4.6 Further Extensions to Positive Dependence	65
4.7 Conclusions	67
5 FDR CONTROLLING PROCEDURES FOR TESTING A-PRIORI ORDERED HYPOTHESES	69
5.1 Introduction	69
5.2 Motivating Examples	70
5.3 Structure of Hypotheses	71

TABLE OF CONTENTS
(Continued)

Chapter	Page
5.4 Procedures for Testing a-priori Ordered Hypotheses	72
5.4.1 Fallback-Type Procedure under PRDS	73
5.4.2 Fallback-Type Procedure under Arbitrary Dependence	77
5.4.3 FDR Control under Independence	78
5.5 Simulation Study	81
5.5.1 Study Design	82
5.5.2 Results of The Simulation Study	83
5.6 Summary and Conclusion	84
6 CONCLUSION AND FUTURE WORK	88
APPENDIX A	90
A.1 Proof of Lemma 4.1.	90
A.2 Proof of Theorem 4.1	90
A.3 Proof of Lemma 4.2.	93
A.4 Proof of Lemma 4.3	93
A.5 Proof of Theorem 4.2	95
A.6 Proof of Theorem 4.3	96
A.7 Proof of Proposition 4.1	100
A.8 Proof of Proposition 4.2	102
A.9 Proof of Theorem 4.4	104
APPENDIX B	106
B.1 Proof of Theorem 5.2	106
B.2 Proof of (5.18)	107
B.3 Proof of (5.19)	108
BIBLIOGRAPHY	110

LIST OF TABLES

Table	Page
1.1 Summary of Outcomes in a Multiple Testing Procedure	4

LIST OF FIGURES

Figure	Page
2.1 A typical microarray gene expression data	22
2.2 A graphical display of the hypotheses in uterine fibroid gene expression data problem.	24
3.1 mdFDR (left), Average Power (right) with the proposed methodology and three variants using Holm, Hochberg and Bonferroni procedures, respectively, in Steps 2 and 3 along with Guo et al. (2010) procedure, under independence among genes and within genes.	45
3.2 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.2$) within genes.	46
3.3 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.5$) within genes.	47
3.4 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.8$) within genes.	48
3.5 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.2$) among genes.	49
3.6 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.5$) among genes.	50
3.7 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.8$) among genes.	51
3.8 Venn diagrams of differentially expressed genes by tumor size.	52
3.9 Venn diagram of enriched pathways by tumor size.	52
5.1 Comparison of the correction factors used in our procedure in Theorem 5.2 and the BY procedure for $m = 200$ hypotheses.	78
5.2 FDR and Average Power for $m = 200, 400$ hypotheses under independence ($\rho = 0$).	84

LIST OF FIGURES
(Continued)

Figure	Page
5.3 FDR and Average Power for $m = 200, 400$ hypotheses under Adjacent-Correlation with $\rho = 0.2$	85
5.4 FDR and Average Power for $m = 200, 400$ hypotheses under Adjacent-Correlation with $\rho = 0.5$	86
5.5 FDR and Average Power for $m = 200, 400$ hypotheses under Adjacent-Correlation with $\rho = 0.8$	87

CHAPTER 1

INTRODUCTION

1.1 Introduction

Scientific experiments are often subject to rigorous statistical analyses that almost always involve simultaneous evaluation of more than one research question. Multiplicity or the occurrence of multiple errors becomes an inherent problem in such analyses. As a consequence, the findings of an experiment can be misleading and the effect of multiplicity needs to be addressed to draw valid scientific conclusions. Multiple testing procedures (MTPs) adjust statistical inference by accounting for multiplicity and avoid declaring an effect when there is none. Multiple testing methodology plays an important role in analysis of data from a varied range of fields like bioinformatics, clinical trials, regression analysis and modeling to mention a few. The importance of MTPs in the analysis of large scale experiments like microarray experiments, genome-wide association studies, fMRI experiments needs special mention. Several interesting and powerful MTPs have been developed for handling different situations in these kinds of data while controlling appropriate error rates.

In single hypothesis testing, the error measure that is controlled is the probability of making a Type I error. But in multiple testing, there are several possible measures of overall Type I error rate. A popular and widely used error rate is the familywise error rate (FWER), which is the probability of making at least one false rejection. It is to be noted that the control of FWER is appropriate when the number of hypotheses tested simultaneously is small or moderate, but it proves to be too conservative when large number of hypotheses are tested simultaneously, which is typically the case in large scale experiments like microarray or fMRI. Benjamini and Hochberg (1995) introduced false discovery rate (FDR) as an appropriate measure to

control while simultaneously testing a large number of hypotheses. FDR is defined as the expected proportion of false rejections among all rejections. FDR allows more hypotheses to be rejected while controlling the proportion of false rejections, thus, opening the way for development of more powerful procedures than those that use FWER as an error measure to control. For a review of multiple testing procedures controlling the FWER, please refer to Dmitrienko et al. (2009). For a review of FDR controlling procedures, refer to Benjamini (2010).

In multiple testing, the hypotheses to be tested are often found to possess some complex structure. This structure can be due to the relative importance of the hypotheses as in dose response studies where the hypotheses corresponding to higher doses are tested prior to those corresponding to a lower dose or due to the formulation of complex stagewise testing in large scale hypotheses testing or due to the natural sequence of occurrence of the hypotheses in real time as in the stream data analysis. Several researchers exploit this inherent structure to develop powerful procedures appropriate to the testing problem. In the FWER framework, the Fixed Sequence procedure [Maurer et al. (1995); Westfall and Krishen (2001); Wiens (2003); Wiens and Dmitrienko (2005)] is introduced for testing a sequence of hypotheses, ordered in advance based on some prior knowledge, while controlling FWER. Recently Qiu et al. (2014) developed more powerful procedures for testing a fixed sequence of hypotheses while controlling FWER. In the FDR framework, Farcomeni and Finos (2013) and Lynch et al. (2014) introduced procedures for controlling FDR while testing a large number of hypotheses that have an inherent structure.

Often, in multiple testing of two-sided hypotheses, it is of interest to make directional decisions once a null hypothesis is rejected. It is important to control Directional errors or Type III errors in addition to Type I errors in such testing. Control of Directional errors in multiple testing set-up is a very challenging problem. A few researchers to discuss the control of directional errors are Shaffer (1980), Finner

(1994) in FWER set-up, Benjamini and Yekutieli (2005), Guo et al. (2010) in FDR set-up.

In this dissertation, we consider multiple testing problems where the hypotheses have some complex structure and sometimes require directional decisions to be made. In many applications of multiple testing such as genomic research, clinical trials, stream data analysis, statistical process control, the hypotheses have such complex inherent structure. This structure may arise from prior knowledge or can be formed by reformulating the underlying problem as in Kropf and Läuter (2002), Kropf et al. (2004), Westfall et al. (2004), Hommel and Kropf (2005), Finos and Farcomeni (2011), Farcomeni and Finos (2013), Lynch et al. (2014) where a fixed sequence structure can be formed among the hypotheses by specifying the testing order of the hypotheses using a data-driven approach. The problem of controlling directional errors may arise in time-course/dose-response experiments or in identifying gene expression patterns/profiles over the ordered categories as in Guo et al. (2010) and Sun and Wei (2011).

1.2 Basic Concepts of Multiple Hypothesis Testing

Consider the problem of simultaneous testing of m null hypotheses $H_i : \theta_i = \theta_{i0}$, $i = 1, 2, \dots, m$ against the corresponding alternative hypotheses denoted by $H'_i : \theta_i \neq \theta_{i0}$, based on the corresponding p -values P_1, P_2, \dots, P_m . Let, m_0 denote the number of true null hypotheses and $m_1 = m - m_0$ denote the number of false null hypotheses. The following is the list of the different kinds of errors that can occur while testing a hypothesis:

- A *Type I error* occurs when a true null hypothesis is rejected.
- A *Type II error* occurs when a false null hypothesis is not rejected.
- A *Type III error* or *Directional error* occurs if, on rejection of a false null hypothesis, a wrong assignment of direction is made, i.e., declare that $\theta_i < \theta_{i0}$ when in reality $\theta_i > \theta_{i0}$ or declare that $\theta_i > \theta_{i0}$ when in reality $\theta_i < \theta_{i0}$.

In any MTP, each hypothesis is either rejected or not rejected. Table 1.1 summarizes the notation for all possible outcomes while simultaneously testing m hypotheses. V denotes the number of falsely rejected hypotheses or Type I errors, S denotes the number of correct rejections and R denotes the total number of hypotheses rejected. In any testing situation, m , is fixed and known, and m_0 , the number of true null hypotheses and m_1 , the number of false null hypotheses are fixed but unknown. All of V , S and R are random but only R is observable. Let D denote the number of Type III errors in an MTP. In the rest of the thesis we denote by I_0 , the set of indices

Table 1.1 Summary and Notation of Outcomes in a Multiple Testing Procedure while Simultaneously Testing m Hypotheses.

	Number of Hypotheses Not Rejected	Number of Hypotheses Rejected	Total
True Null Hypotheses	$m_0 - V$	V	m_0
False Null Hypotheses	$m_1 - S$	S	m_1
Total	$m - R$	R	m

of true null hypotheses and I_1 denotes the set of indices of false null hypotheses with $|I_0| = m_0$ and $|I_1| = m_1$.

1.2.1 Error Measures

In multiple testing, it is important to choose an overall error rate to measure and control Type I errors. A few commonly used Type I error measures are described below:

The *familywise error rate* (FWER) is the probability of making at least one Type I error and is given by,

$$FWER = Pr(V > 0).$$

The *false discovery rate* (FDR) is defined as the expected proportion of falsely rejected hypotheses among all rejected hypotheses and is given by,

$$FDR = E\left(\frac{V}{\max\{R, 1\}}\right).$$

FWER and FDR are related by the following inequality with equality when $m_0 = m$,

$$FDR \leq FWER.$$

In situations where directional decisions need to be made, controlling only Type I error measure is not sufficient and it is necessary to control Type III errors as well. A few error measures are defined in the literature, based on a combination of Type I and Type III errors.

The *mixed directional familywise error rate* (mdFWER) is defined as the probability of making at least one Type I error or Type III error.

$$mdFWER = Pr(V + D > 0). \tag{1.1}$$

The *mixed directional false discovery rate*(mdFDR) is defined as the expected proportion of Type I and Type III errors among all rejections.

$$mdFDR = E\left(\frac{V + D}{\max\{R, 1\}}\right). \tag{1.2}$$

A few other important error measures are as follows: Storey (2002, 2003) introduced the positive false discovery rate (pFDR), which is defined as, $pFDR = E(V/R|R > 0)$. Lehmann and Romano (2005) introduced generalized FWER (k -FWER) which is the

probability of making k or more false rejections for $1 \leq k \leq m$, k -FWER = $Pr(V > k)$. Sarkar (2008) introduced an analogous version of k -FWER for the large scale testing set-up, the generalized FDR (k -FDR), which is the expected proportion of k or more false rejections among all rejections, k -FDR = $E\left(\frac{VI(V \geq k)}{\max\{R, 1\}}\right)$. False discovery proportion (FDP) is defined as the proportion of Type I errors among all rejections, $FDP = V/(R \vee 1)$. Lehmann and Romano (2005) introduced the measure γ -FDP which is, $Pr(FDP > \gamma)$, the probability that the FDP is greater than a given value γ . Efron et al. (2001) used an empirical Bayesian approach to FDR and introduced the local false discovery rate.

1.2.2 Strong and Weak Control

A multiple testing method is said to *strongly control* an error rate if the error rate of the method is less than or equal to a pre-specified level, say α , for any configuration of true and false null hypotheses. A method is said to *weakly control* an error rate if the error rate of the method is controlled when all null hypotheses are true. In applications, strong control is desired as the true configuration of true and false null hypotheses is generally unknown.

1.2.3 Definition of Power

In single hypothesis testing, power of a test is defined as the probability of rejecting a false null hypothesis, which is $1 - Pr(\text{Type II error})$. Evaluation of this probability involves the false null distribution of test statistics. In multiple testing, where several hypotheses are tested simultaneously, various definitions of power are available to measure and compare the performance of MTPs. It is important to choose an appropriate definition of power in addition to a Type I error measure to evaluate the performance of the MTP. A few commonly used concepts of power are described below:

The *minimal power* is defined as the probability of rejecting at least one false null hypothesis,

$$\text{minimal Power} = Pr(S > 0).$$

The *complete power* is defined as the probability of rejecting all false null hypotheses,

$$\text{complete Power} = Pr(S = m_1).$$

The *average power* is defined as expected proportion of rejected false null hypotheses among all false null hypotheses,

$$\text{average Power} = E\left(\frac{S}{m_1}\right).$$

Another concept of power is from the *false non-discovery rate* given by,

$$1 - FNR = 1 - E\left(\frac{m_1 - S}{\max\{m - R, 1\}}\right). \quad (1.3)$$

1.3 Assumptions of MTPs

Several MTPs have been developed in the literature for various scenarios that have different properties. The procedures can be broadly classified according to their distributional assumptions as follows:

- *p-value based Procedures*: These procedures do not make any assumption about the joint distribution of the test statistics and only rely on the univariate p -values.
- *Parametric Procedures*: These procedures make specific assumptions about the distribution of the test statistics, say the joint distribution is a multivariate normal or a multivariate t -distribution.
- *Resampling based Procedures*: These procedures use resampling techniques like bootstrap, permutation, etc., that make fewer assumptions about the data-generating process while still exploiting the dependence structure of the underlying test statistics in multiple testing procedures (for more details see Dmitrienko, Tamhane and Bretz (2009)).

1.3.1 Assumptions on p -values

The basic assumption made about the distribution of the true null p -values is that they are bounded above by the $U(0,1)$ distribution, that is,

$$Pr(P_i \leq u) \leq u, \quad \text{for any } u \in (0, 1) \quad \text{for each } i \in I_0. \quad (1.4)$$

Another assumption made concerns the dependence structure of the p -values. A few common dependence structures of the p -values studied in this thesis while developing MTPs are: independence, positive regression dependence on subset or PRDS (Benjamini and Yekutieli (2001)) and arbitrary dependence. A set of test statistics $\mathbf{T} = \{T_1, \dots, T_m\}$ is said to be positive regression dependent on subset, if for any increasing set U , and for each $i \in I_0$, $Pr(\mathbf{T} \in U | T_i = x)$ is nondecreasing in x . Arbitrary dependence means that the test statistics may have any kind of dependence.

1.4 Procedures based on Ordered p -values

Most common MTPs are based on ordered p -values where the order of testing hypotheses is determined by the order of magnitude of p -values. The procedures are called stepwise methods like single-step, step-up and step-down methods. These stepwise MTPs are described by using a sequence of non-decreasing critical constants $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m$. Let, $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$ be the ordered p -values and $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ be the corresponding null hypotheses.

- A *step-down* procedure starts with the most significant hypothesis $H_{(1)}$ and goes on rejecting hypotheses until an acceptance is observed, that is, it rejects $H_{(1)}, \dots, H_{(r)}$ and does not reject $H_{(r+1)}, \dots, H_{(m)}$ where r is the largest index satisfying

$$P_{(1)} \leq \alpha_1, \dots, P_{(r)} \leq \alpha_r.$$

If no such r exists then the method does not reject any hypothesis.

- A *step-up* procedure starts with the least significant hypothesis $H_{(m)}$ and goes on accepting hypotheses until a rejection is observed, that is, it rejects $H_{(1)}, \dots, H_{(r)}$

and does not reject $H_{(r+1)}, \dots, H_{(m)}$ where r is the largest index satisfying

$$P_{(r)} \leq \alpha_r.$$

- A *single-step* procedure is a stepwise procedure with same critical constants, $\alpha_1 = \alpha_2 = \dots = \alpha_m = c$, that is, reject H_i if $P_i \leq c$ for $i = 1, 2, \dots, m$.

1.4.1 FWER Controlling Procedures

A few popular FWER controlling procedures are discussed here. See Dmitrienko, Tamhane and Bretz (2009) for a review of FWER controlling procedures. The *Bonferroni* procedure is a widely used single-step procedure with critical constants

$$\alpha_i = \frac{\alpha}{m}, \quad i = 1, \dots, m. \quad (1.5)$$

The *Holm* procedure (Holm, 1979) is a step-down procedure that strongly controls FWER under arbitrary dependence with the following critical constants,

$$\alpha_i = \frac{\alpha}{(m - i + 1)}, \quad i = 1, \dots, m. \quad (1.6)$$

The *Hochberg* procedure (Hochberg, 1988) is a step-up procedure with the same critical constants as the Holm procedure and strongly controls FWER under independence and positive dependence (Sarkar, 1998; Sarkar and Chang, 1997).

An important parametric procedure in the FWER framework is the *Dunnett procedure* [Dunnett (1955); Dunnett and Tamhane (1991, 1992)]. This procedure is developed as a powerful procedure for comparing several treatment groups with a common control group. The test assumes that the underlying distribution of the data from the different groups is Normal with equal variance. If θ_i , $i = 1, \dots, m$ and θ_0 denote the means of m treatment groups and control group, respectively, then the hypotheses tested are: $H_{0i} : \theta_i = \theta_0$ against $H_{1i} : \theta_i \neq \theta_0$. The test assumes that the test statistics vector (T_1, \dots, T_m) has a multivariate t -distribution with appropriate degrees of freedom. The details of this procedure are discussed in Chapter 3.

1.4.2 FDR Controlling Procedures

The Benjamini-Hochberg procedure or BH procedure (Benjamini and Hochberg, 1995) is a step-up procedure with the following critical constants,

$$\alpha_i = \frac{i\alpha}{m}, \quad i = 1, 2, \dots, m. \quad (1.7)$$

Benjamini and Hochberg (1995) proved that the BH procedure strongly controls FDR at level $m_0\alpha/m$ under independence of the p -values. Benjamini and Yekutieli (2001) and Sarkar (2002) proved that the BH procedure strongly controls FDR at the same level also under positive regression dependence of the p -values. Benjamini and Yekutieli (2001) also showed that with an appropriate modification in the critical constants the corresponding step-up procedure can strongly control the FDR under arbitrary dependence at level α . This procedure is called the Benjamini-Yekutieli (BY) procedure and the critical constants are,

$$\alpha_i = \frac{i\alpha}{m \sum_{j=1}^m 1/j}, \quad i = 1, 2, \dots, m. \quad (1.8)$$

Storey (2002) introduced an estimation approach to FDR that is opposite to the approach of stepwise methods. In the stepwise methods, the rejection region (critical constants) is determined based on the fixed FDR level, but Storey's approach is to fix the rejection region and estimate the FDR of the rejection region. For some other methods, see Sarkar (2008) and Benjamini (2010).

1.4.3 Directional Errors Controlling Procedures

A few researchers have done some works on the control of directional errors in the multiple testing. Shaffer (1980) proved that Holm's procedure (Holm, 1979) augmented with directional decisions based on sign of test statistics controls mdFWER under independence of test statistics and several conditions on the distribution of the test statistics. Later, Finner (1994) and Liu (1997) independently

proved the same result for Hochberg procedure (Hochberg, 1988) under the same conditions for the distribution of the independent test statistics. Finner (1999) generalized the result by Shaffer (1980) to a large class of stepwise or closed multiple test procedures. Sarkar et al. (2004) gave a simple alternative proof to the method of Shaffer (1980). For a good review and further discussions on the mdFWER control of closed testing methods, refer to Shaffer (2002) and Westfall et al. (2013).

Benjamini and Yekutieli (2005) discussed control of directional errors in an FDR set-up and proved that the BH procedure augmented with directional decisions controls the mdFDR under independence. Guo et al. (2010) developed a two-stage procedure for controlling mdFDR that extends the work of Benjamini and Yekutieli (2005) to multidimensional directional decisions. In addition, some recent results have been obtained in Guo and Romano (2015).

1.5 Procedures based on Structure of Hypotheses

In many applications, hypotheses tested have some inherent structure. For example in clinical trials, the hypotheses pertaining to primary endpoints are more important than those pertaining to secondary endpoints, hence the order of testing hypotheses should be based on this knowledge instead of ordering by observed p -values. Incorporating such prior information into the construction of multiple testing methods leads to more powerful procedures and interpretation becomes more relevant to the problem at hand. Another kind of structure arises in the hypotheses in applications such as stream data analysis or Statistical Process Control where real time decisions are made to regularly monitor the underlying process. In such situations, the hypotheses are naturally ordered by time and require real-time decision making based on partial information. Incorporating the partial information into the construction of multiple testing methods leads to more appropriate procedures to make relevant interpretation.

Several authors developed such procedures that control FWER or FDR while testing such structured hypotheses. To our knowledge, no procedures are developed so far that take the underlying structure of hypotheses into account while making directional decisions and controlling directional errors in addition to Type I errors. The following sections briefly review the FWER and FDR controlling procedures in this context.

1.5.1 FWER Controlling Procedures

The *Fixed-sequence* testing procedure (Maurer et al., 1995; Westfall and Krishen, 2001; Wiens, 2003; Wiens and Dmitrienko, 2005) is proposed for testing hypotheses that have a pre-defined fixed order. This order reflects the relative importance or relevance of the hypotheses, with more important hypotheses tested before the less important ones. The testing starts with the first hypothesis, H_1 and a hypothesis H_i is rejected at the i -th step if,

$$P_j \leq \alpha, \quad j = 1, \dots, i. \quad (1.9)$$

It is proved that the Fixed-sequence procedure strongly controls FWER at level α under arbitrary dependence of p -values. A drawback of this procedure is that once a hypothesis is accepted there is no chance of testing the rest of the hypotheses at all.

The *Fallback* procedure (Wiens, 2003; Wiens and Dmitrienko, 2005, 2010) is introduced for overcoming the drawback of early stopping in fixed-sequence procedure by allowing each hypothesis to be tested. It allocates the overall error rate α among the hypotheses according to their weights $\omega_1, \dots, \omega_m$ where, $\sum_{i=1}^m \omega_i = 1$ and rejects H_i if $P_i \leq \alpha_i$ where,

$$\alpha_i = \begin{cases} \omega_i \alpha + \alpha_{i-1} & \text{if } H_{i-1} \text{ is rejected} \\ \omega_i \alpha & \text{otherwise.} \end{cases} \quad (1.10)$$

Westfall and Krishen (2001) and Dmitrienko et al. (2003) developed procedures for testing a fixed order of families of hypotheses known as gatekeeping procedures where hypotheses or families of hypotheses in a sequence act as gatekeeper for all the hypotheses occurring later in the sequence. Bretz et al. (2009) proposed a graphical approach to construct and compare weighted Bonferroni based closed test procedures such as gatekeeping procedures, Fixed-sequence tests, and Fallback procedures.

Qiu et al. Generalized Fixed Sequence Procedure: More recently, Qiu et al. (2014) introduced Generalized Fixed Sequence procedures whose critical values are defined by using a function of the numbers of rejections and acceptances, and which allows follow-up hypotheses to be tested even if some earlier hypotheses are not rejected. In particular, their Procedure A1 is as follows:

Procedure 1.1 (Qiu et al. (2014) Procedure A1). For $i = 1, \dots, m$, reject H_i if,

$$P_i \leq \frac{\alpha}{m - s_{i-1}}, \quad (1.11)$$

where, s_{i-1} denotes the number of hypotheses rejected when testing H_1, \dots, H_{i-1} .

1.5.2 FDR Controlling Procedures

Very few FDR controlling procedures have been developed in the literature that take the underlying structure of hypotheses into account. By far the procedure by Yekutieli (2008) is the most general procedure developed in this context. Yekutieli (2008) discussed methodology for controlling FDR in large-scale studies that involve testing multiple families of hypotheses arranged in a tree of disjoint subfamilies. Benjamini and Heller (2007), Heller et al. (2009) and Guo et al. (2010) developed procedures to control FDR in special hierarchical structures of hypotheses. Recently, Farcomeni and Finos (2013) developed an FDR controlling fixed sequence procedure that tests each hypothesis in sequence at level α until a stopping condition is reached. Lynch et al.

(2014) developed an FDR controlling fixed sequence procedure that can incorporate more than one acceptance without losing control of FDR. A few important procedures are discussed below.

Lynch et al. Fixed Sequence Procedures: Lynch et al. (2014) proposed multiple testing procedures for controlling FDR while testing a fixed sequence of hypotheses. They introduce the following procedures that allow a fixed number of acceptances. Procedure 1.2 is proved to control FDR under arbitrary dependence of the p -values. Procedure 1.3 is a more powerful procedure that is proved to control FDR under independence of p -values.

Procedure 1.2 (Lynch et al. (2014) Fixed Sequence method stopping on the k^{th} acceptance). Define,

$$\alpha_i = \begin{cases} \frac{\alpha}{k} & \text{if } i = 1, \dots, k \\ \frac{(m-k+1)\alpha}{(m-i+1)k} & \text{if } i = k + 1, \dots, m \end{cases} \quad (1.12)$$

Step 1: If $P_1 \leq \alpha_1$ then reject H_1 ; otherwise accept H_1 . If $k > 1$ or H_1 is rejected, then continue to test H_2 ; otherwise, stop.

Step 2: If $P_i \leq \alpha_i$ then reject H_i ; otherwise accept H_i . If the number of hypotheses accepted so far is less than k then continue to test H_{i+1} ; otherwise, stop.

Procedure 1.3 (Lynch et al. (2014) Fixed Sequence method stopping on the k^{th} acceptance under independence). Define,

$$\alpha_i = \frac{(r_{i-1} + 1)\alpha}{k + (i - k)\alpha} \quad i = 1, \dots, m \quad (1.13)$$

Step 1: If $P_1 \leq \alpha_1$ then reject H_1 ; otherwise accept H_1 . If $k > 1$ or H_1 is rejected, then continue to test H_2 ; otherwise, stop.

Step 2: If $P_i \leq \alpha_i$ then reject H_i ; otherwise accept H_i . If the number of hypotheses accepted so far is less than k then continue to test H_{i+1} ; otherwise, stop.

It is clear from these procedures that though they allow testing of hypotheses in the sequence, they do not allow testing of all the hypotheses and stop when k^{th} acceptance is observed. Choosing a large value for k makes these procedures very conservative. As a result these procedures are not suitable for the application related to real time decision making such as stream data analysis problem.

Farcomeni and Finos Sequential Procedure: The Farcomeni and Finos sequential procedure (Farcomeni and Finos, 2013) based on a possibly data driven ordering of the hypotheses is defined below. This procedure is proved to strongly control FDR under independence and positive regression dependence of p -values

Procedure 1.4 (Farcomeni and Finos (2013) Algorithm). Define, $J(i, \alpha) = i(1 - \alpha)/(2 - \alpha)$, where, $i = 1, \dots, m$.

- Let $i = 1$, $B(1) = I(P_1 < \alpha)$
- while $(B(i) > i - J(i, \alpha))$ and $i \leq m$ **do** $i = i + 1$, $B(i) = \sum_{j=1}^i I(P_j < \alpha)$.
- Let, $u = i - 1$.
- If $u > 0$, reject all hypotheses for $i = 1, \dots, u$ corresponding to $P_i < \alpha$. Do not reject hypotheses for $i > u$ even if $P_i < \alpha$

Farcomeni and Finos (2013) also give a slightly more conservative procedure under arbitrary dependence of the p -values. They show that if α is replaced with,

$$\alpha_D = \alpha / \sum_{j=1}^m \frac{2 - \alpha}{j + 1},$$

in Procedure 1.4, the same procedure controls FDR at level α under arbitrary dependence of p -values.

Though Procedure 1.4 tests the hypotheses as they occur in a sequence, the main drawback is that if $P_1 > \alpha$ then the whole testing stops. Also the correction factor of the critical constant α is pretty big which makes Procedure 1.4 under arbitrary dependence very conservative.

G'Sell et al. (2014): G'Sell et al. (2014) proposed a procedure for a multiple testing setting where the hypotheses are pre-ordered and one is only permitted to reject an initial contiguous block H_1, \dots, H_k , of hypotheses.

Procedure 1.5 (G'Sell et al. (2014) Forward Stop).

- Calculate,

$$\hat{k} = \max \left\{ k \in \{1, \dots, m\} : -\frac{1}{k} \sum_{i=1}^k \log(1 - P_i) \leq \alpha \right\}. \quad (1.14)$$

- Reject hypotheses $H_1, \dots, H_{\hat{k}}$.

This procedure is proved to control FDR when the true null p -values are independently drawn from $U[0, 1]$ distribution. The main drawback of this procedure is that we need to have all the data at hand to calculate \hat{k} . This drawback makes it unsuitable for some applications related to real time decision making.

1.6 Motivation and Thesis Outline

In this dissertation, we focus on developing newer multiple testing procedures, for several applications, that account for inherent structure of the tested hypotheses and control directional errors in addition to Type I errors in some cases as required by the problem of interest. In the following section, we discuss the motivation behind the research.

In fields like drug development, genomics, fMRI studies etc., where a large number of experiments are conducted daily with a varied nature of scientific questions in the background, hypotheses with a stage-wise hierarchical structure arise very commonly. In more complex cases, the hypotheses are grouped into several hierarchical families, and the families are tested in a sequential order. The theoretical and methodological issues related to multiple testing problems in this kind of special structured hypotheses, with appropriate error rate control, require a lot of attention

for investigation and development. A very relevant application arises in the analysis of Uterine Fibroid Gene expression data (Davis et al., 2013) where gene expressions of uterine fibroids and normal uterine tissues are reported. The fibroid samples are from various size categories that correspond to different stages of tumor formation and progression. The question of interest is to discover gene expression patterns in the tumor tissues compared to the normal tissue. We answer this question by formulating multiple testing of hypotheses that identify differentially expressed genes in the fibroids compared to normal tissue, genes specific to different tumor sizes and finally give the gene expression direction in tumor tissues compared to normal tissue. The hypotheses formulated form a family of hypotheses for each gene consisting of one global hypothesis for testing differential expression of the gene and several pairwise hypotheses for testing differential expression in each tumor size compared to the normal tissue. Directional decisions are made once a gene is found to be differentially expressed in a tumor size. This application motivates us to develop newer mdFDR controlling multiple testing procedure that identify genes and expression patterns while controlling overall Type I and Directional errors.

In applications related to clinical trials, dose response study, etc., the sequence of the tested hypotheses is often decided based on their relative importance or by some prior knowledge. For example, in dose-response studies, the hypotheses pertaining to a higher dose are tested before those corresponding to a lower dose; in clinical trials, the hypotheses corresponding to a primary endpoint are tested before those corresponding to secondary endpoint and so on. When the hypotheses are both sided hypotheses in such applications, it is often of interest to decide on direction once the null hypothesis of no difference is rejected. Such decisions may lead to Type III or Directional errors. So while developing a multiple testing procedure for such applications, it is desired to control both Type I and Type III errors. The control of mdFWER in multiple testing is a very challenging problem and the discussion

under dependence of the p -values is the notoriously challenging problem. This also motivates us to consider the simple fixed sequence multiple testing problem and develop mdFWER controlling procedures under independence and different kinds of dependence of the p -values.

In the field of stream data analysis, statistical process control etc., the hypotheses have a natural hierarchy due to time of occurrence of data. While testing the hypotheses one has information only up to that time, which is only partial information. Also as the testing sequence has a large number of hypotheses, it is desired to control FDR. In this dissertation, we develop multiple testing procedures that exploit the intrinsic hierarchy and structure of the hypotheses while allowing all hypotheses to be tested irrespective of acceptances and controlling FDR. These procedures are capable of making real time decisions based on the partial information and conclude whether a process is going out of control or not.

The rest of the dissertation is outlined as follows. In Chapter 2, we present a general mixed directional false discovery rate (mdFDR) controlling procedure for testing of multiple families of hierarchically ordered hypotheses, where later hypotheses can be tested only if their earlier counterparts are rejected with directional decisions made in the final stage. In Chapter 3, we develop a specific methodology based on the general procedure for application to uterine fibroid data discussed in Davis et al. (2013) and present related simulation studies. In Chapter 4, we present new mdFWER controlling procedures for fixed sequence multiple testing. In Chapter 5, we present FDR controlling procedures in the context of real time testing of hypotheses while data is continuously collected as in stream data or statistical process control. In Chapter 6, we summarize the discussions and present several possible future works.

CHAPTER 2

CONTROL OF MDFDR IN TESTING HIERARCHICALLY STRUCTURED FAMILIES OF HYPOTHESES AND DIRECTIONAL DECISIONS

2.1 Introduction

Increasingly it is a commonplace for researchers to conduct large scale genomic studies involving multiple experimental groups along with a control group, also called the normal or the reference group. The goal is to determine features that are differentially expressed in a given experimental group (relative to the reference group) and to determine if a differentially expressed feature is up or down regulated. For example, a toxicologist may be interested in identifying differences in the gene expression profile of spontaneous tumors and chemically induced tumors, relative to normal tissues (Hoenerhoff et al., 2011, 2012; Pandiri et al., 2011, 2012). There is considerable interest among cancer researchers to understand the gene expression profile of tumors according to tumor size (Diaz et al., 2005; Gieseg et al., 2004; Hu et al., 2013; Minn et al., 2007; Riis et al., 2013). For tumor onset and progression, it may be necessary for some genes to express at all stages of tumor growth and development (i.e. express in tumors of all sizes). However, some genes may express only at some specific stages/sizes depending upon their function. For example, some may only be involved during the early stages of tumor formation and others may be necessary for tumor progression. Identification of genes according to tumor stages (or size) may therefore have clinical implications. Accordingly, this has been an active area of research for various cancers over the past decade (Ciarmela et al., 2011; Diaz et al., 2005; Gieseg et al., 2004; Riis et al., 2013). Thus, it is clear there is considerable interest among clinicians and biologists to investigate the expression of genes according to the tumor

size or category. In all such investigations, one is typically interested in performing several pairwise comparisons, of thousands of features, relative to a reference group (e.g. normal tissue). Often researchers are not only interested in determining if a feature is differentially expressed but are also interested in determining whether it is up or down-regulated in the experimental group (relative to the reference group). For simplicity of exposition, throughout this dissertation we shall replace the term “feature” by “gene”.

Multiple testing problems involving multiple pairwise comparisons of high dimensional data along with directional decisions has not received much attention in the literature, yet such testing problems are commonly encountered in practice. When the number of genes is very small (perhaps in tens) several methods have been proposed that control the directional errors as well as the family wise error rate (Finner, 1994, 1999; Liu, 1997; Sarkar et al., 2004; Shaffer, 1980). However, such methods are very conservative when the number of genes is very large as in a microarray data or CpG methylation data. Several ad-hoc methods and strategies are used in the literature when the number of genes is large. For example, some researchers apply multiple testing procedures (e.g. the Benjamini-Hochberg (BH) or Bonferroni procedure) within each pairwise comparison and ignore the fact that they are conducting several pairwise comparisons. Once a differentially expressed gene (DEG) for a pairwise comparison is identified then they are declared to be up or down-regulated by looking at the direction of the fold change (or the test statistic) without accounting for the statistical error associated with such a directional decision. Such strategies result in an inflated overall FDR due to multiple pairwise comparisons and directional decisions. Another important approach for pattern identification in time-course microarray data is given by Sun and Wei (2011), where they formulate a compound decision-theoretic framework for set-wise multiple testing and propose a

data-driven procedure to identify genes that exhibit a specific pattern of differential expression over time.

The only formal methodology available in the literature that controls the mdFDR for the above directional multiple testing problems is the method by Guo et al. (2010), which is designed to make decisions on thousands of features when making multiple pairwise comparisons and deciding on the direction of comparison. Thus, the method controls the false discovery rate when making multiple pairwise comparisons on thousands of genes while also controlling the directional errors committed when falsely declaring a DEG to be up-regulated (or down-regulated) when it is not. Guo et al. (2010) procedure generalizes the procedure of Benjamini and Yekutieli (2005) which was designed for the case when there were only two groups to compare.

While Guo et al. (2010) methodology is useful for making several multiple pairwise comparisons; it is relatively conservative since it relies on the Bonferroni procedure to deal with multiple pairwise comparisons within each significant gene. In this chapter, we develop a general mdFDR controlling testing procedure that allows us to use any mixed directional familywise error rate (mdFWER) controlling procedure in place of the Bonferroni procedure, for conducting pairwise comparisons in high dimensional data that is broadly applicable to a wide range of genomic data including gene expression microarray data, CpG methylation data, RNA-seq data and others. We prove that the procedure controls mdFDR when the underlying test statistics are independent across the genes. Based on this general procedure, we develop a specific methodology in Chapter 3 using the Dunnett's test (Dunnett, 1955; Dunnett and Tamhane, 1991, 1992) which is designed for making comparisons of several experimental groups with the control or the reference group.

The rest of the chapter is organized as follows. Section 2.2 gives a summary of notation and concepts used in this chapter and describes the formulation of the

problem. Section 2.3 introduces a general three-step procedure. Section 2.4 gives some insight on developing specific methodology based on the general procedure. In Section 2.5 we discuss some future work related to this project.

2.2 Background, Notations and Problem Formulation

Figure 2.1 shows a typical microarray gene expression data with genes arranged vertically and tumor samples (categorized by, for example, size of tumor) arranged horizontally. In the uterine fibroid gene expression data explained in Davis et al.

	Sample 1	Sample n
Gene 1	gene expressions		
Gene 2			
.		.	
.		.	
.		.	
Gene m			

Figure 2.1 A typical microarray gene expression data arranged in rows and columns. Rows represent genes and columns represent tissue samples.

(2013), the tumor samples are categorized according to tumor size, tumor location, race and age of women the samples are taken from, etc. The interest here is to identify genes that are differentially expressed across these categories and find the pattern of gene expression across these categories. For example, an important question of interest is identifying genes and pathways that are specific to early stages of tumor formation (i.e. small/tiny tumors), genes that are specific to tumor growth (i.e. medium size tumors) and those genes that are specific to very large tumors which may be necrotic. By identifying such genes, the researcher 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. This kind of analysis requires several steps of analysis that requires multiple testing of families of hypotheses and making multidimensional directional decisions.

Motivated by this we formulate the problem as a three-step multiple testing problem where we first test for significance of each gene for differential expression across categories, then test families of hypotheses corresponding to significant genes and finally conclude on the pattern of gene expression by making directional decisions. The procedure gives rise to the possibility of committing two kinds of errors: *Type I* and *Directional* errors or *Type III* errors. So the overall error measure we want to control is *mixed directional FDR* (mdFDR). For identifying gene expression patterns for different sizes of tumors, the three steps are described as follows.

- Step 1: Identify genes that are differentially expressed in tumor tissues compared to normal tissues. We test a global hypothesis corresponding to each family (gene), testing for difference in expression in tumor samples compared to normal samples while controlling FDR.
- Step 2: Identify in which category (e.g., size of tumor) these genes are differentially expressed compared to normal tissues. For each significant family, we test individually the difference in expression of tumor samples of each size compared to normal tissue.
- Step 3: Identify the direction of expression compared to normal tissue. For the hypotheses that are rejected in Step 2, we decide on direction of expressions of tumor samples (pertaining to different categories) versus normal samples, while controlling mdFWER combining Steps 2 and 3. Overall we aim at controlling mdFDR.

We now introduce some notations and definitions related to the problem. Let m denote the number of genes in the data that has gene expressions for each gene on p categories. Let μ_{ij} denote the mean response corresponding to the i -th category in j -th gene, $i = 1, \dots, p$, $j = 1, \dots, m$. Let q denote the number of hypotheses to be

tested based on a set of pairwise differences of mean responses of p categories. The set of q hypotheses in each gene depend on the problem of interest and a few examples are as follows. A problem of biological interest in the context of uterine fibroid data is to group genes that are differentially expressed in a size category of tumor compared to normal sample, if category “ p ” correspond to normal sample, then we need to test the differences $\theta_{ij} = \mu_{ij} - \mu_{pj}$, $i = 1, 2, \dots, q$ and $j = 1, 2, \dots, m$, here, $q = p - 1$.

Based on the three steps, we introduce the hypotheses to be tested in each step as follows. For each gene j , we have a vector of parameters $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{qj})$. We define the null and alternative “screening hypotheses” to test the significance of each gene as,

$$H_{0screen}^j : \boldsymbol{\theta}_j = \mathbf{0} \text{ against } H_{1screen}^j : \boldsymbol{\theta}_j \neq \mathbf{0}, \quad j = 1, \dots, m. \quad (2.1)$$

For each gene j , the component null and alternative hypotheses are,

$$H_{0i}^j : \theta_{ij} = 0 \text{ against } H_{1i}^j : \theta_{ij} \neq 0, \quad i = 1, \dots, q. \quad (2.2)$$

Figure 2.2 shows a simple graphical representation of the structure of hypotheses in our formulation. Let T_{ij} and P_{ij} , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, m$, denote the test

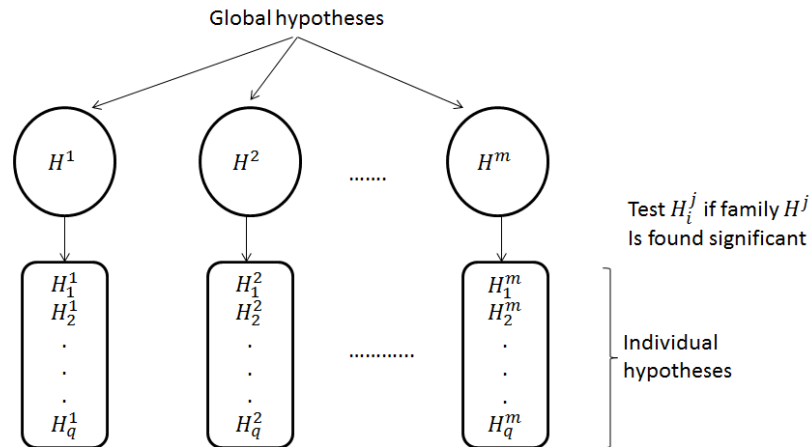


Figure 2.2 A graphical display of the hypotheses in uterine fibroid gene expression data problem.

statistics and the p -values respectively for testing the hypotheses in equation (2.2).

The screening hypotheses form a global test which test whether all parameters θ_{ij} , $i = 1, \dots, q$ are simultaneously 0 or not or equivalently, whether all μ_{ij} , $i = 1, \dots, p$ are equal or not. We denote the p -values for testing screening hypotheses in equation (2.1) as P_{screen}^j . For each family j we denote a vector of p -values, $\mathbf{P}_j = (P_{1j}, P_{2j}, \dots, P_{qj})$ based on the test statistics $\mathbf{T}_j = (T_{1j}, T_{2j}, \dots, T_{qj})$, for testing the component hypotheses in equation (2.2). If H_{0i}^j is rejected we conclude on direction, i.e., declare $\theta_{ij} > 0$ if $T_{ij} > 0$ or declare $\theta_{ij} < 0$ if $T_{ij} < 0$.

Given the screening p -values P_{screen}^j , for every $j = 1, 2, \dots, m$, to carry out the simultaneous testing of the screening hypotheses in equation 2.1, we use the BH-procedure (Benjamini and Hochberg, 1995), as suggested by Guo et al. (2010), that controls the FDR at a given level α . This is a step-up procedure as follows: given ordered p -values $P_{screen}(1) \leq P_{screen}(2) \leq \dots \leq P_{screen}(m)$ and the corresponding null hypotheses $H_{0screen}(1), H_{0screen}(2), \dots, H_{0screen}(m)$, find $R = \max \{1 \leq j \leq m : P_{screen}(j) \leq j\alpha/m\}$ and reject $H_{0screen}(1), \dots, H_{0screen}(R)$ provided the maximum exists, otherwise, accept all the screening hypotheses.

When an $H_{0screen}^j : \boldsymbol{\theta}_j = 0$ is rejected using the BH procedure, further decisions are made on the component hypotheses in equation (2.2) and on rejection, directional decisions are made on the signs of the component θ_{ij} . A Type I error might occur due to wrongly rejecting $H_{0i}^j : \theta_{ij} = 0$ and a directional error might occur due to correctly rejecting $H_{0i}^j : \theta_{ij} = 0$ but wrong assignment of the sign of θ_{ij} . So we need to control Type I as well as Type III (directional) errors. A practical way of doing that would be to use an error rate combining both Type I and Type III errors in the FDR framework and make sure that it is controlled.

An error rate that combines Type I errors and Type III errors in FWER setup is mdFWER (Finner, 1999; Shaffer, 1980), see also equation (1.1). Benjamini and Heller (2008) introduced the error rate Overall False Discovery Rate (OFDR)

in the context of testing of partial conjunction hypotheses. Heller et al. (2009) used this error rate, as an appropriate error measure to control, in their two-stage procedure for identifying differentially expressed gene sets. Inspired by Heller et al. (2009) and Shaffer (1980) we augment directional decisions to the two-stage OFDR controlling procedure and develop a three-stage procedure that controls mdFDR as appropriate overall error measure while identifying significant families, consequently finding significant hypotheses and finally concluding on direction of parameters for those significant hypotheses. The definition of overall mdFDR is given below.

Let $V(j)$ denote the indicator function of at least one Type I error or Directional Error committed while testing family j and the component hypotheses in it, i.e., $V(j)$ is 1 if either $H_{0screen}^j$ is falsely rejected or $H_{0screen}^j$ is correctly rejected but at least one Type I error or Directional error occurs while testing component hypotheses corresponding to gene j ; $V(j)$ is 0 otherwise. Let R denote the number of families discovered. Then mdFDR is formally defined as follows.

Definition 2.1 (mdFDR - mixed directional False Discovery Rate). The expected proportion of Type I and Directional errors among all discovered families.

$$mdFDR = E \left[\frac{\sum_{j=1}^m V(j)}{R \vee 1} \right]. \quad (2.3)$$

2.3 A General Mixed Directional FDR Controlling Procedure

In this section we present a general procedure for testing hierarchically structured families of hypotheses, given in Section 2.2, with multidimensional directional decisions made in the final step, while controlling the mdFDR as the overall error measure. The general procedure is summarized as follows:

Procedure 2.1 (General Hierarchical Procedure Controlling mdFDR).

1. Use BH procedure to find significant genes. Let $P_{screen}(j)$, $j = 1, \dots, m$, be the ordered screening p -values ordered from smallest to largest, and $H_{screen}(j)$

be the corresponding screening hypotheses. Conclude gene j as significant if $P_{screen}(j) \leq \frac{j\alpha}{m}$. Let R denote the number of significant genes.

2. For each significant gene, use any mdFWER controlling procedure such as Holm, Hochberg etc., at level $\frac{R\alpha}{m}$.
3. For the component hypotheses rejected for each significant gene, conclude on direction based on the sign of the test statistic to identify directional pattern.

Remark 2.1. This procedure is general in the sense that any global testing method can be used to obtain the screening p -values in Step 1 of the procedure, any pairwise comparison testing method can be used to obtain the pairwise p -values in Step 2 and any mdFWER controlling procedure can be used in Steps 2-3 of the procedure.

Remark 2.2. Note here that the method of Guo et al. (2010) is a special case of the proposed general procedure in which Bonferroni global test is used for testing the screening hypotheses and Bonferroni method along with additional directional decision works as the mdFWER controlling procedure.

It is important to point out that the goal of this chapter is to identify expression patterns of m genes over p categories which give rise qm hypotheses in total. We treat this as a problem of performing m tests each involving q -dimensional hypotheses. The procedure controls FDR in Step 1 while testing the m screening hypotheses using BH procedure. Each significant family of hypotheses is then tested in Step 2 using an mdFWER controlling procedure. So the control of overall error rate is achieved by controlling the mdFWER for each individual family.

2.3.1 mdFDR Control under Independence

Assumption 2.1 (Independence). *The test statistics vectors $\mathbf{T}_j = (T_{1j}, T_{2j}, \dots, T_{qj})$, $j = 1, \dots, m$ are independent and consequently, the P -value vectors $\{\mathbf{P}_j, j = 1, 2, \dots, m\}$ are independent.*

Theorem 2.1. *Under Assumption 2.1, the mdFDR of Procedure 2.1 is strongly controlled at level α for any procedure controlling mdFWER at level $\frac{R\alpha}{m}$ at Step 2 of Procedure 2.1.*

Proof Let, I_0 denote the set of true null screening hypotheses $H_{0screen}^j$ and I_1 denote set of false $H_{0screen}^j$ with $|I_0| = m_0$ and $|I_1| = m_1$, $m_0 + m_1 = m$. From definition 2.1,

$$mdFDR = E(Q) = E \left[\frac{\sum_{j=1}^m V(j)}{R \vee 1} \right].$$

In the event that $R = r$, $P_{screen}(k) \leq r\alpha/m$ for $k = 1, 2, \dots, r$ and $P_{screen}(k) > (r+1)\alpha/m$ for $k = r+1, \dots, m$. Consequently, we have r number of P_{screen}^j 's that are $\leq r\alpha/m$. Then $mdFDR$ will be equal to,

$$\begin{aligned} E(Q) &= \sum_{r=1}^m \frac{1}{r} \left[\sum_{j=1}^m Pr(V(j) = 1, R = r) \right] \\ &= \sum_{r=1}^m \sum_{j \in I_0} \frac{1}{r} Pr \left(P_{screen}^j \leq \frac{r\alpha}{m}, R^{(-j)} = r - 1 \right) \\ &\quad + \sum_{r=1}^m \sum_{j \in I_1} \frac{1}{r} Pr \left(P_{screen}^j \leq \frac{r\alpha}{m}, \text{Type I or Type III error at } j, R^{(-j)} = r - 1 \right). \end{aligned} \tag{2.4}$$

where, $R^{(-j)}$ denotes the number of screening hypotheses rejected from the set $\{H_1, H_2, \dots, H_{j-1}, H_{j+1}, \dots, H_m\}$, the set of screening hypotheses excluding the j -th one.

Consider the second term in equation (2.4):

$$\begin{aligned} &\sum_{r=1}^m \sum_{j \in I_1} \frac{1}{r} Pr \left(P_{screen}^j \leq \frac{r\alpha}{m}, \text{type I or type III error at } j, R^{(-j)} = r - 1 \right) \\ &= \sum_{r=1}^m \sum_{j \in I_1} \frac{1}{r} Pr \left(P_{screen}^j \leq \frac{r\alpha}{m}, \text{type I or type III error at } j \right) Pr(R^{(-j)} = r - 1) \end{aligned} \tag{2.5}$$

$$\leq \sum_{r=1}^m \sum_{j \in I_1} \frac{1}{r} \frac{r\alpha}{m} Pr(R^{(-j)} = r - 1) \quad (2.6)$$

$$= \frac{m_1 \alpha}{m}. \quad (2.7)$$

The equality in (2.5) follows due to Assumption 2.1. The inequality in (2.6) follows due to the procedure we use which controls the mdFWER at level $\frac{r\alpha}{m}$ and as $j \in I_1$, the probability of making at least one Type I error or directional error in family j is $\leq \frac{r\alpha}{m}$. Summing over all values of r , the equality in (2.7) follows by noting that $\sum_{r=1}^m Pr(R^{(-j)} = r - 1) = 1$.

Next consider the first term in eq (2.4):

$$\begin{aligned} & \sum_{r=1}^m \sum_{j \in I_0} \frac{1}{r} Pr\left(P_{screen}^j \leq \frac{r\alpha}{m}, R^{(-j)} = r - 1\right) \\ &= \sum_{r=1}^m \sum_{j \in I_0} \frac{1}{r} Pr\left(P_{screen}^j \leq \frac{r\alpha}{m}\right) Pr(R^{(-j)} = r - 1) \end{aligned} \quad (2.8)$$

$$\leq \sum_{r=1}^m \sum_{j \in I_0} \frac{1}{r} \frac{r\alpha}{m} Pr(R^{(-j)} = r - 1) \quad (2.9)$$

$$= \frac{m_0 \alpha}{m}. \quad (2.10)$$

The equality in (2.8) follows due to Assumption 2.1. The inequality in (2.9) follows due to the procedure we use, it controls the mdFWER at level $\frac{r\alpha}{m}$. Summing over all values of r , the equality in (2.10) follows by noting that $\sum_{r=1}^m Pr(R^{(-j)} = r - 1) = 1$. The result follows by combining equations (2.7) and (2.10). \blacksquare

Remark 2.3. Note that this theorem is proved under the assumption that the m P -value vectors are independent. This assumption is used only in Step 1 of the general procedure. In the case that we are not sure whether this assumption is satisfied or not, the BH procedure can be replaced by the more conservative BY procedure (Benjamini and Yekutieli, 2001), which shows control of FDR under arbitrary dependence. We do not yet make any assumption about the structure of p -values within a family.

Assumption 2.1 only implies that the p -values across families are independent and which in turn implies that the summary p -values P_{screen}^j are independent.

2.4 Methodology

Several application specific methodology can be developed from this general procedure by choosing appropriate methods to obtain P_{screen}^j and P_{ij} , $i = 1, \dots, q$, $j = 1, \dots, m$ and choosing an appropriate mdFWER controlling procedure for Steps 2 and 3. In this section we discuss a few ways of obtaining these P_{screen}^j and P_{ij} .

2.4.1 Choices of Methods for P_{screen}^j

For testing $H_{0screen}^j$ vs. $H_{1screen}^j$, we have several types of the global testing methods available based on which the screening p -values, P_{screen}^j , can be obtained. A few examples are Bonferroni method, ANOVA, Dunnett method. Here, we discuss how to obtain P_{screen}^j 's for the first two methods, the Dunnett method will be discussed in details in Chapter 3.

Let x_{ik}^j denote the gene expressions, for, $i = 1, \dots, p$ categories, $j = 1, 2, \dots, m$ genes, and $k = 1, 2, \dots, n_i$ samples in category i . Let T_{ij} denote the test statistic for testing the component null hypotheses in equation (2.2). Let $T_{ij} \sim F_{ij}(t, \theta_{ij})$ for some continuous cdf F which is symmetric about 0 under H_{0i}^j with $F_{ij}(t, \theta_{ij}) \geq F_{ij}(t, 0)$ if $\theta_{ij} < 0$ and $F_{ij}(t, \theta_{ij}) \leq F_{ij}(t, 0)$ if $\theta_{ij} > 0$. The two-sided P -value for testing (2.2) is defined as,

$$P_{ij} = 2 \min \{F_{ij}(T_{ij}, 0), 1 - F_{ij}(T_{ij}, 0)\}. \quad (2.11)$$

Bonferroni Adjusted P_{screen} : If we treat $H_{0screen}^j$ as an intersection of the component null hypotheses, that is, $H_{0screen}^j = \cap_{i=1}^q H_{0i}^j$ and $H_{1screen}^j = \cup_{i=1}^q H_{1i}^j$, then we can combine the three component p -values $\{P_{1j}, \dots, P_{qj}\}$ by using Bonferroni

adjusted pooling method and get the Bonferroni adjusted P_{screen} 's as,

$$P_{screen}^j = q \min \{P_{1j}, \dots, P_{qj}\}. \quad (2.12)$$

The procedure to obtain T_{ij} and the assumptions of their distribution need to be discussed here. The data can be assumed to be approximately normal as the data we have are normalized gene expressions. Then the test statistic for testing the component hypotheses (2.2) is given by the two sample t -test statistic,

$$T_{ij} = \frac{\bar{x}_{ij} - \bar{x}_{pj}}{s_{pooled(ip)j} \sqrt{\frac{1}{n_i} + \frac{1}{n_p}}}, \quad (2.13)$$

where, for $i = 1, \dots, q$,

$$s_{pooled(ip)j}^2 = \frac{(n_i - 1)s_{ij}^2 + (n_p - 1)s_{pj}^2}{n_i + n_p - 2}, \quad (2.14)$$

and for $i = 1, \dots, p$,

$$\bar{x}_{ij} = \frac{\sum_{k=1}^{n_i} x_{ij}^k}{n_i}, \quad (2.15)$$

$$s_{ij}^2 = \frac{\sum_{k=1}^{n_i} (x_{ij}^k - \bar{x}_{ij})^2}{n_i - 1}. \quad (2.16)$$

The test statistic T_{ij} has a t -distribution with $(n_i + n_p - 2)$ degrees of freedom and $s_{pooled(ip)j}^2$ is the pooled variance of groups i and p . The corresponding component p -values are given by,

$$P_{ij} = 2 \times (1 - G_t(|T_{ij}|, n_i + n_p - 2)), \quad i = 1, \dots, q, \quad (2.17)$$

where, $G_t(\cdot, n_i + n_p - 2)$ denotes the CDF of t -distribution with $n_i + n_p - 2$ degrees of freedom. This is the procedure, adopted by Guo et al. (2010), for obtaining P_{screen} .

ANOVA P_{screen} : ANOVA gives us a direct method to obtain P_{screen}^j from the data for testing (2.1). The ANOVA global F -test tests the hypotheses $H_0^j : \mu_{1j} = \dots = \mu_{pj}$

vs. H_1^j : at least one μ_{ij} is not equal. Rejection of this null hypothesis is equivalent to rejection of null hypothesis in (2.1). The test statistic for testing (2.1) is given by the ratio of between group variance and within group variance,

$$T_j^{ANOVA} = \frac{\sum_{i=1}^p n_i (\bar{x}_{ij} - \bar{x}_j)^2 / (p - 1)}{\sum_{i=1}^p \sum_{k=1}^{n_i} (x_{ij}^k - \bar{x}_{ij})^2 / (\sum_{i=1}^p n_i - p)}, \quad (2.18)$$

where,

$$\bar{x}_j = \frac{\sum_{i=1}^p \sum_{k=1}^{n_i} x_{ij}^k}{\sum_{i=1}^p n_i}, \quad (2.19)$$

and the other terms are as described in equations (2.14)-(2.16). The null distribution of T_j^{ANOVA} is $F(\cdot, p - 1, \sum_{i=1}^p n_i - p)$, an F -distribution with $(p - 1, \sum_{i=1}^p n_i - p)$ degrees of freedom.

2.4.2 Choices of mdFWER Controlling Methods

The statistical problem in Steps 2 and 3 is to decide how to construct the test statistics for each component hypothesis and obtain the corresponding component p -values that will then be used in an appropriately chosen mdFWER controlling testing procedure to identify significantly expressed categories and direction of expression. For an identified gene, each of the component null hypothesis is tested using the two-sample t -test statistic. The component hypotheses test statistics can be obtained using (2.13) and the corresponding p -values are obtained using (2.11), where, $F_t(\cdot, n_i + n_p - 2)$ denotes the CDF of t -distribution with $(n_i + n_p - 2)$ degrees of freedom, $i = 1, \dots, q$.

We present a few mdFWER controlling multiple testing methods for identification of significantly differentially expressed categories and describe how the methods are implemented. Some of the few available mdFWER controlling procedures are Holm procedure, Hochberg procedure, Fixed Sequence procedure, Bonferroni procedure and Dunnett procedure. The details of the Dunnett procedure are described in Chapter 3. We discuss the rest of the procedures below. For each

procedure, once a null hypothesis is rejected, the sign of the parameter θ_{ij} is decided based on the sign of the test statistic T_{ij} . Holm's procedure and Hochberg's procedure use ordered p -values. Let, $P_j(1) < P_j(2) < \dots < P_j(q)$ denote the ordered P_{ij} and the corresponding component null hypotheses be $H_0^j(1), H_0^j(2), \dots, H_0^j(q)$.

Holm Procedure: Shaffer (1980) proved that the Holm's procedure (Holm, 1979), when augmented with directional decisions, can control mdFWER under independence of the p -values. We use Holm's step-down procedure at level $\frac{R\alpha}{m}$ within each significant gene. Let k be the maximum index such that $P_j(k) \leq \frac{R\alpha}{m(q-k+1)}$, $k = 1, 2, \dots, q$, then reject $H_0^j(1), \dots, H_0^j(k)$ and accept the rest of the hypotheses.

Hochberg Procedure: Liu (1997) and Finner (1999) independently proved that the Hochberg's procedure (Hochberg, 1988), when augmented with directional decisions, can control mdFWER when the p -values are independent. We use Hochberg's step-up procedure at level $\frac{R\alpha}{m}$ within each significant gene to identify significant categories. Let k be the minimum index such that $P_j(k) > \frac{R\alpha}{m(q-k+1)}$, $k = 1, 2, \dots, q$, then reject $H^j(1), \dots, H^j(k-1)$ and accept the rest of the hypotheses.

Fixed Sequence Procedure: Fixed Sequence procedure is commonly used in scenarios where the order of testing of hypotheses is fixed beforehand. We prove in Chapter 4 that the Fixed Sequence procedure augmented with directional can strongly control mdFWER when the p -values are independent or positively dependent. We use Fixed Sequence procedure at level $\frac{R\alpha}{m}$ within each significant gene to identify significant categories. Let k be the minimum index such that $P_{kj} > \frac{R\alpha}{m}$, $k = 1, 2, \dots, q$, then reject $H_{01}^j, \dots, H_{0k-1}^j$ and accept the rest of the hypotheses.

Bonferroni Procedure: Bonferroni procedure is a single step multiple testing procedure and it can strongly controls mdFWER when augmented with directional decisions. We use Bonferroni procedure, as described in Guo et. al. (2010), at level $\frac{R\alpha}{m}$ within each significant gene to identify significant categories. The procedure rejects H_{0i}^j if $P_{ij} \leq \frac{R\alpha}{qm}$.

2.5 Discussion and Conclusion

In this chapter, we developed a general mdFDR controlling procedure that allows testing of hierarchically ordered families of hypotheses while making multidimensional directional decisions. We have shown that Procedure 2.1 controls mdFDR as an overall error measure under the independence of test statistic vectors across families. We have not assumed any specific dependence structure for the test statistics within a family, so the joint distribution of the statistics within a family may have any dependence structure. The simulation studies that we present in Chapter 3 give an indication that we can relax the assumption of independence of the test statistic vectors.

The generality of this procedure makes it a flexible procedure to apply to several practical situations where multidimensional directional decisions are required to make. Although, in Section 2.2 we discuss comparison of gene expressions in each tumor size to the gene expressions in the normal sample, this procedure can be applied to any type of pairwise comparison desired to be tested for each gene. For example, if it is of interest to group genes by the inequalities among the mean responses, we would want to detect the pattern of mean responses in the p categories, known as directional pattern, and see how the mean responses vary across the categories. Some common inequalities are $\mu_{1j} \leq \mu_{2j} \leq \dots \leq \mu_{pj}$ (monotone pattern), $\mu_{1j} \leq \dots \leq \mu_{ij} \geq \mu_{(i+1)j} \geq \dots \mu_{pj}$ (umbrella pattern with peak μ_{ij}). To test for the pattern we need to test the differences of mean response of the categories, $\theta_{ij} = \mu_{i+1j} - \mu_{ij}$, $i = 1, 2, \dots, q$ and $j = 1, 2, \dots, m$ and $q = p - 1$. If the problem of interest is testing all pairwise differences of the p categories, possibly unordered, then $q = p(p - 1)/2$. Based on the question we want to answer from a data, appropriate methodology can be developed from this general procedure.

Hypotheses such as the above can also be tested in the order restricted inference framework as done in Peddada et al. (2003, 2005), a problem with order restricted

inference based methods is that one cannot distinguish between strict inequalities and non-strict inequalities. Although such issues also arise in the present setting because failure to reject null does not imply that the null is true, to a larger extent the problem is reduced by adopting the present strategy. One could consider including a bioequivalence type testing to make sure the null hypothesis is true, but that is beyond the scope of this research.

The general procedure proposed in this chapter provides an interesting view towards the challenging problem of controlling both Type I and Directional errors in multiple testing involving multidimensional parameters. We showed control of overall mdFDR of Procedure 2.1 under assumption of independence across the families. In microarray data, gene expressions are obtained by drawing samples from same subjects. In such cases, there may be dependence in gene expressions from several genes, which leads to dependence of test statistics across families. It will be interesting to theoretically investigate the performance of the proposed procedure under such dependence structures.

CHAPTER 3

METHODOLOGY FOR UTERINE FIBROID DATA

3.1 Introduction

In this chapter, we develop a specific methodology based on the general procedure, developed and discussed in Chapter 2, to analyze a gene expression data (Davis et al., 2013) obtained from the NIEHS Fibroid Growth Study (FGS) (Peddada et al., 2008). Uterine fibroids, also called uterine leiomyoma, are benign smooth muscle tumors which are hormonally mediated. According to some estimates, the cumulative incidence by age 50 of these tumors among Caucasian women exceeds 70% and it is much higher among women of African American descent. The direct and indirect annual cost of fibroids in the US is as high as 34 billion dollars. While several studies investigated the molecular characteristics of these benign tumors relative to normal myometrium (Davis et al. (2013) and references therein) not many have been conducted to identify genes that are specific to fibroid size. However, such studies have been conducted for other tumors (Diaz et al., 2005; Gieseg et al., 2004; Hu et al., 2013; Minn et al., 2007; Riis et al., 2013). Using the gene expression data obtained in the NIEHS FGS and the methodology developed in this chapter, we identify DEGs and pathways that are specific to tumor size of uterine fibroids.

Since the idea is to compare the gene expressions in different sizes of fibroids to the expression in normal uterine tissue, the comparison is of the type where several treatments are compared to a common control. Dunnett procedure (Dunnett, 1955), a single step procedure which uses a multivariate t -distribution of the test statistics to derive the p -values, is specifically introduced for testing hypotheses in this kind of comparison. We incorporate this procedure into the general Procedure 2.1 to obtain a methodology appropriate for analyzing the FGS gene expression data. Not

only that the resulting methodology is practically relevant but as demonstrated in the numerical simulations, the resulting methodology not only controls the mdFDR but is more powerful relative to some potential alternative methods. We report the results of a simulation study evaluating the performance of the proposed procedure under independence and dependence of the underlying test statistics and compare our procedure to the method of Guo et al. (2010). The simulation studies show this methodology to have highest power among other relevant methodologies developed from Procedure 2.1.

Using our methodology we gain deeper insights into molecular characteristics of uterine fibroids according to the tumor size. We have identified several differentially expressed genes and pathways that are specifically enriched according to the tumor size (or stage of growth). While researchers and clinicians who study fibroids are well aware of many of the genes and pathways described in this chapter, we have provided a characterization of these genes and pathways according to the tumor stage. Our data can be further mined to gain deeper insights regarding fibroids.

The rest of the chapter is organized as follows. Section 3.2 describes the gene expression data. Section 3.3 describes the methodology developed, derived from the general procedure introduced in the Chapter 2, for the analysis of data. In Section 3.4, we present a simulation study to compare the specific procedures we consider. In Section 3.5, we present the results of the data analysis. In Section 3.6, we give a brief summary and discussion of the results.

3.2 Uterine Fibroid Gene Expression Data

In their study, Peddada et al. (2008) prospectively tracked growth of fibroids in 72 premenopausal women (38 black and 34 white) over 12 months period and found that growth rates of fibroids were on average much higher in older black women than in older white women. Davis et al. (2013) reported gene expression pattern differences in

tumors and myometrium samples from 12 study participants who underwent surgery during the course of the study. They analyze 52 leiomyoma (fibroid) and 8 myometrial (normal tissue) samples using Affymetrix Gene Chip expression arrays and report genes that were found significant in the comparison of tumor and normal myometrium samples.

The data contains normalized gene expressions of 54675 probe sets from 52 tumor samples and 8 normal myometrium samples, so in total we have 60 tissue samples. The tumor samples are classified by tumor size into three groups: Small (14 samples, volume: 0.08-5.70), Medium (25 samples, volume: 9.0-132.00), Large (13 samples, volume: 240-2016). The three sizes of fibroids and normal samples form four ordered categories of the attribute “size of tumor”. We have data for each gene in an unbalanced one-way format.

3.3 Methodology for FGS Gene Expression Data

Suppose we are interested in comparing “ q ” experimental groups with a reference group (in total, $p = q + 1$ groups) on the basis of the mean expression of “ m ” genes. For example, suppose we are interested in comparing “small”, “medium” and “large” fibroids with a “normal” tissue (also called normal myometrium) from uterus on the basis of “ m ” genes. Our goal is not only to identify differentially expressed genes in any given pairwise comparison but also to determine if the mean expression is up or down-regulated in the tumor tissue compared to the normal myometrium. Our statistical methodology for the FGS gene expression data analysis proceeds in three steps as follows.

1. For each gene we obtain a Dunnett (Dunnett, 1955; Dunnett and Tamhane, 1991, 1992) based screening p -value from all “ q ” pairwise comparisons with the reference group. Apply the BH procedure Benjamini and Hochberg (1995) on these screening p -values to obtain genes that are differentially expressed in at least one pairwise comparison. Suppose we discover R genes in this step. Thus,

there are R genes which are differentially expressed in at least one pairwise comparison with the reference group.

2. For the j^{th} gene discovered in Step 1, we compute Dunnett's p -value P_{ij} for each pairwise comparison i , $i = 1, 2, \dots, q$, with the reference group. If $P_{ij} \leq \frac{R\alpha}{m}$ then we declare that the i^{th} pairwise comparison with the reference group is significant.
3. If a gene j is found to be significant in the i^{th} pairwise comparison with the reference group then we declare it to be up-regulated in the i^{th} group relative to the reference group if $T_{ij} > 0$, otherwise it is declared to be down-regulated. Here T_{ij} denotes the test statistic associated with the j^{th} gene in the i^{th} pairwise comparison.

Specific details of implementation of each step are described in the following sections.

3.3.1 Step 1 - Identifying Differentially Expressed Genes

For this data and analysis of tumor sizes vs. myometrium we have, $m = 54675$, $p = 4$ and $q = 3$. The data we have is gene expressions x_{ik}^j , say, $i = 1, \dots, p$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, n_i$, with, $n_1 = 13$, $n_2 = 25$, $n_3 = 14$ and $n_4 = 8$. The screening and the component null and alternative hypotheses we want to test are as in equations (2.1) and (2.2). All the notations used here are defined in Section 2.4.

Dunnett P_{screen} : Dunnett test (Dunnett, 1955) is a powerful method that is designed specifically for comparison of several treatment groups with a common control group. The test assumes that the underlying distribution of the data from the different groups have same variance and the test statistics are obtained by using a pooled estimate of the variance. This assumption is valid for the Uterine fibroid data as the gene expressions are normalized to have similar means and variances for comparison. The test statistic for testing (2.2) is given by,

$$T_{ij}^{Dunn} = \frac{\bar{x}_{ij} - \bar{x}_{pj}}{s_j \sqrt{\frac{1}{n_i} + \frac{1}{n_p}}}, \quad (3.1)$$

where,

$$s_j^2 = \frac{\sum_{i=1}^p \sum_{k=1}^{n_i} (x_{ij}^k - \bar{x}_{ij})^2}{\sum_{i=1}^p n_i - p} \quad (3.2)$$

is the pooled sample variance. The null distribution of each T_{ij}^{Dunn} is univariate t -distribution with $(\sum_{i=1}^p n_i - p)$ degrees of freedom. The vector of Dunnett test statistics $\mathbf{T}_j^{Dunn} = (T_{1j}^{Dunn}, T_{2j}^{Dunn}, \dots, T_{p-1j}^{Dunn})$ has a $(p-1)$ -variate t -distribution with $\nu = (\sum_{i=1}^p n_i - p)$ degrees of freedom and correlation matrix $\mathbf{R} = (\rho_{ik})_{ik}$, where for $i \neq k$,

$$\rho_{ik} = \sqrt{\frac{n_i}{n_i + n_p}} \sqrt{\frac{n_k}{n_k + n_p}}, \quad i, k = 1, 2, \dots, p-1. \quad (3.3)$$

The Dunnett-adjusted critical value for the two-sided test for $\{T_{ij}^{Dunn}, i = 1, \dots, p-1\}$, denoted by $u_\alpha(p-1, \nu)$, is the quantile of the above $(p-1)$ -variate t -distribution such that,

$$Pr(|T_{1j}^{Dunn}| \leq u_\alpha, \dots, |T_{p-1j}^{Dunn}| \leq u_\alpha) = 1 - \frac{\alpha}{2}, \quad (3.4)$$

or equivalently,

$$Pr\left(\max_{\{i=1, \dots, p-1\}} |T_{ij}^{Dunn}| \leq u_\alpha\right) = 1 - \frac{\alpha}{2}. \quad (3.5)$$

The observed values of T_{ij}^{Dunn} , t_{ij}^{Dunn} , say, are compared to $u_\alpha(p-1, \nu)$ and we reject H_{0i}^j if $|t_{ij}^{Dunn}| > u_\alpha(p-1, \nu)$. For our data, $\nu = 56$, $p = 4$ and $u_{0.05}(3, 56) = 2.657$. For each gene j we have a vector of observed Dunnett test statistics, $\mathbf{t}_{ij}^{Dunn} = (t_{1j}^{Dunn}, t_{2j}^{Dunn}, t_{3j}^{Dunn})$. Let, $t_j^{max} = \max_{i=1,2,3} |t_{ij}^{Dunn}|$. We obtain the screening P -value for testing the screening hypotheses (2.1) as,

$$\begin{aligned} P_{screen}^j &= Pr\left(\max_{\{i=1,2,3\}} |T_{ij}^{Dunn}| > t_j^{max}\right) \\ &= 1 - Pr\left(-t_j^{max} \leq T_{ij}^{Dunn} \leq t_j^{max}, i = 1, 2, 3\right), \end{aligned} \quad (3.6)$$

where, the probability in equation (3.6) is obtained from the CDF of 3-variate t -distribution with ν degrees of freedom and the correlation structure given in (3.3).

Alternately, we can also compare the maximum of absolute value of the test statistics to $u_{0.05}(3, 56)$,

$$T_j^{Dunnnett} = \max \{ |T_{ij}^{Dunn}|, i = 1, 2, 3 \}. \quad (3.7)$$

Reject (2.1) and declare gene j significant if $T_j^{Dunnnett} > u_{0.05}(3, 56)$.

3.3.2 Step 2 - Identifying Significant Pairwise Comparisons

Once the differentially expressed genes are identified, we move on to the next step and answer the question in which tumor sizes, the genes are expressed differentially compared to the normal myometrium. Let $DE \subseteq \{1, 2, \dots, m\}$ denote the set of indices of the identified genes. We denote the number of genes identified by R , that is,

$$R = |DE|. \quad (3.8)$$

In this step, we test the component hypotheses for only identified genes and apply the Dunnett mdFWER controlling procedure for each gene, at level $R\alpha/m$, while identifying the tumor sizes and this procedures will be augmented with directional decisions in Step 3 of the analysis such that the overall error measure controlled for each gene is mdFWER at level $R\alpha/m$. The method is described as follows:

Dunnett Procedure: We use Dunnett procedure as described in Step 1 of section 3.3.1 and obtain the Dunnett-adjusted p -values and call them $\tilde{P}_{ij}^{Dunnnett}$,

$$\tilde{P}_{ij}^{Dunnnett} = 2 \cdot Pr \left(\max_{\{i=1, \dots, q\}} |T_{ij}^{Dunn}| \geq t_{ij}^{Dunn} \right), \quad (3.9)$$

where, T_{ij}^{Dunn} and t_{ij}^{Dunn} are defined in section 3.3.1 while discussing the Dunnett screening procedure. We reject H_{0i}^j if the corresponding adjusted P -value $\tilde{P}_{ij}^{Dunn} \leq \frac{R\alpha}{m}$.

3.3.3 Step 3 - Directional Decisions

Once we find out significantly expressed categories for each differentially expressed gene, we conclude on direction based on the sign of the test statistics. That is, declare $\theta_{ij} > 0$ if $T_{ij}^{Dunn} > 0$ or *vice versa*.

For example, let gene j be identified as significant in Step 1 of the analysis. The corresponding component hypotheses, in equation (2.2), are tested in Step 2. Suppose that the first two component hypotheses are rejected and the third one is not rejected, then we have, $\theta_{i1j} \neq 0$, $\theta_{2j} \neq 0$. That is, $\mu_{1j} \neq \mu_{0j}$ and $\mu_{2j} \neq \mu_{0j}$. In Step 3 of the analysis, we look at the sign of the test statistic corresponding to the rejected component hypothesis and decide on the direction as follows: if $T_{1j} > 0$ then conclude $\mu_{1j} > \mu_{0j}$ and *vice versa*.

3.4 Simulation Study

A simulation study was conducted to evaluate the performance of the specific procedures obtained by combinations of several procedures, described for use in Steps 1 and 2, explained in Section 2.4 and Section 3.3. The procedure with Bonferroni screening procedure and Bonferroni single step procedure in Step 2 is the procedure given by Guo et al. (2010). We address the following questions in the simulation study:

1. How do the procedures compare among themselves in terms of mdFDR and power under independence as well as different types of dependence of test statistics and p -values within a family?

2. How do the procedures compare among themselves in terms of mdFDR and power under independence as well as different types of dependence of hypotheses across families?

For comparison of mdFDR we use definition 2.1 to evaluate mdFDR for a procedure.

For comparison of power we use the concept of average power defined as follows.

Definition 3.1 (Average Power). The average power is defined as the expected proportion of false null hypotheses rejected in Steps 1 and 2 among all rejections in Steps 1 and 2,

$$\text{Average Power} = E \left(\frac{\text{Number of correct decisions in Steps 1 and 2}}{\text{Number of rejections in Steps 1 and 2}} \right). \quad (3.10)$$

3.4.1 Study Design

In our simulation we considered $p = 4$ groups, the first three were taken to be experimental groups and the last group was taken to be the reference group. Thus, all pairwise comparisons are made with the last group. Our simulated microarray chip consisted of $m = 1000$ genes per chip with $n = 10$ chips per group. As often done in simulation studies for microarray data (Dudoit et al., 2003; Efron and Tibshirnai, 2002; Efron et al., 2001; Guo et al., 2010), we generated the expression of each gene in each chip using a normal distribution.

More precisely, for the j^{th} gene, $j = 1, 2, \dots, 1000$, in the s^{th} sample, $s = 1, 2, \dots, 10$ in the g^{th} group, $g = 1, 2, 3, 4$, we generated its expression Z_{gj}^s from a normal distribution with mean value $E(Z_{gj}^s) = \mu_{gj}$ and variance $V(Z_{gj}^s) = 1$. For the reference group (i.e. group 4) we set $\mu_{4j} = 0$ for all j . To create the null data we set $\mu_{gj} = 0$ for $j = 1, 2, \dots, m_0$, $g = 1, 2, 3$ and non-null data were created by generating $\mu_{gj} \sim^{\text{independent}} U(0, 2.5)$, for $g = 1, 2, 3$ and $j = m_0 + 1, m_0 + 2, \dots, m_0 + m_1$, where $m_1 = m - m_0$ and U represents the uniform distribution. It is important to note that, for the non-null means considered here, the standard deviation used in this simulation study is large. Consequently, all methods considered in this simulation

study are expected to have small power. We considered three patterns of correlation structure as follows: (a) Independent gene expressions: the correlation coefficient ρ between any pair of genes and any pair of sample is 0. Thus, the data are completely independent. (b) Gene expressions within sample are dependent: For a given sample s , the correlation coefficient between any pair of genes is ρ but the correlation coefficient between any pair of samples is 0. (c) Gene expressions within genes are dependent: For a given gene j , the correlation coefficient between any pair of samples is ρ but the correlation coefficient between any pair of genes is 0.

We simulated the mdFDR and average power using 1000 simulation runs. Our nominal mdFDR level was taken to be $\alpha = 0.05$. For comparison purposes, we compared the proposed procedure with the Guo et al. (2010) procedure and variants of the proposed procedure in which the mdFWER controlling procedures used in Steps 2 and 3 of the algorithm are respectively, the Holm’s procedure (Holm, 1979), the Hochberg’s procedure (Hochberg, 1988) and the Bonferroni procedure. Although all five procedures control the mdFDR on average at 0.05, the power of our proposed methodology is much higher compared to all four competing procedures both under independence as well dependence structures considered in the simulation.

3.4.2 Results of The Simulation Study

Independence: We summarize the results of the simulation studies in Figure 3.1 for the independence case. The horizontal axis denotes the proportion of truly differentially expressed genes on the array and the vertical axis denotes the average mdFDR (left panel) and average power (right panel). As desired, all five procedures control the mdFDR on average at $\alpha = 0.05$. However, the proposed Dunnett based procedure has highest power compared to all other methods.

Dependence within genes across groups. In this case the components Z_{ij}^s , $i = 1, \dots, p$ are dependent with $Z_{ij}^s \sim N(\mu_{ij}, 1)$ and have a common correlation $\rho =$

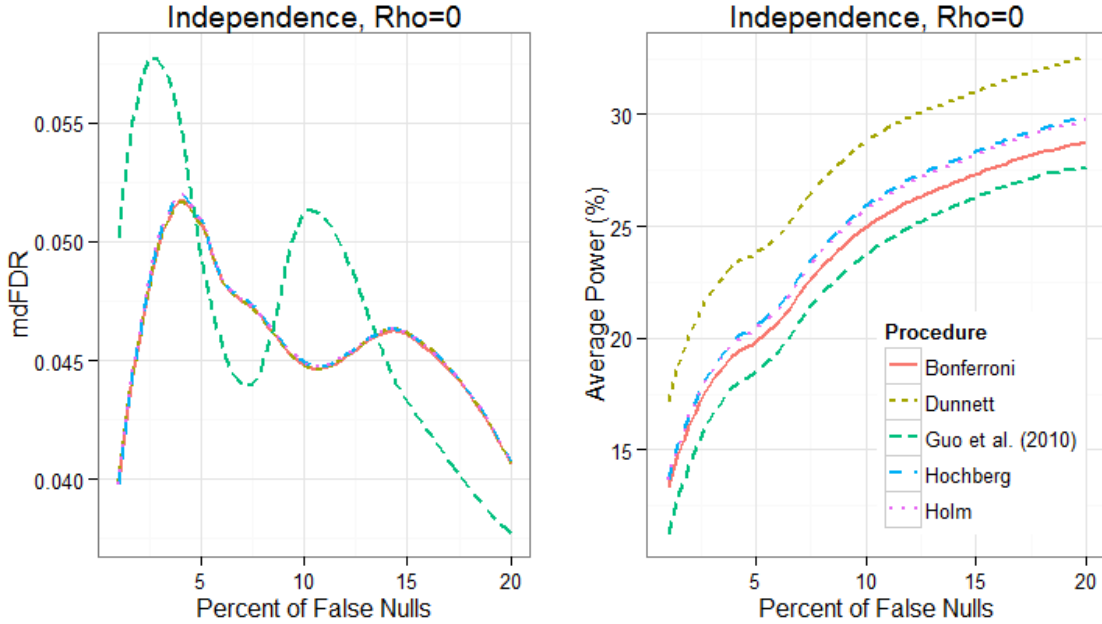


Figure 3.1 mdFDR (left), Average Power (right) with the proposed methodology and three variants using Holm, Hochberg and Bonferroni procedures, respectively, in Steps 2 and 3 along with Guo et al. (2010) procedure, under independence among genes and within genes.

0.2, 0.5, 0.8. The results are summarized in Figures 3.2 - 3.4. All five procedures control the mdFDR at less than $\alpha = 0.05$. Once again, as in the case of independence, the proposed method gains in power compared to the other methods.

Dependence among genes. We next considered the situation where gene expressions are dependent among genes. For this simulation, the components Z_{ij}^s , $j = 1, \dots, m$ are dependent with $Z_{ij}^s \sim N(\mu_{ij}, 1)$ and have a common correlation $\rho = 0.2, 0.5, 0.8$. The results are summarized in Figures 3.5 - 3.7. Again, all five procedures control the mdFDR at less than $\alpha = 0.05$ and as in the case of independence, the proposed method gains in power compared to the other methods.

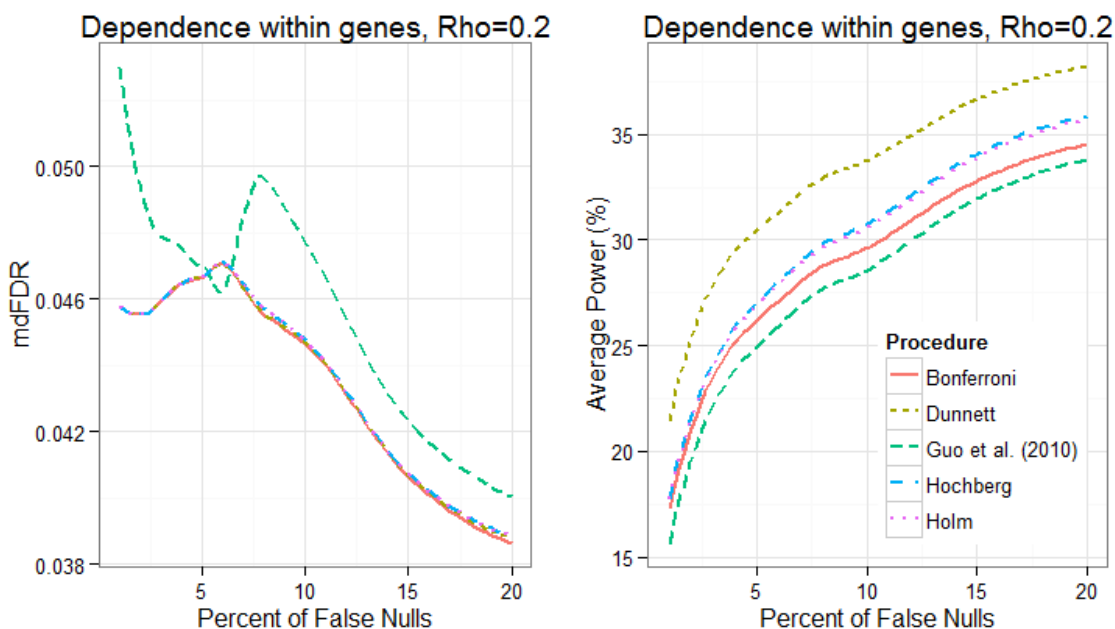


Figure 3.2 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.2$) within genes.

3.5 Results of Data Analysis

In this section, we present the results of the analysis done by applying the suggested methodology to FGS gene expression data. We identified a total of 9553 probe sets to be differentially expressed in at least one pairwise comparison (relative to the normal myometrium) at mdFDR of 0.05. These 9553 probe sets map to 6286 genes. The Venn diagram of the Differentially Expressed Genes (DEGs) by tumor size is in Figure 3.8. Based on the 6286 genes, using the Ingenuity Pathway Analysis (IPA, 2000-2014 QIAGEN), we discovered a total 157 distinct enriched canonical pathways at a false discovery rate of 0.05. The Venn diagram of the number of enriched canonical pathways by tumor size is provided in Figure 3.9. It is not surprising that a third of the DEGs and nearly 30% of all significantly enriched pathways are common to tumors of all sizes since tumor tissues are fundamentally different from the normal myometrium. However, we discover several DEGs and pathways that are

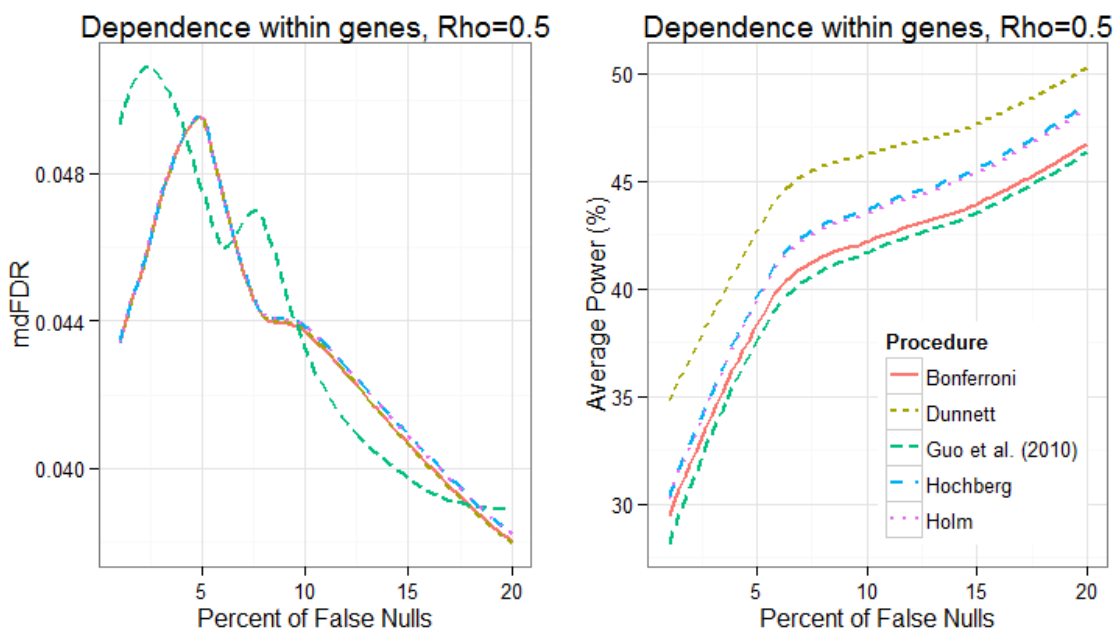


Figure 3.3 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.5$) within genes.

uniquely enhanced according to the tumor size, suggesting that changes in molecular characteristics might be taking place as tumors grow.

Growth factors, vascularization and related pathways: The Netrin signaling pathway is well-known as a versatile pathway with multiple functions. One of its functions is to promote vascular networks and branching of blood vessels (Lejmi et al., 2008) and angiogenesis (Epting et al., 2010). According to our analysis this pathway is uniquely enriched in small tumors only, which suggests that perhaps it is only necessary during the initial stages of tumor onset. The Interleukin-1 (IL-1) pathway is known to induce inflammatory response and the production of prostaglandins and expression of IL-2 which may play a critical role in the fibroid initiation and early development. For example, prostaglandins play a critical role in the promotion of growth factors involved in angiogenesis, such as VEGF, basic fibroblast growth factor (bFGF) and platelet-derived growth factor (PDGF) (Jabbour

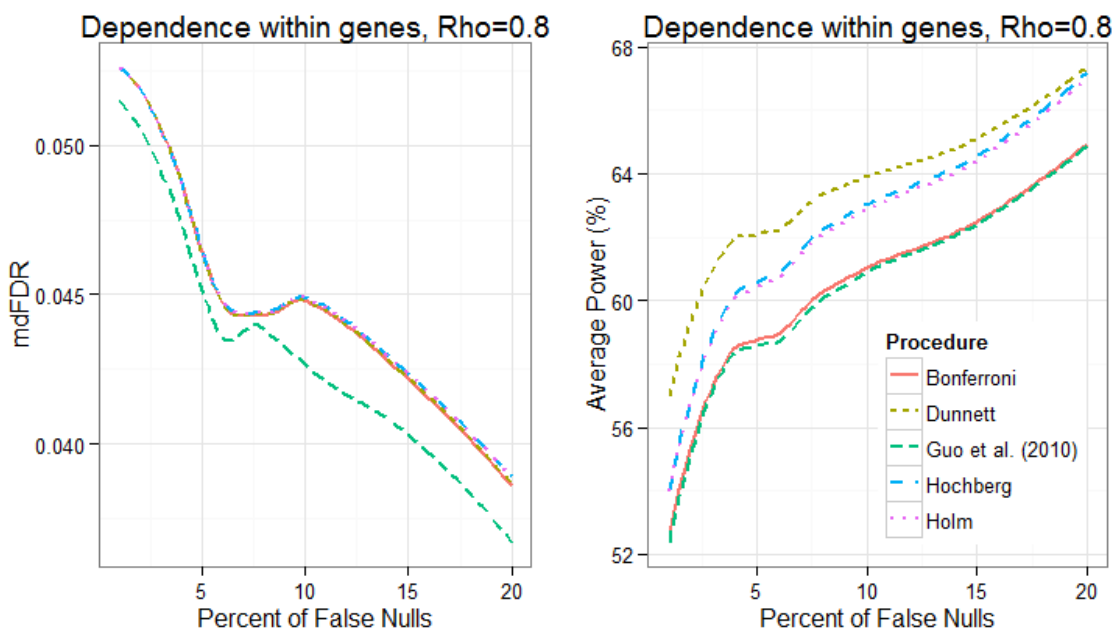


Figure 3.4 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.8$) within genes.

et al., 2006) and development of fibroid requires vascularization and blood supply. Thus, Interleukin-1 (IL-1) pathway likely plays an important role during the initial stages of fibroid development. Interestingly, according to our analysis, this pathway is uniquely enriched in small size tumors but not in the medium or large tumors. Furthermore, the fibroblast growth factors 8 and 20 (FGF8, FGF20) which belong to the Regulation of the Epithelial-Mesenchymal Transition Pathway and are well-known to be involved in vascularization and angiogenesis, are both uniquely down-regulated in small tumors and not differentially expressed in medium or large tumors. Our analysis further implies that the Epithelial-Mesenchymal Transition Pathway was enriched only in small and medium size tumors.

According to Ciarmela et al. (2011), estrogen may promote fibroid growth through up-regulation of epidermal growth factor receptor (EGFR). However, we found EGFR to be down-regulated in fibroids and that too only in the medium

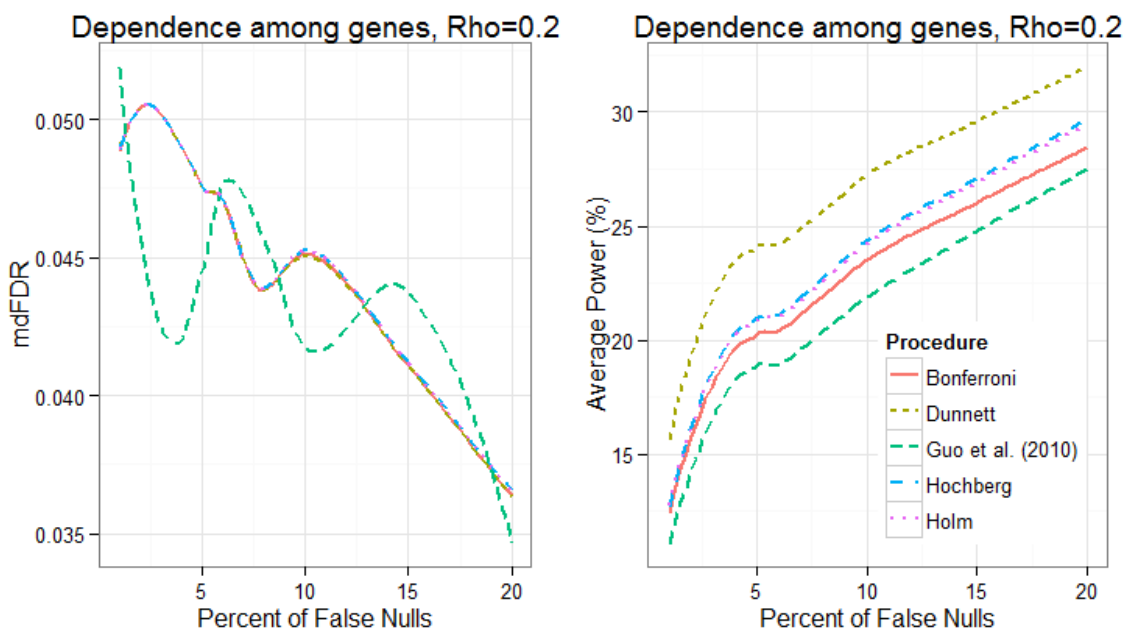


Figure 3.5 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.2$) among genes.

size tumors. Similarly, the fibroblast growth factors (FGF) (acidic and basic) were differentially expressed only in the medium size tumors. The acidic FGF was up-regulated whereas the basic FGF was down regulated. These findings are consistent with Ciarmela et al. (2011) (and references therein, e.g. (Wolanska et al., 2008)) in that they are expressed during tumor progression. Similarly, insulin like growth factor (IGF1) was only differentially expressed (up-regulated) in medium size tumors. Additionally, growth factor signaling pathways such as VEGF, PDGF, TGF β and EGF are uniquely enriched in medium size tumors and not in small or large tumors. While the above results suggest that some growth factors and related pathways are very specific to small and medium tumors i.e. play an important role during the early stages of tumor onset and growth, we discovered several growth factors to be differentially expressed in tumors of all sizes. These included, insulin like growth factor 2 (INS-IGF2), insulin like growth factor binding protein 5 (IGFBP5),

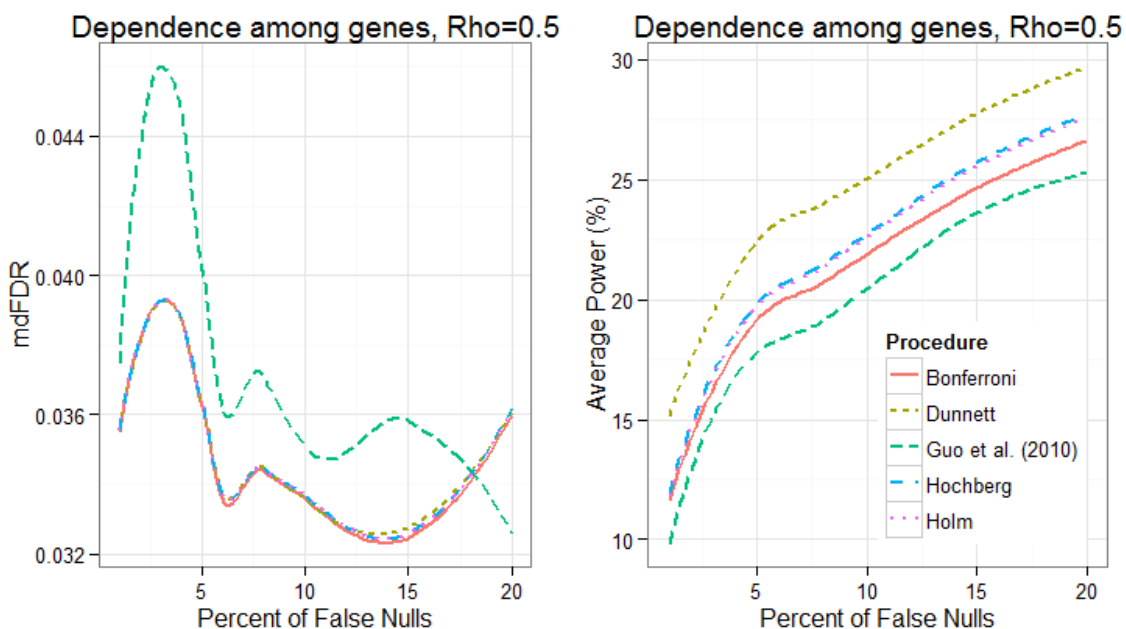


Figure 3.6 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.5$) among genes.

and platelet derived growth factor C (PDGFC) which were up-regulated whereas insulin-like growth factor binding protein 6 (IGFBP6), connective tissue growth factor (CTGF), heparin-binding EGF-like growth factor (HBEGF), transforming growth factor beta receptor II (TGFBR2), fibroblast growth factors 12 and 13 (FGF12, FGF13) were down-regulated. Similarly, growth factor signaling pathways such as the human growth factor (HGF) and IGF-1 were enriched in tumors of all sizes. Thus, it appears that the differential expression of these genes and the enrichment of the above pathways is necessary for tumor onset and progression.

Estrogen and related genes: Fibroids are hormonally mediated and it is also well documented in the literature that accordingly estrogen and progesterone receptors and prostaglandins promote proliferation of fibroids (see (Davis et al., 2013; Talaulikar and Manyonda, 2012)). Not surprisingly, we found the estrogen receptor ESR-1 to be up-regulated in tumors of all sizes. Interestingly, the

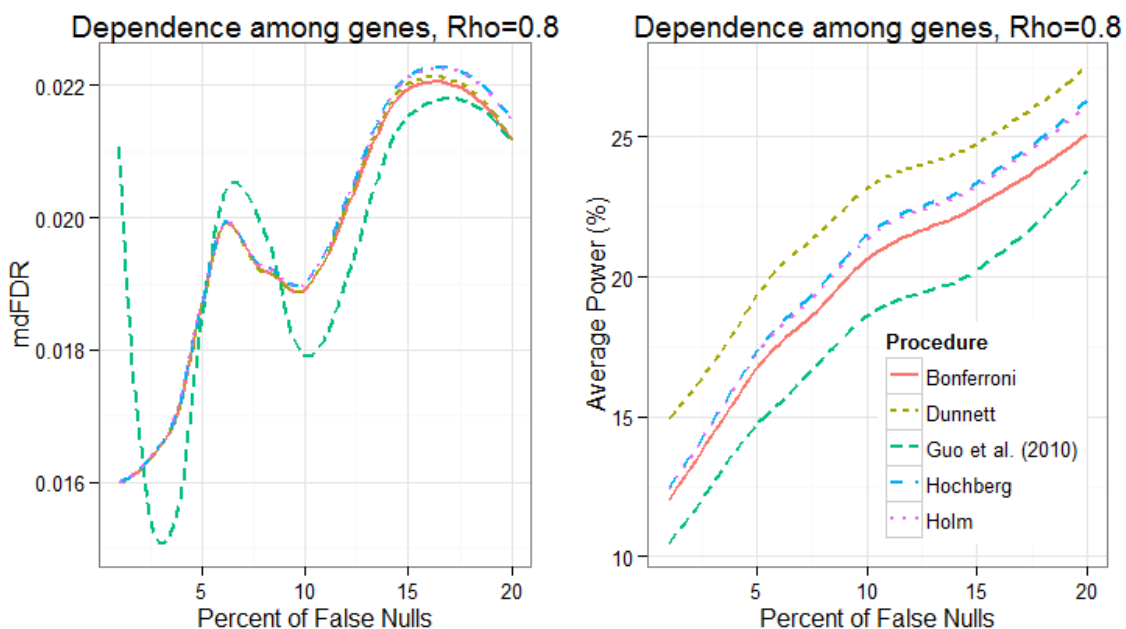


Figure 3.7 mdFDR, Average Power with Dunnett screening and different mdFWER controlling procedures compared to Guo et al. (2010), under dependence ($\rho = 0.8$) among genes.

progesterone receptor (PGR) was up-regulated in only medium size tumors and not differentially expressed in small or large tumors. This suggests that perhaps PGR may not be involved in tumor initiation (i.e. small tumors) but is only involved in growth of the tumor. However, its function ends once the tumor becomes large enough. Most prostaglandins were generally down-regulated in tumors of all sizes. For example, prostaglandin E receptor 3 (PTGER3), prostaglandin F receptor (PTGFR) and prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) (PTGS2) are down-regulated in tumor of all sizes. However, some prostaglandins were differentially expressed according to the size of the tumor. For example, prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase) (PTGS1) was differentially expressed only in the medium sized tumors where it was down-regulated and prostaglandin E synthase 2 (PTGES2) was

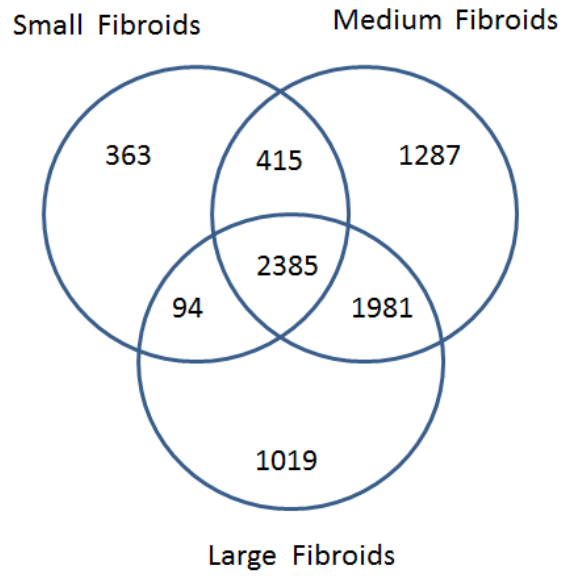


Figure 3.8 Venn diagrams of differentially expressed genes by tumor size.

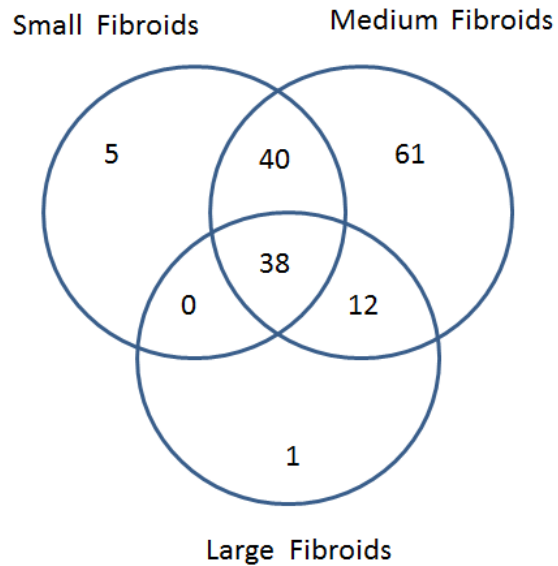


Figure 3.9 Venn diagram of enriched pathways by tumor size.

down-regulated in large tumors only but prostaglandin E receptor 4 (PTGER4) was down-regulated in both medium and large tumors.

Similar to estrogen and progesterone receptors, the α and γ isoforms of peroxisome proliferator-activated (PPAR) receptors have been associated with the regulation of proliferation of uterine fibroids (Houston et al., 2003; Nam et al., 2007). In our data both these isoforms are down-regulated in all tumor sizes compared to the normal myometrium. We also discovered the related retinoid X receptor gamma to be down-regulated in the medium size tumors but was not significant in small or large tumors.

Collagens: There is a vast amount of literature implicating collagens to smooth muscle tumors such as the fibroids (see (Davis et al., 2013), and references therein). Consequently, it is not surprising that several collagens (COL1A1, COL1A2, COL3A1, COL4A1-COL4A4, COL5A2, COL6A3, COL7A1, COL9A2, COL21A1, COL22A1, COL27A1) and extra cellular matrix proteins are differentially expressed in tumors of all sizes. Apart from COL4A3, COL4A4 and COL21A1, which were down-regulated in all tumor size groups, the remaining 11 collagens were up-regulated in tumor samples.

Other genes: Leptin receptor is well-known to be negatively associated with the obesity and obesity is a potential risk factor for fibroids. Interestingly, we discover leptin receptor (LEPR) to be significantly down-regulated in all tumor sizes. As noted earlier, Diaz et al. (Diaz et al., 2005) demonstrated that $\beta 4$ integrin had an increased expression in larger breast tumors and in higher tumor grades. In our fibroid data, however, we notice a down-regulation in $\beta 4$ integrin in medium and large tumors and was not differentially expressed in small tumors.

3.6 Discussion and Conclusion

In this chapter, we offer a statistical methodology for conducting pairwise comparisons in high dimensional data that is broadly applicable to a wide range of genomic data including gene expression microarray data, CpG methylation data, RNA-seq data and others. The proposed methodology not only controls the false discoveries made when making several pairwise comparisons on the basis of high dimensional data, but it controls for directional errors. Although in our algorithm we used Dunnett's procedure in Step 1 to derive the screening p -values, and again Dunnett's procedure in Step 2 to derive the pairwise p -values, they can be replaced by several other procedures in the general procedure and the resulting method will continue to control the mdFDR.

In the simulation studies we see that mdFDR is controlled for all the discussed procedures under independence of test statistics and also under dependence within and across genes. The proposed methodology based on the Dunnett test works best among all the procedures in terms of average power.

Using our methodology we gain deeper insights into molecular characteristics of uterine fibroids according to the tumor size. We have identified several differentially expressed genes and pathways that are specifically enriched according to the tumor size (or stage of growth). While researchers and clinicians who study fibroids are well aware of many of the genes and pathways described in this chapter, we have provided a characterization of these genes and pathways according to the tumor stage. Our data can be further mined to gain deeper insights regarding fibroids.

CHAPTER 4

CONTROL OF DIRECTIONAL ERRORS IN FIXED SEQUENCE MULTIPLE TESTING

4.1 Introduction

Directional errors or Type III errors occur in testing situations with two-sided alternatives when rejections are accompanied by additional directional claims. For example, when testing a null hypothesis $\theta = 0$ against $\theta \neq 0$, rejection of the null hypothesis is often augmented with the decision of whether $\theta > 0$ or $\theta < 0$. In case of testing of a single hypothesis, Type III error or a Directional error is automatically controlled at level α when the Type I error is controlled at the level α . However in the case of simultaneously testing multiple hypotheses, it is often not known whether additional directional decisions can be made without losing control of the mixed directional familywise error rate (mdFWER), the probability of at least one Type I or Type III error. Some methods have been developed in the literature by augmenting additional directional decisions to the existing p -value based stepwise procedures. Shaffer (1980) showed that Holm's procedure (Holm, 1979), augmented with decisions on direction based on the values of test statistics, can strongly control mdFWER under the assumption that the test statistics are independent and under specified conditions on the marginal distributions of the test statistics, but she also showed that counterexamples exist even with two hypotheses. Finner (1994) and Liu (1997) independently proved the same result for Hochberg procedure (Hochberg, 1988). Finner (1999) generalized the result of Shaffer (1980) to a large class of stepwise or closed multiple test procedures under the same assumptions. Some recent results have been obtained in Guo and Romano (2015).

For control of the usual familywise error rate (FWER) (which does not account for the possibility of additional Type III errors), the conventional *fixed sequence* multiple testing procedure that strongly controls the FWER under arbitrary dependence, is known to be a powerful procedure in testing situations with pre-ordered hypotheses (Maurer et al., 1995; Wiens, 2003; Wiens and Dmitrienko, 2005). For reviews on recent relevant developments of fixed sequence multiple testing procedures, see Dmitrienko, Tamhane and Bretz (2009) and Dmitrienko et al. (2013). Indeed, suppose null hypotheses H_1, \dots, H_n are pre-ordered, so that H_{i+1} is tested only if H_1, \dots, H_i have all been rejected. The probability mechanism generating the data is P and H_i asserts that $P \in \omega_i$, some family of data generating distributions. In such case, it is easy to see that each H_i can be tested at level α in order to control the FWER at level α , so that no adjustment for multiplicity is required. The argument is simple and goes as follows. Fix any given P such that at least one H_i is true (or otherwise the FWER is 0 anyway). If H_1 is true, i.e. $P \in \omega_1$, then a Type I error occurs if and only if H_1 is rejected, and so the FWER is just the probability that H_1 is rejected, which is assumed controlled at level α when testing H_1 . If H_1 is false, just let f be the smallest index corresponding to a true null hypothesis, i.e. H_f is true but H_1, \dots, H_{f-1} are all false. In this case, a Type I error occurs if and only if H_f is rejected, which is assumed to be controlled at level α .

In fact, in situations where ordering is not specified, the above result suggests it may be worthwhile to think about hypotheses in order of importance so that potentially false hypotheses are more easily detected. Indeed, as is well-known, when the number n of tested hypotheses is large, control of the FWER is often so stringent that often no rejections can be detected, largely due to the multiplicity of tests and the need to find significance at very low levels (as required, for example, in the Bonferroni method with n large). On the other hand, under a specified ordering, each test is carried out at the same conventional level.

To our knowledge, no one explores the possibility of making additional directional decisions for such fixed sequence procedures. In this chapter, we introduce such fixed sequence procedures augmented with additional directional decisions and discuss its mdFWER control under independence and some dependence. For such directional procedures, its simple fixed sequence structure of the tested hypotheses makes the notoriously challenging problem of controlling the mdFWER under dependence a little easier to handle than stepwise procedures.

The rest of the chapter is organized as follows. Section 4.2 introduces the basic notations and concepts used in this project. In Section 4.3 we discuss mdFWER control under arbitrary dependence. In Section 4.4 we discuss control of mdFWER under independence of test statistics. These results are extended to positive dependence in Sections 4.5 and 4.6. Finally, in Section 4.7 we summarize the results presented in this chapter. The proofs of all the theorems and lemmas presented in this chapter are given in Appendix A.

4.2 Preliminaries

In this section we present the notations and a few assumptions used in this chapter. We consider the problem of testing n two-sided hypotheses, H_1, H_2, \dots, H_n specified as follows:

$$H_i : \theta_i = 0 \quad \text{vs.} \quad H'_i : \theta_i \neq 0, \quad i = 1, 2, \dots, n. \quad (4.1)$$

We assume the hypotheses are ordered in advance, either using some prior knowledge about the importance of the hypotheses or by some other specified criteria, so that H_1 is tested first and H_i is only tested if H_1, \dots, H_{i-1} are all rejected. We also assume that, for each i , a test statistic T_i and p -value P_i are available to test H_i (as a single test). For a rejected hypothesis H_i , we decide on the sign of the parameter θ_i by the sign of the corresponding test statistic, i.e., we conclude $\theta_i > 0$ if $T_i > 0$ and *vice*

versa. The errors that might occur while testing these hypotheses are: Type I and Type III errors. A *Type I error* occurs when a true H_i is falsely rejected. A *Type III error* occurs when a false H_i is correctly rejected but claimed sign of the parameter is wrong. Then, the mdFWER is the probability of making at least a Type I or Type III error, and it is desired that this error rate is no bigger than α for all possible data generating distributions in the model.

We make a few standard assumptions about the test statistics. Let $T_i \sim F_{\theta_i}(\cdot)$ for some continuous cumulative distribution function $F_{\theta_i}(\cdot)$ having parameter θ_i . In general, most of our results also apply through the same arguments when the family of distributions of T_i depends on i , though for simplicity of notation, the notation is suppressed. We assume that F_0 is symmetric about 0, F_{θ_i} is stochastically increasing in θ_i . For testing H_i vs. H'_i , rejections are based on large values of $|T_i|$ and the corresponding two-sided p -value is defined by,

$$P_i = 2 \min\{F_0(T_i), 1 - F_0(T_i)\}, \quad i = 1, \dots, n. \quad (4.2)$$

We assume that the P -values are distributed as Uniform(0,1) when $\theta_i = 0$. Various dependence assumptions between the test statistics are used throughout this chapter.

4.3 The mdFWER Control Under Arbitrary Dependence

A general fixed sequence procedure based on marginal p -values must specify the critical level α_i that is used for testing H_i , in order for the resulting procedure to control the mdFWER at level α . When controlling the FWER without regard to Type III errors, each α_i can be as large as α . However, Theorem 4.1 below shows that by using the critical constant $\alpha_i = \alpha/2^{i-1}$, the mdFWER is controlled at level α . Moreover, we show that these critical constants are not improvable. Formally, the optimal procedure is defined as follows.

Procedure 4.1 (Directional fixed sequence procedure under arbitrary dependence).

- Step 1: If $P_1 \leq \alpha$ then reject H_1 and continue to test H_2 after making directional decision on θ_1 : conclude $\theta_1 > 0$ if $T_1 > 0$ or $\theta_1 < 0$ if $T_1 < 0$. Otherwise, accept all the hypotheses and stop.
- Step i : If $P_i \leq \alpha/2^{i-1}$ then reject H_i and continue to test H_{i+1} after making directional decision on θ_i : conclude $\theta_i > 0$ if $T_i > 0$ or $\theta_i < 0$ if $T_i < 0$. Otherwise, accept the remaining hypotheses H_i, \dots, H_n .

In the following, we discuss the mdFWER control of Procedure 4.1 under arbitrary dependence of the p -values. When testing a single hypothesis, the mdFWER of Procedure 4.1 reduces to the Type I or Type III error rate depending on whether $\theta = 0$ or $\theta \neq 0$, and Procedure 4.1 reduces to the usual p -value based method along with the directional decision for the two-sided test. The following lemma covers this case.

Lemma 4.1. *Consider testing the single hypothesis $H : \theta = 0$ against $H' : \theta \neq 0$ at level α , using the usual p -value based method along with a directional decision. If H is a false null hypothesis, then the Type III error rate is bounded above by $\alpha/2$.*

Generally, when simultaneously testing n hypotheses, by using Lemma 4.1 and mathematical induction, we have the following result holds.

Theorem 4.1. *For Procedure 4.1 defined as above, the following conclusions hold.*

- (i) *This procedure strongly controls the mdFWER at level α under arbitrary dependence of the p -values.*
- (ii) *One cannot increase even one of the critical constants $\alpha_i = \alpha/2^{i-1}$, $i = 1, \dots, n$, while keeping the remaining fixed without losing control of the mdFWER.*

In fact, the proof shows that no strong parametric assumptions are required. However, the rapid decrease in critical values $\alpha/2^{i-1}$ makes rejection of additional

hypotheses difficult. Thus, it is of interest to explore how dependence assumptions can be used to increase these critical constants while maintaining control of the mdFWER. The assumptions and methods will be described in the remaining sections.

Remark 4.1. Instead of Procedure 4.1, let us consider the conventional fixed sequence procedure with the same critical constant α augmented with additional directional decisions, which is defined in Section 4.4 as Procedure 4.2. By using Bonferroni inequality and Lemma 4.1, we can prove that the mdFWER of this procedure is bounded above by $\frac{n+1}{2}\alpha$. Thus, the modified version of the procedure, which has the same critical constant $\frac{2\alpha}{n+1}$, strongly controls the mdFWER at level α under arbitrary dependence of p -values. However, it is unclear if such critical constant can be further improved without losing the control of the mdFWER.

4.4 The mdFWER Control Under Independence

We further make the following assumptions on the distribution of the test statistics.

Assumption 4.1 (Independence). *The test statistics, T_1, \dots, T_n , are mutually independent.*

Of course, it follows that the p -values P_1, \dots, P_n are mutually independent as well.

As will be seen, it will be necessary to make further assumptions on the family of distributions for each marginal test statistic.

Definition 4.1 (Monotone Likelihood Ratio (MLR)). A family of probability density functions $f_\delta(\cdot)$ is said to have monotone likelihood ratio property if, for any two values of the parameter δ , $\delta_2 > \delta_1$ and any two points $x_2 > x_1$,

$$\frac{f_{\delta_2}(x_2)}{f_{\delta_1}(x_2)} \geq \frac{f_{\delta_2}(x_1)}{f_{\delta_1}(x_1)}, \quad (4.3)$$

or equivalently,

$$\frac{f_{\delta_1}(x_1)}{f_{\delta_1}(x_2)} \geq \frac{f_{\delta_2}(x_1)}{f_{\delta_2}(x_2)}. \quad (4.4)$$

Definition 1 means that, for fixed $x_1 < x_2$, the ratio $\frac{f_{\delta}(x_1)}{f_{\delta}(x_2)}$ is non-increasing in δ . Two direct implications of Definition 1 in terms of the cdf $F_{\delta}(\cdot)$ are

$$\frac{F_{\delta_1}(x_2)}{F_{\delta_1}(x_1)} \leq \frac{F_{\delta_2}(x_2)}{F_{\delta_2}(x_1)}, \quad (4.5)$$

and

$$\frac{1 - F_{\delta_1}(x_2)}{1 - F_{\delta_1}(x_1)} \leq \frac{1 - F_{\delta_2}(x_2)}{1 - F_{\delta_2}(x_1)}. \quad (4.6)$$

Assumption 4.2 (MLR Assumption). *The family of marginal distributions of the T_i has monotone likelihood ratio.*

Based on the conventional fixed sequence multiple testing procedure, we define a directional fixed sequence procedure as follows, which is the conventional fixed sequence procedure augmented with directional decisions. In other words, any hypothesis is tested at level α , and as will be seen under the specified conditions, no reduction in critical values is necessary in order to achieve mdFWER control.

Procedure 4.2 (Directional Fixed Sequence Procedure).

- Step 1: If $P_1 \leq \alpha$, then reject H_1 and continue to test H_2 after making a directional decision on θ_1 : conclude $\theta_1 > 0$ if $T_1 > 0$ or $\theta_1 < 0$ if $T_1 < 0$. Otherwise, accept all the hypotheses and stop.
- Step i : If $P_i \leq \alpha$, then reject H_i and continue to test H_{i+1} after making a directional decision on θ_i : conclude $\theta_i > 0$ if $T_i > 0$ or $\theta_i < 0$ if $T_i < 0$. Otherwise, accept the remaining hypotheses, H_i, \dots, H_n .

For Procedure 4.2, in the case of $n = 2$, we derive a simple expression for the mdFWER in Lemma 4.2 below and prove its mdFWER control in Lemma 4.3 by using such simple expression.

Lemma 4.2. *Consider testing two hypotheses $H_1 : \theta_1 = 0$ and $H_2 : \theta_2 = 0$, against both sided alternatives, using Procedure 4.2 at level α . Let $c_1 = F_0^{-1}(\alpha/2)$ and $c_2 = F_0^{-1}(1 - \alpha/2)$. When $\theta_2 = 0$, the following result holds.*

$$mdFWER = \begin{cases} \alpha + F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{(\theta_1,0)}(c_2, c_2) - F_{(\theta_1,0)}(c_2, c_1) & \text{if } \theta_1 > 0 \\ 1 + F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{(\theta_1,0)}(c_1, c_1) - F_{(\theta_1,0)}(c_1, c_2) & \text{if } \theta_1 < 0. \end{cases} \quad (4.7)$$

In the above, $F_{\theta_1, \theta_2}(\cdot, \cdot)$ refers to the joint c.d.f. of (T_1, T_2) . Then, under Assumption 4.1 (independence), (4.7) can be simplified as

$$mdFWER = \begin{cases} \alpha + F_{\theta_1}(c_1) - \alpha F_{\theta_1}(c_2) & \text{if } \theta_1 > 0 \\ 1 + \alpha F_{\theta_1}(c_1) - F_{\theta_1}(c_2) & \text{if } \theta_1 < 0. \end{cases} \quad (4.8)$$

Lemma 4.3. *Under Assumption 4.1 (independence) and Assumption 4.2 (MLR), Procedure 4.2 strongly controls the mdFWER when $n = 2$.*

Generally, for testing any n hypotheses, by using mathematical induction and Lemma 4.3, we also prove the mdFWER control of Procedure 4.2 under the same assumptions as in the case of $n = 2$.

Theorem 4.2. *Under Assumption 4.1 (independence) and Assumption 4.2 (MLR), Procedure 4.2 strongly controls the mdFWER at level α .*

Many families of distributions have the MLR property: normal, uniform, logistic, Laplace, Student's t, generalized extreme value, exponential families of distributions, etc. However, it is also important to know whether or not the above results fail without the MLR assumption. A natural family of distributions to consider without the MLR property is the Cauchy family; indeed, Shaffer (1980) used this family to obtain a counterexample for the directional Holm procedure while testing p -value ordered hypotheses. We now show that Procedure 4.2 fails

to control the mdFWER for this family of distributions with corresponding cdf $F_\theta(x) = 0.5 + \frac{1}{\pi} \arctan(x - \theta)$, even under independence.

Lemma 4.2 can be used to verify the calculation for the case of $n = 2$ with $\theta_1 > 0$ and $\theta_2 = 0$; specifically, see (4.8). Indeed, we just need to show,

$$F_{\theta_1}(-c) = F_0(-c - \theta_1) > \alpha F_{\theta_1}(c) = \alpha F_0(c - \theta_1), \quad (4.9)$$

where c is the $1 - \alpha/2$ quantile of the standard Cauchy distribution, given by $\tan[\pi(1 - \alpha)/2]$. Take $\alpha = 0.05$, so $c = 12.7062$. Then, the above inequality (4.9) is violated for example by $\theta_1 = 100$. The left side is approximately $F(-112.7) \approx 0.002824$ while the right side is

$$0.05 \times F(-87.3) = 0.05 \times 0.0036 = 0.00018.$$

4.5 Extension to Positive Dependence

Clearly, the assumption of independence is of limited utility in multiple testing, as many tests are usually carried out on the same data set. Thus, it is important to generalize the results of the previous section to cover some more general cases. As is typical in the multiple testing literature (Benjamini and Yekutieli (2001); Sarkar (2002); Sarkar and Guo (2010)), assumptions of positive regression dependence will be used.

Before defining the assumptions, for convenience, we introduce several notations below. Among the prior-ordered hypotheses H_1, \dots, H_n , let i_0 denote the index of the first true null hypothesis, n_1 denote the number of all false nulls, and $T_{i_1}, \dots, T_{i_{n_1}}$ denote the corresponding false null test statistics. Specifically, if all H_i 's are false, let $i_0 = n + 1$.

Assumption 4.3. *The false null test statistics along with parameters, $\theta_{i_1} T_{i_1}, \dots, \theta_{i_{n_1}} T_{i_{n_1}}$, are positively regression dependent in the sense of*

$$E \{ \phi(\theta_{i_1} T_{i_1}, \dots, \theta_{i_{n_1}} T_{i_{n_1}}) \mid \theta_{i_k} T_{i_k} \geq u \} \uparrow u,$$

for each $\theta_{i_k} T_{i_k}$ and any (coordinate-wise) non-decreasing function ϕ .

Assumption 4.4. *The first true null statistic, T_{i_0} , is independent of all false null statistics $T_{i_k}, k = 1, \dots, n_1$ with $i_k < i_0$.*

Theorem 4.3. *Under Assumptions 4.2 - 4.4, Procedure 4.2 strongly controls the mdFWER at level α .*

Corollary 4.1. *When all tested hypotheses are false, Procedure 4.2 strongly controls the mdFWER at level α under Assumptions 4.2 - 4.3.*

Remark 4.2. In Theorem 4.3, we note that specifically, when all of the tested hypotheses are false, Assumption 4.4 is automatically satisfied. Generally, consider the case of any combination of true and false null hypotheses where Assumption 4.4 is not imposed. Without loss of generality, suppose $\theta_i > 0, i = 1, \dots, n - 1$ and $\theta_n = 0$, that is, the first $n - 1$ hypotheses are false and the last one is true. Under Assumptions 4.2-4.3, if T_n (or $-T_n$) and T_1, \dots, T_{n-1} are positively regression dependent, then the mdFWER of Procedure 4.2 when testing H_1, \dots, H_n is, for any n , bounded above by

$$\begin{aligned} & \Pr(\text{make at least one Type III error when testing } H_1, \dots, H_{n-1} \text{ or } T_n \notin (c_1, c_2)) \\ & \leq \lim_{\theta_n \rightarrow 0^+} \Pr(\text{make at least one Type III error when testing } H_1, \dots, H_n) \\ & \quad + \lim_{\theta_n \rightarrow 0^+} \Pr(T_n \geq c_2) \\ & \leq \alpha + \alpha/2 = 3\alpha/2. \end{aligned}$$

The first inequality follows from the fact that when $\theta_n \rightarrow 0^+$, H_n can be interpreted as a false null hypothesis with $\theta_n > 0$, and thus, one Type III error is made if H_n is rejected and $T_n \leq c_1$. The second inequality follows from Corollary 4.1 and Lemma 4.1.

Based on the above inequality, a modified version of Procedure 4.2, the directional fixed sequence procedure with the critical constant $2\alpha/3$, strongly controls

the mdFWER at level α under Assumptions 4.2-4.3 and the above additional assumption.

Remark 4.3. In the above remark, further, if we do not make any assumption regarding dependence between the true null statistic T_n and the false null statistics T_1, \dots, T_{n-1} . Then, by Theorem 4.3, the mdFWER of Procedure 4.2 when testing H_1, \dots, H_n is bounded above by,

$$\begin{aligned} & \Pr(\text{make at least one Type III error when testing } H_1, \dots, H_{n-1}) \\ & + \Pr(\text{make Type I error when testing } H_n) \\ & \leq \alpha + \alpha = 2\alpha. \end{aligned}$$

Therefore, an alternative modified version of Procedure 4.2, the directional fixed sequence procedure with the critical constant $\alpha/2$, strongly controls the mdFWER at level α only under Assumptions 4.2-4.3.

4.6 Further Extensions to Positive Dependence

We now develop alternative results to show that Procedure 4.2 can control mdFWER even under certain dependence between the false null and true null statistics. We relax the assumption of independence that the false null statistics are independent of the first true null statistic, and consider a slightly strong version of the conventional positive regression dependence on subset of true null statistics (PRDS) (Benjamini and Yekutieli, 2001), which is given below.

Assumption 4.5. *The false null test statistics, T_1, \dots, T_{i_0-1} and the first true null statistic T_{i_0} , are positive regression dependent in the sense of*

$$E \{ \phi(T_1, \dots, T_{i_0-1}) \mid T_{i_0} \geq u, T_1, \dots, T_j \} \uparrow u, \quad (4.10)$$

for any given $j = 1, \dots, i_0-1$, any given values of T_1, \dots, T_j and any (coordinate-wise) non-decreasing function ϕ .

We firstly consider the case of $n = 2$, that is, while testing two hypotheses, and show control of the mdFWER of Procedure 4.2 when the test statistics are positively regression dependent in the sense of Assumption 4.5.

Proposition 4.1. *Under Assumptions 4.2 and 4.5, the mdFWER of Procedure 4.2 is strongly controlled at level α when $n = 2$.*

Specifically, in the case of bivariate normal distribution, Assumption 4.2 is satisfied and two test statistics T_1 and T_2 are always positively or negatively regression dependent. As in the proof of Proposition 4.1, to show the mdFWER control of Procedure 4.2, we only need to consider the case of $\theta_1 \neq 0$ and $\theta_2 = 0$. Thus, if T_1 and T_2 are negatively regression dependent, we can choose $-T_2$ as the statistic for testing H_2 and Assumption 4.5 is still satisfied. By Proposition 4.1, we have the following corollary holds.

Corollary 4.2. *Under the case of bivariate normal distribution, the mdFWER of Procedure 4.2 is strongly controlled at level α when $n = 2$.*

We now consider the case of three hypotheses. The general case will ultimately be considered, but is instructive to discuss the case separately due to the added multivariate MLR condition, which is described as follows.

Let $f(x|T_1)$ and $g(x|T_1)$ denote the probability density functions of T_2 and T_3 conditional on T_1 , respectively.

Assumption 4.6 (Bivariate Monotone Likelihood Ratio (BMLR)). *For any given value of T_1 , $f(x|T_1)$ and $g(x|T_1)$ have the monotone likelihood ratio (MLR) property in x , i.e., for any $x_2 > x_1$, we have*

$$\frac{f(x_2|T_1)}{g(x_2|T_1)} \geq \frac{f(x_1|T_1)}{g(x_1|T_1)}. \quad (4.11)$$

Proposition 4.2. *Under Assumptions 4.2, 4.3, 4.5, and 4.6, the mdFWER of Procedure 4.2 is strongly controlled at level α when $n = 3$.*

Remark 4.4. In the case of three hypotheses, suppose that the test statistics $T_i, i = 1, \dots, 3$ are trivariate normally distributed with the mean θ_i . Without loss of generality, assume $\theta_i > 0, i = 1, 2$ and $\theta_3 = 0$, that is, H_1 and H_2 are false and H_3 is true. Let $\Sigma = (\sigma_{ij}), i, j = 1, \dots, 3$, denote the variance-covariance matrix of T_i 's. It is easy to see that Assumption 4.2 is always satisfied. Also, when $\sigma_{ij} \geq 0$ for $i \neq j$, Assumption 4.3 and Assumption 4.5 are satisfied. Finally, when $\sigma_{22} = \sigma_{33}$ and $\sigma_{12} = \sigma_{13}$, Assumption 4.6 is satisfied.

Finally, we consider the general case of n hypotheses. Now we must consider the multivariate monotone likelihood ratio property, described as follows. For any given $j = 1, \dots, i_0 - 1$, let $f(x|T_1, \dots, T_{j-1})$ and $g(x|T_1, \dots, T_{j-1})$ denote the probability density functions of T_j and T_{i_0} conditional on T_1, \dots, T_{j-1} , respectively.

Assumption 4.7 (Multivariate Monotone Likelihood Ratio (MMLR)). *For any given values of T_1, \dots, T_{j-1} , $f(x|T_1, \dots, T_{j-1})$ and $g(x|T_1, \dots, T_{j-1})$ have the monotone likelihood ratio (MLR) property in x , i.e., for any $x_2 > x_1$, we have*

$$\frac{f(x_2|T_1, \dots, T_{j-1})}{g(x_2|T_1, \dots, T_{j-1})} \geq \frac{f(x_1|T_1, \dots, T_{j-1})}{g(x_1|T_1, \dots, T_{j-1})}. \quad (4.12)$$

Theorem 4.4. *Under Assumptions 4.2, 4.3, 4.5, and 4.7, the mdFWER of Procedure 4.2 is strongly controlled at level α .*

4.7 Conclusions

In this chapter, we consider the problem of simultaneously testing multiple prior-ordered hypotheses accompanied by directional decisions. The conventional fixed sequence procedure augmented with additional directional decisions are proved to control the mdFWER under independence and some dependence, whereas, this procedure is also shown to be far too liberal to control the mdFWER, if no dependence assumptions are imposed on the test statistics. We hope that the approaches and techniques developed in this chapter will also shed some light on attacking the

notoriously challenging problem of controlling the mdFWER under dependence for p -value ordered stepwise procedures.

CHAPTER 5

FDR CONTROLLING PROCEDURES FOR TESTING A-PRIORI ORDERED HYPOTHESES

5.1 Introduction

In large scale multiple testing problems in applications such as stream data, statistical process control, etc., the tested hypotheses are ordered a-priori by time and it is desired to control False Discovery Rate (FDR) while making real-time decisions about rejecting or accepting a hypothesis and requiring to test all the hypotheses in the sequence irrespective of the number of prior acceptances.

A main feature of the data in these applications is that the data is not static and has a natural hierarchy in the data points that evolve over time and at any instance we only have the information up till that instance. This incomplete information implies that we do not have all the p -values available at the time. This presents several challenges in the development of multiple testing methodology. The existing FDR controlling procedures, such as the Benjamini-Hochberg procedure or any other stepwise procedures have an inherent assumption that all the p -values are available at hand before the method is applied. As a result, these procedures are unable to make real time decisions based on the incomplete information pertaining to the p -values.

A few researchers did some work in the control of FDR in the context of hierarchically ordered hypotheses. Farcomeni and Finos (2013) developed an FDR controlling procedure based on a data-driven ordering of hypotheses where hypotheses are tested until a suitable number are not rejected. Lynch et al. (2014) introduced an FDR controlling procedure in fixed sequence multiple testing which allows more than one acceptance in the fixed sequence testing of hypotheses. Though these procedures incorporate the hierarchy of hypotheses while controlling FDR, they do not allow

testing of each and every hypothesis in the sequence. This motivates us to our goal to develop procedures that exploit the natural hierarchy in the data and control FDR while making real time decisions. In the FWER framework, Wiens (2003) and Wiens and Dmitrienko (2005) developed the Fallback procedure that allows testing of all hypotheses in a sequence, by awarding the later critical constant in case of rejection of a hypothesis. Borrowing the idea from this procedure we develop procedures in the FDR control framework that are suitable for applications requiring real-time decision making.

The rest of the chapter is organized as follows: In Section 5.2, we describe some motivating examples for this research. In Section 5.3, we discuss the unique structure of hypotheses in these applications. We present the proposed methods in Section 5.4. We present a simulation study in Section 5.5 and conclusions in Section 5.6.

5.2 Motivating Examples

A data stream is a sequence of continuously arriving data points obtained as a result of a continuous data generation process that might evolve over time. Some examples are cellular network monitoring data, computer network monitoring data, web click-stream data, readings from stock quotes, etc. Very Often, the stream data is analyzed to detect any problems or failures appearing in the generation process and critical decisions are made based on the information provided by data stream. In such situations it becomes important to ensure that the number of false alarms is statistically controlled at a fixed level, not to overwhelm the users. Take the example of data collected from a cellular network. Here, data points arrive continuously that carry information on attributes like a call connection problem, voice quality problem, data connection failure, etc. These anomalous data points are formed into clusters based on several characteristics, for example, their geographical vicinity. These clusters give an indication of problems at a transmission tower in the corresponding

area. The problem is to identify whether a cluster so formed indicates the need for immediate action to be taken at the corresponding transmission tower, that is, we need to identify whether the cluster is an outlier or not.

Statistical Process Control (SPC) is the branch of Statistics that deals with methods applied to a continuous process, say, manufacturing lines, in order to monitor and control the process. This ensures that the process operates at its full potential and can make as much conforming product as possible with a minimum of waste. SPC emphasizes early detection and prevention of problems, rather than the correction of problems after they have occurred. This requires testing of hypotheses at regular intervals of time while making real time decisions on whether a process is out of control at the time of testing or not. This needs to be handled by an appropriate multiple testing methodology to restrict the error of not identifying the process to be out of control.

These applications motivate us to develop multiple testing methods that can be used to identify anomalous clusters or identifying whether the process is out of control or not, while controlling an appropriate rate of false positives.

5.3 Structure of Hypotheses

As seen in the motivating examples, the hypotheses tested in a sequence in the applications have a very unique structure which is not typically discussed in the existing multiple testing procedures. In most of the multiple testing procedures the testing sequence is such that most of the false null hypotheses occur ahead of the true null hypotheses. This sequence is achieved in the stepwise procedures through p -value ordering. In fixed sequence multiple testing, the order of the hypotheses is fixed based on their relative importance and as a result it is practical to assume that most of the false null hypotheses occur before the true null hypotheses in the sequence. But this is not true in the case of stream data or SPC. The monitoring of the underlying data

generation process is done in real time and the hypotheses tested do not usually have any structure in the occurrence of true null or false null hypothesis.

At any given instance, the purpose is to test whether a process is out of control or not. A scientist, who is monitoring the process, chooses the method to obtain the test statistic and p -value to test the null and alternative hypotheses of the following sort:

H : The underlying process is in control

vs.

H' : The underlying process is out of control.

Whenever the null hypothesis is rejected, remedial measures are immediately taken to rectify the situation and make the process “in control” again. As a result, we expect to see very few false null hypotheses in the testing sequence. The only assumption we can make here is that the probability of a null hypothesis being false is very low. This kind of structure is unique and needs new procedures for multiplicity adjustment.

5.4 Procedures for Testing a-priori Ordered Hypotheses

In this section we present the procedures we develop for testing a-priori ordered hypotheses that allow testing of all the hypotheses while utilizing the partial information available at the time of testing a hypothesis and controlling FDR. In the framework of FWER control, Wiens (2003)’s Fallback procedure and Qiu et al. (2014) Generalized Fixed Sequence Procedure A1 allow testing of all the hypotheses in a sequence while strongly controlling FWER. Motivated by the idea of awarding the critical constants on rejection of a null hypothesis, that is showcased in these procedures for controlling FWER, we develop FDR controlling procedures for applications that require real time decision making.

Suppose, H_1, H_2, \dots, H_m are m null hypotheses ordered a-priori in time. Let, P_1, P_2, \dots, P_m be the p -values for testing the a-priori ordered hypotheses. A conventional method for controlling FDR while testing a-priori ordered hypotheses can be defined as follows:

Definition 5.1 (Conventional Method for a-priori Ordered Hypotheses).

Define the critical constants, $\delta_1, \dots, \delta_m$. Reject H_i if $P_i \leq \delta_i$, for $i = 1, \dots, m$.

5.4.1 Fallback-Type Procedure under PRDS

Let, $\omega_1, \omega_2, \dots, \omega_m$ denote the series of weights such that $\sum_{i=1}^m \omega_i = 1$. Let R_i denote the number of hypotheses rejected before H_i is tested. Note that R_i is a random variable and $R_1 = 0$ with probability 1.

Assumption 5.1 (PRDS Property). *The vector of p -values $\mathbf{P} = (P_1, \dots, P_m)$, is said to be positive regression dependent on a subset (PRDS) I_0 if,*

$$Pr(\mathbf{P} \in \mathbf{C} | P_i) \nearrow \text{ in } P_i \quad (5.1)$$

for each $i \in I_0$, for any increasing set \mathbf{C} and vice versa.

Procedure 5.1 (Fallback-Type Method for controlling FDR under PRDS).

Step 1: Set, $\alpha_1 = \omega_1 \alpha$. Reject H_1 if $P_1 \leq \alpha_1$.

Step 2, \dots, m : If H_{i-1} is rejected, then set $\alpha_i = \omega_i \alpha + \alpha_{i-1}$; otherwise, set $\alpha_i = \omega_i \alpha$.

Reject H_i if $P_i \leq \left(1 + \frac{R_i}{m-i+1}\right) \alpha_i$.

Theorem 5.1. *Under Assumption 5.1, Procedure 5.1 strongly controls FDR at level α .*

Proof Let, U denote the index of the first rejected true null hypothesis. Let, m_0 out of m hypotheses are true null and u_1, u_2, \dots, u_{m_0} denote the indices of the true null hypotheses. Let, V denote the number of true null hypotheses rejected and S denote

the number of false null hypotheses rejected. Then,

$$\text{FDR} = E \left(\frac{V}{V+S} I \{V > 0\} \right) \quad (5.2)$$

The event $\{V > 0\}$ is equivalent to $\cup_{i=1}^{m_0} \{U = u_i\}$. Thus,

$$\text{FDR} = \sum_{i=1}^{m_0} E \left(\frac{V}{V+S} I \{U = u_i\} \right) \quad (5.3)$$

Define,

$$\alpha_{u_i}^* = \alpha \sum_{j=u_{i-1}+1}^{u_i} \omega_j \quad i = 1, \dots, m_0. \quad (5.4)$$

Consider the event $\{U = u_i\}$. This event means that first Type I error occurs by rejecting the true null H_{u_i} . The first $u_i - 1$ hypotheses are either false nulls ($u_i - i$ in number) or accepted true nulls ($i - 1$ in number). Thus, $\{U = u_i\}$ implies,

$$\begin{aligned} V &\leq m - u_i + 1 \\ S &\geq R_{u_i} \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{V}{V+S} &\leq \frac{V}{V+R_{u_i}} \\ &\leq \frac{m - u_i + 1}{m - u_i + 1 + R_{u_i}} \end{aligned} \quad (5.5)$$

For each u_i , define,

$$\beta_r = \frac{m - u_i + 1 + r}{m - u_i + 1} \alpha_{u_i}, \quad r = 0, \dots, u_i - 1.$$

Then, for each u_i we have,

$$E \left(\frac{V}{V+S} I \{U = u_i\} \right)$$

$$\leq E \left(\frac{m - u_i + 1}{m - u_i + 1 + R_{u_i}} I(U = u_i) \right) \quad (5.6)$$

$$\leq E \left(\frac{m - u_i + 1}{m - u_i + 1 + R_{u_i}} I(P_{u_i} \leq ((m - u_i + 1 + R_{u_i})/(m - u_i + 1)) \alpha_{u_i}) \right) \quad (5.7)$$

$$= \sum_{r=0}^{u_i-1} E \left(\frac{m - u_i + 1}{m - u_i + 1 + r} I\{P_{u_i} \leq \beta_r\} \middle| R_{u_i} = r \right) \cdot Pr(R_{u_i} = r) \quad (5.8)$$

The inequality in (5.6) follows due to (5.5). The inequality in (5.7) follows as the event $\{U = u_i\}$ implies the event $\{P_{u_i} \leq ((m - u_i + 1 + R_{u_i})/i(m - i + 1)) \alpha_{u_i}\}$. (5.8) follows as R_{u_i} can only take values between 0 and $u_i - 1$. Then (5.8) becomes,

$$\begin{aligned} &= \sum_{r=0}^{u_i-1} \frac{m - u_i + 1}{m - u_i + 1 + r} Pr(R_{u_i} = r | P_{u_i} \leq \beta_r) Pr(P_{u_i} \leq \beta_r) \\ &\leq \alpha_{u_i} \sum_{r=0}^{u_i-1} Pr(R_{u_i} = r | P_{u_i} \leq \beta_r) \\ &= \alpha_{u_i} \sum_{r=0}^{u_i-1} [Pr(R_{u_i} \geq r | P_{u_i} \leq \beta_r) - Pr(R_{u_i} \geq r + 1 | P_{u_i} \leq \beta_r)] \\ &= \alpha_{u_i} Pr(R_{u_i} \geq 0 | P_{u_i} \leq \beta_0) \\ &\quad + \alpha_{u_i} \left[\sum_{r=1}^{u_i-1} \{Pr(R_{u_i} \geq r | P_{u_i} \leq \beta_r) - Pr(R_{u_i} \geq r | P_{u_i} \leq \beta_{r-1})\} \right] \\ &\leq \alpha_{u_i} \end{aligned} \quad (5.9)$$

The inequality in (5.9) follows as the p -values are stochastically larger than the $U(0, 1)$ distribution. (5.10) follows by Assumption 5.1 and by noting that $\{R_{u_i} \geq r\}$ is a decreasing set. Thus, finally we have,

$$FDR = \sum_{i=1}^{m_0} E \left(\frac{V}{V + S} I\{U = u_i\} \right) \leq \sum_{i=1}^{m_0} \alpha_{u_i} \leq \sum_{i=1}^{m_0} \alpha_{u_i}^* \leq \alpha \sum_{i=1}^m \omega_i = \alpha.$$

This shows strong control of FDR under PRDS assumption of the p -values. ■

Remark 5.1. The critical constants in Procedure 5.1 are larger than the critical constants of the conventional Fallback procedure, so the newly introduced procedure is more powerful than the conventional Fallback procedure.

Remark 5.2. The critical constants utilize the information available before the testing the i^{th} hypothesis by including the term R_i , which is the number of hypotheses rejected before H_i . Thus, the information at hand is utilized which in turn improves the power of the procedure.

Remark 5.3. Note that the critical constants are increasing function of R_i , so it also implies that the idea of the usual Fallback procedure for awarding the rejections is utilized.

Optimality of the Procedure under PRDS: We look at a particular configuration of p -values under PRDS and discuss the optimality of the critical constants of the fallback type procedure by showing that the critical constants cannot be made any larger, while keeping the others constant, without losing control of FDR. The Fallback type procedure will be said to be optimal if it can strongly control FDR with any set of weights ω_i .

Let u_1, u_2, \dots, u_{m_0} denote the indices of the true null hypotheses in the sequence. Let us consider a joint distribution of the p -values such that,

- All the false null hypotheses occur before the true null hypotheses.
- Each false null p -value is 0.
- The true null p -values, $P_{u_i} \sim U(0, 1)$ and $P_{u_i} = p, i = 1, \dots, m_0$.

Let us assume, the weights are such that $\omega_1 \geq \omega_2 \geq \dots \geq \omega_m$, with, $\sum_{i=1}^{u_1} \omega_i = 1$ and $\omega_{u_1+1} = \dots = \omega_m = 0$.

Under such a joint distribution of p -values, for any $u_1, 1 \leq u_1 \leq m$, the first $u_1 - 1$ hypotheses are false nulls whose corresponding p -value is 0 with probability 1 and the remaining $m_0 = m - u_1 + 1$ hypotheses are true nulls. Thus, the sequence of p -values for the a priori ordered hypotheses is, $\mathbf{P} = (0, \dots, 0, p, \dots, p)$. Note that this sequence of p -values is positive regression dependent on a subset of the true null

p -values because, for any increasing set D ,

$$\Pr(\mathbf{P} \in D | P_{u_i} = p) = \begin{cases} 1 & \text{if } (0, \dots, 0, p, \dots, p) \in D, \\ 0 & \text{otherwise.} \end{cases} \quad (5.11)$$

The FDR of the fallback type procedure, under this configuration is exactly,

$$\begin{aligned} &= E\left(\frac{V}{V+S}I(V>0)\right) \\ &= \frac{m-u_1+1}{m} \cdot \frac{m}{m-u_1+1} \alpha_{u_1} \\ &= \alpha \sum_{j=1}^{u_1} \omega_j. \end{aligned} \quad (5.12)$$

Then, the FDR is exactly $= \alpha$ and thus, the critical constant cannot be made any larger without losing control of FDR.

Thus, we show optimality of Procedure 5.1 by showing a configuration where even a single critical constant cannot be made larger, while keeping the others same, without losing strong control of FDR.

5.4.2 Fallback-Type Procedure under Arbitrary Dependence

In this section, we discuss control of FDR under arbitrary dependence of the p -values. We show that with an appropriate modification in the critical constants of Procedure 5.1, we can achieve FDR control under arbitrary dependence of the p -values.

Procedure 5.2 (Fallback-Type Method for controlling FDR under Arbitrary Dependence).

Step 1: Set, $\alpha_1 = \omega_1 \alpha$. Reject H_1 if $P_1 \leq \alpha_1$.

Step 2, ..., m : If H_{i-1} is rejected, then set $\alpha_i = \omega_i \alpha + \alpha_{i-1}$; otherwise, set $\alpha_i = \omega_i \alpha$.

Reject H_i if,

$$P_i \leq \left(1 + \frac{R_i}{m-i+1}\right) \frac{\alpha_i}{c_i},$$

where, $c_i = 1 + \sum_{k=1}^{i-1} (1/(m-i+1+k))$, $i = 2, \dots, m$.

Theorem 5.2. *Procedure 5.2 strongly controls FDR at level α under arbitrary dependence of the p -values.*

The proof of this theorem is provided in Appendix B.

Remark 5.4. The correction factors c_i applied to the critical constants in Procedure 5.1 are very similar to the correction factor applied to the critical constants in the BH procedure in Benjamini and Yekutieli (2001). Figure 5.1 compares the two correction factors. Clearly, the correction factor of Procedure 5.2 is much smaller compared to that of the BY procedure (Benjamini and Yekutieli, 2001).

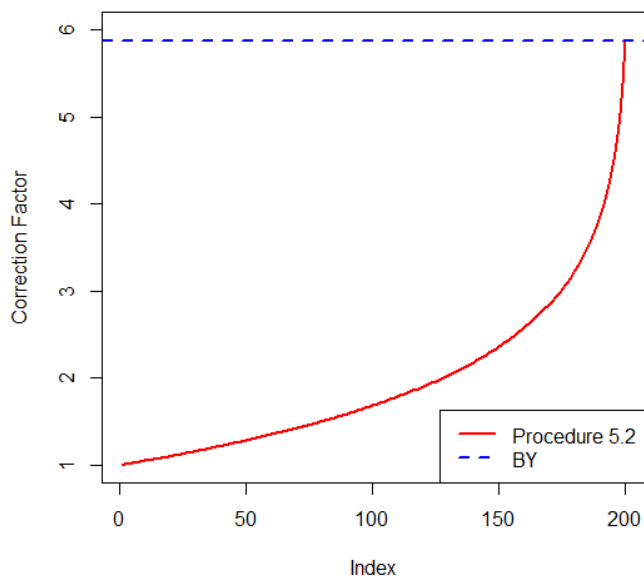


Figure 5.1 Comparison of the correction factors used in our procedure in Theorem 5.2 and the BY procedure for $m = 200$ hypotheses.

5.4.3 FDR Control under Independence

We have already seen that the fallback-type procedure under arbitrary dependence can be improved under the stronger condition of positive regression dependence of the p -values. This motivates us to explore whether an even improved procedure can be

obtained under independence of the p -values. In this section, we develop a procedure, for FDR control while testing a-priori ordered hypotheses under the independence of the p -values.

Procedure 5.3 (Method under Independence).

Define,

$$\alpha_i = \frac{(R_i + 1)\alpha}{m - [\alpha(m - i) + (1 - \alpha)R_i]}, \quad i = 1, \dots, m, \quad (5.13)$$

where, R_i is the number of hypotheses rejected before testing H_i .

If $P_i \leq \alpha_i$ then reject H_i , otherwise, accept H_i , $i = 1, \dots, m$.

Note that, by definition, $R_1 = 0$.

Theorem 5.3. *Procedure 5.3 can strongly control FDR at level α if the p -values are mutually independent.*

The proof of this theorem is inspired by Lynch et al. (2014)'s proof of the FDR control of the procedure allowing k acceptances under independence. Let us introduce a few notations. For $i = 1, \dots, m$, let V_i and S_i , respectively denote the numbers of false rejections and true rejections among the first i rejections and let J_i denote the index of the i^{th} rejected hypothesis. As defined in Lynch et al. (2014), if there are less than i rejections, we define, $V_i = V_{i-1}$, $S_i = S_{i-1}$ and $J_i = m + 1$. For notational convenience, we define, $V_0 = S_0 = J_0 = 0$, $V_0/0 = 0$ and $S_0/0 = 1$.

We use the Lemma 2.2 in Lynch et al. (2014), described below, to prove Theorem 5.3.

Lemma 5.1 (Lemma 2.2 in Lynch et al. (2014)).

The FDR of any fallback type method can be expressed as,

$$FDR = E \left(\sum_{i=1}^m \left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} \right) I\{J_i < m + 1\} \right). \quad (5.14)$$

We now present the proof of Theorem 5.3 which is inspired by the proof of Theorem 2.4 of Lynch et al. (2014).

Proof of Theorem 5.3: We note that for $i = 1, \dots, m$, if there are i rejections, then $i \leq J_i \leq m$. For $i, j = 1, \dots, m$, define,

$$f_i(j) = \begin{cases} \frac{m-j}{m-i} \alpha \frac{S_i}{i}, & \text{if } i < m, \\ 0, & \text{if } i = m \end{cases} \quad (5.15)$$

and $W_i(j) = I\{J_{i-1} \leq j, J_i > j\}$. We use the following two inequalities, proved in Lynch et al. (2014),

$$I\{J_i = j\} = W_i(j-1)I\{P_j \leq \alpha_j\} \quad (5.16)$$

and

$$W_i(j) - W_i(j-1) = I\{J_{i-1} = j\} - I\{J_i = j\}. \quad (5.17)$$

Here, α_j 's are those defined in Procedure 5.3. We next use the following two more equations, the proofs of which are given in Appendix B,

$$E \left(\left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} + f_i(j) - f_{i-1}(j) \right) I\{J_i = j\} \right) \leq E \left(\frac{\alpha}{m-i+1} \frac{S_{i-1}}{i-1} W_i(j-1) \right) \quad (5.18)$$

and

$$\begin{aligned} & E(f_{i-1}(J_{i-1})I\{J_{i-1} < m+1\} - f_{i-1}(J_i)I\{J_i < m+1\}) \\ &= E \left(\sum_{j=i}^m \frac{\alpha}{m-i+1} \frac{S_{i-1}}{i-1} W_i(j-1) \right). \end{aligned} \quad (5.19)$$

Next, from Lemma 5.1, we get the desired result as follows:

$$\text{FDR} = \sum_{i=1}^m E \left(\left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} \right) I\{J_i < m+1\} \right)$$

$$\begin{aligned}
&= \sum_{i=1}^m E \left(\sum_{j=i}^m \left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} + f_i(j) - f_{i-1}(j) \right) I\{J_i = j\} \right) \\
&\quad - \sum_{i=1}^m E (f_{i-1}(J_{i-1}) I\{J_{i-1} < m+1\} - f_{i-1}(J_i) I\{J_i < m+1\}) \\
&\quad + \sum_{i=1}^m E (f_{i-1}(J_{i-1}) I\{J_{i-1} < m+1\} - f_i(J_i) I\{J_i < m+1\}) \\
&\leq \sum_{i=1}^m E (f_{i-1}(J_{i-1}) I\{J_{i-1} < m+1\} - f_i(J_i) I\{J_i < m+1\}) \tag{5.20} \\
&= E \left(\frac{m - J_0}{m} \alpha I\{J_0 < m+1\} - 0 \right) \\
&\leq \alpha. \tag{5.21}
\end{aligned}$$

The inequality in (5.20) follows by (5.18) and (5.19). ■

Remark 5.5. The critical constants in Procedure 5.3 are much larger than the critical constants in Procedure 5.1. As a result, intuitively, Procedure 5.3 is much more powerful than Procedure 5.1.

5.5 Simulation Study

We did a simulation study to compare the performance of the proposed Procedures 5.1 - 5.3 in terms of FDR control and average power for different combinations of true and false null hypotheses. According to our knowledge, the only procedures available in the literature that are able to handle the testing in the applications discussed here are the Bonferroni procedure, Fallback procedure (Wiens, 2003) and Qiu et al. (2014)'s Generalized Fixed Sequence Procedures. These procedures are developed in the framework of FWER control so they can conservatively control FDR. We compare the performance of the proposed procedures, Procedure 5.1 (PRDS), Procedure 5.2 (Arbitrary Dependence), Procedure 5.3 (Independence) to the Wiens (2003)'s

Fallback procedure and Qiu et al. (2014)'s Generalized Fixed Sequence Procedure A1.

5.5.1 Study Design

The hypotheses we want to test, for $i = 1, \dots, m$, are as follows:

$$\begin{aligned}
 H_{0i} : \mu &= \mu_0 \\
 &\text{vs.} \\
 H_{1i} : \mu &= \mu_1,
 \end{aligned} \tag{5.22}$$

where, μ_0 denotes the ‘‘in control’’ mean and μ_1 denotes the ‘‘out of control’’ mean. We simulate test statistics T_1, \dots, T_m from normal distribution, where, m denotes the number of hypotheses tested in a sequence. The true null test statistics, m_0 in number, are from $N(\mu_0, \sigma^2)$ distribution. The false null test statistics, $m_1 = m - m_0$ in number, are from $N(\mu_1, \sigma^2)$ distribution. The test statistics are taken to be correlated with the following variance-covariance matrix which we call Adjacent-Correlation, as only the adjacent test statistics are correlated,

$$\Sigma_{m \times m}^{Adjacent} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 & \cdots & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & \cdots & \rho\sigma^2 & \sigma^2 \end{pmatrix}. \tag{5.23}$$

For the simulation, we generate μ_0 from Uniform distribution on interval $[0, 3]$ and μ_1 from Uniform distribution on interval $[3, 6]$. We fix $\sigma^2 = 1$. We consider the values $\rho = 0, 0.2, 0.5, 0.8$, to account for different degrees of correlation from independence to strong correlation while testing $m = 200, 400$ hypotheses. The weights used in the proposed Fallback-type FDR controlling procedures and the

Fallback procedure (Wiens, 2003) are equal weights. The p -values are calculated as,

$$P_i = Pr_{\mu=\mu_0}(T_i \geq t_i),$$

where, t_i denote the observed values of T_i , $i = 1, \dots, m$.

Order of Hypotheses: We consider a pre-specified order of the hypotheses which conforms to the special structure of hypotheses discussed in Section 5.3. For choosing each null hypothesis, we generate a random number from Bernoulli distribution with outcomes “false” and “true”. The outcome “false” has probability π . Thus, a null hypothesis is set to be false with probability π and set to be true with probability $(1 - \pi)$. Here, π is generally fixed low, say, 0.1, as we expect that the procedure is “in control” most of the time.

5.5.2 Results of The Simulation Study

Independence: We summarize the results of the simulation studies for the independence case in Figure 5.2. The vertical axis denotes the average FDR (left panel) and average power (right panel) for different values of π along the horizontal axis for $m = 200$ and 400 . As desired, all five procedures control the FDR on average at $\alpha = 0.05$. However, the proposed procedure under independence, Procedure 5.3, has the highest power compared to all other methods.

Dependence: We summarize the results of the simulation studies for the dependence case in Figures 5.3-5.5. The correlation we consider is Adjacent-Correlation as in (5.23) with $\rho = 0.2, 0.5, 0.8$. Again, as desired, all four procedures control the FDR on average at $\alpha = 0.05$. However, the proposed procedure under PRDS, Procedure 5.1, has the highest power compared to all other methods.

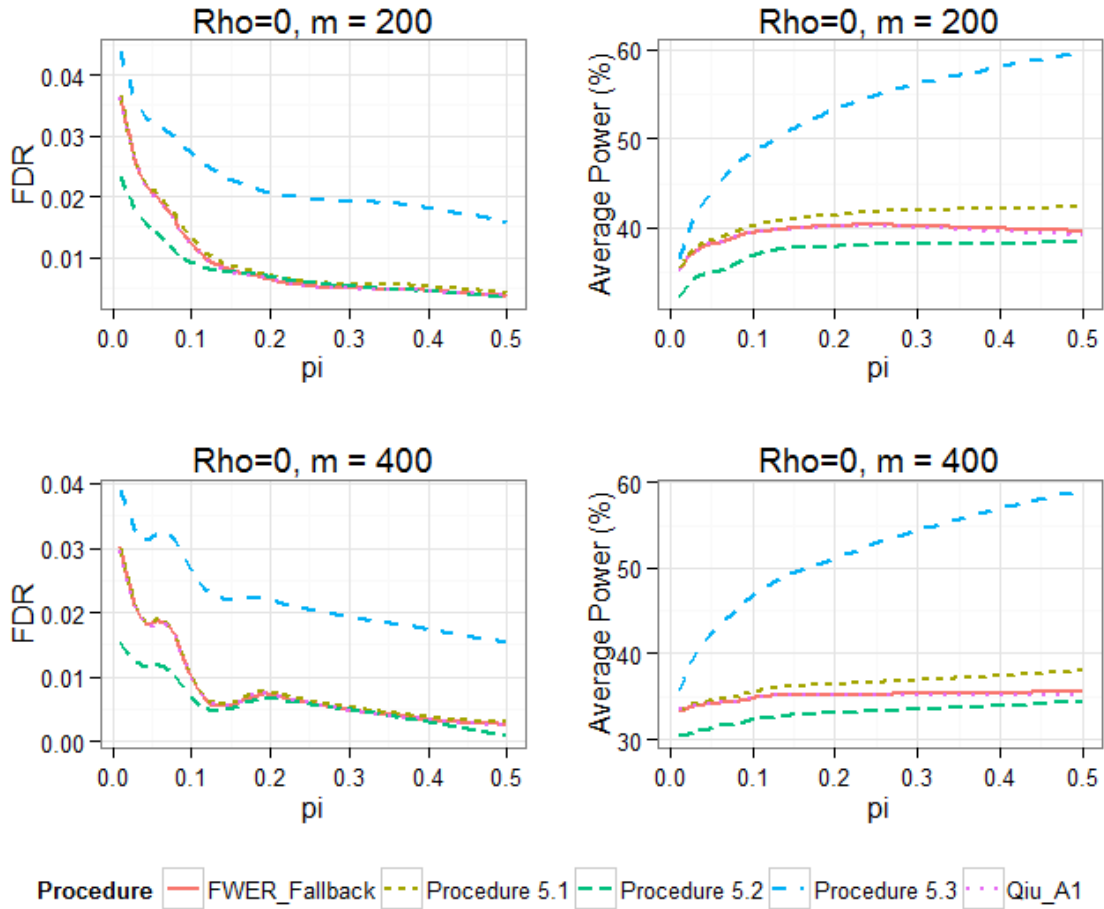


Figure 5.2 FDR and Average Power for $m = 200, 400$ hypotheses under independence ($\rho = 0$).

5.6 Summary and Conclusion

In this chapter, we develop multiple testing procedures that control the FDR and are appropriate for testing a-priori ordered hypotheses in the applications such as network monitoring in stream data, Statistical Process Control, etc., where we need to make real-time decisions based on the partial information available at the time of testing. We developed procedures that cover different dependence scenarios. Our simulation study shows that our proposed procedures always outperform the Fallback procedure (Wiens, 2003) and the Qiu et al. (2014)'s Generalized Fixed Sequence Procedure A1.

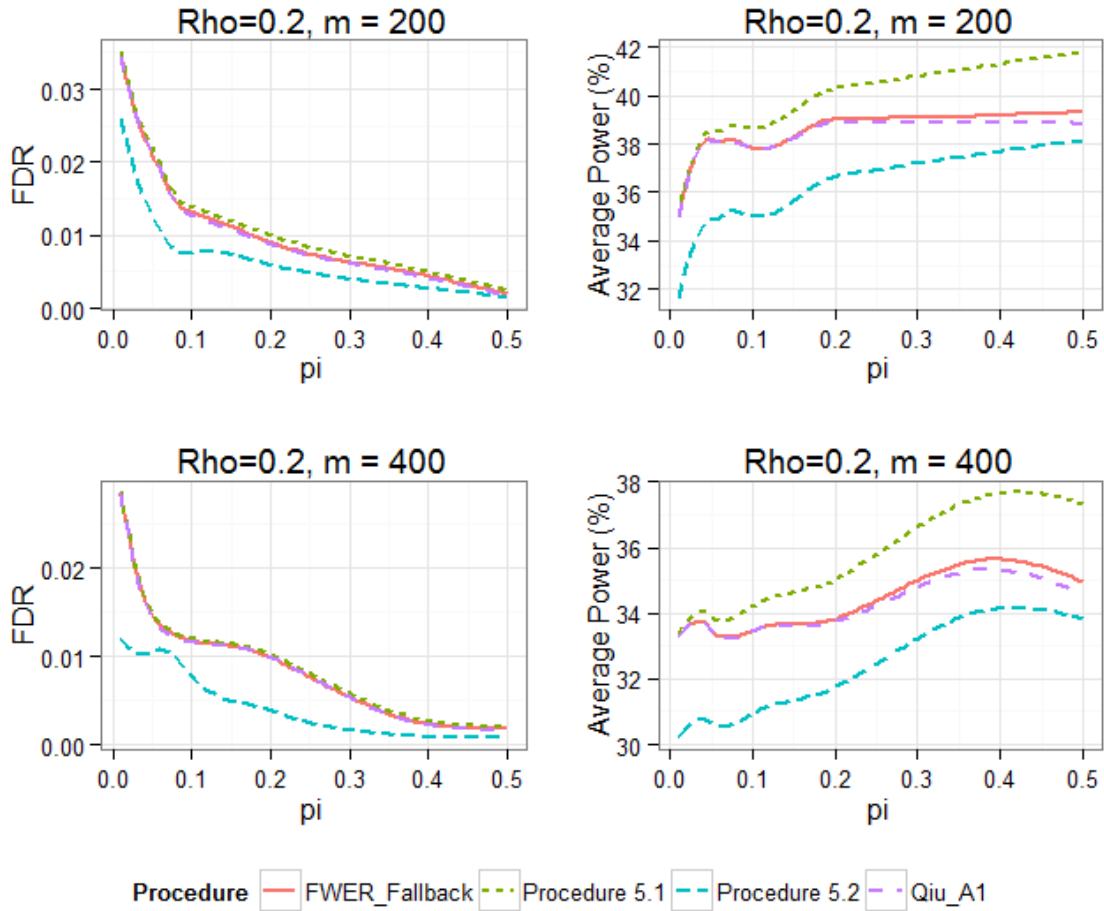


Figure 5.3 FDR and Average Power for $m = 200, 400$ hypotheses under Adjacent-Correlation with $\rho = 0.2$.

But this is expected as both these procedures are developed to control FWER and perform conservatively in FDR framework.

In the simulation study, we considered the correlation structure where only the adjacent test statistics and as a result, only the adjacent p -values are correlated. It would be interesting to see how the procedures perform in other kinds of correlation structures, a few examples would be equal correlation structure, AR(1) correlation structure, etc.

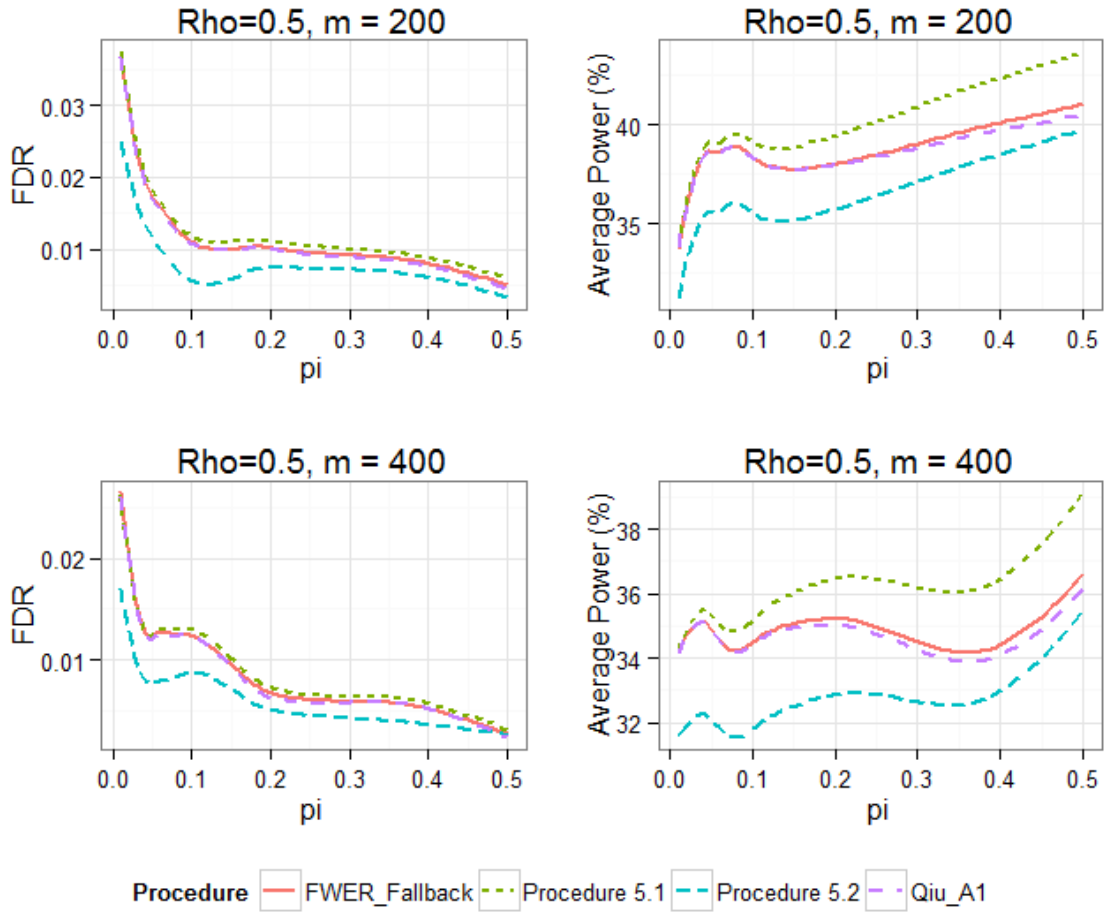


Figure 5.4 FDR and Average Power for $m = 200, 400$ hypotheses under Adjacent-Correlation with $\rho = 0.5$.

The procedures developed in this chapter are for the scenario where, m , the number of hypotheses tested in a sequence is finite. As a future work, it will be interesting to develop procedures that can handle the case when m is not finite.

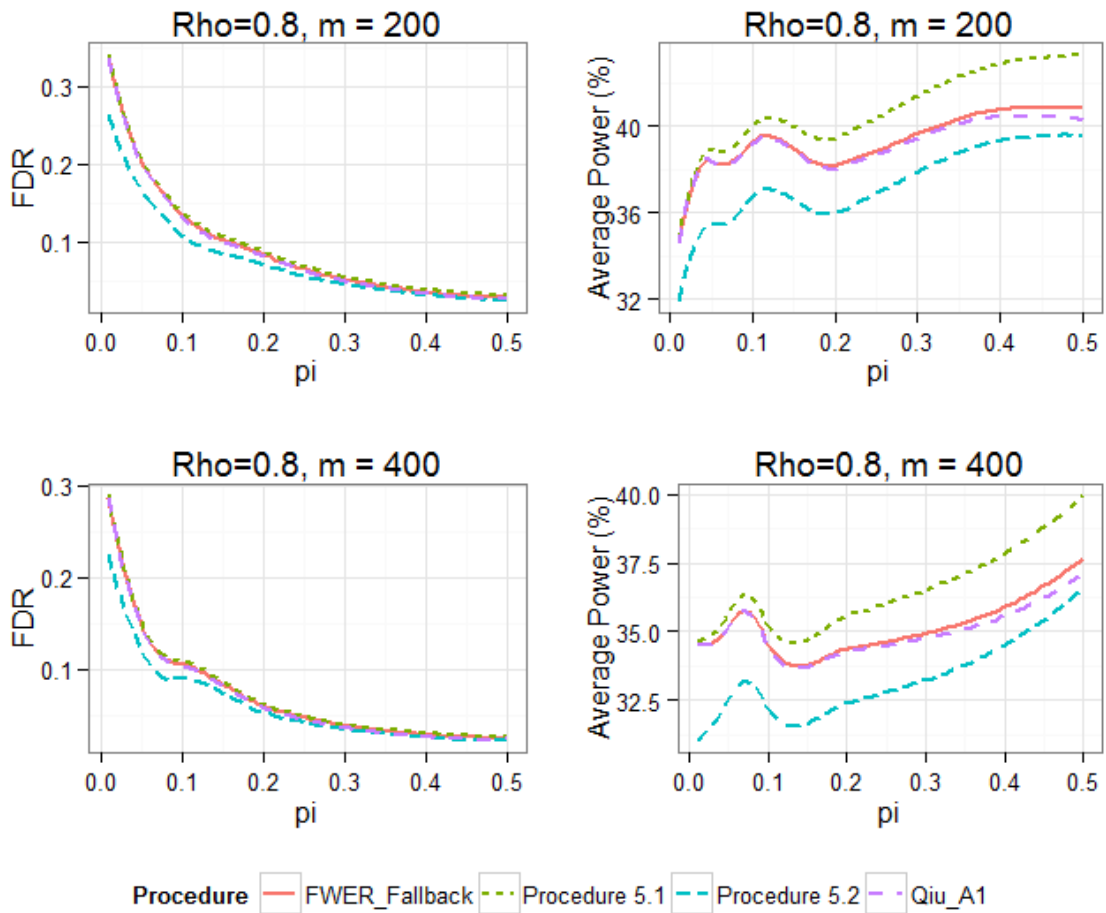


Figure 5.5 FDR and Average Power for $m = 200, 400$ hypotheses under Adjacent-Correlation with $\rho = 0.8$.

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this dissertation, we consider testing hypotheses with several different types of structures that sometimes require making directional decisions based on the problem of interest, and develop novel methods which exploit these inherent complex structures.

In Chapter 2, we present a general procedure for large scale multiple testing of hierarchically structured families of hypotheses while making multidimensional directional decisions with proven control of mdFDR. The approach developed in this chapter can also be used for addressing the problem of variable selection in high-dimensional regression, where data is available on thousands of features/explanatory variables. In large scale data, where the response of different study groups is available for the explanatory variables, the variable selection problem for modeling the response from the different study groups can be formulated as a multiple testing problem with families of hypotheses having similar hierarchical structure discussed in this research. The idea is inspired by the hierarchical approach proposed in Meinshausen (2008), where, variable importance is first tested at the coarsest level, corresponding to the global null hypothesis and the method then tries to attribute any effect to smaller subclusters or even individual variables while controlling FWER. A future direction of research related to the work in Chapter 2 is to explore methods for variable selection for each study groups in the data while controlling FDR.

In Chapter 3, we develop a specific methodology based on the general procedure presented in Chapter 2 and use it for Uterine Fibroid gene expression data analysis to identify important genes and pathways associated with formation and growth of the fibroids. The results we obtained are confounded by the race and age groups of the

subjects in the study. A future direction of research related to this work is to explore methods for testing that remove this confounding effect and gain deeper insights into genes and pathways associated with fibroids.

In Chapter 4, we consider the problem of simultaneously testing multiple pre-ordered hypotheses accompanied by directional decisions. The conventional fixed sequence procedure augmented with additional directional decisions is proved to control the mdFWER under independence and some dependence. A more conservative procedure is developed under arbitrary dependence of test statistics. A future direction of research related to the work in Chapter 4 is to explore the challenge of attacking the notoriously challenging problem of controlling the mdFWER for p -value ordered stepwise procedures. Another direction for future research is to explore the development of fixed sequence type multiple testing methods for group sequential designs in clinical trials where the hypotheses tested at different stages act as gatekeepers for future hypotheses.

Finally in Chapter 5, we present several new FDR controlling procedures for large scale testing of a-priori ordered hypotheses in applications such as stream data, statistical process control, etc. We developed the procedures for testing a sequence of hypotheses where the total number of hypotheses tested in the sequence is fixed. A future direction of research related to the work in Chapter 5 is to explore the problem of testing multiple pre-ordered families of hypotheses. Another direction for future research is to develop generalized fixed sequence procedures for controlling the FDR.

As a concluding remark, we would like to point out that we developed methods in this dissertation that have varied applications in different fields and provide a good direction for the research in development of new and powerful multiple testing procedures.

APPENDIX A

This appendix contains the proofs of the lemmas and theorems stated but not proved in Chapter 4.

A.1 Proof of Lemma 4.1.

Let T and P denote the test statistic and the corresponding p -value for testing H , respectively. When testing H , a Type III error occurs if H is rejected and $\theta T < 0$. Then, the Type III error rate is given by $Pr(P \leq \alpha, \theta T < 0)$.

When $\theta > 0$, we have

$$\begin{aligned} Pr(P \leq \alpha, \theta T < 0) &= Pr(2F_0(T) \leq \alpha, T < 0) \\ &= Pr\left(T \leq F_0^{-1}\left(\frac{\alpha}{2}\right)\right) = F_\theta\left(F_0^{-1}\left(\frac{\alpha}{2}\right)\right) \\ &\leq F_0\left(F_0^{-1}\left(\frac{\alpha}{2}\right)\right) = \frac{\alpha}{2}. \end{aligned}$$

The inequality follows from the assumption that F_θ is stochastically increasing in θ . Similarly, when $\theta < 0$, we can also prove that $Pr(P \leq \alpha, \theta T < 0) \leq \frac{\alpha}{2}$. ■

A.2 Proof of Theorem 4.1

(i). Induction will be used to show that Procedure 1 strongly controls the mdFWER at level α . First, consider the case of $n = 2$. We show control of the mdFWER of Procedure 4.1 in all possible combinations of true and false null hypotheses while testing two hypotheses H_1 and H_2 .

Case I: H_1 is true. Type I or Type III error occurs only when H_1 is rejected.

$$\text{mdFWER} = Pr(P_1 \leq \alpha) \leq \alpha.$$

Case II: Both H_1 and H_2 are false. We have no Type I errors but only Type III errors.

$$\begin{aligned}
\text{mdFWER} &= Pr(\{P_1 \leq \alpha, T_1\theta_1 < 0\} \cup \{P_1 \leq \alpha, P_2 \leq \alpha, T_2\theta_2 < 0\}) \\
&\leq Pr(P_1 \leq \alpha, T_1\theta_1 < 0) + Pr(P_2 \leq \alpha, T_2\theta_2 < 0) \\
&\leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

The first inequality follows from Bonferroni inequality and the second follows from Lemma 4.1.

Case III: H_1 is false and H_2 is true. The mdFWER is bounded above by

$$\begin{aligned}
&Pr(\text{ make Type III error when testing } H_1) \\
&\quad + Pr(\text{ make Type I error when testing } H_2) \\
&\leq Pr(P_1 \leq \alpha, T_1\theta_1 < 0) + Pr(P_2 \leq \alpha/2) \\
&\leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

The first inequality follows from Bonferroni inequality and the second follows from Lemma 4.1 and $P_2 \sim U(0, 1)$ since H_2 is true.

Now assume the inductive hypothesis that the mdFWER is bounded above by α when testing at most $n - 1$ hypotheses by using Procedure 4.1 at level α . In the following, we prove the mdFWER is also bounded above by α when testing n hypotheses H_1, \dots, H_n . Without loss of generality, assume H_1 is a false null (if H_1 is a true null, the desired result directly follows by using the same argument as in Case I of $n = 2$). Then, the mdFWER is bounded above by

$$\begin{aligned}
&Pr(\text{ make Type III error when testing } H_1) \\
&\quad + Pr(\text{ make at least one Type I or Type III errors when testing } H_2, \dots, H_n) \\
&\leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

The inequality follows from the induction assumption, noticing that H_2, \dots, H_n are tested by using Procedure 4.1 at level $\alpha/2$. Thus, the desired result follows.

(ii). We now prove that the critical constants are unimprovable. For instance, when H_1 is true, it is easy to see that the first critical constant, α , is unimprovable. For each given $k = 2, \dots, n$, when $\theta_i > 0, i = 1, \dots, k-1$ and $\theta_k = 0$, that is, $H_i, i = 1, \dots, k-1$ are false and H_k is true, we present a simple joint distribution of the test statistics T_1, \dots, T_k to show that the k th critical constant of this procedure is also unimprovable.

Define $Z_k \sim N(0, 1)$ and $Z_i = \Phi^{-1}(|2\Phi(Z_{i+1}) - 1|), i = 1, \dots, k-1$, where $\Phi(\cdot)$ is the cdf of $N(0, 1)$. Let q_i denote Z_i 's upper $\alpha/2^i$ quantile. It is easy to check that for each $i = 1, \dots, k, Z_i \sim N(0, 1)$. Thus, $-q_i$ is Z_i 's lower $\alpha/2^i$ quantile. In addition, by the construction of Z_i 's, it is easy to see that the event $Z_i \geq q_i$ is equivalent to the event $Z_{i+1} \notin (-q_{i+1}, q_{i+1})$.

Let $T_i = Z_i + \theta_i, i = 1, \dots, k$, thus, $T_i \sim N(\theta_i, 1)$. Then, as $\theta_i \rightarrow 0+$ for $i = 1, \dots, k-1$, we have

$$\begin{aligned}
\text{mdFWER} &= \sum_{j=1}^{k-1} Pr(T_1 \geq q_1, \dots, T_{j-1} \geq q_{j-1}, T_j \leq -q_j) \\
&\quad + Pr(T_1 \geq q_1, \dots, T_{k-1} \geq q_{k-1}, T_k \notin (-q_k, q_k)) \\
&= \sum_{j=1}^{k-1} Pr(Z_1 \geq q_1, \dots, Z_{j-1} \geq q_{j-1}, Z_j \leq -q_j) \\
&\quad + Pr(Z_1 \geq q_1, \dots, Z_{k-1} \geq q_{k-1}, Z_k \notin (-q_k, q_k)) \\
&= \sum_{j=1}^{k-1} Pr(Z_j \leq -q_j) + Pr(Z_k \notin (-q_k, q_k)) \\
&= \sum_{j=1}^{k-1} \frac{\alpha}{2^j} + \frac{\alpha}{2^{(k-1)}} = \alpha.
\end{aligned}$$

Thus, the k th critical constant of Procedure 4.1 is unimprovable and hence each critical constant of Procedure 4.1 is unimprovable under arbitrary dependence. ■

A.3 Proof of Lemma 4.2.

Note that when $\theta_1 > 0$ and $\theta_2 = 0$, we have

$$\begin{aligned}
& \text{mdFWER} \\
&= Pr(P_1 \leq \alpha, \theta_1 T_1 < 0) + Pr(P_1 \leq \alpha, \theta_1 T_1 \geq 0, P_2 \leq \alpha) \\
&= Pr(P_1 \leq \alpha, T_1 < 0) + Pr(P_1 \leq \alpha, T_1 \geq 0, P_2 \leq \alpha, T_2 > 0) \\
&\quad + Pr(P_1 \leq \alpha, T_1 \geq 0, P_2 \leq \alpha, T_2 \leq 0) \\
&= Pr(2F_0(T_1) \leq \alpha) + Pr(2(1 - F_0(T_1)) \leq \alpha, 2(1 - F_0(T_2)) \leq \alpha) \\
&\quad + Pr(2(1 - F_0(T_1)) \leq \alpha, 2F_0(T_2) \leq \alpha) \\
&= Pr(T_1 \leq c_1) + Pr(T_1 \geq c_2, T_2 \geq c_2) + Pr(T_1 \geq c_2, T_2 \leq c_1) \\
&= F_{\theta_1}(c_1) + 1 - F_{\theta_1}(c_2) - F_0(c_2) + F_{(\theta_1, 0)}(c_2, c_2) + F_0(c_1) - F_{(\theta_1, 0)}(c_2, c_1) \\
&= \alpha + F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{(\theta_1, 0)}(c_2, c_2) - F_{(\theta_1, 0)}(c_2, c_1). \tag{A.1}
\end{aligned}$$

Specifically, under Assumption 4.1 (independence), (A.1) can be simplified as,

$$\begin{aligned}
& \alpha + F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{\theta_1}(c_2)F_0(c_2) - F_{\theta_1}(c_2)F_0(c_1) \\
&= \alpha + F_{\theta_1}(c_1) - \alpha F_{\theta_1}(c_2).
\end{aligned}$$

Similarly, when $\theta_1 < 0$ and $\theta_2 = 0$, we can prove that

$$\text{mdFWER} = 1 + F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{(\theta_1, 0)}(c_1, c_1) - F_{(\theta_1, 0)}(c_1, c_2).$$

■

A.4 Proof of Lemma 4.3

By using the same arguments as in Theorem 4.1, we can easily prove control of the mdFWER of Procedure 4.2 in the case of $n = 2$ when H_1 is true or both H_1 and H_2 are false. In the following, we prove the desired result also holds when H_1 is false and H_2 is true.

Note that H_1 is false and H_2 is true imply $\theta_1 \neq 0$ and $\theta_2 = 0$. To show that the mdFWER is controlled for $\theta_1 > 0$ and $\theta_2 = 0$, we only need to show by Lemma 4.2 that $\alpha + F_{\theta_1}(c_1) - \alpha F_{\theta_1}(c_2) \leq \alpha$. This is equivalent to show

$$F_{\theta_1}(c_2)(F_0(c_2) - F_0(c_1)) \leq F_{\theta_1}(c_2) - F_{\theta_1}(c_1). \quad (\text{A.2})$$

For proving (A.2), it is enough to prove the following, as $0 \leq F_0(c_2) \leq 1$,

$$F_{\theta_1}(c_2)(F_0(c_2) - F_0(c_1)) \leq F_0(c_2)(F_{\theta_1}(c_2) - F_{\theta_1}(c_1)). \quad (\text{A.3})$$

Dividing both sides of (A.3) by $F_{\theta_1}(c_2)F_0(c_2)$, we see that we only need to prove,

$$1 - \frac{F_0(c_1)}{F_0(c_2)} \leq 1 - \frac{F_{\theta_1}(c_1)}{F_{\theta_1}(c_2)},$$

which follows directly from (4.5) and Assumption 4.2 (MLR).

Similarly, to show that the mdFWER is controlled for $\theta_1 < 0$ and $\theta_2 = 0$, we only need to show by Lemma 4.2 that $1 + \alpha F_{\theta_1}(c_1) - F_{\theta_1}(c_2) \leq \alpha$. This is equivalent to showing

$$(1 - \alpha)(1 - F_{\theta_1}(c_1)) \leq F_{\theta_1}(c_2) - F_{\theta_1}(c_1).$$

Writing $1 - \alpha$ as $(1 - F_0(c_1)) - (1 - F_0(c_2))$ and writing $F_{\theta_1}(c_2) - F_{\theta_1}(c_1)$ as $(1 - F_{\theta_1}(c_1)) - (1 - F_{\theta_1}(c_2))$, we get that it is equivalent to prove

$$[(1 - F_0(c_1)) - (1 - F_0(c_2))](1 - F_{\theta_1}(c_1)) \leq (1 - F_{\theta_1}(c_1)) - (1 - F_{\theta_1}(c_2)). \quad (\text{A.4})$$

Since $0 \leq 1 - F_0(c_1) \leq 1$, to prove inequality (A.4), it is enough to prove the following,

$$\begin{aligned} & (1 - F_{\theta_1}(c_1)) [(1 - F_0(c_1)) - (1 - F_0(c_2))] \\ & \leq (1 - F_0(c_1)) [1 - F_{\theta_1}(c_1)] - [1 - F_{\theta_1}(c_2)]. \end{aligned} \quad (\text{A.5})$$

Dividing both sides of (A.5) by $(1 - F_{\theta_1}(c_1))(1 - F_0(c_1))$, we see that proving (A.4) is equivalent to showing

$$\frac{1 - F_{\theta_1}(c_2)}{1 - F_{\theta_1}(c_1)} \leq \frac{1 - F_0(c_2)}{1 - F_0(c_1)}, \quad (\text{A.6})$$

which follows directly from (4.6) and Assumption 4.2 (MLR).

By combining the discussion of the above two cases, the desired result follows. ■

A.5 Proof of Theorem 4.2

The proof is by induction on number of hypotheses n . We already proved strong control of the mdFWER for $n = 2$ in Lemma 4.3. Let us assume the result holds for testing any $n = k$ hypotheses, that is, $\text{mdFWER} \leq \alpha$ while testing any k pre-ordered hypotheses. We now argue that it will hold for $n = k + 1$ hypotheses. Without loss of generality, assume H_1 is a false null, as in the proof of Theorem 4.1.

Let $V_{k+1}^{(-1)}$ denote the total number of Type I or Type III errors committed while testing H_2, \dots, H_{k+1} and excluding H_1 . Then, by the inductive hypothesis, the mdFWER while testing the k hypotheses H_2, \dots, H_{k+1} is $\Pr(V_{k+1}^{(-1)} > 0) \leq \alpha$. Then, the mdFWER of testing $k + 1$ hypotheses H_1, \dots, H_{k+1} is defined by

$$\begin{aligned} & \Pr\left(\{P_1 \leq \alpha, T_1\theta_1 < 0\} \cup \{P_1 \leq \alpha, T_1\theta_1 \geq 0, V_{k+1}^{(-1)} > 0\}\right) \\ &= \Pr(P_1 \leq \alpha, T_1\theta_1 < 0) + \Pr(P_1 \leq \alpha, T_1\theta_1 \geq 0) \cdot \Pr(V_{k+1}^{(-1)} > 0) \\ &\leq \Pr(P_1 \leq \alpha, T_1\theta_1 < 0) + \alpha \Pr(P_1 \leq \alpha, T_1\theta_1 \geq 0). \end{aligned} \quad (\text{A.7})$$

The equality follows by Assumption 4.1 (independence) and the inequality follows by the inductive hypothesis. Note that (A.7) is the same as (4.8) under independence, which is equal to the mdFWER of Procedure 4.2 in the case of two hypotheses. So again by applying Lemma 4.3, we get that $\text{mdFWER} \leq \alpha$ for $n = k + 1$. Hence, the proof follows by induction. ■

A.6 Proof of Theorem 4.3

Without loss of generality, we assume $\theta_i > 0$ if $\theta_i \neq 0$ for $i = 1, \dots, n$. Also, if there exists an i with $\theta_i = 0$, by induction, we can simply assume $i_0 = n$. Thus, to prove the mdFWER control of Procedure 4.2, we only need to consider two cases:

- (i) $\theta_i > 0$ for $i = 1, \dots, n$;
- (ii) $\theta_i > 0$ for $i = 1, \dots, n - 1$ and $\theta_n = 0$.

Case (i). Consider the general case of $\theta_i > 0, i = 1, \dots, n$. By Assumption 4.3, the test statistics T_1, \dots, T_n are positively regression dependent. For $j = 1, \dots, n - 1$, let E_{n-j} denote the event of making at least one Type III error when testing H_{j+1}, \dots, H_n using Procedure 4.2 at level α . By using induction, we prove the following two lemmas hold.

Lemma A.1. *Assume the conditions of Theorem 4.3. For $j = 1, \dots, n - 1$, the following inequality holds.*

$$Pr(E_{n-j} | T_1 > c_2, \dots, T_j > c_2) \leq \alpha. \quad (\text{A.8})$$

PROOF OF LEMMA A.1. We prove the result by using reverse induction. When $j = n - 1$, we have

$$\begin{aligned} & Pr(E_{n-j} | T_1 > c_2, \dots, T_j > c_2) \\ &= Pr(T_n < c_1 | T_1 > c_2, \dots, T_{n-1} > c_2) \\ &= \frac{Pr(T_n < c_1) Pr(T_1 > c_2, \dots, T_{n-1} > c_2 | T_n < c_1)}{Pr(T_1 > c_2, \dots, T_{n-1} > c_2)} \\ &\leq Pr(T_n < c_1) \leq \alpha. \end{aligned}$$

The inequality follows from Assumption 4.3.

Assume the inequality (A.8) holds for $j = m$. In the following, we prove that it also holds for $j = m - 1$. Note that

$$Pr(E_{n-m+1} | T_1 > c_2, \dots, T_{m-1} > c_2)$$

$$\begin{aligned}
&= Pr \left(\{T_m < c_1\} \cup \left(\{T_m > c_2\} \cap E_{n-m} \right) \mid T_1 > c_2, \dots, T_{m-1} > c_2 \right) \\
&= Pr (T_m < c_1 \mid T_1 > c_2, \dots, T_{m-1} > c_2) + Pr \left(\{T_m > c_2\} \cap E_{n-m} \mid T_1 > c_2, \dots, T_{m-1} > c_2 \right) \\
&= Pr (T_m < c_1 \mid T_1 > c_2, \dots, T_{m-1} > c_2) \\
&\quad + Pr (T_m > c_2 \mid T_1 > c_2, \dots, T_{m-1} > c_2) Pr (E_{n-m} \mid T_1 > c_2, \dots, T_m > c_2) \\
&\leq Pr (T_m < c_1 \mid T_1 > c_2, \dots, T_{m-1} > c_2) + \alpha Pr (T_m > c_2 \mid T_1 > c_2, \dots, T_{m-1} > c_2) \\
&\leq \alpha.
\end{aligned}$$

Therefore, the desired result follows. Here, the first inequality follows from the assumption of induction and the second follows from Lemma A.2 below. \blacksquare

Lemma A.2. *Assume the conditions of Theorem 4.3. For $j = 1, \dots, n - 1$, the following inequality holds:*

$$Pr (T_j < c_1 \mid T_1 > c_2, \dots, T_{j-1} > c_2) + \alpha Pr (T_j > c_2 \mid T_1 > c_2, \dots, T_{j-1} > c_2) \leq \alpha. \quad (\text{A.9})$$

Specifically, for $j = 1$, we have

$$Pr (T_1 < c_1) + \alpha Pr (T_1 > c_2) \leq \alpha.$$

PROOF OF LEMMA A.2. To prove the inequality (A.9), it is enough to show that

$$Pr (T_j < c_1 \mid T_1 > c_2, \dots, T_{j-1} > c_2) \leq \alpha Pr (T_j < c_2 \mid T_1 > c_2, \dots, T_{j-1} > c_2),$$

which is equivalent to

$$\begin{aligned}
&(1 - \alpha) Pr (T_j < c_2 \mid T_1 > c_2, \dots, T_{j-1} > c_2) \\
&\leq Pr (T_j < c_2 \mid T_1 > c_2, \dots, T_{j-1} > c_2) - Pr (T_j < c_1 \mid T_1 > c_2, \dots, T_{j-1} > c_2).
\end{aligned}$$

Note that

$$1 - \alpha = Pr_{\theta_j=0}(T_j < c_2) - Pr_{\theta_j=0}(T_j < c_1).$$

Thus, the above inequality is equivalent to

$$Pr_{\theta_j=0}(T_j < c_2) - Pr_{\theta_j=0}(T_j < c_1) \leq 1 - \frac{Pr(T_j < c_1 | T_1 > c_2, \dots, T_{j-1} > c_2)}{Pr(T_j < c_2 | T_1 > c_2, \dots, T_{j-1} > c_2)},$$

which in turn is implied by

$$1 - \frac{Pr_{\theta_j=0}(T_j < c_1)}{Pr_{\theta_j=0}(T_j < c_2)} \leq 1 - \frac{Pr(T_j < c_1 | T_1 > c_2, \dots, T_{j-1} > c_2)}{Pr(T_j < c_2 | T_1 > c_2, \dots, T_{j-1} > c_2)}. \quad (\text{A.10})$$

Note that by Assumption 4.2, we have

$$\frac{Pr(T_j < c_1)}{Pr(T_j < c_2)} \leq \frac{Pr_{\theta_j=0}(T_j < c_1)}{Pr_{\theta_j=0}(T_j < c_2)}.$$

Thus, to prove the inequality (A.10), we only need to show that

$$\frac{Pr(T_j < c_1 | T_1 > c_2, \dots, T_{j-1} > c_2)}{Pr(T_j < c_2 | T_1 > c_2, \dots, T_{j-1} > c_2)} \leq \frac{Pr(T_j < c_1)}{Pr(T_j < c_2)},$$

which is equivalent to

$$Pr(T_1 > c_2, \dots, T_{j-1} > c_2 | T_j < c_1) \leq Pr(T_1 > c_2, \dots, T_{j-1} > c_2 | T_j < c_2),$$

which follows from Assumption 4.3. Therefore, the desired result follows. ■

Based on Lemmas A.1 and A.2, we have

$$\begin{aligned} \text{mdFWER} &= Pr(T_1 < c_1) + \sum_{j=2}^n Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) \\ &= Pr(T_1 < c_1) + Pr(T_1 > c_2) \sum_{j=2}^n Pr(T_2 > c_2, \dots, T_{j-1} > c_2, T_j < c_1 | T_1 > c_2) \\ &= Pr(T_1 < c_1) + Pr(T_1 > c_2) Pr(E_{n-1} | T_1 > c_2) \\ &\leq Pr(T_1 < c_1) + \alpha Pr(T_1 > c_2) \\ &\leq \alpha. \end{aligned}$$

Therefore, the mdFWER is controlled at level α for Case (i). Here, the first inequality follows from Lemma A.1 and the second follows from Lemma A.2.

Case (ii). Consider the general case of $\theta_i > 0, i = 1, \dots, n - 1$ and $\theta_n = 0$. Under Assumption 4.3, $T_i, i = 1, \dots, n - 1$ are positively regression dependent and under Assumption 4.4, T_n is independent of T_i 's . Note that

$$\begin{aligned}
& \text{mdFWER} \\
&= \sum_{j=1}^{n-1} Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) \\
&\quad + Pr(T_1 > c_2, \dots, T_{n-1} > c_2, T_n < c_1) + Pr(T_1 > c_2, \dots, T_n > c_2) \\
&= \sum_{j=1}^{n-1} Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) + \alpha Pr(T_1 > c_2, \dots, T_{n-1} > c_2).
\end{aligned}$$

The second equality follows from Assumption 4.4.

For $m = 1, \dots, n - 1$, define

$$\Delta_m = \sum_{j=1}^m Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) + \alpha Pr(T_1 > c_2, \dots, T_m > c_2).$$

Thus, $\text{mdFWER} = \Delta_{n-1}$. By using induction, we prove below that $\Delta_m \leq \alpha$ for $m = 1, \dots, n - 1$.

For $m = 1$, by using Lemma A.2, we have

$$\Delta_1 = Pr(T_1 < c_1) + \alpha Pr(T_1 > c_2) \leq \alpha.$$

Assume $\Delta_m \leq \alpha$. In the following, we show $\Delta_{m+1} \leq \alpha$. Note that

$$\begin{aligned}
\Delta_{m+1} &= \sum_{j=1}^{m+1} Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) \\
&\quad + \alpha Pr(T_1 > c_2, \dots, T_m > c_2, T_{m+1} > c_2) \\
&= \sum_{j=1}^m Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) \\
&\quad + Pr(T_1 > c_2, \dots, T_m > c_2) [Pr(T_{m+1} < c_1 | T_1 > c_2, \dots, T_m > c_2) \\
&\quad\quad + \alpha Pr(T_{m+1} > c_2 | T_1 > c_2, \dots, T_m > c_2)] \\
&\leq \sum_{j=1}^m Pr(T_1 > c_2, \dots, T_{j-1} > c_2, T_j < c_1) + \alpha Pr(T_1 > c_2, \dots, T_m > c_2)
\end{aligned}$$

$$= \Delta_m \leq \alpha. \quad (\text{A.11})$$

The first inequality follows from Lemma A.2 and the second follows from the inductive hypothesis. Thus, $\Delta_m \leq \alpha$ for $m = 1, \dots, n - 1$. Therefore, $\text{mdFWER} = \Delta_{n-1} \leq \alpha$, the desired result.

Combining the arguments of Cases (i) and (ii), the proof of Theorem 4.3 is complete. ■

A.7 Proof of Proposition 4.1

From the proof of Theorem 4.1 and by Lemma 4.1, it is easy to see that we only need to prove the mdFWER control of Procedure 4.2 when H_1 is false and H_2 is true, i.e., $\theta_1 \neq 0$ and $\theta_2 = 0$.

Case I: $\theta_1 > 0$ and $\theta_2 = 0$. By Lemma 4.2, the mdFWER of Procedure 4.2 is controlled at level α if we have the following:

$$F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{(\theta_1,0)}(c_2, c_2) - F_{(\theta_1,0)}(c_2, c_1) \leq 0.$$

After rewriting $F_{(\theta_1,0)}(x, y)$ as $Pr(T_1 \leq x, T_2 \leq y)$ and then dividing through by $Pr(T_1 \leq c_2)$, we get,

$$Pr(T_2 \leq c_2 | T_1 \leq c_2) - Pr(T_2 \leq c_1 | T_1 \leq c_2) \leq 1 - \frac{Pr(T_1 \leq c_1)}{Pr(T_1 \leq c_2)}.$$

Dividing by $Pr(T_2 \leq c_2 | T_1 \leq c_2)$, we get,

$$1 - \frac{Pr(T_2 \leq c_1 | T_1 \leq c_2)}{Pr(T_2 \leq c_2 | T_1 \leq c_2)} \leq \frac{1}{Pr(T_2 \leq c_2 | T_1 \leq c_2)} \left(1 - \frac{Pr(T_1 \leq c_1)}{Pr(T_1 \leq c_2)} \right). \quad (\text{A.12})$$

For proving (A.12), it is enough to prove the following inequality, as $\frac{1}{Pr(T_2 \leq c_2 | T_1 \leq c_2)} \geq 1$.

$$1 - \frac{Pr(T_2 \leq c_1 | T_1 \leq c_2)}{Pr(T_2 \leq c_2 | T_1 \leq c_2)} \leq 1 - \frac{Pr(T_1 \leq c_1)}{Pr(T_1 \leq c_2)}. \quad (\text{A.13})$$

By Assumption 4.2 and (4.5), it follows that $\frac{F_0(c_2)}{F_0(c_1)} \leq \frac{F_{\theta_1}(c_2)}{F_{\theta_1}(c_1)}$, which is equivalent to, $1 - \frac{Pr(T_2 \leq c_1)}{Pr(T_2 \leq c_2)} \leq 1 - \frac{Pr(T_1 \leq c_1)}{Pr(T_1 \leq c_2)}$. Thus, for proving (A.12), it is enough to prove the following:

$$1 - \frac{Pr(T_2 \leq c_1 | T_1 \leq c_2)}{Pr(T_2 \leq c_2 | T_1 \leq c_2)} \leq 1 - \frac{Pr(T_2 \leq c_1)}{Pr(T_2 \leq c_2)}. \quad (\text{A.14})$$

But, (A.14) is equivalent to showing

$$Pr(T_1 \leq c_2 | T_2 \leq c_1) \geq Pr(T_1 \leq c_2 | T_2 \leq c_2),$$

which follows directly from Assumption 4.5.

Case II: $\theta_1 < 0$ and $\theta_2 = 0$. Similarly, by Lemma 4.2, the mdFWER of Procedure 4.2 is controlled at level α if we have the following:

$$1 + F_{\theta_1}(c_1) - F_{\theta_1}(c_2) + F_{(\theta_1,0)}(c_1, c_1) - F_{(\theta_1,0)}(c_1, c_2) \leq \alpha, \quad (\text{A.15})$$

which after some rearrangement and rewriting $1 - \alpha$ as $F_0(c_2) - F_0(c_1)$ gives,

$$(F_0(c_2) - F_{(\theta_1,0)}(c_1, c_2)) - (F_0(c_1) - F_{(\theta_1,0)}(c_1, c_1)) \leq (1 - F_{\theta_1}(c_1)) - (1 - F_{\theta_1}(c_2)). \quad (\text{A.16})$$

Thus, proving (A.15) is equivalent to proving that

$$Pr(T_1 \geq c_1, T_2 \leq c_2) - Pr(T_1 \geq c_1, T_2 \leq c_1) \leq Pr(T_1 \geq c_1) - Pr(T_1 \geq c_2).$$

Dividing through by $Pr(T_1 \geq c_1)$, we get

$$Pr(T_2 \geq c_1 | T_1 \geq c_1) - Pr(T_2 \geq c_2 | T_1 \geq c_1) \leq 1 - \frac{Pr(T_1 \geq c_2)}{Pr(T_1 \geq c_1)}. \quad (\text{A.17})$$

Thus, to prove (A.15), it is enough to prove the following,

$$1 - \frac{Pr(T_2 \geq c_2 | T_1 \geq c_1)}{Pr(T_2 \geq c_1 | T_1 \geq c_1)} \leq 1 - \frac{Pr(T_1 \geq c_2)}{Pr(T_1 \geq c_1)},$$

which is equivalent to proving,

$$\frac{Pr(T_2 \geq c_2 | T_1 \geq c_1)}{Pr(T_2 \geq c_1 | T_1 \geq c_1)} \geq \frac{Pr(T_1 \geq c_2)}{Pr(T_1 \geq c_1)}. \quad (\text{A.18})$$

By Assumption 4.2 and (4.6), it follows that for $\theta_1 < 0$, $\frac{Pr(T_1 \geq c_2)}{Pr(T_1 \geq c_1)} \leq \frac{Pr(T_2 \geq c_2)}{Pr(T_2 \geq c_1)}$. Thus, to prove (A.15), it is enough to prove the following,

$$\frac{Pr(T_2 \geq c_2 | T_1 \geq c_1)}{Pr(T_2 \geq c_1 | T_1 \geq c_1)} \geq \frac{Pr(T_2 \geq c_2)}{Pr(T_2 \geq c_1)}. \quad (\text{A.19})$$

But (A.19) is equivalent to showing

$$Pr(T_1 \geq c_1 | T_2 \geq c_2) \geq Pr(T_1 \geq c_1 | T_2 \geq c_1), \quad (\text{A.20})$$

which follows directly from Assumption 4.5. By combining the arguments of the above two cases, the desired result follows. \blacksquare

A.8 Proof of Proposition 4.2

By Corollary 4.1, without loss of generality, assume that $\theta_i > 0, i = 1, 2$ and $\theta_3 = 0$, that is, H_1 and H_2 are false and H_3 is true. Note that

mdFWER

$$= Pr(T_1 \leq c_1) + Pr(T_1 \geq c_2, T_2 \leq c_1) + Pr(T_1 \geq c_2, T_2 \geq c_2, T_3 \notin (c_1, c_2)). \quad (\text{A.21})$$

In the following, we prove that

$$\begin{aligned} & Pr(T_1 \geq c_2, T_2 \leq c_1) + Pr(T_1 \geq c_2, T_2 \geq c_2, T_3 \notin (c_1, c_2)) \\ & \leq Pr(T_1 \geq c_2, T_3 \notin (c_1, c_2)). \end{aligned} \quad (\text{A.22})$$

To prove (A.22), it is enough to show the following inequality:

$$Pr(T_2 \leq c_1 | T_1) + Pr(T_2 \geq c_2, T_3 \notin (c_1, c_2) | T_1) \leq Pr(T_3 \notin (c_1, c_2) | T_1). \quad (\text{A.23})$$

Note that

$$Pr(T_2 \geq c_2, T_3 \leq c_1 | T_1) = Pr(T_3 \leq c_1 | T_1) - Pr(T_2 < c_2, T_3 \leq c_1 | T_1) \quad (\text{A.24})$$

and

$$\begin{aligned} & Pr(T_2 \geq c_2, T_3 \geq c_2 | T_1) \\ &= 1 - Pr(T_2 < c_2 | T_1) - Pr(T_3 < c_2 | T_1) + Pr(T_2 < c_2, T_3 < c_2 | T_1). \end{aligned} \quad (\text{A.25})$$

In addition, we have

$$Pr(T_3 \notin (c_1, c_2) | T_1) = 1 + Pr(T_3 \leq c_1 | T_1) - Pr(T_3 < c_2 | T_1). \quad (\text{A.26})$$

Thus, in order to show (A.23), by combining (A.24)-(A.26), we only need to prove the following inequality:

$$\begin{aligned} & Pr(T_2 < c_2, T_3 < c_2 | T_1) - Pr(T_2 < c_2, T_3 \leq c_1 | T_1) \\ &\leq Pr(T_2 < c_2 | T_1) - Pr(T_2 \leq c_1 | T_1). \end{aligned} \quad (\text{A.27})$$

Note that (A.27) can be rewritten as

$$\begin{aligned} & Pr(T_2 < c_2, T_3 < c_2 | T_1) \left[1 - \frac{Pr(T_2 < c_2, T_3 \leq c_1 | T_1)}{Pr(T_2 < c_2, T_3 < c_2 | T_1)} \right] \\ &\leq Pr(T_2 < c_2 | T_1) \left[1 - \frac{Pr(T_2 \leq c_1 | T_1)}{Pr(T_2 < c_2 | T_1)} \right]. \end{aligned} \quad (\text{A.28})$$

Thus, to prove (A.27), it is enough to show

$$1 - \frac{Pr(T_2 < c_2, T_3 \leq c_1 | T_1)}{Pr(T_2 < c_2, T_3 < c_2 | T_1)} \leq 1 - \frac{Pr(T_2 \leq c_1 | T_1)}{Pr(T_2 < c_2 | T_1)}. \quad (\text{A.29})$$

That is,

$$\frac{Pr(T_2 \leq c_1 | T_1)}{Pr(T_2 < c_2 | T_1)} \leq \frac{Pr(T_2 < c_2, T_3 \leq c_1 | T_1)}{Pr(T_2 < c_2, T_3 < c_2 | T_1)}. \quad (\text{A.30})$$

By Assumption 4.6 (BMLR), we have

$$\frac{Pr(T_2 \leq x_2 | T_1)}{Pr(T_3 \leq x_2 | T_1)} \geq \frac{Pr(T_2 \leq x_1 | T_1)}{Pr(T_3 \leq x_1 | T_1)}. \quad (\text{A.31})$$

By (A.31), to prove (A.30), it is enough to show

$$\frac{Pr(T_3 \leq c_1 | T_1)}{Pr(T_3 < c_2 | T_1)} \leq \frac{Pr(T_2 < c_2, T_3 \leq c_1 | T_1)}{Pr(T_2 < c_2, T_3 < c_2 | T_1)}. \quad (\text{A.32})$$

That is,

$$Pr(T_2 < c_2 | T_3 < c_2, T_1) \leq Pr(T_2 < c_2 | T_3 < c_1, T_1). \quad (\text{A.33})$$

The inequality (A.33) holds under Assumption 4.5. Therefore, the inequality (A.22) holds.

Based on (A.21)-(A.22) and Proposition 1, we have

$$\text{mdFWER} = Pr(T_1 \leq c_1) + Pr(T_1 \geq c_2, T_3 \notin (c_1, c_2)) \leq \alpha.$$

Thus, the desired result follows. ■

A.9 Proof of Theorem 4.4

By Corollary 4.1, without loss of generality, assume that $\theta_i > 0, i = 1, \dots, n-1$ and $\theta_n = 0$, that is, $H_i, i = 1, \dots, n-1$ are false and H_n is true. Note that

$$\begin{aligned} & \text{mdFWER} \quad (\text{A.34}) \\ &= \sum_{j=1}^{n-1} Pr(T_1 \geq c_2, \dots, T_{j-1} \geq c_2, T_j \leq c_1) + Pr(T_1 \geq c_2, \dots, T_{n-1} \geq c_2, T_n \notin (c_1, c_2)). \end{aligned}$$

In the following, we prove that

$$\begin{aligned} & Pr(T_1 \geq c_2, \dots, T_{n-2} \geq c_2, T_{n-1} \leq c_1) + Pr(T_1 \geq c_2, \dots, T_{n-1} \geq c_2, T_n \notin (c_1, c_2)) \\ & \leq Pr(T_1 \geq c_2, \dots, T_{n-2} \geq c_2, T_n \notin (c_1, c_2)). \quad (\text{A.35}) \end{aligned}$$

To prove (A.35), it is enough to show the following inequality:

$$\begin{aligned} & Pr(T_{n-1} \leq c_1 | T_1, \dots, T_{n-2}) + Pr(T_{n-1} \geq c_2, T_n \notin (c_1, c_2) | T_1, \dots, T_{n-2}) \\ & \leq Pr(T_n \notin (c_1, c_2) | T_1, \dots, T_{n-2}). \end{aligned} \tag{A.36}$$

By using the same argument as in proving (A.23) in the case of three hypotheses, we can prove that the inequality (A.36) holds under Assumptions 4.5 and 4.7. Then, by combining (A.34) and (A.35), we have

$$\begin{aligned} & \text{mdFWER} \tag{A.37} \\ & \leq \sum_{j=1}^{n-2} Pr(T_1 \geq c_2, \dots, T_{j-1} \geq c_2, T_j \leq c_1) + Pr(T_1 \geq c_2, \dots, T_{n-2} \geq c_2, T_n \notin (c_1, c_2)). \end{aligned}$$

Note that the right-hand side of (A.37) is the mdFWER of Procedure 4.2 when testing H_1, \dots, H_{n-2}, H_n . By induction and Proposition 1, the mdFWER is bounded above by α , the desired result.

APPENDIX B

This appendix contains the proofs of the lemmas and theorems stated but not proved in Chapter 5.

B.1 Proof of Theorem 5.2

Proof As noted in the proof of Theorem 5.1, we have,

$$\text{FDR} = \sum_{i=1}^{m_0} E \left(\frac{V}{V+S} I \{U = u_i\} \right) \quad (\text{B.1})$$

For each u_i , let us define,

$$\beta_r = \left(\frac{m - u_i + 1 + r}{m - u_i + 1} \right) \cdot \frac{\alpha_{u_i}}{c_{u_i}}, \quad r = 0, \dots, u_i - 1.$$

Then for each $i = 1, \dots, m_0$ we have,

$$\begin{aligned} & E \left(\frac{V}{V+S} I \{U = u_i\} \right) \\ & \leq E \left(\frac{m - u_i + 1}{m - u_i + 1 + R_{u_i}} I (U = u_i) \right) \end{aligned} \quad (\text{B.2})$$

$$\leq E \left(\frac{m - u_i + 1}{m - u_i + 1 + R_{u_i}} I (P_{u_i} \leq \gamma_{u_i}) \right) \quad (\text{B.3})$$

$$= \sum_{r=0}^{u_i-1} E \left(\frac{m - u_i + 1}{m - u_i + 1 + r} I (P_{u_i} \leq \beta_r) \middle| R_{u_i} = r \right) Pr (R_{u_i} = r) \quad (\text{B.4})$$

The inequality in (B.2) follows due to (5.5). The inequality in (B.3) follows as the event $\{U = u_i\}$ implies the event $\{P_{u_i} \leq \gamma_{u_i}\}$. (B.4) follows as R_{u_i} can take values between 0 and $u_i - 1$. Then (B.4) becomes,

$$\begin{aligned} & = \sum_{r=0}^{u_i-1} \frac{m - u_i + 1}{m - u_i + 1 + r} Pr (R_{u_i} = r, P_{u_i} \leq \beta_r) \\ & = \sum_{r=0}^{u_i-1} \frac{m - u_i + 1}{m - u_i + 1 + r} [Pr (R_{u_i} \geq r, P_{u_i} \leq \beta_r) - Pr (R_{u_i} \geq r + 1, P_{u_i} \leq \beta_r)] \end{aligned}$$

$$\begin{aligned}
&\leq Pr(P_{u_i} \leq \beta_0) + \sum_{r=1}^{u_i-1} \frac{m - u_i + 1}{m - u_i + 1 + r} Pr(R_{u_i} \geq r, \beta_{r-1} \leq P_{u_i} \leq \beta_r) \\
&\leq Pr(P_{u_i} \leq \beta_0) + \sum_{r=1}^{u_i-1} \frac{m - u_i + 1}{m - u_i + 1 + r} Pr(\beta_{r-1} \leq P_{u_i} \leq \beta_r) \\
&= Pr(P_{u_i} \leq \beta_0) + \sum_{r=1}^{u_i-1} \frac{m - u_i + 1}{m - u_i + 1 + r} Pr(P_{u_i} \leq \beta_r) \\
&\quad - \sum_{r=0}^{u_i-2} \frac{m - u_i + 1}{m - u_i + 1 + (r + 1)} Pr(P_{u_i} \leq \beta_r) \\
&= \sum_{r=0}^{u_i-2} \frac{m - u_i + 1}{(m - u_i + 1 + r)(m - u_i + 1 + (r + 1))} Pr(P_{u_i} \leq \beta_r) \\
&\quad + \frac{m - u_i + 1}{m} Pr(P_{u_i} \leq \beta_{u_i-1}) \\
&\leq \alpha_{u_i}
\end{aligned}$$

Thus, finally we have,

$$FDR = \sum_{i=1}^{m_0} E \left(\frac{V}{V+S} I\{U = u_i\} \right) \leq \sum_{i=1}^{m_0} \alpha_{u_i} \leq \sum_{i=1}^{m_0} \alpha_{u_i}^* \leq \alpha \sum_{i=1}^m \omega_i = \alpha.$$

This shows strong control of FDR under general dependence of the p -values. ■

B.2 Proof of (5.18)

To prove (5.18), we separately consider the cases when $j \in I_0$ and when $j \in I_1$.

Case I: $j \in I_0$: In this case, $S_i = S_{i-1}$ and $V_i = V_{i-1} + 1$ when $J_i = j$. Also, $V_{i-1} + S_{i-1} = i - 1$. When $i < m$, the left hand side of 5.18, after some algebra, becomes,

$$\begin{aligned}
&E \left(\left(\frac{S_{i-1}}{i(i-1)} + \frac{(m-j)\alpha S_{i-1}}{(m-1)(i-1)} \frac{2i-m-1}{i(m-i+1)} \right) I\{J_i = j\} \right) \\
&= E \left(\frac{S_{i-1}}{(i-1)(m-i+1)} \left(\frac{(m-i)(m-i+1) + \alpha(m-j)(2i-m-1)}{i(m-i)} \right) W_i(j-1) \right) \\
&\quad \cdot Pr(P_j \leq \alpha_j)
\end{aligned}$$

$$\begin{aligned}
&\leq E \left(\frac{S_{i-1}}{(i-1)(m-i+1)} \left(\frac{m - (\alpha(m-i) + (1-\alpha)(i-1))}{i} \right) W_i(j-1) \right) Pr(P_j \leq \alpha_j) \\
&\leq E \left(\frac{\alpha}{m-i+1} \frac{S_{i-1}}{i-1} W_i(j-1) \right) \tag{B.5}
\end{aligned}$$

The first equality follows from (5.16) and the fact that both S_{i-1} and $W_i(j-1)$ are only dependent on the first $j-1$ hypotheses and hence independent of P_j . The first inequality follows by noting that $m-j \leq m-i$ and some algebra. The second inequality follows as the p -values are stochastically larger than $U(0,1)$ distribution.

Case II: $j \in I_1$: In this case, $S_i = S_{i-1} + 1$ and $V_i = V_{i-1}$ when $J_i = j$. Also, $V_{i-1} + S_{i-1} = i-1$. Then the left hand side of 5.18, after some algebra, becomes,

$$\begin{aligned}
&E \left(\left(\frac{-V_{i-1}}{i(i-1)} + \frac{(m-j)(2i-m-1)\alpha}{i(m-i)(m-i+1)} \frac{S_{i-1}}{i-1} + \frac{(m-j)\alpha}{i(m-i)} \right) I\{J_i = j\} \right) \\
&= E \left(\left(\frac{S_{i-1}((m-i)(m-i+1) + \alpha(m-j)(2i-m-1))}{(m-i+1)(i-1)i(m-i)} + \frac{(m-j)\alpha - (m-i)}{i(m-i)} \right) W_i(j-1) \right) \\
&\quad \cdot Pr(P_j \leq \alpha_j) \\
&\leq E \left(\frac{S_{i-1}}{(i-1)(m-i+1)} \left(\frac{m - (\alpha(m-i) + (1-\alpha)(i-1))}{i} \right) W_i(j-1) \right) Pr(P_j \leq \alpha_j) \\
&\leq E \left(\frac{\alpha}{m-i+1} \frac{S_{i-1}}{(i-1)} W_i(j-1) \right)
\end{aligned}$$

The first equality follows from (5.16) and by the fact that both S_{i-1} and $W_i(j-1)$ are only dependent on the first $j-1$ hypotheses and hence independent of P_j . The first inequality follows as $m-j \leq m-i$. The second inequality follows as the p -values are stochastically larger than $U(0,1)$ distribution. ■

B.3 Proof of (5.19)

We start from the left hand side of (5.19)

$$E(f_{i-1}(J_{i-1})I\{J_{i-1} < m+1\} - f_{i-1}(J_i)I\{J_i < m+1\})$$

$$\begin{aligned}
&= E \left(\sum_{j=i-1}^m f_{i-1}(j) I\{J_{i-1} = j\} - \sum_{j=i}^m f_{i-1}(j) I\{J_i = j\} \right) \\
&= E \left(\sum_{j=i}^m f_{i-1}(j) (I\{J_{i-1} = j\} - I\{J_i = j\}) + f_{i-1}(i-1) I\{J_{i-1} = i-1\} \right) \\
&= E \left(\sum_{j=i}^m f_{i-1}(j) (W_i(j) - W_i(j-1)) + f_{i-1}(i-1) W_i(i-1) \right) \\
&= E \left(\sum_{j=i-1}^m f_{i-1}(j) W_i(j) - \sum_{j=i}^m f_{i-1}(j) W_i(j-1) \right) \\
&= E \left(\sum_{j=i}^m (f_{i-1}(j-1) - f_{i-1}(j)) W_i(j-1) \right) \\
&= E \left(\sum_{j=i}^m \frac{\alpha}{m-i+1} \frac{S_{i-1}}{i-1} W_i(j-1) \right)
\end{aligned}$$

The third equality follows from (5.17). The last equality follows from definition of $f_{i-1}(j)$. ■

BIBLIOGRAPHY

- Y. Benjamini. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B*, 72:405–416, 2010.
- Y. Benjamini and R. Heller. False discovery rates for spatial signals. *Journal of the American Statistical Association*, 102:1272–1281, 2007.
- Y. Benjamini and R. Heller. Screening for partial conjunction hypotheses. *Biometrics*, 64:1215–1222, 2008.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57:289–300, 1995.
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29:1165–1188, 2001.
- Y. Benjamini and D. Yekutieli. False discovery rateadjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100:71–81, 2005.
- F. Bretz, W. Maurer, W. Brannath, and M. Posch. A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine*, 28: 586–604, 2009.
- P. Ciarmela, M. S. Islam, F. M. Reis, P. C. Gray, E. Bloise, F. Petraglia, W. Vale, and M. Castellucci. Growth factors and myometrium: biological effects in uterine fibroid and possible clinical implications. *Human Reproduction Update*, 17: 772–790, 2011.
- B. J. Davis, J. I. Risinger, G. V. R. Chandramouli, P. R. Bushel, D. D. Baird, and S. D. Peddada. Gene expression in uterine leiomyoma from tumors likely to be growing (from black women over 35) and tumors likely to be non-growing (from white women over 35). *PLOS ONE*, 8:e63909, 2013.
- L. K. Diaz, M. Cristofanilli, X. Zhou, K. L. Welch, T. L. Smith, Y. Yang, N. Sneige, A. A. Sahin, and M. Z. Gilcrease. $\beta 4$ integrin subunit gene expression correlates with tumor size and nuclear grade in early breast cancer. *Modern Pathology*, 18:1165–1175, 2005.
- A. Dmitrienko, W. W. Offen, and P. H. Westfall. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in Medicine*, 22:2387–2400, 2003.
- A. Dmitrienko, A. C. Tamhane, and F. Bretz, editors. *Multiple testing problems in pharmaceutical statistics*. CRC Press, Boca Raton, FL, 2009.

- A. Dmitrienko, R. D'Agostino, and M. Huque. Key multiplicity issues in clinical drug development. *Statistics in Medicine*, 32:1079–1111, 2013.
- S. Dudoit, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18:71–103, 2003.
- C. W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50:1096–1121, 1955.
- C. W. Dunnett and A. C. Tamhane. Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine*, 10:939–947, 1991.
- C. W. Dunnett and A. C. Tamhane. A step-up multiple test procedure. *Journal of the American Statistical Association*, 87:162–170, 1992.
- B. Efron and R. Tibshirani. Empirical Bayes methods and false discovery rates for microarrays. *Genetics Epidemiology*, 23:70–86, 2002.
- B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96:1151–1160, 2001.
- D. Epting, B. Wendik, K. Bennewitz, C. T. Dietz, W. Driever, and J. Kroll. The Rac1 regulator ELMO1 controls vascular morphogenesis in zebrafish. *Circulation research*, 107:45–55, 2010.
- A. Farcomeni and L. Finos. FDR control with pseudo-gatekeeping based on a possibly data driven order of the hypotheses. *Biometrics*, 69:606–613, 2013.
- H. Finner. Two-sided tests and one-sided confidence bounds. *Annals of Statistics*, 22:1502–1506, 1994.
- H. Finner. Stepwise multiple test procedures and control of directional errors. *Annals of Statistics*, 27:274–289, 1999.
- L. Finos and A. Farcomeni. k-FWER control without multiplicity correction, with application to detection of genetic determinants of multiple sclerosis in Italian twins. *Biometrics*, 67:174–181, 2011.
- M. Gieseg, M. Man, N. Gorski, S. Madore, E. Kaldjian, and W. Leopold. The influence of tumor size and environment on gene expression in commonly used human tumor lines. *BMC Cancer*, 4:35, 2004.
- M. G. G'Sell, S. Wager, A. Chouldechova, and R. Tibshirani. Sequential selection procedures and false discovery rate control. *arXiv:1309.5352v2*, 2014.
- W. Guo and J. Romano. On stepwise control of directional errors under independence and some dependence. *Journal of Statistical Planning and Inference*, 2015.

- W. Guo, S. K. Sarkar, and S. D. Peddada. Controlling false discoveries in multidimensional directional decisions, with applications to gene expression data on ordered categories. *Biometrics*, 66:485–492, 2010.
- R. Heller, E. Manduchi, G. Grant, and W. Ewens. A flexible two-stage procedure for identifying gene sets that are differentially expressed. *Bioinformatics*, 25:929–942, 2009.
- Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75:800–802, 1988.
- M. J. Hoenerhoff, A. R. Pandiri, S. A. Lahousse, H. H. Hong, T. V. Ton, T. Masinde, S. S. Auerbach, K. Gerrish, P. R. Bushel, K. R. Shockley, S. D. Peddada, and R. C. Sills. Global gene profiling of spontaneous hepatocellular carcinoma in B6C3F1 mice: similarities in the molecular landscape with human liver cancer. *Toxicologic Pathology*, 39:678–699, 2011.
- M. J. Hoenerhoff, A. R. Pandiri, S. A. Snyder, H. H. Hong, T. V. Ton, S. D. Peddada, K. Shockley, K. Witt, P. Chan, C. Rider, L. Kooistra, A. Nyska, and R. C. Sills. Hepatocellular carcinomas in B6C3F1 mice treated with ginkgo biloba extract for two years differ from spontaneous liver tumors in cancer gene mutations and genomic pathways. *Toxicologic Pathology*, 41:826–841, 2012.
- S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.
- G. Hommel and S. Kropf. Testing for differentiation in gene expression using a data-driven order or weights for hypotheses. *Biometrical Journal*, 47:554–562, 2005.
- K. D. Houston, J. A. Copland, R. R. Broaddus, M. M. Gottardis, S. M. Fischer, and C. L. Walker. Inhibition of proliferation and estrogen receptor signaling by peroxisome proliferator-activated receptor γ ligands in uterine leiomyoma. *Cancer Research*, 63:1221–1227, 2003.
- J. Hu, J. W. Locasale, J. H. Bielas, J. O’Sullivan, K. Sheahan, L. C. Cantley, M. G. VanDerHeiden, and D. Vitkup. Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nature Biotechnology*, 31:522–529, 2013.
- H. N. Jabbour, K. J. Sales, O. P. Smith, S. Battersby, and S. C. Boddy. Prostaglandin receptors are mediators of vascular function in endometrial pathologies. *Molecular and Cellular Endocrinology*, 252:191–200, 2006.
- S. Kropf and J. Läuter. Multiple tests for different sets of variables using a data-driven ordering of hypotheses, with an application to gene expression data. *Biometrical Journal*, 44:789–800, 2002.

- S. Kropf, J. Läuter, M. Eszlinger, K. Krohn, and R. Paschkeb. Nonparametric multiple test procedures with data-driven order of hypotheses and with weighted hypotheses. *Journal of Statistical Planning and Inference*, 125:31–47, 2004.
- E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. *Annals of Statistics*, 33:1138–1154, 2005.
- E. Lejmi, L. Leconte, S. Pdron-Mazoyer, S. Ropert, W. Raoul, S. Lavalette, I. Bouras, J. G. Feron, M. Maitre-Boube, F. Assayag, C. Feumi, M. Alemany, T. X. Jie, T. Merkulova, M. Poupon, M. M. Ruchoux, G. Tobelem, F. Sennlaub, and J. Plout. Netrin-4 inhibits angiogenesis via binding to neogenin and recruitment of Unc5B. *Proceedings of the National Academy of Sciences*, 105:12491–12496, 2008.
- W. Liu. Control of directional errors with step-up multiple tests. *Statistics & Probability Letters*, 31:239–242, 1997.
- G. Lynch, W. Guo, and S. K. Sarkar. The control of the false discovery rate in fixed sequence multiple testing. 2014. Under revision.
- W. Maurer, L. Hothorn, and W. Lehmacher. *Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses*, volume 6, pages 3–18. Fischer-Verlag, Stuttgart, Germany, 1995.
- N. Meinshausen. Hierarchical testing of variable importance. *Biometrika*, 95:265–278, 2008.
- A. J. Minn, G. P. Gupta, D. Padua, P. Bos, D. X. Nguyen, D. Nuyten, B. Kreike, Y. Zhang, Y. Wang, H. Ishwaran, J. A. Foekens, M. VanDeVijver, and J. Massagué. Lung metastasis genes couple breast tumor size and metastatic spread. *Proceedings of the National Academy of Sciences*, 104:6740–6745, 2007.
- D. H. Nam, S. Ramachandran, D. K. Song, K. Y. Kwon, D. S. Jeon, S. J. Shin, S. H. Kwon, S. D. Cha, I. Bae, and C. C. Cho. Growth inhibition and apoptosis induced in human leiomyoma cells by treatment with the PPAR γ ligand ciglitizone. *Molecular Human Reproduction*, 13:829–836, 2007.
- A. R. Pandiri, R. C. Sills, M. J. Hoenerhoff, S. D. Peddada, T. V. Ton, H. H. Hong, G. P. Flake, D. E. Malarkey, I. P. Pogribny, N. J. Walker, and M. D. Boudreau. Aloe vera non-decolorized whole leaf extract-induced large intestinal tumors in F344 rats share similar molecular pathways with human sporadic colorectal tumors. *Toxicologic Pathology*, 39:1065–1074, 2011.
- A. R. Pandiri, R. C. Sills, V. Ziglioli, T. V. Ton, H. H. Hong, S. A. Lahousse, K. E. Gerrish, S. S. Auerbach, K. R. Shockley, P. R. Bushel, S. D. Peddada, and M. J. Hoenerhoff. Differential transcriptomic analysis of spontaneous lung tumors 1 in B6C3F1 mice: 2 comparison to human non-small cell lung cancer. *Toxicologic Pathology*, 40:1141–1159, 2012.

- S. D. Peddada, E. K. Lobenhofer, L. Li, C. A. Afshari, C. R. Weinberg, and D. M. Umbach. Gene selection and clustering for time-course and doseresponse microarray experiments using order-restricted inference. *Bioinformatics*, 19: 834–841, 2003.
- S. D. Peddada, S. Harris, J. Zajd, and E. Harvey. Oriogen: order restricted inference for ordered gene expression data. *Bioinformatics*, 21:3933–3934, 2005.
- S. D. Peddada, S. K. Laughlin, K. Miner, J-P. Guyon, K. Haneke, H. L. Vahdat, R. C. Semelka, A. Kowalik, D. Armao, B. Davis, and D. D. Baird. Growth of uterine leiomyomata among premenopausal black and white women. *Proceedings of the National Academy of Sciences*, 105:19887–19892, 2008.
- Z. Qiu, W. Guo, and G. Lynch. On generalized fixed sequence procedures for controlling the FWER. 2014. Under revision.
- M. L. H. Riis, X. Zhao, F. Kaveh, H. S. Vollan, A. Nesbakken, H. K. Solvang, T. Lüders, I. R. K. Bukholm, and V. N. Kristensen. Gene expression profile analysis of T1 and T2 breast cancer reveals different activation pathways. *ISRN Oncology*, 2013:12, 2013.
- S. K. Sarkar. Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Annals of Statistics*, 26:494–504, 1998.
- S. K. Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *Annals of Statistics*, 30:239–257, 2002.
- S. K. Sarkar. On methods controlling the false discovery rate. *Sankhya*, 70A:135–168, 2008.
- S. K. Sarkar and C-K. Chang. The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, 92:1601–1608, 1997.
- S. K. Sarkar and W. Guo. Procedures controlling the k -FDR using bivariate distributions of the null p -values. *Statistica Sinica*, 20:1227–1238, 2010.
- S. K. Sarkar, P. K. Sen, and H. Finner. *On two results in multiple testing*, volume 47, pages 89–99. Institute of Mathematical Statistics, 2004.
- J. P. Shaffer. Control of directional errors with stagewise multiple test procedures. *Annals of Statistics*, 8:1342–1347, 1980.
- J. P. Shaffer. Multiplicity, directional (Type III) errors, and the null hypothesis. *Psychological Methods*, 7:356–369, 2002.
- J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B*, 64:479–498, 2002.

- J. D. Storey. The positive false discovery rate: a Bayesian interpretation and the q-value. *Annals of Statistics*, 31:2013–2035, 2003.
- W. Sun and Z. Wei. Multiple testing for pattern identification, with applications to microarray time-course experiments. *Journal of the American Statistical Association*, 106:73–88, 2011.
- V. S. Talaulikar and I. Manyonda. Progesterone and progesterone receptor modulators in the management of symptomatic uterine fibroids. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 165:135–140, 2012.
- P. H. Westfall and A. Krishen. Optimally weighted, fixed sequence and gate-keeper multiple testing procedures. *Journal of Statistical Planning and Inference*, 99: 25–41, 2001.
- P. H. Westfall, S. Kropf, and L. Finos. Weighted FWE-controlling methods in high-dimensional situations. *Recent Developments in Multiple Comparison Procedures*, 47:143–154, 2004.
- P. H. Westfall, F. Bretz, and R. D. Tobias. Directional error rates of closed testing procedures. *Statistics in Biopharmaceutical Research*, 5:345–355, 2013.
- B. L. Wiens. A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics*, 2:211–215, 2003.
- B. L. Wiens and A. Dmitrienko. The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics*, 15:929–942, 2005.
- B. L. Wiens and A. Dmitrienko. On selecting a multiple comparison procedure for analysis of a clinical trial: fallback, fixed sequence, and related procedures. *Statistics in Biopharmaceutical Research*, 2:22–32, 2010.
- M. Wolanska, E. Bańkowska-Guszczyn, and S. Jaworski. Fibroblast growth factor gene expression in uterine leiomyomas. *Ginekologia Polska*, 79:555–559, 2008.
- D. Yekutieli. Hierarchical false discovery rate-control methodology. *Journal of the American Statistical Association*, 103:309–316, 2008.