12-31-2018

FWER controlling procedures in simultaneous and selective inference

Li Yu
New Jersey Institute of Technology

Recommended Citation

Yu, Li, "FWER controlling procedures in simultaneous and selective inference" (2018). *Dissertations*. 1650.
https://digitalcommons.njit.edu/dissertations/1650

ABSTRACT

FWER CONTROLLING PROCEDURES IN SIMULTANEOUS AND SELECTIVE INFERENCE

by
Li Yu

With increasing complexity of research objectives in clinical trials, a variety of relatively complex and less intuitive multiple testing procedures (MTPs) have been developed and applied in clinical data analysis. In order to make testing strategies more explicit and intuitive to communicate with non-statisticians, several flexible and powerful graphical approaches have recently been introduced in the literature for developing and visualizing newer MTPs. Nevertheless, some theoretical as well as methodological issues still remain to be fully addressed. This dissertation addresses several important issues arising in graphical approaches and related selective inference problems. It consists of three parts.

In the first part of this dissertation, a generalized graphical approach is introduced, which allows one to reject more than one hypothesis at each step. This overcomes a main drawback of existing graphical approaches in which only one rejection is allowed at each step. Through some clinical examples, the proposed approach is illustrated to be more flexible and computationally efficient than existing graphical approaches. Theoretically, it is shown that the generalized graphical approach strongly controls the FWER under arbitrary dependence. To show the FWER control of the proposed method, as a by-product, a new concept of a multivariate critical value function is introduced and based on this function, the sequential rejection principle (Goeman and Solari, 2010) is generalized from the case of univariate critical value function to that of multivariate.

In the second part of this dissertation, a new graphical approach for general logically related multiple hypotheses testing is developed. By re-assigning critical

values between testable and non-testable hypotheses, all local critical values can be made fully used. Theoretically, it is shown that the proposed graphical approach strongly controls the FWER at level α under arbitrary dependence, by employing the generalized sequential rejection principle developed in the first part of this dissertation. Through some clinical examples, it demonstrates that the proposed graphical approach is more flexible and computationally efficient than entangled graphical approach for testing general logically related hypotheses (Maurer and Bretz, 2013).

In the third part of this dissertation, several powerful MTPs based on the very recently introduced ideas and methods of selective inference are proposed, which can be applied in large scale data analysis, such as microarray study, genomewise association study (GWAS), etc. By further developing the idea of independent filtering (Bourgon et al., 2010; Dai et al., 2012; Du and Zhang, 2014; Ignatiadis et al., 2016), where hypotheses are splitted into two blocks by selection process, three two-stage MTPs, adaptive two-stage Bonferroni procedure, selective parallel gatekeeping procedure and data-driven weighted selective procedure, are proposed. The proposed MTPs can not only exploit information of selected hypotheses more explicitly by estimating the true null proportion, but also deal with non-selected hypotheses. In order to exploit information of each null hypothesis more explicitly, the proposed procedures are further generalized from two blocks to multiple blocks. Theoretically, it is shown that the proposed MTPs strongly control the FWER at level α. Under independence, the proposed procedures are evaluated through extensive simulation studies.

FWER CONTROLLING PROCEDURES IN SIMULTANEOUS AND SELECTIVE INFERENCE

by
Li Yu

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

December 2018

APPROVAL PAGE

FWER CONTROLLING PROCEDURES IN SIMULTANEOUS AND SELECTIVE INFERENCE

Li Yu

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

---

Dr. Ji Meng Loh, Committee Member Date
Associate 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. Antai Wang, Committee Member Date
Associate Professor, New Jersey Institute of Technology, Department of Mathematical Sciences

---

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

BIOGRAPHICAL SKETCH

Author: Li Yu

Degree: Doctor of Philosophy

Date: December 2018

Undergraduate and Graduate Education:

- Doctor of Philosophy in Applied Statistics,
  New Jersey Institute of Technology, Newark, NJ, 2018
- Master of Science in General Biology
  Rutgers University, Newark, NJ, 2013
- Bachelors of Science in Biotechnology
  Zhe Jiang Forestry University, China, 2011

Major: Mathematical Sciences

Presentations and Publications:

Yu L. (2017). A Generalized Graphical Approach for Sequentially Rejective Multiple Testing Procedures, Joint Statistical Meeting, Baltimore, MD.

Yu L., Guo W. and Qiu Z. (2018). A New Graphical Approach with Generalized Sequential Rejection Principle to Control the Familywise Error Rate, (In Preparation)

Yu L. and Guo W. (2018). A New Graphical Approach for Hierachically Restricted Multiple Hypotheses Testing with the Control of Familywise Error Rate, (In Preparation)

Qiu Z., Yu L. and Guo W. (2018). A Family-based Graphical Approach for Testing Hierarchically Ordered Families of Hypotheses. (Submitted)

I dedicate my dissertation work to my parents without whom I would not be where I am now.

ACKNOWLEDGMENT

There are many people I would like to express my sincere gratitude for their contribution during my time at NJIT. Without them, this thesis would not have been possible.

First, my special thanks to my thesis advisor Dr. Wenge Guo. It has been an amazing experience working with Wenge in the past four years. I thank Wenge for his immense patience and unconditional support. Wenge has been an inspiration through his enthusiasm and encouragement to pursue exciting ideas, and his practicality and guidance. One important idea I learned from him is that always having a big picture in mind, which is helpful not only in my academic research, but also in my personal life, even in my future career. I'm so grateful having such a tremendous mentor during my PhD study.

Besides my advisor, I would like to thank the rest of my dissertation committee members: Dr. Antai Wang, Dr. Ji Meng Loh, Dr. Sundar Subramanian and Dr. Zhi Wei, for their serving on my committee. Thanks for their insightful comments and suggestions.

I would also like to thank my fellow colleagues and friends for helps you provided for me, for making my experience at NJIT exciting and fun. Thank you Dr. Xieyang Jia, Dr. Pejman Sanaei, Dr. Luan Gan, Mrs. Yan Zhang and Mr. Zhongcheng Lin.

Last but not the least, I would like to thank my family for their love and encouragement. To my parents, thanks for supporting me in all my pursuits. I am so grateful for the way you raised me up, without your efforts, I can not have the opportunity to pursue my graduate study at USA. To my mother-in-law, thanks for taking care of Hannah and for all the sacrifices that you have made on my behalf. To my beloved husband Mr. Shu Jia, thanks for your endless support. To my beloved daughter Hannah Jia, thanks for your warm smile. Your smile is a sunshine that cheers me up everyday.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

1.1 Introduction

In current clinical trials practice, the study objectives become increasingly complex, such as different doses to compare, multiple endpoints to investigate, several treatment groups to explore, etc. Instead of using single hypothesis testing to explore one research question, multiple hypotheses testing are used to investigate multiple study objectives simultaneously. However, multiplicity issue arises in the sense that type I error rate is inflated with the increasing number of tested hypotheses. Challenges in the filed of multiple hypotheses testing are that how to address the multiplicity issue appropriately and how to control overall error rates. Various multiple testing procedures (MTPs) have been developed to overcome the issue with the control of proper error rates. Among the proposed MTPs in the past decades, Bonferroni-based sequentially rejective MTPs are well developed to explicitly address the problem of multiple testing. But it turns out that the decision tables of these MTPs may be long and abstract, which make them difficult to present to non-statisticians clearly and intuitively. Therefore, in the first two parts of this dissertation, we aim to develop new MTPs which are more flexible and efficient to present to the clinical team and can also take complex logical structures among hypotheses into account. In the last part of this dissertation, we focus on developing MTPs in large scale multiple hypotheses testing. The arising problem of large scale multiple hypotheses testing, such as microarray study, genomewise association study (GWAS), is that the number of testing hypotheses is large. We aim to develop powerful MTPs by using the very recently developed ideas and methods of selective inference to deal with the problem of large scale multiple testing, which can greatly reduce the number of tested hypotheses through a selection process.

1.1.1 Basic Concepts in Multiple Hypotheses Testing

Consider a problem of testing m hypotheses, $H_1, \ldots, H_m$, simultaneously. We first introduce some important concepts related to multiple testing and MTPs.

Error Rates and Power The overall error measurments in multiple testing are not unqiue. It is important to choose a suitable error rate before developing any MTP. In the following, we first describe two commonly used error measurements in multiple testing: familywise error rate (FWER) and false discovery rate (FDR).

- *Familywise error rate* (FWER) is defined as the probability of making at least one false rejection, which is given by:

$$\text{FWER} = \Pr\left(\text{at least one false rejection}\right).$$

- *False discovery rate* (FDR) is defined as the expected proportion of false rejections among all rejected hypotheses, which is given by:

$$\text{FDR} = \text{E}\left(\frac{\text{Number of false rejections}}{\text{Total number of rejections}}\right).$$

When dealing with small scale multiple testing problems in clinical trials, FWER is commonly used to measure type I error rate. There are two types of FWER control. One is *weak* control, which controls the FWER only when all null hypotheses are true, the other is *strong* control, which controls the FWER under any combination of true and false null hypotheses. In this dissertation, we only consider strong control of FWER for our proposed MTPs, due to unknown numbers of true and false null hypotheses. Moreover, strong control of FWER is mandatory by regulators in all confirmatory clinical trials (CPMP, 2002).

When dealing with large scale multiple testing problems, such as in genome wide association study (GWAS), controlling the FWER is often too conservative to detect any false null hypothesis, thus the FDR is commonly used to measure the error rate. Besides the FWER and FDR, several other error measurements are also introduced in the literature, such as comparisonwise error rate (CWER), perfamily error rate (PFER), generalized familywise error rate (k-FWER), etc.

In addition to the type I error rate control, we also need to evaluate the power performance of a MTP. Various definitions of overall power are available in multiple testing. We describe two commonly used power definitions in the literature as follows.

- The *minimal power* is the probability of rejecting at least one false null hypothesis,
$$\text{minimal power} = \Pr\left(\text{reject at least one false null}\right).$$
- The *average power* is the expected proportion of rejected false null hypotheses among all false null hypotheses,
$$\text{average power} = \mathrm{E}\left(\frac{\text{Number of rejected false nulls}}{\text{Total number of false nulls}}\right).$$

1.1.2 Assumptions on p-values

Under true null hypotheses, the distributions of marginal p-values are assumed to be bounded above by $U(0,1)$, i.e.,

$$Pr\{P_i \leq p\} \leq p, \text{ for any } p \in (0,1) \text{ and } i \in I_0,$$

where I_0 is the indices of true null hypotheses.

In multiple testing, several commonly used joint dependence structures of the p-values are considered, such as independence, positive regression dependence on subset (PRDS) and arbitarily dependent. In this dissertation, we assume the p-values are arbitrary dependent, that is, we do not make any assumptation on the joint dependence of the p-values.

1.1.3 Closure Principle

Closure principle is a fundamental principle in multiple testing which is widely used for developing FWER controlling procedures (Marcus, Peritz and Gabriel, 1976). The closure principle states that an individual hypothesis can be rejected in the context of multiple testing if and only if all intersection hypotheses including this hypothesis are rejected by valid local tests in the context of single hypothesis testing. Any MTPs derived by the closure principle are termed as closed test procedures. In order to make a decision on a single hypothesis in multiple testing, by using a closed test procedure, we often need to perform a large number of local tests, which results in a high demand of computation.

1.1.4 Sequential Rejection Principle

Goeman and Solari (2010) proposed a sequential rejection principle of the FWER control, which is used for constructing FWER controlling procedures. It can be considered as equally fundamental as the closure principle. This principle emphasizes on the sequential aspect of testing in the sense that rejections are made sequentially, and the rejection of hypotheses at the current step is based on the rejections made in previous steps.

In order to present this principle formally, we first introduce some notations. A sequentially rejective procedure is defined on a random and measureable univariate successor function $\mathcal{N}$ with a variable $\mathcal{R}$, where $\mathcal{N}(\mathcal{R})$ is a collection of rejections that can be made in the next step after rejecting a collection of hypotheses $\mathcal{R}$ in the previous steps. Relying on the successor function $\mathcal{N}$, a sequentially rejective procedure is defined with iteratively rejecting a collection of null hypotheses at each step. If at the end of step i, we have the collection of rejections $\mathcal{R}_i$, then at the end of step $i+1$, we have the collection of rejections $\mathcal{R}_{i+1} = \mathcal{R}_i \cup \mathcal{N}(\mathcal{R}_i)$. Let $\mathcal{R}_\infty = \lim_{i\to\infty} \mathcal{R}_i$ be the final collection of rejections.

A sequential rejection principle (Goeman and Solari, 2010) states sufficient conditions on $\mathcal{N}$ to control the FWER strongly for sequentially rejective procedures. We denote M as a submodel of a set of statistical models, $\mathcal{M}$, i.e., $M \in \mathcal{M}$. A probability measure P_M is defined on the whole outcome space Ω. $\mathcal{H}$ is a collection of all tested hypotheses. According to the probability measure P_M, for any null hypothesis H in $\mathcal{H}$, we say that H is true if $M \in H$; otherwise, we say that H is false. A set of true null hypotheses and a set of false null hypotheses are defined as follows:

$$\mathcal{T}(M) = \{H \in \mathcal{H} : M \in H\};\ \mathcal{F}(M) = \mathcal{H} \backslash \mathcal{T}(M).$$

For simplicity, we let $\mathcal{T} = \mathcal{T}(M)$ and $\mathcal{F} = \mathcal{F}(M)$.

Theorem 1 (Sequential Rejection Principle)**.** *Suppose that for every $\mathcal{R} \subseteq \mathcal{S} \subset \mathcal{H}$, almost everywhere,*

$$\mathcal{N}(\mathcal{R}) \subseteq \mathcal{N}(\mathcal{S}) \cup \mathcal{S}, \tag{1.1}$$

and at each single step, for every $M \in \mathcal{M}$,

$$P_M(\mathcal{N}(\mathcal{F}) \subseteq \mathcal{F}) \geq 1 - \alpha, \tag{1.2}$$

then for every $M \in \mathcal{M}$, we have

$$P_M(\mathcal{R}_\infty \subseteq \mathcal{F}) \geq 1 - \alpha. \tag{1.3}$$

The sequential rejection principle states that as long as the conditions in (1.1) and (1.2) are satisfied for the successor function, the sequentially rejective procedure strongly controls the FWER. Condition in (1.1) means that the successor function is monotonic, and condition in (1.2) guarantees the FWER control in the critical case in which all false null hypotheses have been rejected and none of the true ones.

1.2 Multiple Testing Procedures

In multiple testing, most common MTPs can be divided into two general classes, one is hypotheses are ordered based on some prior information, such as clinical importance of tested hypotheses, and the other is hypotheses are ordered based on p-values

1.2.1 Procedures for Testing Pre-ordered Hypotheses

For some clinical applications, the hypotheses often have some inherent structure, such as hypotheses corresponding to primary endpoints are claimed more important than those corresponding to secondary endpoints. Thus, primary hypotheses should be tested first. Fixed sequence procedures are developed for hierarchically ordered hypotheses $H_1, H_2, \ldots, H_m$, where the order is pre-specified, usually based on clinical importance. Several fixed sequence procedures were developed in the literature. The conventional fixed sequence procedure (Maurer et al., 1995; Westfall and Krishen, 2001) and the fallback procedure (Wiens, 2003) are widely used in clinical trials. And the fallback procedure is further developed by Wiens and Dmitrienko (2005), Dmitrienko, Wiens and Westfall (2006), Hommel, Bretz and Maurer (2007), Hommel and Bretz (2008) and Bretz et al. (2009).

Suppose that the tested hypotheses $H_1, \ldots, H_m$ are pre-ordered and the correspond p-values $P_1, \ldots, P_m$ are available.

- *Conventional fixed sequence procedure.* The conventional fixed sequence procedure is proposed for testing multiple hypotheses that have a pre-specified fixed order. The test starts with the first hypothesis H_1, and reject all hypotheses $H_1, \ldots, H_j$, where $1 \leq j \leq m$ is the largest index satisfying

  $$P_1 \leq \alpha, \ldots, P_j \leq \alpha.$$

  The procedure strongly controls the FWER at lelvel α under any dependence of p-values. However, one drawback of this procedure is that once an acceptance is observed, the rest hypotheses have no chance to be tested.

- *Fallback procedure.* The fallback procedure is proposed to overcome the shortcoming of earlying stopping in the conventional fixed sequence procedure.

It allows each hypothesis to be tested at least at its local critical value. Initially, the overall critical value α is allocated to m hypotheses as local critical values, $\alpha_i, i = 1, \ldots, m$, with $\sum_{i=1}^m \alpha_i = \alpha$. Let α_i^* be the updated critical value for testing H_i and $\alpha_1^* = \alpha_1$. Hypothesis H_i is rejected if $P_i \leq \alpha_i^*$, where

$$\alpha_i^* = \begin{cases} \alpha_i + \alpha_{i-1}^*, & \text{if } H_{i-1} \text{ is rejected} \\ \alpha_i, & \text{otherwise.} \end{cases} \tag{1.4}$$

With the propagation of critical values, the latter hypotheses may have more chance to be rejected with higher critical values. The fallback procedure is pretty general and the conventional fixed sequence procedure can be regarded as its special case, when $\alpha_1 = \alpha$ and $\alpha_i = 0$ for $i = 2, \ldots, m$.

1.2.2 Procedures Based on Ordered p-values

Stepwise methods, such as step-down, step-up and single-step, are based on ordered p-values. The stepwise MTPs can be described by using a sequence of non-decreasing critical constants $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_m$. Let $P_1, P_2, \ldots, P_m$ be marginal p-values of tested hypotheses $H_1, H_2, \ldots, H_m$. Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$ be the ordered p-values with the corresponding null hypotheses $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$.

- Step-down procedure. A step-down procedure starts with the most significant hypothesis $H_{(1)}$, and gradually steps down to the least significant hypothesis $H_{(m)}$. The procedure goes on rejecting hypotheses as long as the corresponding p-value $P_{(i)} \leq \alpha_i$. The test stops until no more rejection can be made. That is, reject $H_{(1)}, \ldots, H_{(r)}$, where $1 \leq r \leq m$ is the largest index satisfying

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

  And the rest hypotheses $H_{(r+1)}, \ldots, H_{(m)}$ are accepted. If $P_{(1)} > \alpha_1$, accept all m null hypotheses.

  Holm procedure. One typical example of a step-down procedure is Holm procedure (Holm, 1979). It controls the FWER under arbitrary dependence. The critical values of Holm procedure are updated as $\alpha_i = \frac{\alpha}{m-i+1}$, for $i = 1, \ldots, m$.

- Step-up procedure. A step-up procedure starts with the least significant hypothesis $H_{(m)}$, and gradually steps up to the most significant hypothesis $H_{(1)}$. The procedure goes on accepting hypotheses as long as the corresponding p-value $P_{(i)} > \alpha_i$. The test stops until a rejection is observed. That is, reject

$H_{(1)}, \ldots, H_{(r)}$, and do not reject $H_{(r+1)}, \ldots, H_{(m)}$, where $1 \leq r \leq m$ is the largest index such that

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

If $P_{(m)} \leq \alpha_m$, we reject all m hypotheses.

Hochberg procedure. One typical example of a step-up procedure is Hochberg procedure (Hochberg, 1988). It controls the FWER under positive dependence. The critical values of Hochberg procedure are updated as $\alpha_i = \frac{\alpha}{m-i+1}$, for $i = 1, \ldots, m$.

Benjamini-Hochberg procedure. The other typical example of a step-up procdure is Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), short for BH procedure. It is a commonly used FDR controlling procedure. The critical values of BH procedure are updated as $\alpha_i = \frac{i\alpha}{m}$, for $i = 1, \ldots, m$.

- Single-step procedure. A single-step procedure is a special stepwise procedure with the same critical values, i.e., $\alpha_1 = \alpha_2 = \ldots = \alpha_m = c$, where c is a constant. Hypothesis H_i is rejected if and only of $P_i \leq c$, for $i = 1, \ldots, m$.

  Bonferroni procedure. The Bonferroni procedure is one typical and widely used single-step procedure, which controls the FWER under arbitrary dependence. Its critical constant is $\alpha_i = \frac{\alpha}{m}$, for all $i = 1, \ldots, m$.

1.2.3 Graphical Approaches

The aforementioned MTPs, such as the Holm procedure, conventional fixed sequence procedure and fallback procedure, are widely used in clinical trials. However, the decision tables of these MTPs are often long and abstract, which make them difficult to be presented to non-statisticians clearly and intuitively.

Bretz et al. (2009) and Burman et al. (2009) independently proposed to use graphical tools to visualize Bonferroni-type sequentially rejective procedures. The graphical approaches provide a mean for specifying, communicating, and assessing different hypothesis testing strategies. Moreover, a variety of testing strategies can be demonstrated and compared to tailor a suitable test scheme to specific study objectives.

Basically, in the graphical approach, the testing strategy is defined by a graph that shows each null hypothesis located at a vertex, and depicts the relationships between null hypotheses via directed edges with an arrowheads. The overall critical value α is initially allocated to each vertex with initial critical value (local critical value), say α_i for hypothesis H_i. The number along each directed edge is termed as transition coefficient (weight), which indicates the fraction of the preserved local critical value to be shifted along that edge to the receiving hypothesis, when the hypothesis at the tail end of the edge is significant. For example, g_{jl} indicates the proportion of α_j from H_j can be passed to H_l if H_j is rejected. If at each step, a hypothesis H_j is rejected, the graph will be updated by removing the rejected hypothesis H_j. And all local critical values and transition coefficients will be updated as well based on the following rules:

$$
\begin{aligned}
&I \rightarrow I \backslash \{j\}, \\
&\alpha_l = \begin{cases} \alpha_l + \alpha_j g_{jl}, & \text{if } l \in I, \\ 0, & \text{otherwise,} \end{cases} \\
&g_{lk} = \begin{cases} \frac{g_{lk} + g_{lj} g_{jk}}{1 - g_{lj} g_{jl}}, & \text{if } l, k \in I, l \neq k, g_{lj} g_{jl} \neq 1, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

I is the indices of the remaining hypotheses. Initially, $I = \{1, \ldots, m\}$. Such graphical approach satisfies the following regularity conditions on critical values and transition coefficients:

$$\sum_{l=1}^{m} \alpha_l \leq \alpha, \tag{1.5}$$

$$0 \leq g_{lk} \leq 1, g_{ll} = 0, \text{ for } l, k = 1, \ldots, m. \tag{1.6}$$

$$\sum_{k=1}^{m} g_{lk} \leq 1, \text{ for } l = 1, \ldots, m. \tag{1.7}$$

Bretz et al. (2009) shows that the graphical approach strongly controls the FWER at level α if conditions in (1.5), (1.6) and (1.7) are satisfied.

At the same time, graphical tools have been extensively used to visualize different testing strategies in various clinical applications, such as testing of non-inferiority and superiority (Hung and Wang, 2010; Guilbaud, 2011; Lawrence, 2011), assessing of composite endpoints and their components (Huque et al., 2011; Rauch and Beyersmann, 2013), and subgroup analyses (Bretz et al., 2011a). The original graphical approaches mostly focus on Bonferroni-based MTPs, but researchers have also extended to weighted Simes' or parametric tests based graphical approaches (Bretz et al., 2011b; Maurer et al., 2011; Millen and Dmitrienko, 2011; Xi et al., 2016; Lu, 2016). For complex study objectives in clinical trials, the approaches have been further extended to adaptive designs (Sugitani et al., 2013, 2014; Klinglmueller et al., 2014), group sequential designs (Maurer and Bretz, 2013b; Xi and Tamhane, 2015) and families of hypotheses (Kordzakhia and Dmitrienko, 2013; Maurer and Bretz, 2014), etc. The power performance of the graphical approaches is also considered in Bretz et al. (2011a). The implementation of graphical approaches in SAS and R are described in Bretz et al. (2011a, b).

1.3 Multiple Testing Procedures for Logically Related Hypotheses

In many clinical applications, tested hypotheses often have some logical relationships. For example, there are usually multiple endpoints of interest in clinical trials and these endpoints are generally classified as primary, secondary and sometimes tertiary endpoints which forms a natural hierarchical structure. The hypotheses corresponding to the secondary or tertiary endpoints are testable if some logical conditions in the hypotheses corresponding to primary endpoints are satisfied. Various logical relationships often exist among tested hypothese or among families of hypotheses. To deal with such logically related hypotheses testing problems, several general methods were developed in the literature.

1.3.1 Gatekeeping Procedures

A common logical structure in clinical trials is that the hypotheses to be tested are grouped into multiple families, and these families are tested in a sequential manner. Maurer, Hothorn and Lehmacher (1995) and Bauer et al. (1998) introduced a convenient and efficient way called gatekeeping strategy in which primary family is tested first, and whether secondary family is testable contingent upon the testing results within the primary family. That is, the primary family serves as a gatekeeper for the secondary one.

Several gatekeeping strategies are commonly used in clinical trials, including serial gatekeeping, parallel gatekeeping and tree-structured gatekeeping strategies.

- *Serial gatekeeping.* Westfall and Krishen (2001) proposed a serial gatekeeping strategy, in which each family can be tested using any FWER controlling method if and only if all hypotheses in the previous families are statistically significant. Thus, the logical condition for the current family to be testable is that all hypotheses are rejected in the previous families.
- *Parallel gatekeeping.* Dmitrienko, Offen and Westfall (2003) introduced a parallel gatekeeping strategy, in which the current family can be tested if and only if at least one hypothesis in the previous family is statistically significant, that is, in parallel gatekeeping strategy, the logical condition for the current family to be testable is that at least one hypothesis is rejected in the previous family.
- *Tree-structured gatekeeping.* Tree-structured gatekeeping procedure was proposed by Dmitrienko, Wiens and Tamhane (2007), which is a hybrid procedure unifying serial gatekeeping strategy and parallel gatekeeping strategy. In this strategy, the tested hypothese are formulated as a tree structure, where each node represents a null hypothesis. Instead of exhibiting a simple sequential structure in the decision making process, the procedure is based on a decision tree with multiple branches. Tree-structured gatekeeping procedure is derived based on the closure principle and uses weighted Bonferroni procedure for all intersection hypotheses.

With increasing complexity of hierachically ordered families of hypotheses, an extension of the tree-structured gatekeeping procedure, mixture procedure proposed by Dmitrienko and Tamhane (2011, 2013), were also developed to deal with the problems of multiple families of hypotheses. Instead of using the Bonferroni procedure, the mixture procedure uses more powerful tests for all intersection hypotheses, such as Simes' test. Moreover, it can cover more general logical restrictions than the tree-structured approach. In order to avoid challenging computational issues caused by the closure principle, Dmitrienko et al. (2006), Guilbaud (2007) and Dmitrienko, Tamhane and Wiens (2008) developed simple stepwise approaches in dealing with gatekeeping strategies. Particularly, Dmitrienko, Tamhane and Wiens (2008) introduced a general multistage gatekeeping procedure, which includes simple stepwise approaches of Dmitrienko et al. (2006) and Guilbaud (2007) as its special cases. A key property of the multistage gatekeeping procedure is that the unused critical values of the current family can be passed to the subsequent family. In order to quantify the amount of critical value for the current family that will be transferred to the subsequent family, a new concept of error rate function is introduced in Dmitrienko, Tamhane and Wiens (2008), which is formally defined as follows:

Error rate function. Consider a single family, $F = \{H_1, \ldots, H_m\}$, for any $I \subseteq \{1, \ldots, m\}$, the error rate function is defined as follows:

$$e(I) = \sup_{H_I} \Pr\left(\bigcup_{i \in I} \{\text{reject } H_i\} | H_I\right), \tag{1.8}$$

where $H_I = \cap_{i \in I} H_i$. And $e(I)$ is the maximum probability of making at least one Type I error in the subfamily $\{H_i, i \in I\}$. Generally, an exact expression of $e(I)$ is difficult to drive and thus a computable upper bound $e^*(I)$ is often used to replace it. For example, for the conventional Bonferroni procedure, the upper bound of its error rate function is

$$e^*(I) = \frac{\alpha |I|}{m}, \tag{1.9}$$

where $|I|$ is the cardinality of the set I.

Due to the stepwise shortcut, the multistage gatekeeping procedure is apparently more straightforward and easier to explain to clinicians in practice. However, in terms of dealing with complex logical conditions, multistage gatekeeping procedure is still not as flexible as the mixture procedure (Dmitrienko and Tamhane, 2011, 2013).

1.3.2 Graphical Visualization

Several graphical approaches were also developed for testing logically related hypotheses:

- *Tree structured method.* Meinshausen (2008) formulated a variable selection problem in high-dimensional regression as a tree-structured hypotheses testing problem and developed a hierarchy method. It can be regarded as a special case of the original graphical approach (Bretz et al., 2009), where descendant hypotheses in the same layer are allocated with equal critical values if their parent hypotheses are rejected, otherwise they are non-testable.
- *Entangled graphical approach.* For increasing complex of logical restrictions among tested hypotheses, proper visualization and presentation of logically related MTPs should be taken into account. Entangled graphical approach has been proposed to perform MTPs for testing logical restricted hypotheses (Maurer and Bretz, 2013a). Each graph represents one logical restriction among all tested hypotheses, and within each graph, one can use the algorithm introduced in the original graphical approach by Bretz et al. (2009) to make rejection or acceptance decisions regarding tested hypotheses. However, a drawback of the entangled graphical approach is that with the increasing number of logical restrictions, one needs to construct and update a large number of graphs, which makes this method computationally inefficient.
- *Superchain procedure.* Superchain procedure (Kordzakhia and Dmitrienko, 2013) was developed to sequentially test logically structured hypotheses via graphical approaches. Each family is presented as a vertex and local significant levels are propagated via transition coefficients between families instead of hypotheses. However, this approach tests all families of hypotheses simultaneously at each step which is not suitable in some clinical trial settings, such as families of hypotheses having hierachical structure.
- *Family-based graphical approach.* The aforementioned gatekeeping strategies are often either non-intuitive or less flexible when addressing increasingly complex logical relationships among families of hypotheses. Qiu, Guo and Yu (2017) proposed a family-based graphical approach to solve such probelms, in which equally important families are grouped in the same layer and a directed and weighted graph is used to develop family-based gatekeeping strategy where

each node corresponds to a family of hypotheses and two simple updating rules are used for updating the critical value of each family and the transition coefficient between any two families. The proposed approach can be used to easily derive and visualize different gatekeeping strategies. However, this approach is not suitable for dealing with multiple testing probelms with general logical restrictions among families of hypotheses.

1.4 Selective Inference

With the increasing number of tested hypotheses, one natural testing strategy is to first reduce the number of tested hypotheses by some selection process, and then to simultaneously test the selected hypotheses. The main advantage of this strategy is to greatly reduce the severe effect of high dimensions. However, the first screening or selection stage must be properly accounted for in order to maintain some type of error rate control.

1.4.1 Procedures after Selection/Screening

Benjamin and Yekutieli (2005) introduced a new approach for constructing multiple selective confidence intervals after screening a large number of parameters with the control of false coverage rate (FCR), which is the expected proportion of nonovering confidence interval among all constructed confidence intervals. It is the first time that the concept of selective inference was formally introduced in the literature. Benjamini (2010) and Taylor and Tibshirani (2015) also demonstrated imporatnce of selectvie inference in the era of big data. The construction methods of selection adjusted multiple confidence intervals with the control of FCR were further developed by Weinstein, Fithian and Benjamin (2014), Weinstein and Yekutieli (2016) and Peng et al. (2017).

In high-dimensional regression analysis, several novel breakthroughs have been recently made by Berk et al. (2013), Barber and Candes (2015), Lee et al. (2016), Fithian et al. (2014). All of these works emphasize on how to perform valid post selection inference after model building. Following these works, many selective inference/post selection inference methods have been developed for a variety of model selection algorithms (Barber and Candes, 2016; Fithian et al.,2015; Tian and Taylor, 2015a, b; Yekutieli, 2012; Tian and Taylor, 2018; Tian, Loftus and Taylor, 2018; Panigrahi and Taylor, 2018; Taylor and Tibshirani, 2018).

Selective inference is also a hot topic in multiple hypotheses testing recently. Several selective inference methods have been developed in large scale multiple hypotheses testing (Skol et al., 2006; Bourgon et al., 2010; Benjamini and Bogomolov, 2014; Heller et al., 2016; Guo and Romano, 2017). Among these selective inference methods, an interesting development is independent filtering method, where selection statistic and testing statistic are chosen to be independent when the corresponding hypothesis is true (Bourgon et al., 2010; Dai et al., 2012; Du and Zhang, 2014; Ignatiadis et al., 2016; Guo and Romano, 2017; Heller et al., 2018), which has several nice properties: (i). it completely removes the selection effect; (ii). it reduces the multiplicity effect; (iii). it does not "waste" data while carry out to the selection testing. In this dissertation, we further develop the ideas and methods of independent filtering.

1.4.2 Data-driven Weighted MTPs

In multiple hypotheses testing, importance of null hypotheses is often different. It is natural to assign different weights to different hypotheses. Traditionally, the weights are pre-specified by some prior knowledge. However, without the prior knowledge at hand, how can one specify the weights? Roeder and Wasserman (2009) and Poisson et al. (2012) introduced data-driven weighted multiple testing procedures, where weights for null hypotheses are constructed by exploiting information in the data.

It is interesting to note that the aferomentioned screening methods can be regarded as a special case of general data-driven weighted methods. For data-driven weighted methods, there are still some important questions remain to be answered, such as, how to ensure the control of the FWER or FDR when constructing and assigning weights in a data-driven way? How to ensure such methods are more powerful than the conventional FWER or FDR controlling procedures? etc. Very recently, by employing side information to construct weights which are independent of the test statistics under the corresponding null hypotheses, several Bonferroni-based and Benjamini-Hochberg based data-driven weighted methods have been developed to increase power while still controlling the FWER and FDR, respectively (Fino and Salmaso, 2007; Ignatiadis, et al., 2016; Li and Barber, 2016; Lei and Fithian, 2016; Ignatiadis and Huber, 2017).

1.5 Motivation and Outline

Although existing sequentially rejective, weighted Bonferroni-based MTPs in the literature can often address multiplicity issues propoerly, the drawbacks of such procedures are that the decision tables or results are often abstract and difficult to present to clinical teams. Even though the original graphical approach (Bretz et al., 2009) provided a simple and clear graphical visualization to demonstrate the underlying testing strategies, a main shortcoming of this graphical approach is that it only allows one rejection at each step. As a result, too many graphs need to be generated and updated in the whole testing process with the increasing number of null hypotheses. In this dissertation, we first develop a new graphical approach with proven control of the FWER, which allows one to reject more than one hypothesis at each step.

Moreover, with complex logically structured hypotheses, to our knowlegde, there is no MTP exists in the literature, which can deal with general logical relationships among tested hypotheses, and at the same time, which has a simple and clear visualization to non-statisticians. The existing gatekeeping procedures can only be used to test families of hypotheses with simple logical restrictions. The existing graphical approaches provide a way to visualize complex testing strategies clearly and intuitively, but almost all of them can not deal with general logically related hypothese. The only exception is the entangled graphical approach, but it needs to generate and update too many graphs with the increasing complexity of logical relationships and it has low power performance in the sense that critical values from non-testable (logical restricted) hypotheses can not be fully used. In this dissertation, we propose a new graphical approach, which can deal with general logical restrictions among tested hypotheses as well as fully using critical values from non-testable hypotheses.

For large scale multiple hypothesis testing, most existing selective inference procedures emphasize on performing valid inference after selection, when the selection ruels are given. However, in practice, how to choose an appropriate selection rule to lead to more powerful testing procedures remains open. In this dissertation, we propose several powerful selective inference procedures with strong control of the FWER. Moreover, the proposed procedures appropriately take selection effect into account.

The graphical approaches mentioned in this dissertation are developed based on weighted Bonferroni procedures. The weights are usually pre-specified. However, without having any prior knowledge of the weights, one can only construct the weights based on data. Several data-driven weighted MTPs were developed very recently. The main drawback of these MTPs is that it is complicated to construct the weights. In this dissertation, we also develop simple ways to construct data-driven weights and thus develop more powerful weighted MTPs, which can be regarded as extensions of our proposed selective inference procedures.

The rest of this dissertation is outlined as follows: in Chapter 2, we propose a new graphical approach with proven FWER control by generalizing the sequential rejection principle, where one can reject more than one hypothesis at each step. Through some clinical examples, we illustrate that the proposed approach is more flexible and computationally efficient than the existing graphical approaches. In Chapter 3, we develop a new graphical approach for dealing with general logical relationships. In Chapter 4, we propose several powerful selective inference procedures with proven FWER control. In Chapter 5, extensive simulation studies are conducted for the proposed procedures in Chapter 4. In Chapter 6, we summarize and discuss the future works we are planning to do.

CHAPTER 2

A NEW GENERALIZED GRAPHICAL APPROACH

2.1 Introduction

In this chapter, we focus on developing a new sequentially rejective, weighted Bonferroni based graphical approach, which can reject more than one hypothesis at each step. The existing Bonferroni-based sequentially rejective MTPs include Holm procedure (Holm, 1979), fixed sequence procedure (Wiens, 2003), Bonferroni-adjusted gatekeeping procedures (Bauer et al., 1998; Westfall and Krishen, 2001; Dmitrienko, Offen and Westfall, 2003), etc. However, the decision tables of these MTPs may be long and non-visualized, which make them difficult to present to non-statisticians clearly and intuitively. Bretz et al. (2009) proposed to use graphical tools to describe and develop Bonferroni-type sequentially rejective procedures, which clearly demonstrate relationships between tested hypotheses via directed edges. In this dissertation, we term this graphical approach as the original graphical approach. Applying the original graphical approach, different testing strategies can be visualized to investigate and thus tailor proper multiple test procedures to specific clinical trial objectives. It strongly controls the FWER at level α under arbitrary dependence. The original graphical approach have been further extended by many authors, including Hung and Wang (2009), Bretz et al. (2011a, b), Maurer et al. (2011), etc. Bretz, Maurer and Maca (2014) provide a general review of the development of graphical approaches in Chapter 14 of Young and Chen (2015).

One main drawback of the original graphical approach (Bretz et al., 2009) and its extensions is that they can only reject one hypothesis at each test step. It is natural to consider generally how can one reject more than one hypothesis at each test step using graphical approaches. Since for a large number of tested hypotheses, if using the original graphical approach, we may need to iteratively generate many graphs and present all of theses graphs at the end of the test. For example, suppose we are interested in testing eight null hypotheses, where the relationships among theses null hypotheses are illustrated in Figure 2.1. Using the original graphical approach, if all eight hypotheses can be finally rejected, one needs to update eight different graphs and present all theses graphs to clinical teams to demonstrate the testing process of the underlying MTP. As one can see, the original graphical approach is not computationally efficient.

In this chapter, the main goal is to develop a more flexible and efficient graphical approach, which can reject more than one hypothesis at each step and strongly controls the FWER as well. We term it as a generalized graphical approach throughout our dissertation. The elements in the proposed graphical display is the same as the original graphical approach as described in Section 1.2.3, where each hypothesis is located in a vertex, and the relationships among null hypotheses are mapped by direct edges with associated transition coefficients. However, in the proposed graphical approach, we test all hypotheses that have positive local critical values, and reject all hypotheses H_i if its p-value $p_i \leq \alpha_i$. Consider the motivating example in Figure 2.1, hypotheses H_1 and H_2 are tested at level α_1 and α_2, simultaneously, such that $\alpha_1 + \alpha_2 = \alpha$. If H_1 and H_2 are both rejected, α_1 and α_2 will be splitted equally and passed to H_3, H_4 and H_5 relying on the edges. If H_3, H_4 and H_5 are rejected in the subsequence at their updated local critical values, their critical values are passed down further. The test stops until no more hypotheses can be rejected. We will revisit the motivating example in Section 2.5 for more details.

Moreover, in theoretical aspect, regarding strong control of the FWER of the proposed method, we also develop an alternative way through which we can show its FWER control. Many existing MTPs using the original graphical approach belong to a subclass of weighted Bonferroni-based closed testing procedures (Marcus, Eric and Gabriel, 1976), thus strongly control the FWER via the closure principle. However,

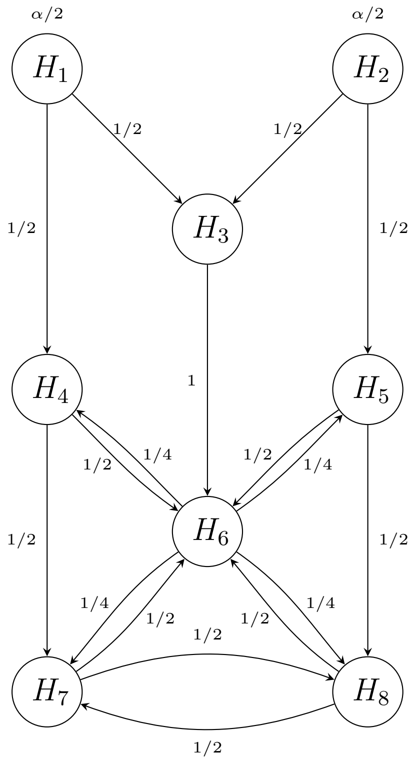


Figure 2.1 Graphical illustration of the motivating example, with $m = 8$, initial allocation $\boldsymbol{\alpha} = \{\alpha/2, \alpha/2, 0, 0, 0, 0, 0, 0\}$, and overall level $\alpha = 0.05$.

with a different aspect of view, one can show that the Bonferroni-based sequentially rejective multiple testing procedures strongly control the FWER by the sequential rejection principle (Goeman and Solari, 2010). Inspired by the original sequential rejection principle, we show the FWER control of the proposed method by the generalized sequential rejection principle. Theoretical details are shown in Section 2.4.

The advantages of the proposed generalized graphical approach are: compared to the original graphical approach, more than one hypothesis can be rejected at each step, which makes it computationally efficient. Moreover, with the reduction of total updated graphs, the presentation to clinical teams is more clear. And as a by-product, we generalize the sequential rejective principle and provide an alternative way to show the FWER control of some MTPs in clinical trials.

The rest of this chapter is organized as follows: in Section 2.2, we introduce some general notations used in this chapter. The generalized graphical approach is formalized with a simple iterative algorithm to perform a MTP with a graphical display in Section 2.3. The FWER control of the proposed approach is proved in Section 2.4. In Section 2.5, we demonstrate clinical trial examples to clearly illustrate the proposed graphical approach. In Section 2.6, some concluding remarks are given.

2.2 Preliminaries

In this section, we introduce some general notations used in this chapter. Consider a multiple testing problem with m null hypotheses $H_1, ..., H_m$. Let $\mathcal{H} = \{H_1, ..., H_m\}$ be a collection of the m null hypotheses, and $I = \{1, ..., m\}$ be the corresponding index set. Suppose $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)$ be the pre-speficied allocation of the overall critical value α to each individual hypothesis at the beginning of test, and $\mathbf{G}$=(g_{ij}) be an initial $m \times m$ transition coefficient matrix, where the transition coefficient g_{ij}, indicates the proportion of the critical value of hypothesis H_i that will be passed down to hypothesis H_j if H_i is rejected. In this dissertation, we assume true null p-values are always stochastically greater than or equal to a uniform distribution on [0,1]. That is, for $p \in [0,1]$,

$$\Pr\{P_i \leq p\} \leq p, i \in I_0, \tag{2.1}$$

where I_0 is the indices of true null hypotheses.

At the beginning of test, the initial critical values $\boldsymbol{\alpha}$ and the transition coefficient matrix $\mathbf{G}$=(g_{ij}) are given. All raw p-values $p_1, ..., p_m$ are observed, where p_i denotes the observed p-value of $P_i, i = 1, \ldots, m$.

2.3 The Proposed Graphical Approach

In this section, we introduce a generalized graphical approach, different from the original one (Bretz et al., 2009), which can reject more than one hypothesis at each step. In this graphical approach, instead of updating a critical value and transition coefficient at each step as in the original graphical approach, we update a critical value function and transition coefficient function based on a set of rejections. Define $\alpha_l(\hat{\mathbf{R}}_{i-1})$ be a function based on $\hat{\mathbf{R}}_{i-1}$ for hypothesis $H_l, l = 1, ..., m$, at step $i = 1, ..., m$, where $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$ is a vector of rejection sets, and $\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1}$ are nonempty, mutually exclusive sets of rejections at step $1, ..., i-1$, respectively. We name $\alpha_l(\hat{\mathbf{R}}_{i-1})$ as a multivariate critical value function throughout the dissertation. And $g_{lk}(\hat{\mathbf{R}}_{\mathbf{i-1}})$ as a multivariate transition coefficient function. Denote that $\hat{\mathcal{R}}_0 = \emptyset$, $\alpha_l(\hat{\mathcal{R}}_0) = \alpha_l$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}$. We assume that

$$\sum_{l=1}^{m} \alpha_l \leq \alpha, \tag{2.2}$$

$$0 \leq g_{lk} \leq 1, g_{ll} = 0, \text{ for } l, k = 1, \ldots, m, \tag{2.3}$$

$$\sum_{k=1}^{m} g_{lk} \leq 1, \text{for every } l = 1, ..., m. \tag{2.4}$$

At the beginning of each step i, we have an updated indices of the remaining hypotheses I_i. Initially, we have $I_1 = I = \{1, ..., m\}$. For each $\hat{\mathcal{R}}_i$ at step i, let the corresponding indices be J_i with $J_0 = \emptyset$. With all the above notations, the following algorithm defines a sequentially rejective procedure:

Algorithm 1.

1. *For $j \in I_1$, test every H_j at level α_j. If $p_j \leq \alpha_j$, reject H_j. Then we have $\hat{\mathcal{R}}_1 = \{H_j \in \mathcal{H} : p_j \leq \alpha_j, \text{for any } j \in I_1\}$; If $\hat{\mathcal{R}}_1 = \emptyset$, then stop.*

Update the graph:

$$I_1 \to I_2 = I_1 \backslash J_1,$$

$$\alpha_l(\hat{\mathcal{R}}_1) = \begin{cases} \alpha_l + \sum\limits_{j \in J_1} \alpha_j g_{jl}, & \textit{if } l \in I_2, \\ 0, & \textit{otherwise}, \end{cases} \tag{2.5}$$

$$g_{lk}(\hat{\mathcal{R}}_1) = \begin{cases} \frac{g_{lk} + \sum\limits_{j \in J_1} g_{lj} g_{jk}}{1 - \sum\limits_{j \in J_1} \sum\limits_{q \in J_1} g_{lj} g_{jq} - \sum\limits_{j \in J_1} g_{lj} g_{jl}}, & \textit{if } l, k \in I_2, l \neq k, \\ 0, & \textit{otherwise}. \end{cases} \tag{2.6}$$

***i*(*i* ≥ 2)**. *For $j \in I_i$, test every H_j at level $\alpha_j(\hat{\mathbf{R}}_{i-1})$. If $p_j \leq \alpha_j(\hat{\mathbf{R}}_{i-1})$, reject H_j. Then we have $\hat{\mathcal{R}}_i = \{H_j \in \mathcal{H} \backslash \cup_{k=1}^{i-1} \hat{\mathcal{R}}_k : p_j \leq \alpha_j(\hat{\mathbf{R}}_{i-1}), \text{ for all } j \in I_i\}$; If $\hat{\mathcal{R}}_i = \emptyset$, stop.*

Update the graph:

$$I_i \to I_{i+1} = I_i \backslash J_i,$$

$$\alpha_l(\hat{\mathbf{R}}_i) = \begin{cases} \alpha_l(\hat{\mathbf{R}}_{i-1}) + \sum\limits_{j \in J_i} \alpha_j(\hat{\mathbf{R}}_{i-1}) g_{jl}(\hat{\mathbf{R}}_{i-1}), & \textit{if } l \in I_{i+1}, \\ 0, & \textit{otherwise}, \end{cases} \tag{2.7}$$

$$g_{lk}(\hat{\mathbf{R}}_i) = \begin{cases} \frac{g_{lk}(\hat{\mathbf{R}}_{i-1}) + \sum\limits_{j \in J_i} g_{lj}(\hat{\mathbf{R}}_{i-1}) g_{jk}(\hat{\mathbf{R}}_{i-1})}{1 - \sum\limits_{j \in J_i} \sum\limits_{q \in J_i} g_{lj}(\hat{\mathbf{R}}_{i-1}) g_{jq}(\hat{\mathbf{R}}_{i-1}) - \sum\limits_{j \in J_i} g_{lj}(\hat{\mathbf{R}}_{i-1}) g_{jl}(\hat{\mathbf{R}}_{i-1})}, & \textit{if } l, k \in I_{i+1}, l \neq k, \\ 0, & \textit{otherwise}. \end{cases} \tag{2.8}$$

Remark 1. *In any case for updating the transition coefficient function, if we encounter the case of the denominators of the right-sides in (2.6) and (2.8) being equal to zero, we simply set* $g_{lk}(\hat{\mathbf{R}}_i) = 0$*, for every* $i = 1, ..., m$.

Remark 2. *The updating rules of the critical value function indicate that all critical values of rejected hypotheses will pass to hypotheses via directed edges. Moreover, in order to standardize the transition coefficient function at each step, in the denominators of the right-sides in (2.6) and (2.8), we substract product of any two transition coefficient functions corresponding to rejected hypotheses if they are connected with each other.*

By removing all rejected hypotheses at the end of each step, not only the critical value functions, but also the transition coefficient functions need to be updated.

Proposition 1. *Under the assumptions stated in (2.2), (2.3), and (2.4) for initial critical values and transition coefficients, the critical value functions and transition coefficient functions at each step* $i = 1, \ldots, m$*, have the following properties:*
For transition coefficient functions,

$$1.\ \sum_{k \in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1,\ \textit{for } i = 1, ..., m,\ \textit{and } I_i \subseteq \{1, \ldots, m\}, \tag{2.9}$$

$$2.\ 0 \leq g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1, g_{kk}(\hat{\mathbf{R}}_{i-1}) = 0,\ \textit{for } i = 1, ..., m. \tag{2.10}$$

For critical value functions,

$$3.\ \sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha,\ \textit{for } i = 1, ..., m,\ \textit{and } I_i \subseteq \{1, \ldots, m\}, \tag{2.11}$$

$$4.\ \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}),\ \textit{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \tag{2.12}$$

$$\textit{and for } i = 1, ..., m.$$

The proof of Proposition 1 is deferred to Appendix A.

2.3.1 Relationship between the Proposed and the Original Approaches

In this section, we explore the relationship between the proposed graphical approach and the original one. We first define some notations. Let $\tilde{\mathcal{R}}_s = \{H_{i_1}, \ldots, H_{i_s}\}$ for $1 \leq s \leq m$ denote a rejection set of the original graphical approach, including s rejected hypotheses, and the order of rejected hypotheses is from H_{i_1} to H_{i_s}. Let $\tilde{J}_s = \{i_1, \ldots, i_s\}$ be the corresponding indices. Note that $\tilde{\mathcal{R}}_0 = \emptyset$.

Based on $\tilde{\mathcal{R}}_s$, we recursively define the critical value functions and the transition coefficient functions for our proposed method as follows:

For $s = 0$,

$$\alpha_l(\tilde{\mathcal{R}}_0) = \alpha_l, \text{ and } g_{lk}(\tilde{\mathcal{R}}_0) = g_{lk}. \tag{2.13}$$

For $1 \leq s \leq m$,

$$\alpha_l(\tilde{\mathcal{R}}_s) = \begin{cases} \alpha_l(\tilde{\mathcal{R}}_{s-1}) + \alpha_{i_s}(\tilde{\mathcal{R}}_{s-1}) g_{i_s l}(\tilde{\mathcal{R}}_{s-1}), & \text{if } l \in I \backslash \tilde{J}_s, \\ 0, & \text{otherwise,} \end{cases} \tag{2.14}$$

$$g_{lk}(\tilde{\mathcal{R}}_s) = \begin{cases} \frac{g_{lk}(\tilde{\mathcal{R}}_{s-1}) + g_{l i_s}(\tilde{\mathcal{R}}_{s-1}) g_{i_s k}(\tilde{\mathcal{R}}_{s-1}))}{1 - g_{l i_s}(\tilde{\mathcal{R}}_{s-1}) g_{i_s l}(\tilde{\mathcal{R}}_{s-1})}, & \text{if } l, k \in I \backslash \tilde{J}_s, l \neq k, \\ 0, & \text{otherwise.} \end{cases} \tag{2.15}$$

Based on the critical value functions and transition coefficient functions defined in (2.13), (2.14) and (2.15), the following algorithm describes an original graphical approach with more than one rejection at each step.

Algorithm 2.

1. *For $j \in I_1$, test every H_j at level $\alpha_j(\tilde{\mathcal{R}}_0)$. If $p_j \leq \alpha_j(\tilde{\mathcal{R}}_0)$, reject H_j. Then we have $\hat{\mathcal{R}}_1 = \{H_j \in \mathcal{H} : p_j \leq \alpha_j(\tilde{\mathcal{R}}_0),$ for any $j \in I_1\}$, with $\mathcal{R}_1 = \hat{\mathcal{R}}_1$; If $\hat{\mathcal{R}}_1 = \emptyset$, then stop.*

Update the graph:

$$
\begin{aligned}
& I_1 \to I_2 = I_1 \backslash J_1, \\
& \alpha_l(\tilde{\mathcal{R}}_0) \to \alpha_l(\tilde{\mathcal{R}}_{|\mathcal{R}_1|}), \ \textit{for } l \in I_2 && (2.16) \\
& g_{lk}(\tilde{\mathcal{R}}_0) \to g_{lk}(\tilde{\mathcal{R}}_{|\mathcal{R}_1|}), \ \textit{for } l, k \in I_2, l \neq k. && (2.17)
\end{aligned}
$$

$\boldsymbol{i(i \geq 2)}$. *For* $j \in I_i$, *test every* H_j *at level* $\alpha_j(\tilde{\mathcal{R}}_{|\mathcal{R}_{i-1}|})$. *If* $p_j \leq \alpha_j(\tilde{\mathcal{R}}_{|\mathcal{R}_{i-1}|})$, *reject* H_j. *Then we have* $\hat{\mathcal{R}}_i = \{H_j \in \mathcal{H} \backslash \mathcal{R}_{i-1} : p_j \leq \alpha_j(\tilde{\mathcal{R}}_{|\mathcal{R}_{i-1}|}),\ \textit{for all } j \in I_i\}$ *with* $\mathcal{R}_i = \mathcal{R}_{i-1} \cup \hat{\mathcal{R}}_i$; *If* $\hat{\mathcal{R}}_i = \emptyset$, *stop.*

Update the graph:

$$
\begin{aligned}
& I_i \to I_{i+1} = I_i \backslash J_i, \\
& \alpha_l(\tilde{\mathcal{R}}_{|\mathcal{R}_{i-1}|}) \to \alpha_l(\tilde{\mathcal{R}}_{|\mathcal{R}_i|}), \ \textit{for } l \in I_{i+1} && (2.18) \\
& g_{lk}(\tilde{\mathcal{R}}_{|\mathcal{R}_{i-1}|}) \to g_{lk}(\tilde{\mathcal{R}}_{|\mathcal{R}_i|}), \ \textit{for } l, k \in I_{i+1}, l \neq k. && (2.19)
\end{aligned}
$$

In Section 2.4, we will provide an alternative proof to show that it strongly controls the FWER at level α.

Remark 3. For the original graphical approach, since the order of rejections does not affect the final rejection results in the sense that if hypotheses H_1 and H_2 are both rejected by the original graphical approach, no matter either H_1 or H_2 is rejected first, the overall rejection set remains the same. Therefore, the critical value functions and transition coefficient functions in Algorithm 2 are unique based on a set of all previous rejections.

Remark 4. In Algorithm 2, we extend the original graphical approach to more general case with more than one hypothesis can be rejected at each step. In this case, we need to update the univariate critical value functions and the transition coefficient functions iteratively based on each individual rejection, which means that if we reject k hypotheses at step i, then the transition coefficient functions need to update k times based on the algorithm of the original graphical apptoach. There is no closed form for updating the transition coefficient function.

Remark 5. The generalized graphical approach propose a way for updating the transition coefficient function when rejecting more than one hypothesis at each step. Due to construction of the updating rule for transition coefficient function, the order of rejections affects the testing results. Thus we introduce an alternative concept of critical value function – multivariate critical value function $\alpha_l(\hat{\mathbf{R}}_i)$, which is defined on the collections of rejection of the first i steps, $\hat{\mathbf{R}}_i = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_i)^T$.

2.4 Main Theoretical Results

In this section, we show that the proposed graphical approach strongly controls the FWER under arbitrary dependence. Goeman and Solari (2010) introduce a sequential rejection principle to show the FWER control for sequentially rejective procedures, which provides an alternate besides the closure principle, and avoids the high demand of computation caused by using the closure principle. A sequentially rejective procedure is defined based on a univariate successor function $\mathcal{N}$. The sequential rejection principle states that a sequentially rejective procedure strongly controls the FWER as long as the univariate successor function $\mathcal{N}$ satisfies the following two conditions almost everywhere:

$$\text{Condition 1: } \mathcal{N}(\mathcal{R}) \subseteq \mathcal{N}(\mathcal{S}) \cup \mathcal{S}, \text{ for every } \mathcal{R} \subseteq \mathcal{S} \subset \mathcal{H}; \tag{2.20}$$

$$\text{Condition 2: } P_M(\mathcal{N}(\mathcal{F}(\mathcal{M})) \subseteq \mathcal{F}(\mathcal{M})) \geq 1 - \alpha. \tag{2.21}$$

Since

$$\mathcal{N}(\mathcal{R}) = \{H_j \in \mathcal{H} \backslash \mathcal{R} : p_j \leq \alpha_j(\mathcal{R})\},$$

the monotonicity of the univariate successor function in (2.20) is equivalent to the monotonicity of the univariate critical value function. The original graphical approach described in Section 2.3.1 is based on a univariate critical value function, we can apply the sequential rejection principle to show strong control of the FWER in the following:

An alternative proof of the FWER control of the original graphical approach. For condition 1 in (2.20), we have $\mathcal{N}(\mathcal{R}) = \{H_j \in \mathcal{H}\backslash\mathcal{R} : p_j \leq \alpha_j(\mathcal{R})\}$. Since $\mathcal{R} \subseteq \mathcal{S} \subset \mathcal{H}$, by monotonicity of the critical value function, we have

$$\alpha_j(\mathcal{R}) \leq \alpha_j(\mathcal{S}).$$

If $p_j \leq \alpha_j(\mathcal{R})$, then $p_j \leq \alpha_j(\mathcal{S})$. Hence all rejected hypotheses in $\mathcal{N}(\mathcal{R})$ must also belong to $\mathcal{N}(\mathcal{S})$. Thus $\mathcal{N}(\mathcal{R}) \subseteq \mathcal{N}(\mathcal{S}) \cup \mathcal{S}$.

For condition 2 in (2.21), we let $\mathcal{R} = \mathcal{F}(M)$ and $\mathcal{T}(\mathcal{M}) = \mathcal{H}\backslash\mathcal{F}(\mathcal{M})$ is the set of all true null hypothese, we have

$$\begin{aligned} & P_M(\text{reject at least one true null hypothesis}) \\ = \; & P_M(\cup_{j \in \mathcal{T}(\mathcal{M})}\{p_j \leq \alpha_j(\mathcal{R})\}) \\ \leq \; & \sum_{j \in \mathcal{T}(\mathcal{M})} P_M(p_j \leq \alpha_j(\mathcal{R})) \\ \leq \; & \sum_{j \in \mathcal{T}(\mathcal{M})} \alpha_j(\mathcal{R}) \leq \sum_{j \in \mathcal{H}} \alpha_j(\mathcal{R}) \leq \alpha. \end{aligned} \tag{2.22}$$

The first inequality in (2.22) is due to the Bonferroni inequality. The second inequality in (2.22) holds under the assumption that all true null p-values are stochastically greater than or equal to a uniform distribution on [0,1]. The third inequality in (2.22) holds since $\alpha_j(\mathcal{R}) > 0$. Thus, it is a valid sequentially rejective procedure with strong control of the FWER. □

Note that the proposed graphical approach is based on a multivariate critical value function as described in Section 2.3, we need to extend the original sequential rejection principle from the case of univariate critical value function to that of multivariate. We first define a random and measureable multivariate successor function $\mathcal{N}(\hat{\mathbf{R}})$, such that,

$$\mathcal{N}(\hat{\mathbf{R}}_{i-1}) = \hat{\mathcal{R}}_i = \{H_j \in \mathcal{H} \backslash \bigcup_{k=1}^{i-1} \hat{\mathcal{R}}_k : p_j \leq \alpha_j(\hat{\mathbf{R}}_{i-1})\}$$

is a collection of rejected hypotheses that can be made at step i after rejecting a collection of hypotheses $\hat{\mathbf{R}}_{i-1}$ in the previous $i-1$ steps. For convenience, in the rest of this dissertation, we denote $\mathcal{T} = \mathcal{T}(M)$ and $\mathcal{F} = \mathcal{F}(M)$. Let $\mathcal{F}_i$ be a collection of rejected false null hypotheses at step i, $i = 1, ..., m$. And $\mathbf{F}_i = (\mathcal{F}_1, ..., \mathcal{F}_i)^T$ be a vector of collections of all rejected false null hypotheses at the end of step i. Based on this multivariate successor function, we develop a generalized sequential rejection principle.

Theorem 2 (Generalized Sequential Rejection Principle)**.** *Suppose* $\hat{\mathcal{R}}_i \subseteq \mathcal{S}_i \subset \mathcal{H}$ *for every i=1, ..., m,* $\hat{\mathcal{R}}_1, \ldots, \hat{\mathcal{R}}_i$ *are all mutually exclusive, non-empty sets, and* $\mathcal{S}_1, \ldots, \mathcal{S}_i$ *are also all mutually exclusive, non-empty sets, almost everywhere, if*

$$\mathcal{N}(\hat{\mathbf{R}}_i) \subseteq \mathcal{N}(\mathbf{S}_i) \bigcup_{k=1}^{i} \mathcal{S}_k \quad \textit{for all } i = 1, \ldots, m, \tag{2.23}$$

and at each single step i, for every $M \in \mathcal{M}$,

$$P_M(\mathcal{N}(\mathbf{F}_i) \subseteq \mathcal{F}) \geq 1 - \alpha, \tag{2.24}$$

then for every $M \in \mathcal{M}$*, we have*

$$P_M(\cup_{k=1}^{\infty} \hat{\mathcal{R}}_k \subseteq \mathcal{F}) \geq 1 - \alpha. \tag{2.25}$$

Proof. By the condition of the single step FWER control in (2.24), we have $P_M(\mathcal{N}(\mathbf{F}_i) \subseteq \mathcal{F}) \geq 1 - \alpha$. If the event $\{\mathcal{N}(\mathbf{F}_i) \subseteq \mathcal{F}\}$ is realized, which means at step $i+1$, we did not reject any true null hypotheses. We will show that $\bigcup_{k=1}^{i} \hat{\mathcal{R}}_k \subseteq \mathcal{F}$. Clearly $\hat{\mathcal{R}}_0 = \emptyset \subseteq \mathcal{F}$. By mathematical induction, suppose that $\bigcup_{k=1}^{i} \hat{\mathcal{R}}_k \subseteq \mathcal{F}$, we will show that $\bigcup_{k=1}^{i+1} \hat{\mathcal{R}}_k \subseteq \mathcal{F}$. Since $\hat{\mathcal{R}}_1, \hat{\mathcal{R}}_2, ..., \hat{\mathcal{R}}_i$ are all mutually exclusive sets, we have $\hat{\mathcal{R}}_1 \subseteq \mathcal{F}_1$,$\hat{\mathcal{R}}_2 \subseteq \mathcal{F} \backslash \mathcal{F}_1$,...,$\hat{\mathcal{R}}_i \subseteq \mathcal{F} \backslash \bigcup_{k=1}^{i-1} \mathcal{F}_k$, then by monotonicity condition in (2.23), we have

$$\mathcal{N}(\hat{\mathbf{R}}_i) \subseteq \mathcal{N}(\mathbf{F}_i) \bigcup_{k-1}^{i} \mathcal{F}_k, \text{ where } \mathbf{F}_i = (\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_i)^T.$$

Thus, $\bigcup_{k=1}^{i+1} \hat{\mathcal{R}}_k \cap \mathcal{T} = \hat{\mathcal{R}}_{i+1} \cap \mathcal{T} = \mathcal{N}(\hat{\mathbf{R}}_i) \cap \mathcal{T} \subseteq \{\mathcal{N}(\mathbf{F}_i) \cup \mathcal{F}_i\} \cap \mathcal{T} = \emptyset$. Therefore, $\bigcup_{k=1}^{i+1} \hat{\mathcal{R}}_k \subseteq \mathcal{F}$ for all $i = 1, ..., m$. Hence, $P_M(\bigcup_{k=1}^{i} \hat{\mathcal{R}}_k \subseteq \mathcal{F}) \geq P_M(\mathcal{N}(\mathbf{F}_i) \subseteq \mathcal{F}) \geq 1 - \alpha$. By Dominated Convergence Theorem,

$$\begin{aligned} \lim_{i \to \infty} P_M(\bigcup_{k=1}^{i} \hat{\mathcal{R}}_k \subseteq \mathcal{F}) &= P_M(\lim_{i \to \infty} \bigcup_{k=1}^{i} \hat{\mathcal{R}}_k \subseteq \mathcal{F}) \\ &= P_M(\bigcup_{k=1}^{\infty} \hat{\mathcal{R}}_k \subseteq \mathcal{F}) \geq \lim_{i \to \infty} (1 - \alpha) = 1 - \alpha. \end{aligned}$$

□

Remark 6. *The conditionin (2.23) states that all multivariate successor function is monotonic. Since the successor function is defined upon critical value functions, this condition is equivalent to the monotonicity of multivariate critical value function. The condition in (2.24) guarantees the FWER control in the critical case in which all false null hypotheses have been rejected and none of the true ones.*

Remark 7. *Bretz et al. (2009) showed that the graphs together with the updating rules of critical values and transition coefficients are equivalent to a short-cut for a closed test procedure (Hommel, Bretz and Maurer, 2009), thus their graphical approach controls the FWER in the strong sense. In this dissertation, we use an alternative way to show the FWER control for the proposed graphical approach, which is based on the generalized sequential rejection principle. Note that the successor function in the sequential rejection principle (Goeman and Solari, 2010) is a univariate function, as well as the critical value function (by definition of successor function). At each step,* *the local critical value for a hypothesis is defined by a unique critical value function. However, the critical value function in our approach is multivariate. For the original graphical approach that no more than one hypothesis can be rejected at each step, we can apply the sequential rejection principle instead of the closure principle to show its FWER control. For our suggested generalized graphical approach that more than one hypothesis can be rejected at each step or other sequentially rejective procedures that the order of the rejections of null hypotheses in previous steps affects the rejections in the next step, the generalized sequential rejection principle should be applicable to show their FWER control.*

By Theorem 2, we can show that the proposed graphical approach strongly controls the FWER under arbitrary dependence.

Theorem 3 (FWER control of the generalized graphical approach)**.** *The generalized graphical approach strongly controls the FWER at level α under arbitrary dependence.*

For the proof of Theorem 3, see Appendix A.

Even though the sequential rejection principle (Goeman and Solari, 2010) is widely used for developing MTPs with proper control of the FWER, however, for some MTPs, it may not be applicable, such as generalized Bartroff-Lai procedure (Bartroff and Song, 2016). The generalized Bartroff-Lai procedure is a sequential step-down procedure, where the successor/rejection function $\rho(\mathcal{R}, \mathcal{A}, n)$ is trivariate with previous rejection set $\mathcal{R}$, previous acceptance set $\mathcal{A}$ and the sample size of the current testing step n. Clearly, the sequential rejection principle does not work for this procedure. However, applying the proposed generalized sequential rejection principle, we reduce the multivariate successor function to a trivariate function in this case. For each n, as long as the successor function satisfies monotonicity condition and the single step FWER control, the generalized Bartroff-Lai procedure strongly controls the FWER (Bartroff and Song, 2016).

2.5 Clinical Trial Examples

In this section, we apply the proposed graphical approach and the original graphical approach to two different clinical trial examples.

2.5.1 Example 2.1

In this section, we demonstrate the motivating example in Figure 2.1. The overall critical value $\alpha = 0.05$ is allocated to eight tested hypotheses, $\boldsymbol{\alpha} = \{\alpha/2, \alpha/2, 0, 0, 0, 0, 0, 0\}$, and the transition coefficient matrix is

$$\mathbf{G} = \left\{ \begin{array}{cccccccc} 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{array} \right\}.$$

Assume the observed unadjusted p-values $p_1, \ldots, p_8$ are 0.001, 0.002, 0.018, 0.011, 0.009, 0.03, 0.015, 0.021. Applying Algorithm 1, the test is done in the following steps:

Step 1: we test all eight null hypotheses at their local critical values. Since $p_1 = 0.001 < 0.025 = \alpha/2$, $p_2 = 0.002 < 0.025 = \alpha/2$, we reject H_1 and H_2, so $\hat{\mathcal{R}}_1 = \{H_1, H_2\}$. Update the graph (see Figure 2.2) and the critical value functions:

$$\begin{aligned} \alpha_3(\hat{\mathcal{R}}_1) &= 1/2 \times \alpha/2 + 1/2 \times \alpha/2 = \alpha/2, \\ \alpha_4(\hat{\mathcal{R}}_1) &= 1/2 \times \alpha/2 = \alpha/4, \\ \alpha_5(\hat{\mathcal{R}}_1) &= 1/2 \times \alpha/2 = \alpha/4, \end{aligned}$$

while the transition coefficient functions remain the same.

Step 2: we test the remaining six hypotheses. Since $p_3 = 0.018 < \alpha/2$, $p_4 = 0.011 < \alpha/4$, $p_5 = 0.009 < \alpha/4$, we reject hypotheses H_3, H_4 and H_5, so $\hat{\mathcal{R}}_2 = \{H_3, H_4, H_5\}$. Update the graph in Figure 2.3, as well as the critical value functions and transition coefficient functions according to Algorithm 1:

$$\begin{aligned}
\alpha_6(\hat{\mathbf{R}}_2) &= \alpha/2 + 1/2 \times \alpha/4 + 1/2 \times \alpha/4 = 3\alpha/4, \\
\alpha_7(\hat{\mathbf{R}}_2) &= 1/2 \times \alpha/4 = \alpha/8, \\
\alpha_8(\hat{\mathbf{R}}_2) &= 1/2 \times \alpha/4 = \alpha/8, \\
g_{67}(\hat{\mathbf{R}}_2) &= g_{68}(\hat{\mathbf{R}}_2) = 1/2, \\
g_{76}(\hat{\mathbf{R}}_2) &= g_{86}(\hat{\mathbf{R}}_2) = 1/2, \\
g_{78}(\hat{\mathbf{R}}_2) &= g_{87}(\hat{\mathbf{R}}_2) = 1/2.
\end{aligned}$$

Step 3: we test the rest three hypotheses. Since $p_6 = 0.03 < 3\alpha/4$, we reject hypothesis H_6, so $\hat{\mathcal{R}}_3 = \{H_6\}$. Update the graph in Figure 2.4 and the critical value functions and transition coefficient functions:

$$\begin{aligned}
\alpha_7(\hat{\mathbf{R}}_3) &= \alpha_8(\hat{\mathbf{R}}_3) = \alpha/8 + 1/2 \times 3\alpha/4 = \alpha/2, \\
g_{78}(\hat{\mathbf{R}}_3) &= g_{87}(\hat{\mathbf{R}}_3) = 1.
\end{aligned}$$

Step 4: Test H_7 and H_8 at their corresponding updated critical values. Since $p_7 = 0.015 < \alpha/2, p_8 = 0.021 < \alpha/2$, we reject H_7 and H_8, so $\hat{\mathcal{R}}_4 = \{H_7, H_8\}$.

Overall, it takes four steps to complete the whole test, where all hypotheses are rejected. However, if we apply the original graphical approach, it will take eight steps to complete the whole test.

2.5.2 Example 2.2

We next discuss the example depicted in Figure 2.5, in which one aims to test nine hypotheses by using the generalized graphical approach, where H_1, H_2, H_3 are three primary hypotheses, H_4, H_5, H_6 are three secondary hypotheses and H_7, H_8, H_9 are three tertiary hypotheses. The original critical value α is equally distributed to H_1, H_2 and H_3 such that the initial allocation of overall critical value is $\boldsymbol{\alpha} = \{\alpha/3, \alpha/3, \alpha/3, 0, 0, 0, 0, 0, 0\}$, where $\alpha = 0.05$, and the transition coefficient matrix is

$$\mathbf{G} = \left\{ \begin{matrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{matrix} \right\}.$$

Assume the observed unadjusted p-values $p_1, \ldots, p_9$ are 0.008, 0.011, 0.006, 0.014, 0.03, 0.013, 0.015, 0.001, 0.016. Applying Algorithm 1, the test is done in the following steps:

Step 1: we test all nine hypotheses. Since $p_1 = 0.008 < 0.0167 = \alpha/3$, $p_2 = 0.011 < 0.0167 = \alpha/3$, $p_3 = 0.006 < 0.0167 = \alpha/3$, we reject H_1, H_2 and H_3, i.e. $\hat{\mathcal{R}}_1 = \{H_1, H_2, H_3\}$. Update the graph in Figure 2.6 and the critical values function:

$$\alpha_4(\hat{\mathcal{R}}_1) = \alpha_5(\hat{\mathcal{R}}_1) = \alpha_6(\hat{\mathcal{R}}_1) = \alpha/3,$$

while the transition coefficient functions among the remaining hypotheses are still the same.

Step 2: we next test the remaining six hypotheses at their local critical values. Since $p_4 = 0.014 < 0.0167 = \alpha/3, p_6 = 0.013 < 0.0167 = \alpha/3$, we reject H_4 and H_6, i.e. $\hat{\mathcal{R}}_2 = \{H_4, H_6\}$. The updated graph is in Figure 2.7. We update the critical value functions and transition coefficient functions as:

$$\begin{aligned}
\alpha_5(\hat{\mathbf{R}}_2) &= \alpha_7(\hat{\mathbf{R}}_2) = \alpha_8(\hat{\mathbf{R}}_2) = \alpha/3, \\
g_{57}(\hat{\mathbf{R}}_2) &= g_{58}(\hat{\mathbf{R}}_2) = g_{75}(\hat{\mathbf{R}}_2) = g_{78}(\hat{\mathbf{R}}_2) \\
&= g_{79}(\hat{\mathbf{R}}_2) = g_{97}(\hat{\mathbf{R}}_2) = g_{98}(\hat{\mathbf{R}}_2) \\
&= g_{89}(\hat{\mathbf{R}}_2) = 1/2.
\end{aligned}$$

Step 3: we then test H_5, H_7, H_8 and H_9 at their local critical values. Since $p_7 = 0.015 < \alpha_7(\hat{\mathbf{R}}_2), p_8 = 0.001 < \alpha_8(\hat{\mathbf{R}}_2)$, we reject H_7 and H_8, i.e. $\hat{\mathcal{R}}_3 = \{H_7, H_8\}$. The updated graph is in Figure 2.8. Update the critical value function and transition coefficient function of the rest hypotheses as:

$$\begin{aligned}
\alpha_5(\hat{\mathbf{R}}_3) &= 2\alpha/3, \\
\alpha_9(\hat{\mathbf{R}}_3) &= \alpha/3, \\
g_{59}(\hat{\mathbf{R}}_3) &= g_{95}\hat{\mathbf{R}}_3) = 1.
\end{aligned}$$

Step 4: finally we test the rest hypotheses and reject H_5, H_9.

Overall, it takes four steps to test and reject all nine hypotheses by using the proposed graphical approach. However, if we apply the original graphical approach to the same example, it needs nine steps to complete testing.

2.6 Discussion

In this chapter, we introduce a flexible and efficient graphical approach to contruct, visualize and perform sequentially rejective, weighted Bonferroni-based multiple test procedures. Compared to the original graphical approach, the proposed generalized graphical approach is more flexible and efficient. In terms of efficacy, the proposed

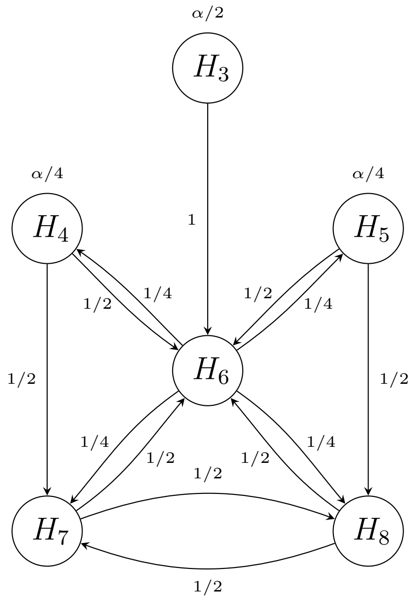


Figure 2.2 Graphical illustration of Example 2.1 at step 2.

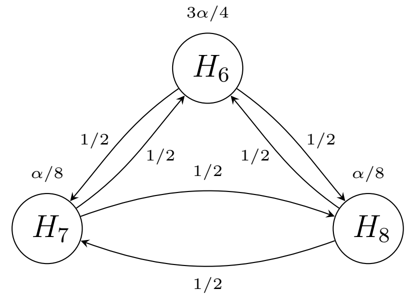


Figure 2.3 Graphical illustration of Example 2.1 at step 3.

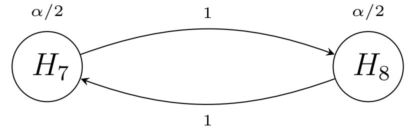


Figure 2.4 Graphical illustration of Example 2.1 at step 4.

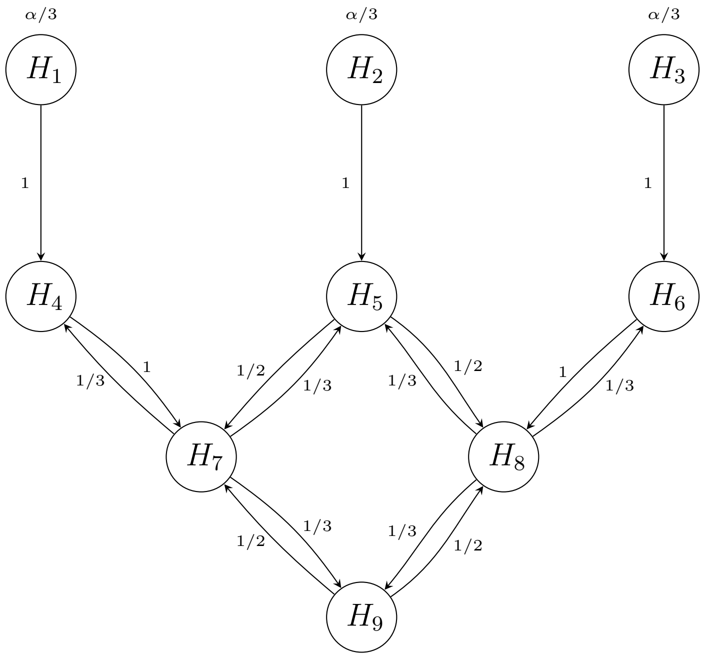


Figure 2.5 Graphical illustration of Example 2.2, with $m = 9$, initial allocation $\boldsymbol{\alpha} = \{\alpha/3, \alpha/3, \alpha/3, 0, 0, 0, 0, 0, 0\}$, and overall level $\alpha = 0.05$.

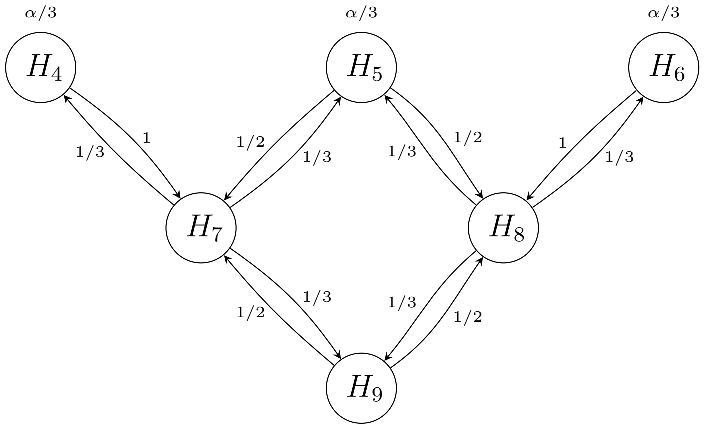


Figure 2.6 Updated graph of Example 2.2 at step 2.

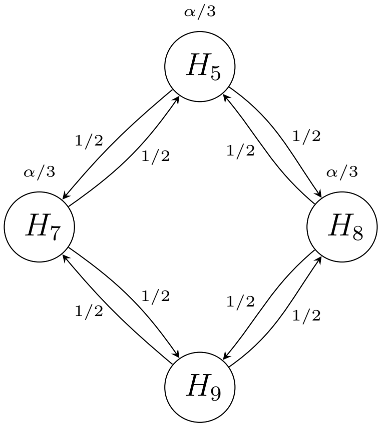


Figure 2.7 Updated graph of Example 2.2 at step 3.

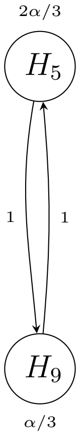


Figure 2.8 Updated graph of Example 2.2 at step 4.

generalized graphical approach is able to reject more than one hypothesis at each step which remarkably streamlines the testing process. Even though the extended original graphical approach can also reject more than one hypothesis at each step, there is no closed form for updating critical value functions and transition coefficient functions once for a rejection set at each step. Instead, one needs to update them for each rejected hypothesis iteratively. Regarding the flexibility, hypotheses with positive local critical values have chances to be rejected no matter wherever they are in the graphs by the proposed approach. Specifically, if we only allow one hypothesis to be rejected at each step, then the proposed graphical approach reduces to the original one. As a by-product, we generalize the sequential rejection principle of Goeman and Solari (2010) from the univariate case to the multivariate case, and develop a generalized sequential rejection principle. By using this principle, we show the FWER control of the generalized graphical approach.

Even though Burman et al. (2009) proposed a recycling-based graphical approach for Bonferroni-based MTPs, which can reject more than one hypothesis at each step, their graphical approach is not as general as our proposed graphical approach. All tested hypotheses are displayed as different sequence according to some clinical relationships among hypotheses in a graph, and only hypotheses located at the top of each sequence have chances to be rejected. However, for our proposed graphical approach, hypotheses with positive local critical values have chances to be rejected no matter wherever they are in the graphs. Moreover, transition coefficients among hypotheses are not as general as they are in our proposed graphical approach. With the increased number of tested hypotheses, it becomes complicated and tedious to diaplay all possible sequences/paths of tested hypotheses.

In the future, we plan to implement the proposed graphical approach in R package, such that users can use the proposed approach more conventionally and quickly.

CHAPTER 3

A GRAPHICAL APPROACH FOR LOGICALLY RELATED MULTIPLE HYPOTHESES TESTING

3.1 Introduction

In this chapter, we focus on developing a new graphical approach for logically related multiple hypotheses testing. With the increasing complexity of study objectives in clinical trials, one arising question is that how to reflect the complex research objectives properly, such as how to formulate multiple structured families of hypotheses for a specific clinical study. Various MTPs, which are used for testing structured families of hypotheses, have been developed with strong control of the FWER. For instance, gatekeeping procedures are developed for testing multiple families of hypotheses with special logical structures (Hommel, Bretz and Maurer, 2007; Guilard, 2007; Dmitrienko, Tamhane and Wiens, 2008; Dmitrienko and Tamhane, 2011; Dmitrienko and Tamhane, 2013). The conventional fixed sequence procedures are further extended to accommodate to logically structured multiple families of hypotheses (Kim, Entsuah and Shults, 2011). However, these MTPs are usually applied to hierarchically ordered families of hypotheses, not applicable for any general logically related hypotheses testing.

As discussed in Chapter 2, Bretz et al. (2009) proposed the original graphical approach to describe MTPs. However, Maurer and Bretz (2013) claimed that this graphical approach can not work efficiently for logically related hypotheses. To solve the issue, they proposed an entangled graphical approach (Maurer and Bretz, 2013), which memorizes all logical relationships among tested hypotheses in the sense that the origin of the propagated critical value is memorized in subsequent tests. In this graphical approach, each logical relationship is taken into account in an individual graph, and hypotheses in each individual graph are tested by using the same algorithm of the original graphical approach (Bretz et al., 2009). Moreover, each individual graph is updated step by step by removing rejected hypotheses. The overall rejections of all individual graphs are the final testing results of the study. It is shown that the proposed entangled graphs are equivalent to the default graphs introduced in Burman et al. (2009), where each logical relationship is illustrated as a sequence of logically related hypotheses in a default graph. Meinshausen (2008) formulated a variable selection problem in high-dimensional regression as a tree-structured hypotheses testing problem and developed a hierarchy method. It can be regarded as a special case of the original graphical approach (Bretz et al., 2009), where descendant hypotheses in the same layer are allocated with equal critical values if their parent hypotheses are rejected, otherwise they are non-testable.

The entangled graphical approach provides one solution to deal with general logical relationships among tested hypotheses. However, even when testing four hypotheses with two constraints (see the case study in Maurer and Bretz, 2013), the entangled graph with both solid and dashed edges at each iteration seems to be complicated and non-intuitive. If more logical relationships among hypotheses there are, one can image how complicated the graphs will be if using the entangled graphical approach. Moreover, for each individual graph, no more than one hypothesis can be rejected at each step, thus it is also not computationally efficient.

In this chapter, the main goal is to develop an efficient and flexible graphical approach for testing any logically related multiple hypotheses, in which all logical relationships can be visualized in one graph. For non-testable hypotheses, which are hypotheses with some unsatisfied logical constraints, we represent them as dashed circles instead of original solid circles in graphs. The dashed circle is changed to a solid one if the logical constraint(s) is satisfied. For example, suppose we are interested in testing eight hypotheses in Figure 3.1, where hypotheses H_4 and H_5 are testable if and only if H_1 and H_2 are rejected, respectively; H_7 and H_8 are testable if and only if H_4 and H_5 are rejected, respectively; and H_6 is testable if and only if H_3 is rejected. Hypotheses $H_4, \ldots, H_8$ are all non-testable at the initial step, thus they are all represented as dashed circles.

In order to make full use of positive critical values assigned to non-testable hypotheses, we propose a re-assignment rule. In Figure 3.1, initially, $\alpha_7 = \alpha_8 = \alpha/4$, these critical values need to be re-assigned to testable hypotheses via edges from non-testable to testable hypotheses, which increase the chance of rejections. However, the re-assigned critical values are temporarily "borrowed" to testable hypothese; after each test step, they need to be returned to their original non-testable hypotheses. Utilizing the re-assignement rule at the beginning of each step, one can take full use of thoses critical values assgined to non-testable hypotheses. We will demonstrate the re-assignment rule in details in Section 3.2.1.

The proposed graphical approach in this chapter is able to reject more than one hypothesis at each step by employing the generalized graphical approach introduced in Chapter 2, thus increases the efficacy of multiple testing strategies. By applying the generalized sequential rejection principle, we can show that the proposed graphical approach strongly controls the FWER at level α. Theoretical details are presented in Section 3.4.

The advantages of the proposed graphical approach for logically related multiple hypotheses testing are: (i). compared to the entangled graphical approach, logical relationships can be presented in an individual graph transparently, instead of several individual graphs, which makes it clear and simple to communicate with clinical teams. Moreover, the proposed re-assignment rule guarantees full use of assigned critical values, especially for critical values of non-testable hypotheses. (ii) compared to the original graphical approach in Bretz et al. (2009), the proposed graphical approach can work efficiently for logically related multiple hypotheses testing. Moreover, more than one hypothesis can be rejected at each step, which

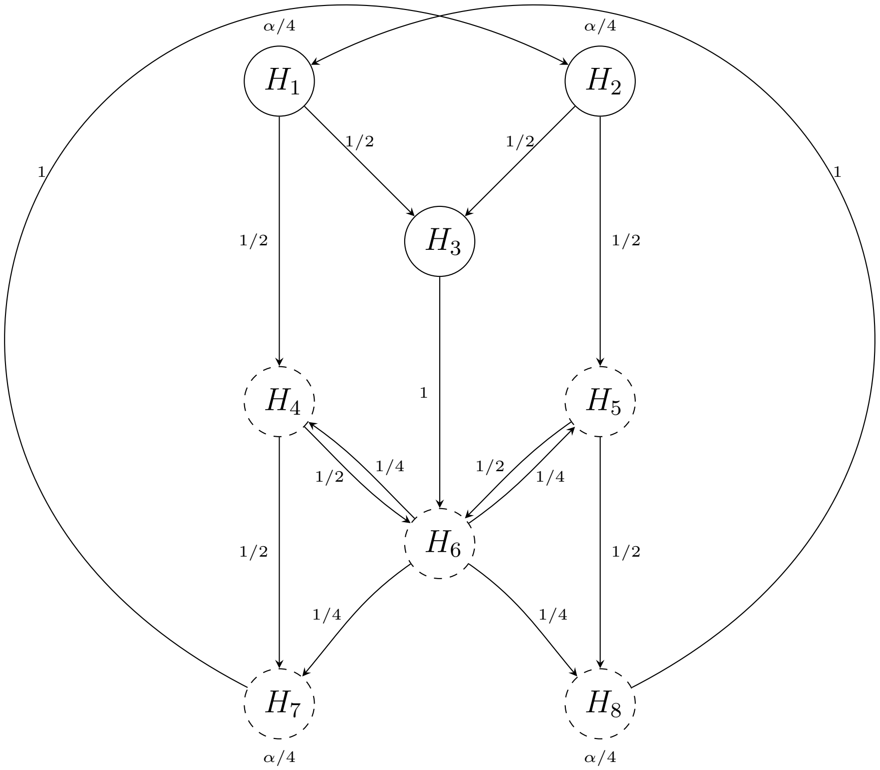


Figure 3.1 Graphical illustration of Example 3.1, with $m = 8$, and initial allocation $\boldsymbol{\alpha} = \{\alpha/4, \alpha/4, 0, 0, 0, 0, \alpha/4, \alpha/4\}$, and overall level $\alpha = 0.05$.

makes it computationally efficient. Furthermore, with reduction of total updated graphs, the graphical presentation to clinical teams is much more explicit. (iii) the proposed graphical approach seperates the logical relationships and the clinical importance of tested hypotheses, which is usually not distinguished in many existing MTPs.

The rest of this chapter is organized as follows: in Section 3.2, we introduce some general notations, assumptations and definitions used in this chapter. The graphical approach for logically related multiple hypothese testing is introduced with a simple and iterative algorithm to perform a MTP in Section 3.3. The FWER control of the proposed approach is showed in Section 3.4. In Section 3.5, we demonstrate clinical trial examples to clearly illustrate the proposed graphical approach. In Section 3.6, conclusion and further discussion are given.

3.2 Preliminaries

In this section, we introduce some general notations, assumptions and definitions used in this chapter. Consider a logically related multiple testing problem with m null hypotheses $H_1, ..., H_m$. Let $\mathcal{H} = \{H_1, ..., H_m\}$ be a collection of m null hypotheses, and $I = \{1, ..., m\}$ be the corresponding indices. Suppose $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)$ be the pre-speficied allocation of the overall critical value α to each individual hypothesis at the beginning of test. We assume all true null p-values are stochastically greater than or equal to uniform distribution on [0,1]. Initially, the critical values $\boldsymbol{\alpha}$ and transition coefficient matrix $\mathbf{G}$=(g_{ij}) are given. All raw p-values are observed, such that $\mathbf{p} = (p_1, ..., p_m)$.

Different from the original graphical approach proposed by Bretz et al. (2009), we employ the graphical approach introduced in Chapter 2, where a multivariate critical value function and transition coefficient function based on a set of rejections are presented. Given a function $\alpha_l(\hat{\mathbf{R}}_{i-1})$ defined on $\hat{\mathbf{R}}_{i-1}$ for hypothesis $H_l, l \in I_i$, at step $i, i = 1, ..., m$, where $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$ is a vector of rejections and $\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1}$ are the nonempty, mutually exclusive sets of rejections at step $1, ..., i-1$, respectively. The transition coefficient at step i is defined by a corresponding transition coefficient function $g_{lk}(\hat{\mathbf{R}}_{i-1})$. Denote $\hat{\mathcal{R}}_0 = \mathcal{N}(\mathcal{R}_0) = \emptyset$, $\alpha_l(\hat{\mathcal{R}}_0) = \alpha_l$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}$. We assume that

$$\sum_{l=1}^{m} \alpha_l \leq \alpha, \tag{3.1}$$

$$0 \leq g_{lk} \leq 1, g_{ll} = 0, \text{ for } l, k = 1, \ldots, m, \tag{3.2}$$

$$\sum_{k=1}^{m} g_{lk} = 1, \text{for every } l = 1, ..., m. \tag{3.3}$$

At the beginning of each step i, we have an updated indices of the remaining hypotheses I_i, and an updated set of the remaining testable hypotheses T_i, where $T_i \subseteq I_i$ for all $i = 1, ..., m$. Initially, we have $I_1 = I = \{1, ..., m\}$. For each $\hat{\mathcal{R}}_i$ at step i, let the corresponding indices be J_i with $J_0 = \emptyset$.

3.2.1 The Re-assignment Rule

In this section, we introduce a re-assignment rule to ensure positive critical values of the non-testable hypotheses are fully used. If at the beginning of any testing step $i, i = 1, ..., m$, for any $f \in I_i \backslash T_i$, we have $\alpha_f(\hat{\mathbf{R}}_{i-1}) > 0$, then apply the re-assignment rule, which is stated as follows.

Definition 1 (Re-assignment Rule (RAR))**.** *If* $\alpha_f(\hat{\mathbf{R}}_{i-1}) > 0$*, for* $f \in I_i \backslash T_i$*, and* $g_{fl}(\hat{\mathbf{R}}_{i-1}) > 0$*, for* $l \in T_i$*, where* $i = 1, ..., m$*, then*

$$b_{fl}(\hat{\mathbf{R}}_{i-1}) = g_{fl}(\hat{\mathbf{R}}_{i-1}) \times \alpha_f(\hat{\mathbf{R}}_{i-1})$$

is re-assigned critical value function to hypothesis H_l *from the non-testable hypothesis* H_f *at the beginning of step* i*. Note that* $b_{fl}(\hat{\mathcal{R}}_0) = b_{fl}$*.*

After re-assignment, the critical values of all non-testable hypotheses become zero. We construct a re-assignment table which illustrates details of overall re-assignment at the beginning of each step. The first big column in the table is non-testable hypotheses with positive critical values, while the second big column in the table is testable hypotheses that "borrowed" critical values from thoses hypotheses listed in the first column. For Example 3.1 displayed in Figure 3.1, at the beginning of step 1, we have the re-assignment table, see Table 3.1.

Table 3.1 Re-assignment Table of Example 3.1 at Step 1.

	$H_{(\text{test})}$	
$H_{(\text{nontest})}$	H_1	H_2
H_7	0	$\alpha/4$
H_8	$\alpha/4$	0

Note: $H_{(\text{test})}$ are testable hypotheses that "borrow" critical values from non-testable hypotheses, while $H_{(\text{nontest})}$ are non-testable hypotheses with positive critical values.

Remark 8. *For simplicity of illustration, we assume there exist direct edges between non-testable and testable hypotheses. In practice, the connection between a non-testable and a testable hypothesis is not always direct, there may have some "bridges", i.e. other non-testable hypotheses, to connect them. In this case, the eventually re-assigned critical value from the original non-testable hypothesis to the testable hypothesis is the product of all associated transition coefficients along this path multiplied by the critical value function of the original non-testable hypothesis. The re-assignment only affects the local critical value of each individual hypothesis, but has no effect on the sum of the critical values over all hypotheses at any step i.*

3.3 The Proposed Graphical Approach

In this section, we present our proposed graphical approach for logically related multiple hypotheses testing. With RAR introduced in Section 3.2.1, using the same multivariate critical value function and transition coefficient function defined in Section 2.3, the following algorithm defines a sequentially rejective procedure for testing logically related hypotheses.

Algorithm 3. *:*

Step 1.

a). If $\alpha_f > 0$, for any $f \in I_1 \backslash T_1$, apply RAR. Fill in the corresponding re-assignment table.

b). For $j \in T_1$, if $p_j \leq \alpha_j + \sum_f b_{fj}$, then reject H_j, thus we have $\hat{\mathcal{R}}_1 = \{H_j \in \mathcal{H} : p_j \leq \alpha_j + \sum_f b_{fj}$, for all $j \in T_1\}$; If $\hat{\mathcal{R}}_1 = \emptyset$, stop testing.

c). Return the "borrowed" critical values to their original hypotheses according to the re-assignment table.

d). Update the graph:

$$I_1 \rightarrow I_2,$$

$$\alpha_l(\hat{\mathcal{R}}_1) = \begin{cases} \alpha_l + \sum\limits_{j \in J_1} \alpha_j g_{jl}, & \textit{if } l \in I_2, \\ 0, & \textit{otherwise}, \end{cases} \tag{3.4}$$

$$g_{lk}(\hat{\mathcal{R}}_1) = \begin{cases} \dfrac{g_{lk} + \sum\limits_{j \in J_1} g_{lj} g_{jk}}{1 - \sum\limits_{j \in J_1} \sum\limits_{q \in J_1} g_{lj} g_{jq} - \sum\limits_{j \in J_1} g_{lj} g_{jl}}, & \textit{if } l, k \in I_2, \textit{ and } l \neq k, \\ 0, & \textit{otherwise}. \end{cases} \tag{3.5}$$

Step $i (i \geq 2)$.

a). For $|T_i| \geq 1$, if $\alpha_f(\hat{\mathbf{R}}_{i-1}) > 0$, for any $f \in I_i \backslash T_i$, apply RAR. Fill in the corresponding re-assignment table.

b). For $j \in T_i$, if $p_j \leq \alpha_j(\hat{\mathbf{R}}_{i-1}) + \sum_f b_{fj}(\hat{\mathbf{R}}_{i-1})$, then reject H_j, thus we have $\hat{\mathcal{R}}_i =$

$\{H_j \in \mathcal{H} : p_j \leq \alpha_j(\hat{\mathbf{R}}_{i-1}) + \sum_f b_{fj}(\hat{\mathbf{R}}_{i-1}),$ *for all* $j \in T_i\}$*; If* $\hat{\mathcal{R}}_i = \emptyset$*, stop testing.*

c). Return the "borrowed" critical values to their original hypotheses according to the re-assignment table.

d). Update the graph:

$$I_i \to I_{i+1},$$

$$\alpha_l(\hat{\mathbf{R}}_i) = \begin{cases} \alpha_l(\hat{\mathbf{R}}_{i-1}) + \sum\limits_{j \in J_i} \alpha_j(\hat{\mathbf{R}}_{i-1}) g_{jl}(\hat{\mathbf{R}}_{i-1}), & \textit{if } l \in I_{i+1}, \\ 0, & \textit{otherwise}, \end{cases} \tag{3.6}$$

$$g_{lk}(\hat{\mathbf{R}}_i) = \begin{cases} \frac{g_{lk}(\hat{\mathbf{R}}_{i-1}) + \sum\limits_{j \in J_i} g_{lj}(\hat{\mathbf{R}}_{i-1}) g_{jk}(\hat{\mathbf{R}}_{i-1})}{1 - \sum\limits_{j \in J_i} \sum\limits_{q \in J_i} g_{lj}(\hat{\mathbf{R}}_{i-1}) g_{jq}(\hat{\mathbf{R}}_{i-1}) - \sum\limits_{j \in J_i} g_{lj}(\hat{\mathbf{R}}_{i-1}) g_{jl}(\hat{\mathbf{R}}_{i-1})}, & \textit{if } l, k \in I_{i+1} \textit{ and } l \neq k, \\ 0, & \textit{otherwise.} \end{cases} \tag{3.7}$$

Logical relationships and clinical importance are usually involved together in a multiple testing problem; by using RAR, one can distinguish these two concepts, and make efficiently use of the critical values that assigned to non-testable hypotheses. At any step of Algorithm 3, the updates of critical value function and transition coefficient function apply for both testable and non-testable hypotheses.

Proposition 2. *Under the assumptions stated in (3.1), (3.2), and (3.3) for initial critical values and transition coefficients, at each step* $i = 1, \ldots, m$*, the critical value function and transition coefficient function defined in Algorithm 3 have the following properties:*

For transition coefficient function,

$$1. \sum_{k \in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1}) = 1, \textit{ for } i = 1, ..., m, \textit{ and } I_i \subseteq \{1, \ldots, m\}, \tag{3.8}$$

$$2.\ 0 \leq g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1, g_{kk}(\hat{\mathbf{R}}_{i-1}) = 0, \textit{ for } i, l, k = 1, ..., m, \tag{3.9}$$

For critical value function,

$$3.\ \sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha,\ \textit{for } i = 1, ..., m,\ \textit{and } I_i \subseteq \{1, \ldots, m\}, \quad (3.10)$$

$$4.\ \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}),\ \textit{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \quad (3.11)$$

$$\textit{and for } i = 1, ..., m.$$

The proof of Proposition 2 is deferred to Appendix B.

Remark 9. *For Algorithm 3, if we encounter the case of the denominators of the right-sides in (3.5) and (3.7) being equal to zero, we always set $g_{lk}(\hat{\mathbf{R}}_i)$ to be zero. At any step, the critical value defined by the critical value function of one specific hypothesis is fully allocated to its related hypotheses via the transition coefficients defined by the transition coefficient function, i.e. $\sum_{k \in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1}) = 1$. This also aims to make full advantage of the initial critical values of logically related hypotheses.*

3.4 Main Theoretical Results

In this section, we will show that the proposed graphical approach strongly controls the FWER at level α. The MTPs generated by this graphical approach with Algorithm 3 are sequentially rejective procedures. Emphasizing on the sequential aspect of this approach, in Theorem 4, we show the FWER control of the proposed approach by utilizing the generalized sequential rejection principle introduced in Chapter 2.

Theorem 4 (FWER control of the proposed graphical approach)**.** *The graphical approach for logically related multiple hypotheses testing strongly controls the FWER at level α under arbitrary dependence.*

For the proof of Theorem 4, see Appendix B.

3.5 Clinical Trial Examples

In this section, we apply the proposed graphical approach to two different clinical trial examples, where there exists some logical relationships among tested hypotheses.

3.5.1 Example 3.1

In this section, we revisit Example 3.1 displayed in Figure 3.1. The overall critical value $\alpha = 0.05$ is allocated to eight hypotheses, $\boldsymbol{\alpha} = \{\alpha/4, \alpha/4, 0, 0, 0, 0, \alpha/4, \alpha/4\}$, and the transition coefficient matrix is

$$\mathbf{G} = \left\{ \begin{array}{cccccccc} 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right\}.$$

Assume the observed unadjusted p-values $p_1, \ldots, p_8$ are 0.001, 0.002, 0.018, 0.011, 0.009, 0.016, 0.015, 0.012. Applying Algorithm 3, the test is done in the following steps:

Step 1:

a). Since the non-testable hypotheses H_7 and H_8 have positive local critical values, we apply the re-assignment rule first. See Table 3.1 for details of re-assignment. Thus the critical values of H_1 and H_2 are changed from $\alpha/4$ to $\alpha/2$.

b). We test all eight hypotheses at their local critical values. Since $p_1 = 0.001 < 0.025 = \alpha/2$, $p_2 = 0.002 < 0.025 = \alpha/2$, we reject H_1 and H_2, i.e. $\hat{\mathcal{R}}_1 = \{H_1, H_2\}$.

c). Before updating the critical value function and transition coefficient function in Algorithm 3, we first return the "borrowed" critical values to non-testable hypotheses H_7 and H_8.

d). The updated graph is displayed in Figure 3.2. We also update the critical value functions and transition coefficient functions as follows:

$$\begin{aligned}
\alpha_3(\hat{\mathcal{R}}_1) &= 1/2 \times \alpha/4 + 1/2 \times \alpha/4 = \alpha/4, \\
\alpha_4(\hat{\mathcal{R}}_1) &= 1/2 \times \alpha/4 = \alpha/8, \\
\alpha_5(\hat{\mathcal{R}}_1) &= 1/2 \times \alpha/4 = \alpha/8, \\
\alpha_7(\hat{\mathcal{R}}_1) &= \alpha_8(\hat{\mathcal{R}}_1) = \alpha/4, \\
g_{73}(\hat{\mathcal{R}}_1) &= g_{75}(\hat{\mathcal{R}}_1) = g_{83}(\hat{\mathcal{R}}_1) = g_{84}(\hat{\mathcal{R}}_1) = 1/2.
\end{aligned}$$

The other transition coefficient functions among the rest hypotheses remain the same. According to the rejection set $\hat{\mathcal{R}}_1$, logical conditions of H_4 and H_5 are fulfilled.

Step 2:

a). Since the re-assignment rule is still applicable, the critical value of non-testable hypotheses H_7 is equally re-assigned to H_3 and H_5 and the critical value of non-testable hypotheses H_8 is equally re-assigned to H_3 and H_4. See Table 3.2 for the re-assignment of critical values at the beginning of step 2. The critical values of H_3, H_4 and H_5 are changed from $\alpha/4, \alpha/8, \alpha/8$ to $\alpha/2, \alpha/4, \alpha/4$, respectively.

b). We test the remaining six hypotheses. Since $p_3 = 0.018 < 0.029 = \alpha/2$, $p_4 = 0.011 < \alpha/4$, $p_5 = 0.009 < \alpha/4$, thus we reject hypotheses H_3, H_4 and H_5, i.e. $\hat{\mathcal{R}}_2 = \{H_3, H_4, H_5\}$.

c). Return the "borrowed" critical values to H_7 and H_8.

d). The updated graph is displayed in Figure 3.3. We also update the critical value functions and transition coefficient functions:

$$\begin{aligned}
\alpha_6(\hat{\mathbf{R}}_2) &= \alpha/4 + 1/2 \times \alpha/8 + 1/2 \times \alpha/8 = 3\alpha/8,\\
\alpha_7(\hat{\mathbf{R}}_2) &= \alpha/4 + 1/2 \times \alpha/8 = 5\alpha/16,\\
\alpha_8(\hat{\mathbf{R}}_2) &= \alpha/4 + 1/2 \times \alpha/8 = 5\alpha/16,\\
g_{67}(\hat{\mathbf{R}}_2) &= g_{68}(\hat{\mathbf{R}}_2) = g_{76}(\hat{\mathbf{R}}_2) = g_{78}(\hat{\mathbf{R}}_2)\\
&= g_{86}(\hat{\mathbf{R}}_2) = g_{87}(\hat{\mathbf{R}}_2) = 1/2.
\end{aligned}$$

According to the rejection set $\hat{\mathcal{R}}_2$, the logical conditions of H_7 and H_8 are fulfilled.

Step 3:

The remaining three hypotheses are all testable. Since $p_6 = 0.016 < 3 \times \alpha/8, p_7 = 0.015 < 5\alpha/16, p_8 = 0.012 < 5\alpha/16$, we reject the remaining hypotheses, so $\hat{\mathcal{R}}_3 = \{H_6, H_7, H_8\}$.

Overall, it takes three steps to complete the whole test, and all eight hypotheses are rejected.

3.5.2 Example 3.2

In this example, we consider simultaneously testing six hypotheses $H_1, \ldots, H_6$, with two logical restrictions, see Figure 3.4. The hypothesis H_5 is testable if and only if H_1 or H_2 is rejected. And hypothesis H_6 is testable if and only if H_3 or H_4 is rejected. The initial critical value is allocated as $\boldsymbol{\alpha} = \{\alpha/4, \alpha/4, 0, 0, \alpha/6, \alpha/3\}$, and overall level $\alpha = 0.05$. We assume the observed raw p-values $p_1, \ldots, p_6$ are 0.001, 0.006, 0.015, 0.01, 0.001, 0.028. Applying Algorithm 3, the test is done in the following steps:

Step 1:

a). Since for non-testable hypotheses H_5 and H_6, their corresponding local critical values $\alpha_5 > 0$ and $\alpha_6 > 0$, thus we apply RAR. The critical value of H_5, α_5

is re-assigned proportionally to H_1 and H_2 based on their transition coefficients g_{51} and g_{52}; Similarly, the critical value of H_6, α_6 is re-assigned proportionally to H_3 and H_4 based on their transition coefficients g_{63} and g_{64}. The re-assignment is illustrated in Table 3.3. After re-assignment, test H_1, H_2, H_3 and H_4 at level $\alpha/4 + 1/3 \times \alpha/6 = 11\alpha/36$, $\alpha/4+1/3\times\alpha/6 = 11\alpha/36$, $1/2\times\alpha/3 = \alpha/6$ and $1/2\times\alpha/3 = \alpha/6$, respectively.

b). Test all six hypotheses. Since $p_1 = 0.001 < 11\alpha/36, p_2 = 0.006 < 11\alpha/36$, we reject H_1 and H_2, i.e. $\hat{\mathcal{R}}_1 = \{H_1, H_2\}$.

c). Return the "borrowed" critical values to H_5 and H_6.

d). The updated graph is displayed in Figure 3.5, where H_5 becomes testable. We also update the graph as well as the critical value functions and transition coefficient functions by Algorithm 3:

$$\begin{aligned}
\alpha_3(\hat{\mathcal{R}}_1) &= \alpha_4(\hat{\mathcal{R}}_1) = 3/4 \times \alpha/4 = 3\alpha/16, \\
\alpha_5(\hat{\mathcal{R}}_1) &= \alpha/6 + 1/4 \times \alpha/4 + 1/4 \times \alpha/4 = 7\alpha/24, \\
\alpha_6(\hat{\mathcal{R}}_1) &= \alpha/3, \\
g_{56}(\hat{\mathcal{R}}_1) &= 2/5, \\
g_{53}(\hat{\mathcal{R}}_1) &= g_{54}(\hat{\mathcal{R}}_1) = 3/10, \\
g_{36}(\hat{\mathcal{R}}_1) &= g_{46}(\hat{\mathcal{R}}_1) = 1, \\
g_{63}(\hat{\mathcal{R}}_1) &= g_{64}(\hat{\mathcal{R}}_1) = 1/2.
\end{aligned}$$

Step 2:

a). Since H_6 is still non-testable, at the beginning of step 2, we apply RAR. Re-assign $\alpha_6(\hat{\mathcal{R}}_1)$ to H_3 and H_4 according to the associated transition coefficients. The details of the re-assignment is listed in Table 3.4.

b). Test H_3, H_4 and H_5 at level $17\alpha/48, 17\alpha/48$ and $7\alpha/24$, respectively. Since $p_3 = 0.015 < 17\alpha/48, p_5 = 0.001 < 7\alpha/24$, we reject H_3 and H_5, i.e $\hat{\mathcal{R}}_2 = \{H_3, H_5\}$.

c). We first return the "borrowed" critical value to H_6.

d). The updated graph is displayed in Figure 3.6, we also update the critical value functions and transition coefficient functions:

$$\begin{aligned}
\alpha_4(\hat{\mathbf{R}}_2) &= 3\alpha/16 + 3/10 \times 7\alpha/24 = 11\alpha/40, \\
\alpha_6(\hat{\mathbf{R}}_2) &= \alpha/3 + 3\alpha/16 + 2/5 \times 7\alpha/24 = 17\alpha/40, \\
g_{46}(\hat{\mathbf{R}}_2) &= g_{64}(\hat{\mathbf{R}}_2) = 1.
\end{aligned}$$

Step 3:

Since the remaining two hypotheses are both testable, we test them at their local critical values. Since $p_4 = 0.01 < 11\alpha/40$, we reject H_4, i.e $\hat{\mathcal{R}}_3 = \{H_4\}$.

Step 4:

We further test H_6 at level $28\alpha/40$. Since $p_6 = 0.028 < 28\alpha/40$, we finally reject H_6.

Overall, it takes four steps to reject all six tested hypotheses.

Table 3.2 Re-assignment Table of Example 3.1 at Step 2.

	$H_{(\text{test})}$		
$H_{(\text{nontest})}$	H_3	H_4	H_5
H_7	$\alpha/8$	0	$\alpha/8$
H_8	$\alpha/8$	$\alpha/8$	0

3.6 Discussion

In this chapter, we propose a flexible and efficient graphical approach for logically related multiple hypotheses testing, which is a further development of the generalized graphical approach in Chapter 2. In this graphical approach, we introduce different graphical representations of hypotheses with and without logical restrictions in one

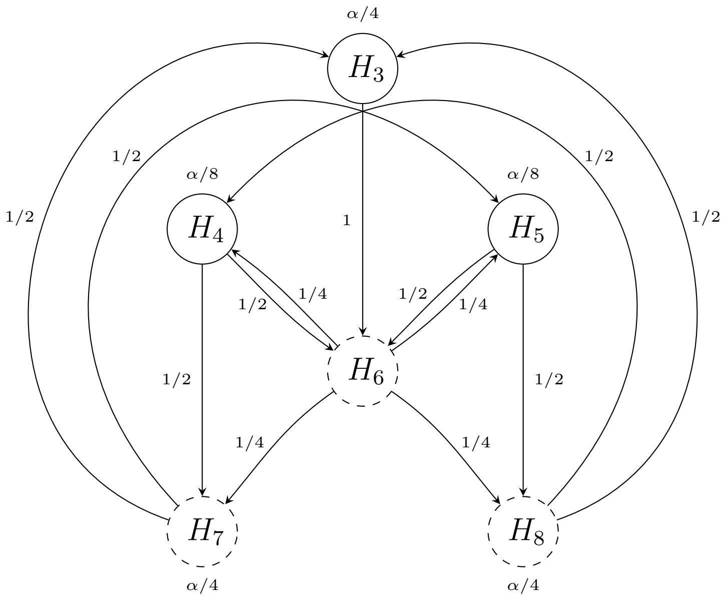


Figure 3.2 Graphical illustration of Example 3.1 at step 2.

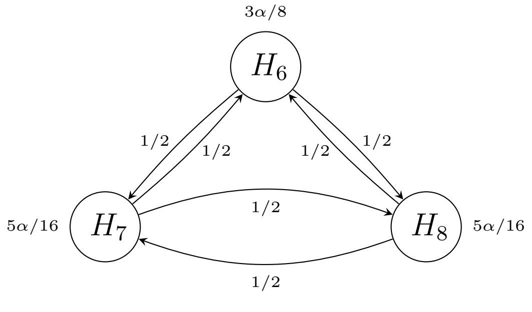


Figure 3.3 Graphical illustration of Example 3.1 at step 3.

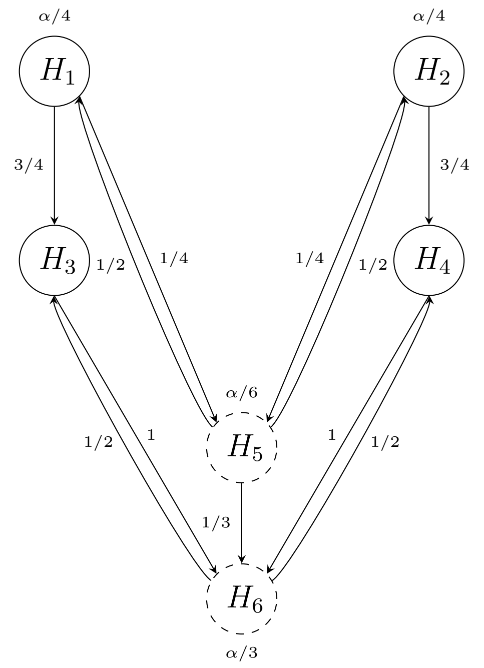


Figure 3.4 Graphical illustration of Example 3.2, with $m = 6$ and initial allocation $\boldsymbol{\alpha} = \{\alpha/4, \alpha/4, 0, 0, \alpha/6, \alpha/3\}$.

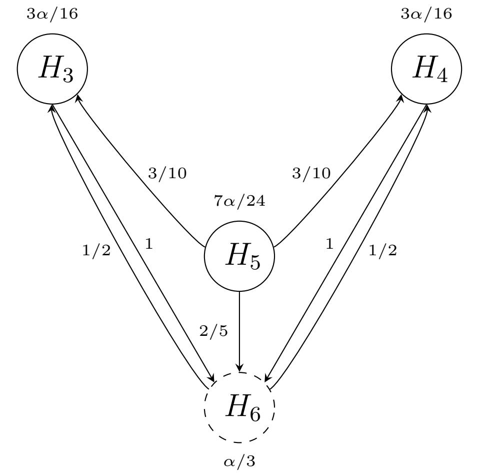


Figure 3.5 Updated graph of Example 3.2 at step 2.

Table 3.3 Re-assignment Table of Example 3.2 at Step 1.

	$H_{(\text{test})}$			
$H_{(\text{nontest})}$	H_1	H_2	H_3	H_4
H_5	$\alpha/18$	$\alpha/18$	0	0
H_6	0	0	$\alpha/6$	$\alpha/6$

Table 3.4 Re-assignment Table of Example 3.2 at Step 2.

	$H_{(\text{test})}$	
$H_{(\text{nontest})}$	H_3	H_4
H_6	$\alpha/6$	$\alpha/6$

graphical display. A re-assignment rule is included in the graph such that one can make full use of the assigned critical values from non-testable hypotheses. Also, the allocation of the overall critical value is based on the clinical importance of hypotheses, which is independent of logical relationships among tested hypotheses. Compared with the original graphical approach, the proposed graphical approach has at least two advantages. First, the proposed graphical approach can be used to efficiently perform MTPs for any general logically related hypotheses. Second, the proposed graphical approach can reject more than one hypothesis at each step, which is computationally efficient. Compared with the entangled graphical approach, the proposed graphical

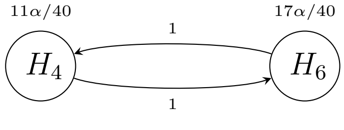


Figure 3.6 Updated graph of Example 3.2 after step 3.

approach can display and perform a MTP in one graph at each step, which is more simple and intuitive to present to clinical teams. For future research, a main work is to implement the proposed graphical approach in R package, which can help users to conduct logically related hypotheses testing more conventionally and quickly.

CHAPTER 4

SELECTIVE INFERENCE PROCEDURES IN LARGE SCALE HYPOTHESES TESTING

4.1 Introduction

In this chapter, we focus on developing powerful selective inference methods in large scale multiple hypotheses testing. When testing a large number of hypotheses simultaneously, one natural testing strategy is to first reduce the number of tested hypotheses by some selection process, and then to simultaneously test the selected hypotheses. The methods developed based on this idea is named as a two-stage procedure or filtering method or selective inference procedure(Benjamin and Yekutieli, 2005; Fithian et al., 2015; Barber and Candes, 2015; Benjamini and Bogomolov, 2014; Heller et al., 2016; etc). When selection effects are taken into account, conditional inference is often used. Therefore, instead of marginal p-values, conditional p-values should be used to measure statistical significance of every selected hypothesis. However, the conditional p-values are often difficult to derive. Whether a selective inference procedure can perform well depends on its conditional p-values have good statistical properties, which in turn depends on appropriate choices of selection and testing statistics and the selection threshold.

Independent filtering methods (Bourgon et al., 2010, Dai et al., 2012, Du and Zhang, 2014, Ignatiadis et al., 2016) were proposed such that the selection and testing statistics are chosen to be independent when the corresponding null hypothesis is true. Instead of using conditional p-values, marginal p-values can be used for independent filtering methods. However, several questions arise about how to best apply the independent filtering methods: How to choose the selection threshold? How to deal with the non-selected hypotheses? How to develop powerful procedures such that filtering information is beneficial?

The existing independent filtering methods can only deal with selected hypotheses. And even for selected hypotheses, the information is not fully exploited. Under the assumptation of independence, we propose three powerful two-stage MTPs, which can not only exploit information from selected hypotheses more explicitly by estimating the true null proportion, but also exploit information from the non-selected hypotheses.

By employing the idea of adaptive procedures, we first propose an adaptive two-stage Bonferroni procedure, where the proportion of true nulls among selected hypotheses is estimated from the data and the estimate is incorporated into two-stage Bonferroni procedure. The adaptive two-stage Bonferroni procedure is generally more powerful than the corresponding two-stage Bonferroni procedure introduced in Guo and Romano (2017). A simple selection rule is used for selecting which hypotheses H_i are to be tested at the second stage. Given a fixed threshold t, H_i is selected iff $U_i \geq t$, where U_i is the selection statistics. Let $S = \{i \in \{1, \ldots, m\} : U_i \geq t\}$ denote the indices of selected hypotheses, with $|S|$ be the number of selected hypotheses. We estimated true null proportion within S, which is denoted by $\hat{\pi}_S$. At the second stage, we apply an adaptive Bonferroni procedure on the selected hypotheses S for which hypothesis H_i is rejected iff the corresponding p-value $P_i \leq \frac{\alpha}{\hat{\pi}_S |S|}$.

The existing filtering methods only exploit the information contained in the selected hypotheses, which means the information contained in non-selected hypotheses are discarded. However, the non-selected hypotheses may have some useful information. By a selection process, we split hypotheses into two hierarchically ordered blocks/families, F_1 and F_2. Based on the hierarchy order of two blocks, we apply the idea of parallel gatekeeping strategy (Dmitrienko, Offen and Westfall, 2003) and develop a selective parallel gatekeeping procedure. While based on the importance of two blocks, we apply the idea of data-driven weights (Fino and Salmaso, 2007; Roeder and Wasserman, 2009; Poisson et al., 2012; Ignatiadis et al., 2016; Li and Barber, 2016; Lei and Fithian, 2016; Ignatiadis and Huber, 2017) and develop a data-driven weighted selective procedure.

The main idea of selective parallel gatekeeping procedure is that critical values of the current family can be passed down to the subsequent families if at least one significant result is obtained in the current family. Apply the conventional Bonferroni procedure to test hypotheses of F_1 at level α, and let $|R_1|$ denote the number of rejected hypotheses in F_1, then a fraction of the critical value of F_1, $\frac{\alpha}{|S|}|R_1|$ will be passed to F_2 to further exploit significant results within F_2.

The main idea of data-driven weighted selective procedure is that different weights W_i are assigned to different families $F_i, i = 1, 2$ to measure their importance. For construction of data-driven weights, our basic idea is that a family with higher proportion of false nulls is assigned with a higher weight, such that we have higher chance to obtain significant results. Moreover, weights can be further improved by incorporating into the information of proportions of true nulls.

The proposed two-stage MTPs divide hypotheses into two blocks, in order to exploit information from each null hypothesis more explicitly, we generalize them from two blocks to multiple blocks. We first propose a blockwise adaptive two-stage Bonferroni procedure. Inspired by Kim and Schliekelman (2015), tested hypotheses are ordered according to the selection/filtering staitstics U_i, then are divided into K (pre-defined) blocks. We then estimate true null proportion $\hat{\pi}_{B_l}$ for B_l, where $B_l, l = 1, \ldots, K$ is the index set of hypotheses in such block. Within each block, we apply adaptive Bonferroni procedure, for which hypothesis H_i is rejected iff the corresponding p-value $P_i \leq \frac{\alpha}{\hat{\pi}_{B_l}|B_l|K}$.

When we have more than two hierarchy ordered blocks, a blockwise selective parallel garekeeping procedure is naturally developed. The basic idea is that for ordered null hypotheses, we assign different weight w_l for $B_l = 1, \ldots, K-1$, where $w_l = \frac{\sum_{j=1}^{K-1}|B_j|}{\lambda \sum_{j=1}^{l}|B_j|}$. λ is a correction factor to ensure the FWER control of the proposed procedure. Such weight construction is followed in Kim and Schliekelman (2015). We have $w_1 > \ldots > w_K = 0$. We then apply weighted Bonferroni procedure from B_1 to B_K. The critical values are also allocated in this order if any rejections made, i.e. $\alpha_{l+1} = w_{l+1}\alpha + \frac{\alpha_l |R_l|}{|B_l|}$, where R_l is denoted as a collection of rejections made within B_l. Thus hypotheses in lower weighted block may have a chance to be rejected with such allocation of critical values.

When we have more than two blocks, based on their importance, it is natural to propose a blockwise data-driven weighted selective procedure, in which different weights W_l are assigned to different $B_l = 1, \ldots, K$. For construction of data-driven weights, our basic idea is similar as the data-driven weighted selective procedure. After ordering null hypotheses according to the selection statistics U_i, a block with higher proportion of false nulls is assigned with a higher weight, such that we have higher chance to obtain significant results. Moreover, weights can be further improved by incorporating into the information of proportions of true nulls.

The rest of the chapter is organized as follows: in Section 4.2, we formulate our problem. In Section 4.3, we introduce three different selective procedures as well as their theoretical results on FWER control. In Section 4.4, we further generalize the proposed three selective inference MTPs described in Section 4.3. In Section 4.5, a concluding remark is given.

4.2 Problem Formulation

In this section, we introduce the multiple testing problem that we focus on throughout this chapter. Assume that for $i = 1, \ldots, m$, a sample of size n_i from a normal population with unknown mean μ_i and variance $\sigma_i^2 = 1$. The data

$$X_{ij} \overset{iid}{\sim} N(\mu_i, \sigma_i^2), \text{ for } j = 1, \ldots, n_i.$$

The m samples are assumed to be mutually independent. For simplicity, we assume $n_i = n$. For $i = 1, \ldots, m$, consider testing hypotheses

$$H_i : \mu_i = 0 \;\; vs. \;\; H_i^{'} : \mu_i \neq 0.$$

Let $I = \{1, \ldots, m\}$ be the index set of the tested hypotheses. And let U_i and T_i denote the the selection and testing statistics, respectively, which are defined as follows:

$$\begin{aligned} U_i &= \sum_{j=1}^{n_i} X_{ij}^2, & (4.1) \\ T_i &= \frac{\overline{X_i}}{\hat{\sigma}_i/\sqrt{n_i}}, & (4.2) \end{aligned}$$

where $\overline{X}_i$ and $\hat{\sigma}_i^2$ are respectively the sample mean and (unbiased) sample variance for the ith sample, i.e., $\overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{i,j}$ and $\hat{\sigma}_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_{i,j} - \overline{X}_i)^2$. The statistics U_i is first used to "select" which of the hypotheses to "test" in the second stage, at which point the statistics T_i is used. It is showed that under true null hypothesis H_i, U_i follows χ_n^2 and T_i follows t_{n-1}, and U_i is independent of T_i (Lehmann and Romano, 2005).

For each $H_i, i = 1, \ldots, m$, we calculate the p-value P_i based on T_i such that

$$P_i = \Pr_{H_i}\{|T_i| \geq t_i\},$$

where t_i is the observed value of the T-statistics T_i. Let $P_{(1)}, \ldots, P_{(m)}$ be the ordered values of $P_1, \ldots, P_m$ and $H_{(1)}, \ldots, H_{(m)}$ be the corresponding null hypotheses.

4.3 Proposed Two-stage Selective Inference MTPs

In order to improve existing two-stage MTPs, we propose three different MTPs in this section. We first develop an adaptive two-stage Bonferroni procedure by incorporating into the information of the proportion of true nulls among selected hypotheses. We then take non-selected hypotheses into account as well, by using a parallel gatekeeping strategy, where selected hypotheses serve as a gatekeeper for non-selected hypotheses, we develop a selective parallel gatekeeping procedure to exploit more information among non-selected hypotheses. Finally, we develop a data-driven weighted selective procedure by combining ideas of weighted procedures and adaptive procedures.

4.3.1 Adaptive Two-stage Bonferroni Procedure

In order to improve existing two-stage Bonferroni procedures, we employ the adaptive Bonferroni procedure among selected hypotheses. Adaptive Bonferroni procedures estimate the proportion of true null hypotheses and then incorporate it into the conventional Bonferroni procedure to derive more powerful testing procedures. Similarly, by incorporating into the information of the proportion of true nulls among selected hypotheses S, we can develop more powerful MTPs than exisitng two-stage MTPs. The key point is how to estimate the true null proportion among S. Several estimators of true null proportion have been introduced in the literature (Schweder and Spjotvoll, 1982; Benjamini and Hochberg, 2000; Storey et al., 2004; Benjamini, Krieger and Yekutieli, 2006; Meinshausen and Rice, 2006; Sarkar, 2008; Blanchard and Roquain, 2009; Sarkar, Guo and Finner, 2012). We use widely-used Storey-type estimator for the proposed adaptive two-stage Bonferroni procedure in this chapter:

$$\hat{\pi}_S = \frac{Q_S(\lambda)}{(1-\lambda)|S|}, \tag{4.3}$$

where λ is a fixed constant with $0 < \lambda < 1$, $|S|$ is the number of selected hypotheses, and $Q_S(\lambda) = \sum_{k \in S} I_{\{P_k > \lambda\}}$ is the number of selected hypotheses such that the corresponding p-values exceed λ.

Based on the pre-defined selection threshold and the conservative Stroey-type estimator $\hat{\pi}_S$, an adaptive two-stage Bonferroni procedure is defined as follows:

Definition 2 (Adaptive two-stage Bonferroni procedure)**.**

1. For a fixed selection threshold t, we have $S = \{i \in \{1, \ldots, m\} : U_i > t\}$.

2. For a fixed $\lambda \in (0, 1)$, calculate $Q_S(\lambda) = \sum_{i \in S} I_{\{P_i > \lambda\}}$, and then calculate $\hat{\pi}_S$ based on (4.3).

3. Reject $H_{(1)}, \ldots, H_{(r)}$, where:

$$r = max\{i = 1, \ldots, |S| : P_{(i)} \leq \frac{\alpha}{\hat{\pi}_S |S|}\}.$$

No rejection if such maximum does no exist.

Before we show the FWER control of the proposed procedure, we first introduce the following lemma (Benjamini et al., 2006):

Lemma 1. *If $Y \sim Bin(N, p)$, then $E\{(Y+1)^{-1}\} < \{(N+1)p\}^{-1}$.*

Let S_0 be the index set of true null hypotheses among S, and $Q_{S_0}(\lambda) = \sum_{i \in S_0} I_{\{P_i > \lambda\}}$ be the number of true null hypotheses among S such that the corresponding p-values exceed λ. Note that $\Pr_{H_{i0}}\{P_i > \lambda\} = 1 - \lambda$, by the independence of true null p-values, we have

$$Q_{S_0}(\lambda) \sim \text{Bin}(|S_0|, 1 - \lambda).$$

By Lemma 1, we have the following result.

Theorem 5. *The above defined adaptive two-stage Bonferroni procedure strongly controls the FWER at level α under the assumption that pairs of (U_i, T_i) are independent.*

Proof. For given selection statistics $U_i, i = 1, \ldots, m$, we have $S = \{i \in \{1, \ldots, m\} : U_i > t\}$, which is the index set of the selected hypotheses. We denote cFWER as the conditional familywise error rate on selection such that

$$\text{cFWER} = \Pr\left(\text{reject at least one true null } | U_1, \ldots, U_m\right).$$

Let S_0 be the index set of true null hypotheses among S, and let $Q_S^{(-i)}(\lambda) = \sum\limits_{k \in S \setminus \{i\}} I_{\{P_k > \lambda\}}$, and $Q_{S_0}^{(-i)}(\lambda) = \sum\limits_{k \in S_0 \setminus \{i\}} I_{\{P_k > \lambda\}}$. Similar to (4.3), define $\hat{\pi}_S^{(-i)} = \frac{Q_S^{(-i)}(\lambda)}{(1-\lambda)|S|}$. When H_i is true, by independence of true null p-values, we have

$$Q_{S_0}^{(-i)}(\lambda) \sim \text{Bin}(|S_0| - 1, 1 - \lambda). \tag{4.4}$$

We first consider the cFWER,

$$\begin{aligned}
\text{cFWER} &= \Pr\left(\bigcup_{i \in S_0} \{P_i \leq \frac{\alpha}{\hat{\pi}_S |S|}\} | U_1, \ldots, U_m\right) \\
&\leq \sum_{i \in S_0} \Pr\left(P_i \leq \frac{\alpha}{\hat{\pi}_S |S|} | U_1, \ldots, U_m\right) \\
&\leq \sum_{i \in S_0} \Pr\left(P_i \leq \frac{\alpha}{\hat{\pi}_S^{(-i)} |S|} | U_1, \ldots, U_m\right) \\
&\leq \sum_{i \in S_0} E\left(\frac{\alpha}{\hat{\pi}_S^{(-i)} |S|} | U_1, \ldots, U_m\right) \\
&= \sum_{i \in S_0} E\left(\frac{\alpha(1-\lambda)}{Q_S^{(-i)}(\lambda) + 1} | U_1, \ldots, U_m\right) \\
&\leq \sum_{i \in S_0} E\left(\frac{\alpha(1-\lambda)}{Q_{S_0}^{(-i)}(\lambda) + 1} | U_1, \ldots, U_m\right) \\
&= \alpha(1-\lambda) \sum_{i \in S_0} E\left(\frac{1}{Q_{S_0}^{(-i)}(\lambda) + 1} | U_1, \ldots, U_m\right) \\
&< \alpha(1-\lambda) \sum_{i \in S_0} \frac{1}{(|S_0|)(1-\lambda)} = \alpha \frac{|S_0|}{|S_0|} = \alpha.
\end{aligned} \tag{4.5}$$

where the first inequality in (4.5) is due to Bonferroni inequality. The second inequality holds since $Q_S^{(-i)}(\lambda) \leq Q_S(\lambda)$, according to (4.3), $\hat{\pi}_S^{(-i)} \leq \hat{\pi}_S$. The third inequality is because of pariwise independence of (U_i, T_i), independence of U_i and T_i under H_i, and independence of P_i and $\hat{\pi}_S^{(-i)}$. The fourth inequality holds since $Q_{S_0}^{(-i)}(\lambda) \leq Q_S^{(-i)}(\lambda)$. Due to Lemma 1 and (4.4), we have the inequality in (4.5).

Therefore, FWER $= E\,(\text{cFWER}) \leq \alpha$. The desire result is proved. □

4.3.2 Selective Parallel Gatekeeping Procedure

In this section, we introduce a selective parallel gatekeeping procedure such that information of non-selected hypotheses is further exploited. Basically, tested hypotheses are divided into two families F_1 and F_2 by a selection process, such that

$$F_1 = \{H_i : i \in S\} \text{ and } F_2 = \mathcal{H} \backslash F_1.$$

Using the parallel gatekeeping strategy, F_1 serves as a gatekeeper of F_2. And within F_1, we apply the conventional Bonferroni procedure. Based on the testing results of F_1, the critical values of rejected hypotheses are passed to F_2 to further test hypotheses within F_2.

A selective parallel gatekeeping procedure is defined as follows:

Definition 3 (Selective parallel gatekeeping procedure)**.**

1. Based on a fixed selection threshold t, we have $S = \{i \in \{1, \ldots, m\} : U_i > t\}$. Construct $F_1 = \{H_i : i \in S\}$ and $F_2 = \mathcal{H} \backslash F_1$.

2. Simultaneously test all hypotheses in F_1 using the conventional Bonferroni procedure at level $\alpha_1 = \alpha$. Let $\mathcal{R}_1$ denote the rejection set, such that

$$\mathcal{R}_1 = \{H_i \in F_1 : P_i \leq \frac{\alpha}{|S|}\}.$$

3. If $|R_1| = 0$, stop test. Otherwise, update the critical value of F_2, such that

$$\alpha_2 = \frac{\alpha |R_1|}{|S|}.$$

Simultaneously test all hypotheses in F_2 using any FWER controlling procedure at level α_2.

Remark 10. *Consider using the multistage gatekeeping procedure introduced in Dmitrienko, Tamhane and Wiens (2008) for sequentially testing families F_1 and F_2.*

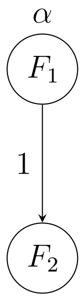


Figure 4.1 Graphical illustration of selective parallel gatekeeping procedure.

If applying the conventional Bonferroni procedure to test the hypotheses within F_1 *at level* α*, then hypotheses within* F_2 *can be tested at level* $\alpha_2 = \alpha - e^*(A_1) = \frac{\alpha|R_1|}{|S|}$*, where* A_1 *is the acceptance set within* F_1 *and* $e^*(A_1) = \frac{\alpha|A_1|}{|S|}$ *is the upper bound of the error function of the conventional Bonferroni procedure.*

The graphical illustration of such procedure is shown in Figure 4.1. For the proposed procedure, we have the following result.

Theorem 6. *The above defined selective parallel gatekeeping procedure strongly controls the FWER at level* α *under the assumption that pairs of* (U_i, T_i) *are independent.*

Proof. For given selection statistics $U_i, i = 1, \ldots, m$, we have $S = \{i \in \{1, \ldots, m\} : U_i > t\}$, which is the index set of the selected hypotheses. Then $F_1 = \{H_i : i \in S\}$ and $F_2 = \mathcal{H} \backslash F_1$ are defined. Following closely the proof in Qiu et al. (2017), we first define an event

$$E_i(x) = \{\text{at least one false rejection in } F_i \text{ at level } x\}, i = 1, 2.$$

Then the FWER can be expressed as follows:

$$\begin{aligned} \text{cFWER} &= \Pr\big(E_1(\alpha_1) \cup E_2(\alpha_2) | U_1 \ldots, U_m\big) \\ &= \Pr\big(E_1(\alpha_1) | U_1, \ldots, U_m\big) + \Pr\big(\bar{E}_1(\alpha_1) \cap E_2(\alpha_2) | U_1, \ldots, U_m\big). \quad (4.6) \end{aligned}$$

Let $e^*(\cdot)$ as an upper bound of error rate function $e(\cdot)$. By the definition of the error rate function in (1.8), we know that

$$\Pr\big(E_1(\alpha_1)|U_1,\ldots,U_m\big) \le e(T_1) \le e^*(T_1),$$

where T_1 is the set of true null hypotheses in F_1.

Let A_1 denote as the acceptance set within F_1. For the second term in (4.6), if $\bar{E}_1(\alpha_1)$ is realized, i.e., all rejected hypotheses in F_1 are false, then $T_1 \subseteq A_1$. By monotonicity of $e^*(\cdot)$ (Dmitrienko, Tamhane and Wiens, 2008), we have $e^*(T_1) \le e^*(A_1)$ for $T_1 \subseteq A_1$. Therefore,

$$\alpha_2 = \alpha_1 - e^*(A_1) \le \alpha_1 - e^*(T_1),$$

which further implies

$$\bar{E}_1(\alpha_1) \cap E_2(\alpha_2) \subseteq E_2(\alpha_1 - e^*(T_1)).$$

Thus, by (4.6), we have

$$\begin{aligned} \text{cFWER} &\le e^*(T_1) + \Pr\big(E_2(\alpha_1 - e^*(T_1))|U_1,\ldots,U_m\big) \\ &\le e^*(T_1) + \alpha_1 - e^*(T_1) = \alpha_1 = \alpha. \end{aligned} \tag{4.7}$$

The last inequality of (4.7) holds since $\Pr\big(E_2(\alpha_1 - e^*(T_1))|U_1,\ldots,U_m\big) \le \alpha_1 - e^*(T_1)$. Therefore, FWER $= E\,(\text{cFWER}) \le \alpha$. $\square$

4.3.3 Data-driven Weighted Selective Procedure

The aforementioned two families F_1 and F_2 are determined by a selection process, with corresponding index set S and $I\backslash S$, respectively. Note that there are different proportions of true and false nulls between S and $I\backslash S$, we consider to construct different weights by using the proportion information of S and $I\backslash S$ and develop a new weighted selective MTP. In most cases, we do not have prior knowledge about the proportions and the resulting weights. Therefore, we plan to develop a data-driven weighted selective MTP by estimating the proportions of true and false nulls. For related recent works, see Ignatiadis et al. (2016) and Ignatiadis and Huber (2017). They developed alternative data-driven weighted methods under independent filtering.

Ideal case. Suppose we know the true and false null proportions among selected hypotheses S, which are denoted as $\pi_0(S)$ and $\pi_1(S)$, and those among non-selected hypotheses $I\backslash S$, which are denoted as $\pi_0(I\backslash S)$ and $\pi_1(I\backslash S)$. Since the higher the false null proportion within a family, the more significant results can be obtained; we then assign a higher weight to that family. Therefore, the weights W_1' and W_2' for S and $I\backslash S$ are constructed as

$$W_1' = \frac{\pi_1(S)}{\pi_1(S) + \pi_1(I\backslash S)} \text{ and } W_2' = \frac{\pi_1(I\backslash S)}{\pi_1(S) + \pi_1(I\backslash S)}, \tag{4.8}$$

such that $W_1' + W_2' = 1$. Then, the same weight is assigned to each hypothesis within a family. That is, hypotheses within S are assigned with the same weight W_1' and hypotheses within $I\backslash S$ are assigned with the same weight W_2'. In order to ensure the sum of the weights of all tested hypotheses is equal to m, we modify W_1' and W_2' as follows:

$$W_1'' = \frac{mW_1'}{|S|} \text{ and } W_2'' = \frac{mW_2'}{m - |S|}, \tag{4.9}$$

such that

$$\sum_{i \in S} W_1'' + \sum_{j \in I\backslash S} W_2'' = m.$$

So far the weights are constructed only based on the proportions of false nulls within S and $I\backslash S$. By employing the idea of adaptive procedures, we can further improve the weights by incorporating into the information of the proportions of true nulls. Therefore, our final weights for S and $I\backslash S$ are constructed as follows:

$$W_1 = \frac{W_1''}{\pi_0(S)} = \frac{m\pi_1(S)}{\pi_0(S)(\pi_1(S) + \pi_1(I\backslash S))|S|} \tag{4.10}$$

and

$$W_2 = \frac{W_2''}{\pi_0(I\backslash S)} = \frac{m\pi_1(I\backslash S)}{\pi_0(I\backslash S)(\pi_1(S) + \pi_1(I\backslash S))(m - |S|)}. \tag{4.11}$$

And for each hypothesis H_i, the weight is assigned as follows:

$$\tilde{W}_i = \begin{cases} W_1, & if\ i \in S, \\ W_2, & if\ i \in I\backslash S. \end{cases}$$

Finally, we can construct weighted procedures based on the above weights for testing all m hypotheses simultaneously.

We need to point out that the above weights are derived under the ideal case of known proportions of true and false nulls. However, in practice, $\pi_0(S)$, $\pi_1(S)$, $\pi_0(I\backslash S)$ and $\pi_1(I\backslash S)$ are often unknown. With a data at hand, several methods were proposed to estimate true null proportions in the literature (Schweder and Spjotvoll, 1982; Benjamini and Hochberg, 2000; Storey et al., 2004; Benjamini, Krieger and Yekutieli, 2006; Meinshausen and Rice, 2006; Sarkar, 2008; Blanchard and Roquain, 2009; Sarkar, Guo and Finner, 2012), which in turn, can also estimate false null proportions as well. Then, we can use the estimated true/false null proportions to replace the true/false null proportions in (4.10) and (4.11) to construct data-driven weights. In the following, we introduce methods to estimate the true/false null proportions.

Estimation of True/False Null Proportions As we can see in (4.10), W_1 is constructed by two parts, one is W_1'', the other is $\pi_0(S)$; moreover, W_1'' is constructed by false null proportions $\pi_1(S)$ and $\pi_1(I\backslash S)$. In order to ensure independence between such two parts, we estimate false null proportions $\pi_1(S)$ and $\pi_1(I\backslash S)$, which are contributed to construct W_1'', based on selection statistics U_i, and $\pi_0(S)$, which is contributed to constructed W_1 in (4.10), based on testing statistics T_i. The same argument holds for constructing W_2 in (4.11).

Based on selection statistics U_i and testing statistics T_i, we can calculate their respective p-values with a given data. Let $\tilde{P}_i$ be a p-value of hypothesis H_i based on U_i, and let P_i be a p-value of hypothesis H_i based on T_i:

$$\tilde{P}_i = \Pr_{H_i}\{U_i \geq u_0\} \text{ and } P_i = \Pr_{H_i}\{|T_i| \geq t_0\},$$

where u_0 and t_0 are observed values of U_i and T_i for H_i, respectively.

(i). Estimation of true null proportions based on testing statistics T_i.

By using Storey-type estimator, as we stated in (4.3) in Section 4.3.1, we first estimate true null proportions among selected hypotheses S and non-selected hypotheses $I\backslash S$ based on P_i, which are denoted as $\hat{\pi}_0(S)$ and $\hat{\pi}_0(I\backslash S)$ as follows:

$$\hat{\pi}_0(S) = \frac{Q_S(\lambda)}{(1-\lambda)|S|} \text{ and } \hat{\pi}_0(I\backslash S) = \frac{Q_{I\backslash S}(\lambda)}{(1-\lambda)(m-|S|)}, \tag{4.12}$$

where $Q_S(\lambda) = \sum\limits_{k\in S} I_{\{P_k>\lambda\}}, Q_{I\backslash S}(\lambda) = \sum\limits_{k\in I\backslash S} I_{\{P_k>\lambda\}}$, and λ is a fixed constant.

(ii). Estimation of true null proportions based on selection statistics U_i.

We then estimate true null proportions among S and $I\backslash S$ based on $\tilde{P}_i$, which are denoted as $\tilde{\pi}_0(S)$ and $\tilde{\pi}_0(I\backslash S)$. Thus the corresponding estimated false null proportions, denoted as $\tilde{\pi}_1(S)$ and $\tilde{\pi}_1(I\backslash S)$ can be derived as

$$\tilde{\pi}_1(S) = 1 - \tilde{\pi}_0(S) \text{ and } \tilde{\pi}_1(I\backslash S) = 1 - \tilde{\pi}_0(I\backslash S).$$

Let $\lambda_0 = \Pr(U_i > t)$ and $\lambda_1 < \lambda_0$ be a fixed constant. The expected number of true null hypotheses among S with p-values exceed λ_1 can be expressed as follows:

$$m_0(S)\Pr\left(\tilde{P}_i > \lambda_1 | \tilde{P}_i < \lambda_0\right) = m_0(S)\frac{\lambda_0 - \lambda_1}{\lambda_0}, \tag{4.13}$$

where $m_0(S)$ is the number of true null hypotheses among S. The expected number of true nulls in S can be approximated by the observed number of $\tilde{P}_k$ exceeding λ_1 in S:

$$\sum_{k \in S} I_{\{\tilde{P}_k > \lambda_1\}} \equiv Q_S(\lambda_1). \tag{4.14}$$

Combining (4.13) and (4.14), the estimated number of true null hypotheses $\hat{m}_0(S)$ among S can be derived:

$$\hat{m}_0(S) = \frac{Q_S(\lambda_1)}{\frac{(\lambda_0 - \lambda_1)}{\lambda_0}}, \tag{4.15}$$

thus, the true null proportion among S can be estimated,

$$\tilde{\pi}_0(S) = \frac{\hat{m}_0(S)}{|S|} = \frac{Q_S(\lambda_1)}{\frac{(\lambda_0 - \lambda_1)}{\lambda_0}|S|}. \tag{4.16}$$

Similarly, the expected number of true null hypotheses among $I \backslash S$ with p-values exceed a fixed contant $\lambda_2 > \lambda_0$ can be expressed as follows:

$$m_0(I \backslash S)\Pr\left(\tilde{P}_i > \lambda_2 | \tilde{P}_i > \lambda_0\right) = m_0(I \backslash S)\frac{1 - \lambda_2}{1 - \lambda_0}, \tag{4.17}$$

where $m_0(I \backslash S)$ is the number of true null hypotheses among $I \backslash S$. Similarly, the expected number of true nulls in $I \backslash S$ can be approximated by the observed number of $\tilde{P}_k$ exceeding λ_2 in $I \backslash S$:

$$\sum_{k \in I \backslash S} I_{\{\tilde{P}_k > \lambda_2\}} \equiv Q_{I \backslash S}(\lambda_2). \tag{4.18}$$

Combining (4.17) and (4.18), the estimated number of true null hypotheses $\hat{m}_0(I\backslash S)$ among $I\backslash S$ can be derived:

$$\hat{m}_0(I\backslash S) = \frac{Q_{I\backslash S}(\lambda_2)}{\frac{1-\lambda_2}{1-\lambda_0}}, \tag{4.19}$$

thus, the true null proportion among $I\backslash S$ can be estimated,

$$\tilde{\pi}_0(I\backslash S) = \frac{\hat{m}_0(I\backslash S)}{(m-|S|)} = \frac{Q_{I\backslash S}(\lambda_2)}{\frac{1-\lambda_2}{1-\lambda_0}(m-|S|)}. \tag{4.20}$$

Therefore, false null proportions based on selection statistics U_i can be estimated as:

$$\begin{aligned} \tilde{\pi}_1(S) &= 1-\tilde{\pi}_0(S) = 1-\frac{Q_S(\lambda_1)}{\frac{(\lambda_0-\lambda_1)}{\lambda_0}|S|}, & (4.21)\\ \tilde{\pi}_1(I\backslash S) &= 1-\tilde{\pi}_0(I\backslash S) = 1-\frac{Q_{I\backslash S}(\lambda_2)}{\frac{1-\lambda_2}{1-\lambda_0}(m-|S|)}. & (4.22) \end{aligned}$$

By replacing false null proportions $\pi_1(S)$ and $\pi_1(I\backslash S)$ with $\tilde{\pi}_1(S)$ and $\tilde{\pi}_1(I\backslash S)$ in (4.8), and replacing true null proportions $\pi_0(S)$ and $\pi_0(I\backslash S)$ with $\hat{\pi}_0(S)$ and $\hat{\pi}_0(I\backslash S)$ in (4.10) and (4.11), our data-driven weights are constructed as stated in (4.10) and (4.11). Now we are ready to define a data-driven weighted selective procedure as follows:

Definition 4 (Data-driven weighted selective procedure)**.**

1. For a fixed selection threshold t, we have $S = \{i \in \{1,\ldots,m\} : U_i > t\}$ and $\lambda_0 = Pr(U_i > t)$. And for fixed $\lambda \in (0,1)$, calculate $\hat{\pi}_0(S)$ and $\hat{\pi}_0(I\backslash S)$ based on (4.12).

2. For fixed λ_1 and $\lambda_2 \in (0,1)$, with $\lambda_1 < \lambda_0 < \lambda_2$, calculate $\tilde{\pi}_0(S)$ and $\tilde{\pi}_0(I\backslash S)$ based on (4.16), (4.20), respectively, and then compute $\tilde{\pi}_1(S)$ and $\tilde{\pi}_1(I\backslash S)$.

3. We then have observed values w_1 and w_2 respectively for W_1 and W_2 according to (4.10) and (4.11).

4. For each hypothesis H_i among all m hypotheses, assign a weight as follows:

$$\tilde{w_i} = \begin{cases} w_1, & if\ i \in S, \\ w_2, & if\ i \in I \backslash S. \end{cases}$$

5. Reject H_i if $p_i \leq \frac{\tilde{w}_i \alpha}{m}$, for all $i = 1, \ldots, m$.

Remark 11. *The turning parameter λ in Storey-type estimator is often set to be $\lambda = 0.5$, see Storey et al. (2004). When determining values of the turning parameters λ_1 and λ_2 in estimation of true null proportions $\tilde{\pi}_0(S)$ and $\tilde{\pi}_0(I \backslash S)$ based on $\tilde{P}_i$, we set*

$$Pr\left(\tilde{P}_i > \lambda_1 | \tilde{P}_i < \lambda_0\right) = \frac{\lambda_0 - \lambda_1}{\lambda_0} = 0.5,$$

thus $\lambda_1 = 0.5\lambda_0$, such that about 50% true null p-values in S are used to estimate the proportion of true nulls in S. Similarily, we set

$$Pr\left(\tilde{P}_i > \lambda_2 | \tilde{P}_i > \lambda_0\right) = \frac{1 - \lambda_2}{1 - \lambda_0} = 0.5,$$

thus $\lambda_2 = 0.5 + 0.5\lambda_0$, such that about 50% true null p-values in $I \backslash S$ are used to estimate the proportion of true nulls in $I \backslash S$. And the value of λ_0 is determined by the selection threshold, the more strict the selection threshold, the lower value of λ_0 is, and vice versa.

For the proposed procedure, we have the following result.

Theorem 7. *The above defined data-driven weighted selective procedure strongly controls the FWER at level α under the assumption that pairs of (U_i, T_i) are independent.*

Proof. For given selection statistics $U_i, i = 1, \ldots, m$, we have $S = \{i \in \{1, \ldots, m\} : U_i > t\}$, which is the index set of the selected hypotheses. We denote cFWER as the conditional familywise error rate on selection, such that

$$\text{cFWER} = \Pr\left(\text{reject at least one true null } | U_1, \ldots, U_m\right).$$

Let I_0 be the index set of true null hypotheses and S_0 be the index set of true null hypotheses among S. Let

$$Q_S^{(-i)}(\lambda) = \sum_{k \in S \backslash \{i\}} I_{\{P_k > \lambda\}}, \text{ and } Q_{S_0}^{(-i)}(\lambda) = \sum_{k \in S_0 \backslash \{i\}} I_{\{P_k > \lambda\}}.$$

Also let

$$Q_{I \backslash S}^{(-j)}(\lambda) = \sum_{k \in (I \backslash S) \backslash \{j\}} I_{\{P_k > \lambda\}}, \text{ and } Q_{I_0 \backslash S_0}^{(-j)}(\lambda) = \sum_{k \in (I_0 \backslash S_0) \backslash \{j\}} I_{\{P_k > \lambda\}}.$$

When H_i and H_j are true, where $i \in S_0$ and $j \in I_0 \backslash S_0$, by independence of true null p-values, we have

$$Q_{S_0}^{(-i)}(\lambda) \sim \text{Bin}(|S_0| - 1, 1 - \lambda) \tag{4.23}$$

and

$$Q_{I_0 \backslash S_0}^{(-j)}(\lambda) \sim \text{Bin}(|I_0| - |S_0| - 1, 1 - \lambda). \tag{4.24}$$

We can find the upper bound of cFWER as follows:

$$\begin{aligned}
\text{cFWER} &\leq \Pr\left(\bigcup_{i\in S_0}\{P_i \leq \frac{W_1\alpha}{|S|}\}|U_1,\ldots,U_m\right) \\
&\quad + \Pr\left(\bigcup_{j\in I_0\backslash S_0}\{P_j \leq \frac{W_2\alpha}{m-|S|}\}|U_1,\ldots,U_m\right) \\
&= \Pr\left(\bigcup_{i\in S_0}\{P_i \leq \frac{W_1''\alpha}{\hat{\pi}_0(S)\cdot|S|}\}|U_1,\ldots,U_m\right) \\
&\quad + \Pr\left(\bigcup_{j\in I_0\backslash S_0}\{P_j \leq \frac{W_2''\alpha}{\hat{\pi}_0(I\backslash S)\cdot(m-|S|)}\}|U_1,\ldots,U_m\right) \\
&= \Pr\left(\bigcup_{i\in S_0}\{P_i \leq \frac{W_1''\alpha(1-\lambda)}{Q_S(\lambda)}\}|U_1,\ldots,U_m\right) \\
&\quad + \Pr\left(\bigcup_{j\in I_0\backslash S_0}\{P_j \leq \frac{W_2''\alpha(1-\lambda)}{Q_{I\backslash S}(\lambda)}\}|U_1,\ldots,U_m\right). \qquad (4.25)
\end{aligned}$$

Let

$$\begin{aligned}
A_1 &= \Pr\left(\bigcup_{i\in S_0}\{P_i \leq \frac{W_1''\alpha(1-\lambda)}{Q_S(\lambda)}\}|U_1,\ldots,U_m\right) \\
&\quad + \Pr\left(\bigcup_{j\in I_0\backslash S_0}\{P_j \leq \frac{W_2''\alpha(1-\lambda)}{Q_{I\backslash S}(\lambda)}\}|U_1,\ldots,U_m\right),
\end{aligned}$$

then we have

$$
\begin{aligned}
A_1 \ &\leq \ \sum_{i\in S_0}\Pr\left(P_i \leq \frac{W_1''\alpha(1-\lambda)}{Q_S^{(-i)}(\lambda)}|U_1,\ldots,U_m\right) \\
&\quad + \sum_{j\in I_0\backslash S_0}\Pr\left(P_j \leq \frac{W_2''\alpha(1-\lambda)}{Q_{I\backslash S}^{(-j)}(\lambda)}|U_1,\ldots,U_m\right) \\
&\leq \ \sum_{i\in S_0}E\left(\frac{W_1''\alpha(1-\lambda)}{Q_S^{(-i)}(\lambda)}|U_1,\ldots,U_m\right) \\
&\quad + \sum_{j\in I_0\backslash S_0}E\left(\frac{W_2''\alpha(1-\lambda)}{Q_{I\backslash S}^{(-j)}(\lambda)}|U_1,\ldots,U_m\right) \\
&\leq \ \sum_{i\in S_0}E\left(\frac{W_1''\alpha(1-\lambda)}{Q_{S_0}^{(-i)}(\lambda)}|U_1,\ldots,U_m\right) \\
&\quad + \sum_{j\in I_0\backslash S_0}E\left(\frac{W_2''\alpha(1-\lambda)}{Q_{I_0\backslash S_0}^{(-j)}(\lambda)}|U_1,\ldots,U_m\right), \qquad (4.26)
\end{aligned}
$$

where the first inequality in (4.26) is due to Bonferroni inequality. The second inequality holds since $Q_S^{(-i)}(\lambda) \leq Q_S(\lambda)$ and $Q_{I\backslash S}^{(-j)}(\lambda) \leq Q_{I\backslash S}(\lambda)$. The third inequality is because of independence of P_i and $Q_S^{(-i)}$ and independence of P_j and $Q_{I\backslash S}^{(-j)}$, which are due to pariwise independence of (U_i, T_i), and independence between P_i and W_1'', and between P_j and W_2'', which are due to pariwise independence of (U_i, T_i) and independence of U_i and T_i. The last inequality holds since $Q_{S_0}^{(-i)}(\lambda) \leq Q_S^{(-i)}(\lambda)$, and $Q_{I_0\backslash S_0}^{(-j)}(\lambda) \leq Q_{I\backslash S}^{(-j)}(\lambda)$.

Let

$$A_2 = \sum_{i\in S_0}E\left(\frac{W_1''\alpha(1-\lambda)}{Q_{S_0}^{(-i)}(\lambda)}|U_1,\ldots,U_m\right) + \sum_{j\in I_0\backslash S_0}E\left(\frac{W_2''\alpha(1-\lambda)}{Q_{I_0\backslash S_0}^{(-j)}(\lambda)}|U_1,\ldots,U_m\right)$$

be the upper bound in (4.26), then we have the following result:

$$\begin{aligned}
A_2 &= \alpha(1-\lambda)\left(\sum_{i\in S_0} E\left(W_1''|U_1,\ldots,U_m\right)\cdot E\left(\frac{1}{Q_{S_0}^{(-i)}(\lambda)+1}|U_1,\ldots,U_m\right)\right) \\
&\quad + \alpha(1-\lambda)\left(\sum_{j\in I_0\backslash S_0} E\left(W_2''|U_1,\ldots,U_m\right)\cdot E\left(\frac{1}{Q_{I_0\backslash S_0}^{(-j)}(\lambda)+1}|U_1,\ldots,U_m\right)\right) \\
&< \alpha(1-\lambda)\left(\sum_{i\in S_0}\frac{E\left(W_1''|U_1,\ldots,U_m\right)}{(|S_0|)(1-\lambda)} + \sum_{j\in I_0\backslash S_0}\frac{E\left(W_2''|U_1,\ldots,U_m\right)}{(m-|S_0|)(1-\lambda)}\right) \\
&= \alpha E\left(W_1''+W_2''\right) = \alpha, \qquad (4.27)
\end{aligned}$$

where the first equality in (4.27) holds since independence of W_1'' and $Q_{S_0}^{(-i)}(\lambda)$, and independence of W_2'' and $Q_{I_0\backslash S_0}^{(-j)}(\lambda)$. Due to (4.23), (4.24) and Lemma 1, we have the last inequality in (4.27).

Therefore, FWER $= E\,(\text{cFWER}) \leq \alpha$. The desire result is proved. □

4.4 Data-driven Weighted Multiple Testing Procedures

In this section, we generalize the aforementioned proposed three procedures from the case of two blocks to multiple blocks to further develop more practical and powerful MTPs. The blockwise adaptive Bonferroni procedure in Section 4.4.1 is a generalization of the adaptive two-stage Bonferroni procedure. In Section 4.4.2, we generalize the selective parallel gatekeeping procedure to blockwise selective parallel gatekeeping procedure. In Section 4.4.3, the proposed blockwise data-driven weighted procedure is a generalization of the data-driven weighted selective procedure.

4.4.1 Blockwise Adaptive Two-stage Bonferroni Procedure

Motivated by Kim and Schliekelman (2015), we propose a blockwise adaptive two-stage Bonferroni procedure. Suppose that the m hypotheses are ordered by the selection statistics U_i, which attempts to ordering the hypotheses from the most promising hypothesis to the least promising hypothesis. We then divide the ordered hypotheses into K (fixed) blocks. Let B_l be the index set of null hypotheses in block $l = 1, \ldots, K$,

$$\begin{aligned} |B_l| &= \left\lfloor \frac{m}{K} \right\rfloor, l = 1, \ldots, K-1; && (4.28) \\ |B_K| &= m - (K-1) \times |B_l|. && (4.29) \end{aligned}$$

Within each block, we apply adaptive Bonferroni procedure and reject the null hypothesis $H_i, i \in B_l$ if

$$p_i \leq \frac{\alpha}{\hat{\pi}_{B_l}.|B_l|.K}.$$

And we use the similar way to estimate $\hat{\pi}_{B_l}$ as in adaptive two-stage Bonferroni procedure as follows:

$$\hat{\pi}_{B_l} = \frac{Q_{B_l}(\lambda) + 1}{(1-\lambda)|B_l|}, \tag{4.30}$$

where λ is a fixed constant and $Q_{B_l}(\lambda) = \sum_{i \in B_l} I_{\{P_i > \lambda\}}$ is the number of hypotheses in block l such that the corresponding p-values exceed λ.

We define the blockwise adaptive two-stage Bonferroni procedure as follows:

Definition 5 (Blockwise adaptive two-stage Bonferroni procedure)**.**

1. For fixed K blocks, calculate $\hat{\pi}_{B_l}, l = 1, \ldots, K$ based on (4.30).

2. Within each $B_l = 1, \ldots, K$, apply adaptive Bonferroni procedure at level $\frac{\alpha}{\hat{\pi}_{B_l}.|B_l|}$. For $i \in B_l$, reject H_i if

$$p_i \leq \frac{\alpha}{\hat{\pi}_{B_l}.|B_l|.K}.$$

Remark 12. *When the number of blocks $K = 2$, then the proposed procedure is similar to the adaptive two-stage Bonferroni procedure as we proposed before.*

Moreover, the true null proportion $\hat{\pi}_{B_l}$ for each block is estimated from the data, thus the proposed procedure can be regarded as a data-driven weighted procedure, where $w_l = \frac{1}{\hat{\pi}_{B_l}}$ for block l.

By Lemma 1, we have the following result.

Theorem 8. *The above defined blockwise adaptive two-stage Bonferroni procedure strongly controls the FWER at level α.*

Proof. Following closely the proof in Section 4.3.1, we show the FWER control for general case as follows. Let $N_l = l * \left\lfloor \frac{m}{K} \right\rfloor$. For given $B_l = 1, \ldots, K$, we have $B_l = \{i \in \{1, \ldots, m\} : U_{(N_{l-1}+1)} \leq U_{(i)} \leq U_{(N_l)}\}$, which is the index set of hypotheses in B_l, where $U_{(i)}$ is ordered selection statistics. We denote cFWER$_1$ as the conditional familywise error rate on selection statistics for B_1 such that

$$\text{cFWER}_1 = \Pr\Big(\text{reject at least one true null in } B_1 \; |U_1 \ldots, U_m\Big) .$$

Let B_{10} be the index set of true null hypotheses among B_1, and let $Q_{B_1}^{(-i)}(\lambda) = \sum\limits_{k \in B_1 \setminus \{i\}} I_{\{P_k > \lambda\}}$, and $Q_{B_{10}}^{(-i)}(\lambda) = \sum\limits_{k \in B_{10} \setminus \{i\}} I_{\{P_k > \lambda\}}$. Similar to (4.30), define $\hat{\pi}_{B_1}^{(-i)} = \frac{Q_{B_1}^{(-i)}(\lambda)}{(1-\lambda)|B_1|}$. When H_i is true, by independence of true null p-values, we have

$$Q_{B_{10}}^{(-i)}(\lambda) \sim \text{Bin}(|B_{10}| - 1, 1 - \lambda). \tag{4.31}$$

We first consider the $cFWER_1$,

$$\begin{aligned}
\text{cFWER}_1 &= \Pr\left(\bigcup_{i\in B_{10}} \{P_i \le \frac{\alpha}{\hat{\pi}_{B_1}|B_1|K}\}|U_1,\ldots,U_m\right) \\
&\le \sum_{i\in B_{10}} \Pr\left(P_i \le \frac{\alpha}{\hat{\pi}_{B_1}|B_1|K}|U_1,\ldots,U_m\right) \\
&\le \sum_{i\in B_{10}} \Pr\left(P_i \le \frac{\alpha}{\hat{\pi}^{(-i)}_{B_1}|B_1|K}|U_1,\ldots,U_m\right) \\
&\le \sum_{i\in B_{10}} E\left(\frac{\alpha}{\hat{\pi}^{(-i)}_{B_1}|B_1|K}|U_1,\ldots,U_m\right) \\
&= \sum_{i\in B_{10}} E\left(\frac{\alpha(1-\lambda)}{(Q^{(-i)}_{B_1}(\lambda)+1)K}|U_1,\ldots,U_m\right) \\
&\le \sum_{i\in B_{10}} E\left(\frac{\alpha(1-\lambda)}{(Q^{(-i)}_{B_{10}}(\lambda)+1)K}|U_1,\ldots,U_m\right) \\
&= \frac{\alpha(1-\lambda)}{K}\sum_{i\in B_{10}} E\left(\frac{1}{Q^{(-i)}_{B_{10}}(\lambda)+1}|U_1,\ldots,U_m\right) \\
&< \frac{\alpha(1-\lambda)}{K}\sum_{i\in B_{10}} \frac{1}{(|B_{10}|)(1-\lambda)} = \frac{\alpha|B_{10}|}{K|B_{10}|} = \frac{\alpha}{K}, \qquad (4.32)
\end{aligned}$$

where the first inequality in (4.32) is due to Bonferroni inequality. The second inequality holds since $Q^{(-i)}_{B_1}(\lambda) \le Q_{B_1}(\lambda)$, according to (4.30), $\hat{\pi}^{(-i)}_{B_1} \le \hat{\pi}_{B_1}$. The third inequality is because of pariwise independence of (U_i, T_i), independence of U_i and T_i under H_i, and independence of P_i and $\hat{\pi}^{(-i)}_{B_1}$. The fourth inequality holds since $Q^{(-i)}_{B_{10}}(\lambda) \le Q^{(-i)}_{B_1}(\lambda)$. Due to Lemma 1 and (4.31), we have the last inequality in (4.32).

Therefore, $\text{cFWER}_1 = E(\text{cFWER}_1) \le \frac{\alpha}{K}$. Similarily, we can show $\text{FWER}_i \le \frac{\alpha}{K}$, for $i = 2,\ldots,K$. Thus,

$$\text{FWER} \le \sum_{i=1}^{K} \text{FWER}_i \le \sum_{i=1}^{K} \frac{\alpha}{K} = \alpha. \qquad (4.33)$$

Therefore, the desire result is proved. □

4.4.2 Blockwise Selective Parallel Gatekeeping Procedure

We proposed a blockwise selective parallel gatekeeping procedure, which is an extension of selective parallel gatekeeping procedure. We further exploit information of each hypotheses more explicitly by splitting tested hypotheses into multiple blocks. Suppose that the m hypotheses are ordered by the selection statistics U_i, which attempts to ordering the hypotheses from the most promising hypothesis to the least promising hypothesis. We then divide the ordered hypotheses into K (fixed) blocks. Let B_l be the index set of null hypotheses in $B_l = 1, \ldots, K$.

Hypotheses in the first $K-1$ blocks are in family F_1, while the remaining hypotheses in family F_2, which is regarded as the Kth block. Initially, each B_l is assigned with critical value $\alpha_l = w_l \alpha, l = 1, \ldots, K-1$, and $\alpha_K = 0$. Hypotheses are tested from lower ranked block to higher ranked block, i.e., from B_1 to B_K. Critical values are also allocated in such order if hypotheses are rejected. Following the weight construction in Kim and Schliekelman (2015),

$$w_l = \frac{\sum_{j=1}^{K-1} |B_j|}{\lambda \sum_{j=1}^{l} |B_j|} \tag{4.34}$$

is the weight of B_l, where λ is the correction factor. Thus $w_1 > \ldots > w_{K-1} > w_K = 0$.

Note that in order to determine the correction factor λ, we set

$$\sum_{l=1}^{K} w_l \alpha = \sum_{l=1}^{K} \frac{\sum_{j=1}^{K-1} |B_j|}{\lambda \sum_{j=1}^{l} |B_j|} \alpha = \alpha, \tag{4.35}$$

such that λ can be solved from (4.35).

We define a blockwise selective parallel gatekeeping procedure as follows:

Definition 6 (Blockwise selective parallel gatekeeping procedure)**.**

1. *For a fixed K, calculate w_l for $l = 1, \ldots, K-1$ based on (4.34).*

2. *Test each hypothesis H_i in B_1. Reject H_i if $p_i \leq \frac{w_1 \alpha}{|B_1|}$. Rejections are in $\mathcal{R}_1 = \{i \in B_1 : p_i \leq \frac{w_1 \alpha}{|B_1|}\}$. Update the critical value of B_2:*

$$\alpha_2 = w_2\alpha + \frac{\alpha_1|R_1|}{|B_1|}.$$

3. Test each hypothesis H_i *in* $B_l (2 \leq i \leq K-1)$*. Reject* H_i *if* $p_i \leq \frac{\alpha_i}{|B_l|}$*. Rejections are in* $\mathcal{R}_l = \{i \in B_l : p_i \leq \frac{\alpha_l}{|B_l|}\}$*. Update the critical value of* B_{l+1}*:*

$$\alpha_{l+1} = w_{l+1}\alpha + \frac{\alpha_l|R_l|}{|B_l|}.$$

4. Test each hypothesis H_i *in* B_K*. Reject* H_i *if* $p_i \leq \frac{\alpha_K}{|B_K|}$*. Rejections are in* $\mathcal{R}_K = \{i \in B_K : p_i \leq \frac{\alpha_K}{|B_K|}\}$*.*

Remark 13. *The constructed weights* $w_1, \ldots, w_K$ *are monotonic decreasing, thus hypotheses in lower ranked block have a higher chance to reject. Moreover, with the allocation of critical values, one can have more rejections in lower weighted blocks. We divide hypotheses of* F_1 *into* $K-1$ *blocks, which is more explicit than the selective parallel gatekeeping procedure introduced in Section 4.3.2, where no further specification among* F_1*.*

For the proposed procedure, we have the following result.

Theorem 9. *The above defined blockwise selective parallel gatekeeping procedure strongly controls the FWER at level* α*.*

Proof. In selective parallel gatekeeping procedure, we showed that FWER $\leq \alpha$, that is the case when we have two blocks. Following closely the proof in Qiu et al. (2017), we use mathematical induction to show the cFWER control for general case.

Assume that when $n = K, K \geq 2$,

$$\text{cFWER}(\alpha_1, \ldots, \alpha_k) \leq \sum_{l=1}^{K} \alpha_i \leq \alpha, \tag{4.36}$$

In the following, we show that (4.36) holds when $n = K + 1$, that is

$$\text{cFWER}(\alpha_1, \ldots, \alpha_{k+1}) \leq \sum_{l=1}^{K+1} \alpha_i \leq \alpha. \tag{4.37}$$

We first define an event

$$E_1 = \{\text{at least one false rejection in } B_1\},$$

$$E_2 = \{\text{at least one false rejection in all blocks except } B_1\}.$$

Then the cFWER can be expressed as follows:

$$\begin{aligned} \text{cFWER}(\alpha_1, \ldots, \alpha_{K+1}) &= \Pr\big(E_1 \cup E_2 | U_1, \ldots, U_m\big) \\ &= \Pr\big(E_1 | U_1, \ldots, U_m\big) + \Pr\big(\bar{E}_1 \cap E_2 | U_1, \ldots, U_m\big) \end{aligned} \tag{4.38}$$

Let $e^*(\cdot)$ as an upper bound of error rate function $e(\cdot)$. By the definition of the error rate function, we know that

$$\Pr\big(E_1 | U_1, \ldots, U_m\big) \leq e(T_1) \leq e^*(T_1),$$

where T_1 is the set of true null hypotheses in B_1.

Let A_1 denote as the acceptance set within B_1. After testing hypotheses in B_1, $\alpha_1 - e^*(A_1)$ will be transferred to $B_2, \ldots, B_n$. Let α_l^* be the updated significant level for $B_l \geq 2$, such that

$$\alpha_l^* = \alpha_l + \alpha_{l-1} - e_1^*(A_l). \tag{4.39}$$

Then

$$
\begin{aligned}
& \mathrm{cFWER}(\alpha_1, \ldots, \alpha_{K+1}) \\
\leq\ & e_1^*(T_1) + \mathrm{cFWER}(\alpha_2^*, \ldots, \alpha_{K+1}^*) \\
\leq\ & e_1^*(T_1) + \sum_{l=2}^{K+1} \alpha_i^* \\
\leq\ & e_1^*(T_1) + \alpha_1 - e_1^*(T_1) + \sum_{l=2}^{K+1} \alpha_i \\
=\ & \alpha_1 + \sum_{l=2}^{K+1} \alpha_i = \sum_{l=1}^{K+1} w_i \alpha = \alpha, \qquad (4.40)
\end{aligned}
$$

where the first inequality in (4.40) holds by the definition of erro rate function. The second inequality in (4.40) is due to the assumption in (4.36). And the last inequality in (4.40) holds by the updating rule of α_l^* in (4.39).

Therefore, FWER $= E\,(\mathrm{cFWER}) \leq \alpha$. The desire result is proved. $\square$

4.4.3 Blockwise Data-driven Weighted Selective Procedure

Unlike the blockwise selective parallel gatekeeping procedure, where no critical value is initially assigned to F_2, in this section, we propose a blockwise data-driven weighted selective procedure, which is an extension of the data-driven weighted selective procedure. By the similar estimation of true null proportion stated in (4.16), (4.20), for all K blocks, we estimate their true null proportion $\tilde{\pi}_{B_{l0}}$, such that

$$\tilde{\pi}_{B_{l0}} = \frac{\hat{m}_{B_{l0}}}{|B_l|} = \frac{Q_{B_l}(\lambda_l)}{\frac{\lambda_0 - \lambda_l}{\lambda_0}|B_l|}, \text{ for } l = 1, \ldots, K, \qquad (4.41)$$

where $\hat{m}_{B_{l0}}$ is the estimated number of true null hypotheses among B_l, and λ_l is the fixed constant for B_l. Thus we can estimate false null proportion $\tilde{\pi}_{B_{l1}} = 1 - \tilde{\pi}_{B_{l0}}$. Therefore, similar to (4.10), we construct the weights as follows:

$$W_l = \frac{W_l'}{\hat{\tilde{\pi}}_{B_{l0}}}, \qquad (4.42)$$

where $W_l' = \frac{m\tilde{\pi}_{B_{l1}}}{|B_l|\left(\sum\limits_{l=1}^{K}\tilde{\pi}_{B_{l1}}\right)}$. And for each hypothesis H_i, the weight is assigned as follows:

$$\tilde{W}_i = W_l, \text{if } i \in B_l.$$

Finally, we can construct weighted procedures based on the above weights for testing all m hypotheses simultaneously.

We propose a blockwise data-driven weighted selective procedure as follows:

Definition 7 (Blockwise data-driven weighted selective procedure)**.**

1. For fixed K blocks, calculate $\hat{\pi}_{B_{l0}}, l = 1, \ldots, K$ based on (4.41). Then calculate w_l for B_l based on (4.42).

2. Apply weighted Bonferroni procedure at level $\frac{\tilde{w}_i\alpha}{m}$, reject H_i if

$$p_i \leq \frac{\tilde{w}_i\alpha}{m}.$$

For the proposed procedure, we have the following result.

Theorem 10. *The above defined blockwise data-driven weighted selective procedure strongly controls the FWER at level α.*

Proof. Similar to the proofs in Section 4.3.3, we show the FWER control of blockwise data-driven weighted selective procedure as follows. Let $N_l = l * \left\lfloor \frac{m}{K} \right\rfloor$. For given $B_l = 1, \ldots, K$, we have $B_l = \{i \in \{1, \ldots, m\} : U_{(N_{l-1}+1)} \leq U_{(i)} \leq U_{(N_l)}\}$, which is the index set of hypotheses in B_l, where $U_{(i)}$ is ordered selection statistics. We denote $cFWER_l$ as the conditional familywise error rate on selection statistics for B_l, such that

$$cFWER_l = \Pr\left(\text{reject at least one true null in } B_l \mid U_1, \ldots, U_m\right)$$

Let I_0 be the index set of true null hypotheses and B_{l0} be the index set of true null hypotheses among B_l. Let

$$Q_{B_l}^{(-i)}(\lambda) = \sum_{k \in B_l \setminus \{i\}} I_{\{P_k > \lambda\}}, \text{ and } Q_{B_{l0}}^{(-i)}(\lambda) = \sum_{k \in B_{l0} \setminus \{l\}} I_{\{P_k > \lambda\}}.$$

When H_i is true, where $i \in B_{l0}$, by independence of true null p-values, we have

$$Q_{B_{l0}}^{(-i)}(\lambda) \sim \text{Bin}(|B_{l0}| - 1, 1 - \lambda), \text{for } l = 1, \ldots, K. \tag{4.43}$$

We can find the upper bound of cFWER as follows:

$$\begin{aligned}
\text{cFWER} &\leq \sum_{l=1}^{K} \Pr\left(\bigcup_{i \in B_{l0}} \{P_i \leq \frac{W_l \alpha}{|B_l|}\} | U_1, \ldots, U_m \right) \\
&= \sum_{l=1}^{K} \Pr\left(\bigcup_{i \in B_{l0}} \{P_i \leq \frac{W_l'' \alpha}{\hat{\pi}_{B_{l0}} \cdot |B_l|}\} | U_1, \ldots, U_m \right) \\
&= \sum_{l=1}^{K} \Pr\left(\bigcup_{i \in B_{l0}} \{P_i \leq \frac{W_l'' \alpha (1-\lambda)}{Q_{B_l}(\lambda) + 1}\} | U_1, \ldots, U_m \right) \\
&\leq \sum_{l=1}^{K} \Pr\left(\bigcup_{i \in B_{l0}} \{P_l \leq \frac{W_l'' \alpha (1-\lambda)}{Q_{B_l}^{(-i)}(\lambda) + 1}\} | U_1, \ldots, U_m \right) \\
&\leq \sum_{l=1}^{K} \sum_{i \in B_{l0}} E\left(\frac{W_l'' \alpha (1-\lambda)}{Q_{B_l}^{(-i)}(\lambda) + 1} | U_1, \ldots, U_m \right) \\
&\leq \sum_{l=1}^{K} \sum_{i \in B_{l0}} E\left(\frac{W_l'' \alpha (1-\lambda)}{Q_{B_{l0}}^{(-i)}(\lambda) + 1} | U_1, \ldots, U_m \right),
\end{aligned} \tag{4.44}$$

where the first inequality in (4.44) is due to Bonferroni inequality. The second inequality holds since $Q_{B_l}^{(-i)}(\lambda) \leq Q_{B_l}(\lambda)$. The third inequality is because of independence of P_i and $Q_{B_l}^{(-i)}$, which is due to pariwise independence of (U_i, T_i), and independence between P_i and W_l'', which is due to pariwise independence of (U_i, T_i) and independence of U_i and T_i. The fourth inequality holds since $Q_{B_{l0}}^{(-i)}(\lambda) \leq Q_{B_l}^{(-i)}(\lambda)$.

Let

$$B=\sum_{l=1}^{K}\sum_{i\in B_{l0}}E\left(\frac{W_l''\alpha(1-\lambda)}{Q_{B_{l0}}^{(-i)}(\lambda)+1}|U_1,\ldots,U_m\right),$$

then we have

$$\begin{aligned}
B &= \alpha(1-\lambda)\sum_{l=1}^{K}\left(\sum_{i\in B_{l0}}E\left(W_l''|U_1,\ldots,U_m\right)\cdot E\left(\frac{1}{Q_{B_{l0}}^{(-i)}(\lambda)+1}|U_1,\ldots,U_m\right)\right)\\
&< \alpha(1-\lambda)\sum_{l=1}^{K}\left(\sum_{i\in B_{l0}}\frac{E\left(W_l''|U_1,\ldots,U_m\right)}{(|B_{l0}|)(1-\lambda)}\right)\\
&= \frac{\alpha(1-\lambda)|B_{l0}|}{(1-\lambda)|B_{l0}|}\sum_{l=1}^{K}E\left(W_l''|U_1,\ldots,U_m\right)\\
&= \alpha\sum_{l=1}^{K}E\left(W_l''\right)=\alpha, \qquad (4.45)
\end{aligned}$$

where the third equality in (4.45) holds since independence of W_l'' and $Q_{B_{l0}}^{(-i)}(\lambda)$. Due to (4.43) and Lemma 1, we have the last inequality in (4.45).

Therefore, FWER $= E\,(\text{cFWER}) \leq \alpha$. The desire result is proved. □

Remark 14. *Compared to the data-driven weighted selective procedure in Section 4.3.3, the proposed procedure in this section is more explicit in exploting information by splitting hypotheses into more than two blocks. Furthermore, if* W_K *is non-negative, even though it may be small, we can still test those hypotheses and have significant results.*

4.5 Summary

Although independent filtering methods have been commonly used in high-dimensional data analysis, such as microarray data, genomic data, etc, their statistical properties are not well understood. In this chapter, we explore independent filtering procedures where tested hypothese are divided into two blocks by a selection process and propose three different powerful MTPs. We first develop a powerful adaptive two-stage Bonferroni procedure by incorporating into the information of the proportion of true nulls among selected hypotheses, with the proven of FWER control. We then take non-selected hypotheses into account as well, by using a parallel gatekeeping strategy, where F_1 serves as a gatekeeper for F_2, and develop a powerful selective parallel gatekeeping procedure to exploit more information among F_2, with the proven of FWER control. Finally, we develop a powerful data-driven weighted selective procedure by combining ideas of weighted procedures and adaptive procedures, with the proven of FWER control.

In order to exploit information more sufficiently than above proposed three procedures, we further develop alternative data-driven weighted procedures, which generalize the case from two blocks to multiple blocks. By using a similar idea in Kim and Schliekelman (2015), we first order hypotheses from the most promising hypotheses to the least promising hypotheses by the selection statistics U, and then divide the ordered hypotheses into K (fixed) blocks. By incorporating into the information of the proportion of true nulls within each block l, we first develop a blockwise adaptive two-stage Bonferroni procedure, with the proven of FWER control. We then allow critical values pass from lower rank block to higher rank blocks with at least one rejection within each block, and develop a blockwise selective parallel gatekeeping procedure, with the proven of FWER control. Finally, we develop a blockwise data-driven weighted selective procedure by combining ideas of weighted procedures and adaptive procedures within each block l, with the proven of FWER control. By such data-driven weight construction in the proposed blockwise procedures, the constructed weights ensure more promising hypotheses in lower ranked blocks are assigned with higer weights, which will result in more rejections.

CHAPTER 5

SIMULATION STUDIES OF SELECTIVE INFERENCE MULTIPLE TESTING PROCEDURES

5.1 Introduction

In this chapter, we perform simulation studies for the proposed procedures in Chapter 4. We aim to explore the difference between the proposed procedures and the existing MTPs, by comparing the performance of the proposed procedures with those exisitng MTPs, with respect to the FWER control and average power. The existing MTPs to be compared are two-stage Bonferroni procedure in Guo and Romano (2017), and the conventional Bonferroni procedure. In Section 5.2-5.4, we explore the performance of the proposed procedures under independent data structure, while in Section 5.5, we explore the performance of the proposed procedures under different types of dependent structures. In Section 5.6, a concluding remark is given.

5.2 Numerical Study for Adaptive Two-stage Bonferroni Procedure

In this simulation study, we consider simultaneously testing $H_i : \mu_i = 0$ vs. $H_i^{'} : \mu_i \neq 0$, for $i = 1, \ldots, m$, at level $\alpha = 0.05$. Each simulated data set is obtained by generating $m = 100$ independent normal random samples $N(\mu_i, \sigma^2)(i = 1, \ldots, m)$, and sample size $n = 10$. Among the 100 hypotheses, that is among $100\mu_i^{\prime}s$, $100\pi_0$ are equal to 0, and the remaining are equal to $\mu_i > 0$, where π_0 is the proportion of $\mu_i = 0$. And the population variance σ^2 is drawn from $U(0.5, 1.5)$ to ensure that the population variance is unknown. For the adaptive two-stage Bonferroni procedure, the two-stage Bonferroni procedure and the conventional Bonferroni procedure, we use one-sample t-statistics as our testing statistic. For the proposed adptive two-stage Bonferroni procedure and the two-stage Bonferroni procedure, we use the sum of squares as the selection statistic. The selection threshold we choose is a sequence of numbers from 6 to 20. Guo and Romano (2017) proposed a method to derive an asymptotically optimal selection threshold. Based on the same selection statistic, sum of squares, the selection threshold can be expressed as

$$t = \chi^2_n(1-\beta),$$

where $\beta = m^{-(1-\gamma)}$ indicates roughly how much proportion of hypotheses can be selected. The number of selected hypotheses is $\beta m = m^{\gamma}$, roughly. Particularly, let $d = \frac{\log(m)}{n} = \frac{\log(100)}{10} \approx 0.46$. According to the relationship between d and γ (see Figure 4.1 in Guo and Romano, 2017), the optimal γ in our simulation setting is $\gamma \approx 0.68$. Thus, we can compute $\beta \approx 0.229$. Hence the optimal selection threhold $t \approx 12.9$. Since in our numerical study, we explore the average power and estimated FWER with different selection thresholds, we extend the range of our selection threshold about $\pm 0.5 \times 12.9$ from 12.9, which is approximate $(6, 20)$.

For the proposed adaptive two-stage Bonferroni procedure, we use a Storey-type estimator to estimate the true null proportion $\hat{\pi}_0$ among the selected hypotheses as stated in Section 4.3.1, with tuning parameter $\lambda = 0.5$.

Our simulation is repeated for 1,000 times. The simulated FWER is the proportion of times that at least one false rejection occurring among the 1,000 replications and the simulated power is the average of the proportions of correct rejections over 1,000 runs. As seen from Figure 5.1, the proposed procedure performs best in terms of average power, while still controlling the FWER at level α as seen from Figure 5.2.

In Figure 5.3, for different true null proportion π_0, we compare the estimated true null proportion among selected hypotheses $\hat{\pi}_0$ to the selection threshold. As we can see, with the increasing selection threshold, $\hat{\pi}_0$ is decreasing. Thus the more strict the selection threshold, the lower chance that true null hypotheses can be selected.

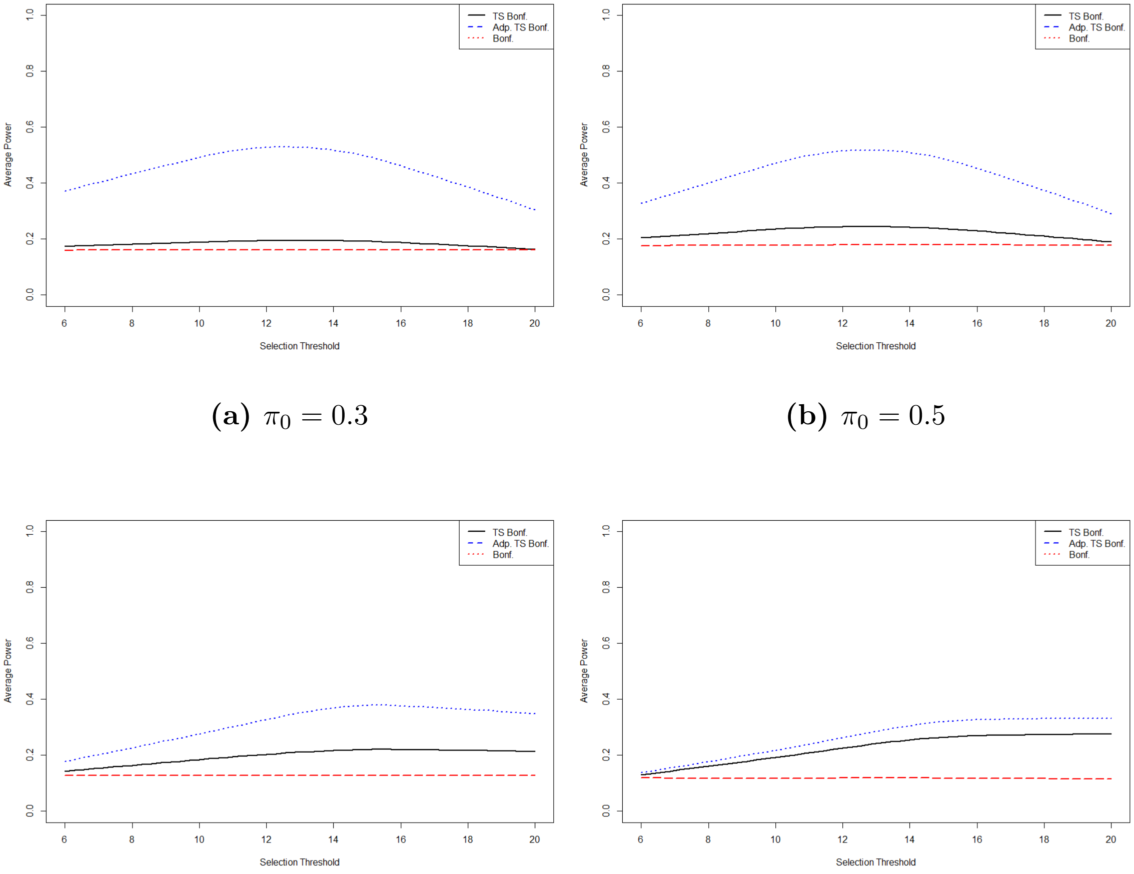
One interesting observation from Figure 5.1 when $\mu_1 = 1$ is that at the optimal selection threshold $t \approx 12.9$ (calculated according to the method in Guo and Romano, 2017), the maximum average power is achieved when $\pi_0 = 0.3$. And with the increased true null proportion, π_0, the value of the optimal selection threshold is increasing.

5.3 Numerical Study for Selective Parallel Gatekeeping Procedure

We also conduct a simulation study to compare the performance of the proposed selective parallel gatekeeping procedure with several existing MTPs with respect to the FWER control and average power. The procedures to be compared are two-stage Bonferroni procedure and the conventional Bonferroni procedure.

In this simulation study, we consider simultaneously testing $H_i : \mu_i = 0$ vs. $H_i^{'} : \mu_i \neq 0$, for $i = 1, \ldots, m$, at level $\alpha = 0.05$. Each simulated data set is obtained by generating $m = 100$ independent normal random samples $N(\mu_i, \sigma^2)(i = 1, \ldots, m)$, and equal sample size is n. Among the 100 hypotheses, that is among $100\mu_i's$, $100\pi_0$ are equal to 0, and the remaining are equal to $\mu_i = 1$, where π_0 is the proportion of $\mu_i = 0$. And the population variance σ^2 is drawn from $U(0.5, 1.5)$ to ensure that the population variance is unknown. For the three procedures, we use one-sample t-statistics as our testing statistic. For the proposed selective parallel gatekeeping procedure and the two-stage Bonferroni procedure, we use the sum of squares as the selection statistic . The selection threshold we choose is a sequence of numbers from 6 to 30. The range of the selection threshold is broader than that for the adaptive two-stage Bonferroni procedure. The reason we extend the range from $(6, 20)$ to $(6, 30)$ is that the larger the selection threshold, the more hypotheses are non-selected. And the more non-selected hypotheses, the more proper to use the selective parallel gatekeeping procedure.

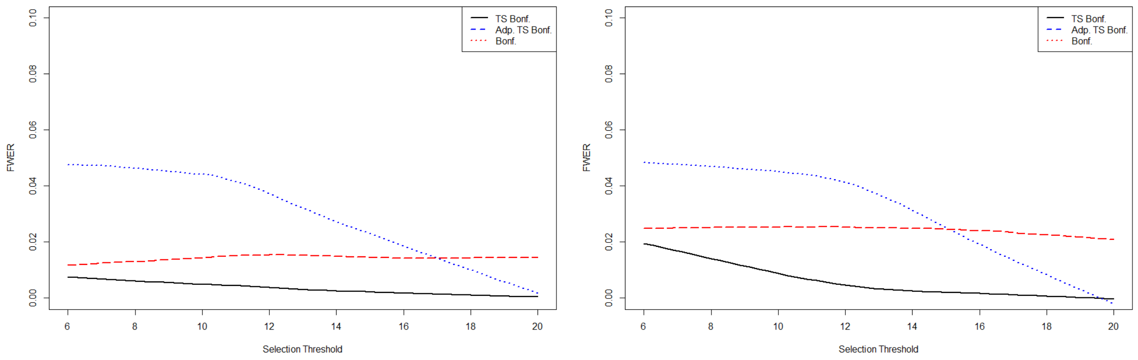
For the selected hypotheses S, we perform the conventional Bonferroni procedure, and for the non-selected hypotheses $I \backslash S$, we perform the adaptive Bonferroni



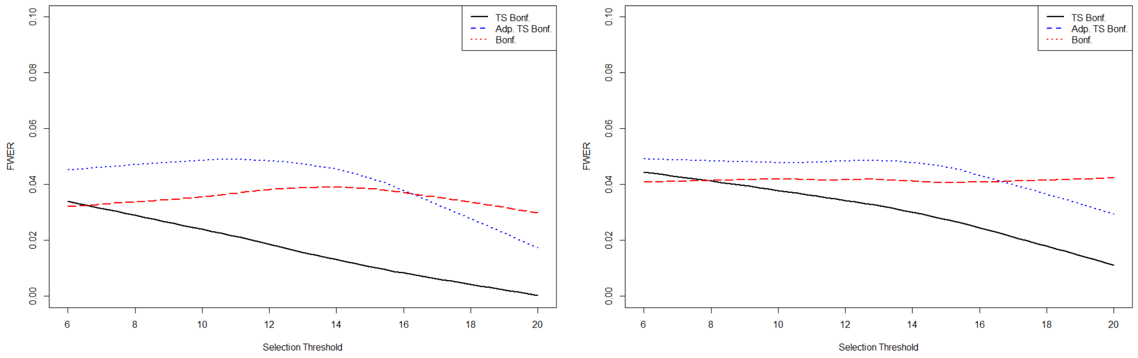
(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.1 Average power of the proposed adaptive two-stage Bonferroni procedure (Adp. TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to one. Here, the value of selection threshold t is from 6 to 20, $m = 100, n = 10, \alpha = 0.05$.

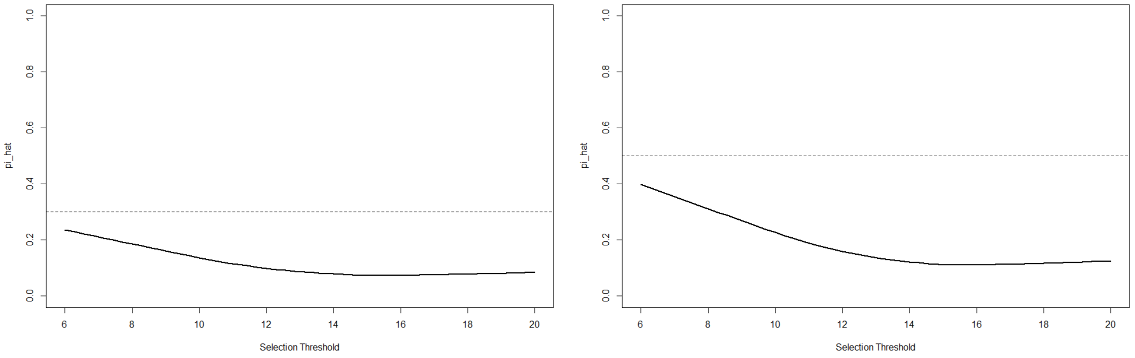


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

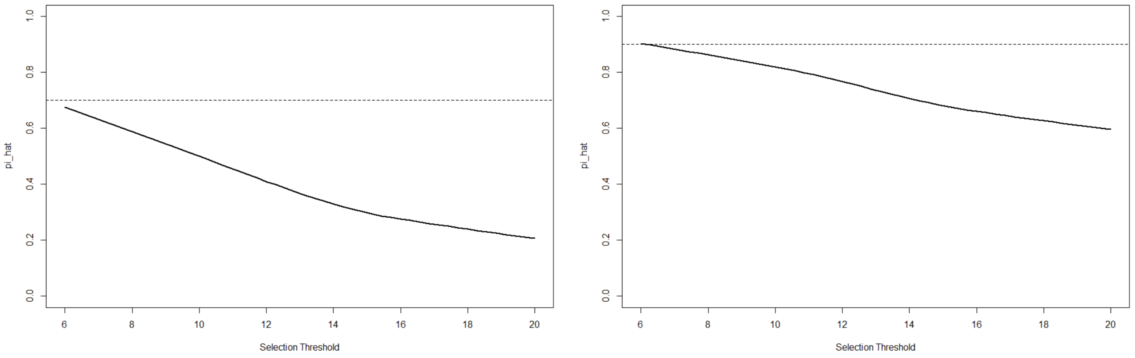


(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.2 Estimated FWER of the proposed adaptive two-stage Bonferroni procedure (Adp. TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to one. Here, the value of selection threshold t is from 6 to 20, $m = 100, n = 10, \alpha = 0.05$.



(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$



(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.3 Estimated true null proportion $\hat{\pi}_0$ of the proposed adaptive two-stage Bonferroni procedure with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to one. Here, the value of selection threshold t is from 6 to 20, $m = 100, n = 10, \alpha = 0.05$.

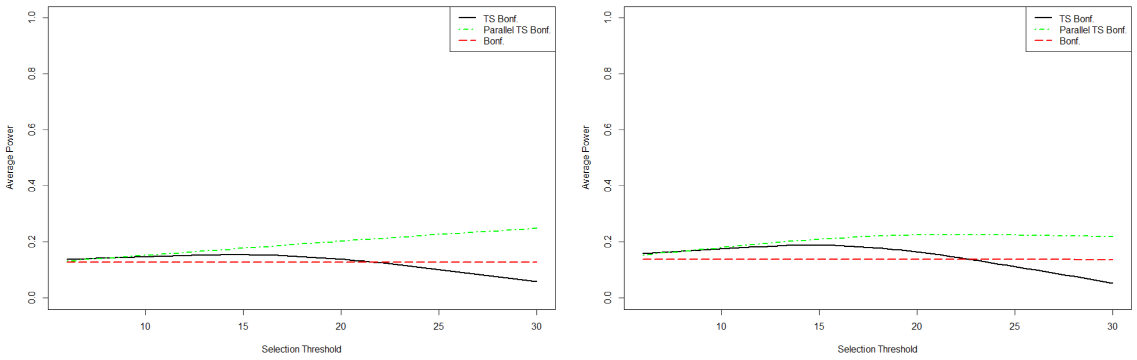
procedure. We use a Storey-type estimator to estimate the true null proportion $\hat{\pi}_0(I\backslash S) = \frac{Q_{I\backslash S}(\lambda)}{(m-|S|)(1-\lambda)}$, with tuning parameter $\lambda = 0.5$.

Our simulation is repeated for 1,000 times. Scenario 1 is when $n = 10$, and Scenario 2 is when $n = 15$. The simulated FWER is the proportion of times that at least one false rejection occurring among the 1,000 replications and the simulated power is the average of the proportions of correct rejections over 1,000 runs. As seen from Figure 5.4, when sample size $n = 10$, the average power of the proposed procedure almost perfoms the best, except when the selection threhold is around 30 at $\pi_0 = 0.9$. This may be because when the selection threhold is too strict, few hypotheses is selected, thus few critical values are passed to F_2 to conduct adaptive Bonferroni procedure. Figure 5.5 shows that the proposed selective procedure controls the FWER at level α.

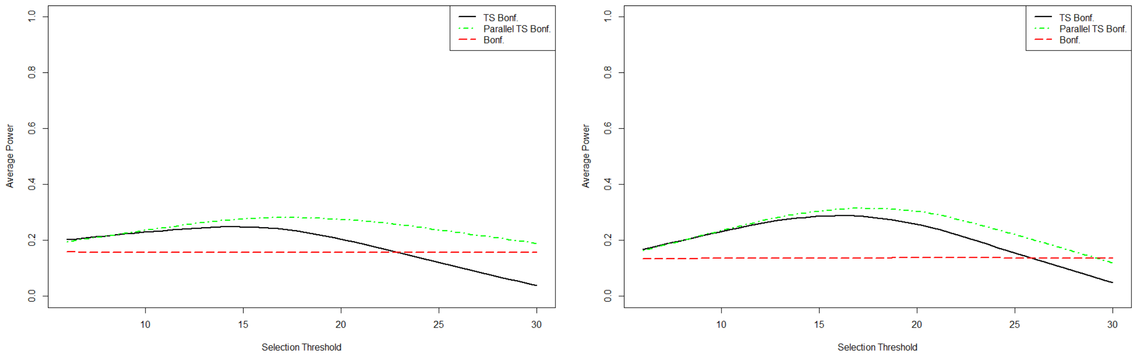
When the sample size is changed from $n = 10$ to $n = 15$, as seen from Figure 5.6, the proposed procedure perfoms best in terms of average power compared to the two-stage Bonferroni procedure and the conventional Bonferroni procedure, and controls the FWER at level α, as shown in Figure 5.7. Moreover, more strict the selection threshold, higher difference of the average power between the proposed procedure and the two-stage Bonferroni procedure among almost all scenarios.

5.4 Numerical Study for Data-driven Weighted Selective Procedure

We finally conduct a simulation study to compare the performance of the proposed data-driven weighted selective procedure with the proposed adaptive two-stage Bonferroni procedure, the selective parallel gatekeeping procedure and several existing MTPs, with respect to the FWER control and average power. The existing MTPs to be compared are two-stage Bonferroni procedure and the conventional Bonferroni procedure.

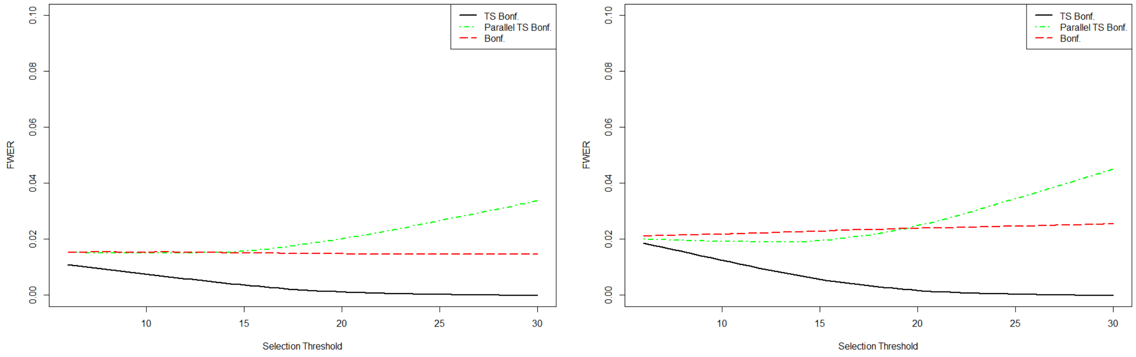


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

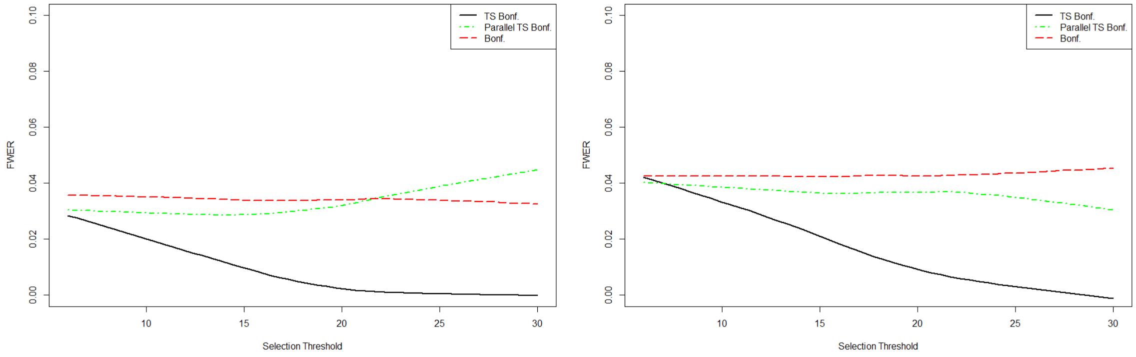


(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.4 Average power of the proposed selective parallel gatekeeping procedure (Parallel TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to 1. Here, the value of selection threshold t is from 6 to 30, $m = 100, n = 10, \alpha = 0.05$.

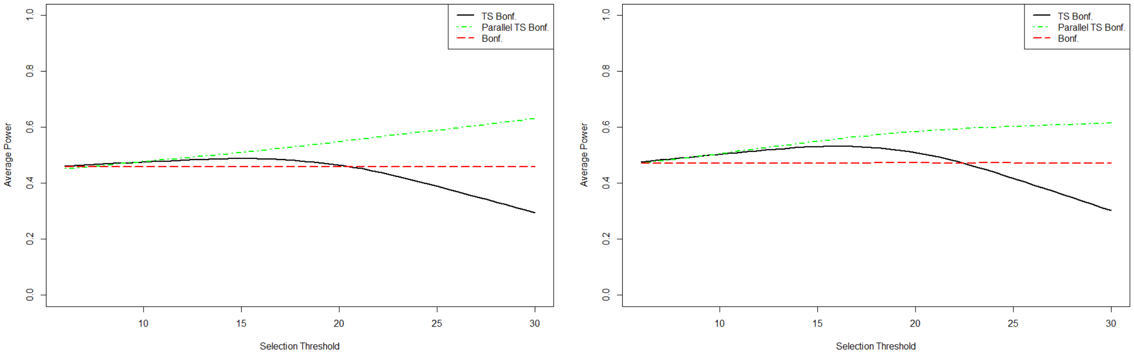


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

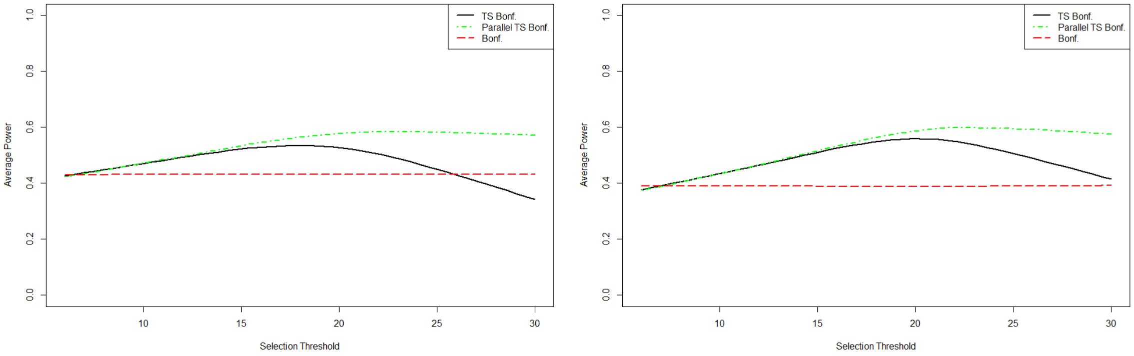


(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.5 Estimated FWER of the proposed selective parallel gatekeeping procedure (Parallel TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to 1. Here, the value of selection threshold t is from 6 to 30, $m = 100, n = 10, \alpha = 0.05$.

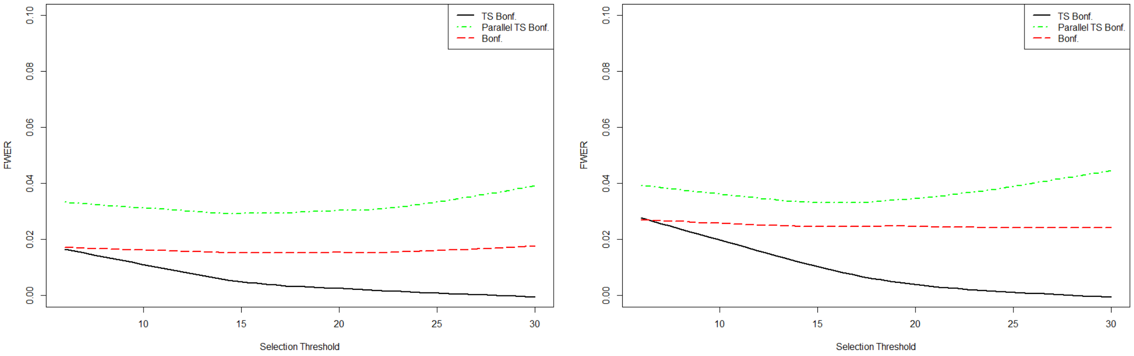


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

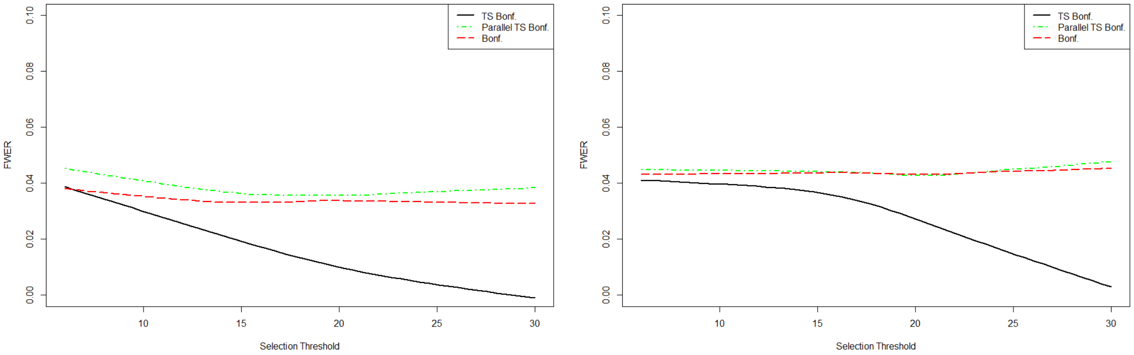


(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.6 Average power of the proposed selective parallel gatekeeping procedure (Parallel TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to 1. Here, the value of selection threshold t is from 6 to 30, $m = 100, n = 15, \alpha = 0.05$.



(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$



(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.7 Estimated FWER of the proposed selective parallel gatekeeping procedure (Parallel TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to 1. Here, the value of selection threshold t is from 6 to 30, $m = 100, n = 15, \alpha = 0.05$.

In this simulation study, we consider simultaneously testing $H_i : \mu_i = 0$ vs. $H_i^{'} : \mu_i \neq 0$, for $i = 1, \ldots, m$, at level $\alpha = 0.05$. Each simulated data set is obtained by generating $m = 100$ independent normal random samples $N(\mu_i, \sigma^2)(i = 1, \ldots, m)$ of equal sample size $n = 15$. Among the 100 hypotheses, that is among $100\mu_i^{'}s$, $100\pi_0$ are equal to 0, and the remaining are equal to $\mu_i = 1$, where π_0 is the proportion of $\mu_i = 0$. And the population variance σ^2 is drawn from $U(0.5, 1.5)$ to ensure that the population variance is unknown. For all compared five procedures, we use one-sample t-statistics as our testing statistic. For our proposed three procedure and the two-stage Bonferroni procedure, we use the sum of squares as the selection statistic. The selection threshold we choose is a sequence of numbers from 6 to 40. The range of the selection threshold is broader than that for the selective parallel gatekeeping procedure. The reason we extend the range from $(6, 30)$ to $(6, 40)$ is that the larger the selection threshold, the more hypotheses are non-selected. And the more non-selected hypotheses, the more proper to exploit the information among them. Based on t, λ_0 can be computed as $\lambda_0 = \Pr(U_i > t)$. Thus, values of the other two turning parameters $\lambda_1 = 0.5\lambda_0$ and $\lambda_2 = 0.5 + 0.5\lambda_0$ are determined.

The weights W_1 and W_2 for S and $I\backslash S$ are computed according to (4.10) and (4.11). If the number of selection $|S| = 0$ or $|S| = m$, we perform the conventional Bonferroni procedure to all m hypotheses, i.e., $W_2 = 1$ or $W_1 = 1$. If we encounter $W_1 = W_2 = 0$, we also perform the conventional Bonferroni procedure to all m hypotheses. Moreover, if we encounter $W_1 = 0$ and $W_2 \neq 0$, we perform the adaptive Bonferroni procedure to $I\backslash S$ and vice versa.

Our simulation is repeated for 1,000 times. The simulated FWER is the proportion of times that at least one false rejection occurring among the 1,000 replications and the simulated power is the average of the proportions of correct rejections over 1,000 runs. As seen from Figure 5.8, among a certain range of selection threshold, the data-driven weighted selective procedure almost performs best in terms of average power with the compared procedures, except when $\pi_0 = 0.9$. When the selection threshold becomes strict, the average power of the two-tage Bonferroni procedure and the adaptive two-stage Bonferroni procedure drops dramatically. The proposed data-driven weighted selective procedure performs better than the two-tage Bonferroni procedure and the adaptive two-stage Bonferroni procedure in this situation, but getting close to the conventional Bonferroni procedure with the increased true null proportion. When the selection threshold becomes strict, the selective parallel gatekeeping procedure performs almost best in terms of average power with the compared procedures. And the adaptive two-stage Bonferroni procedure performs best when the selection threshold is mild. When μ_i of false null hypothesis H_i varies such that $\mu_i \sim U(0.5, 1.5)$, among a certain range of selection threshold, the selective parallel gatekeeping procedure almost performs best in terms of average power with the compared procedures, see Figure 5.10. In Figure 5.9 and Figure 5.11 we show that the proposed three procedures control the FWER at level α in both cases, i.e. when $\mu_i = 1$ and $\mu_i \sim U(0.5, 1.5)$.

Overall, by comparing the average power performance among all five procedure, there is no such procedure which is always more powerful than the others. In different practical situation, one needs to choose a most suitable procedure.

5.5 Dependent Structure

In this section, we aim to explore the performance of the proposed procedures in terms of average power and FWER control under dependent structure. Two types of dependent structures are considered, which are equal correlation in Section 5.5.1 and block dependence in Section 5.5.2.

5.5.1 Equal Correlation

In this simulation, under equal correlation of testing statistics among all tested hypotheses, we aim to explore the difference between the proposed selective inference procedures with two-stage Bonferroni and conventional Bonferroni procedure by comparing the average power and FWER control vs. correlation coefficient (ρ).

Our simulation runs 1000 times, where $B = 1000$. And we assign different values to correlation coefficient ρ from 0 to 0.95. The whole sample size $n = 15$. For each fixed B and t, we will simulate a data as follows to test $m = 100$ two-sided hypotheses:

$$H_i : \mu = 0 \text{ vs. } H_i^{'} : \mu \neq 0.$$

The sample is from a multivariated normal population, that is

$$X_i \sim MVN(\tilde{\mu}, \Sigma),$$

where the correlation $\rho_{ij} = \rho$ btween H_i and H_j for all $i, j = 1, \ldots, m$, and $i \neq j$. Among the 100 hypotheses, that is among $100\mu_i^{\prime}s$, $100\pi_0$ are equal to 0, and the remaining are equal to $\mu_i = 1$, where π_0 is the proportion of $\mu_i = 0$. And the population variance σ^2 is drawn from $U(0.5, 1.5)$ to ensure that the population variance is unknown.

In our simulation, we use an alternative way to simulate our data which follows the same distribution mentioned above:

$$X_i = \sqrt{\rho} Z_0 + \sqrt{1-\rho} Z_i + \mu_i,$$

where $Z_0, Z_i \sim N(0, \sigma^2)$, for $i = 1, \ldots, m$. For all compared five procedures, we use one-sample t-statistics as our testing statistic. For our proposed three procedure and the two-stage Bonferroni procedure, we use the sum of squares as the selection statistic. The selection threshold we choose is $t = c(15, 20, 25, 40)$, which gurantees $50\%, 20\%, 5\%$ and $< 1\%$ of hypotheses are selected, respectively.

For known $\mu_i = 1$, as seen from Figure 5.12, when we select $50\%(t = 15)$, $20\%(t = 20)$ and $5\%(t = 25)$ of tested hypotheses, adaptive two-stage Bonferroni procedure performs best regarding to average power for $\pi_0 = 0.5$. However, the FWER of adaptive two-stage Bonferroni procedure is out of control, see Figure 5.13. When the number of selected hypotheses is less than $1\%(t = 40)$ and $\pi_0 = 0.5$, the average power of selective parallel gatekeeping procedure performs best with the control of FWER in a certain range of ρ, see Figures 5.12 and 5.13. Similarily, we observe the same pattern when μ_i of false null hypotheses is unknwn with a uniform distribution, i.e. $\mu_i \sim U(0.5, 1.5)$ for both $\pi_0 = 0.5$, see Figure 5.14 and Figure 5.15.

From Figure 5.16 and Figure 5.17, we also investigate the average power performance and estimated FWER with different selection threshold t. Adaptive two-stage Bonferroni procedure performs best when t is mild, however its estimated FWER is out of control. When t becomes strict, selective parallel gatekeeping procedure almost performs best, and its estimated FWER is around level $\alpha = 0.05$, with a slightly greater than 0.05 at certain points. This may be due to the repeated time of simulation is only $1,000$. However, for lower true null proportion, data-driven weighted selective procedure performs best when t is strict with the control of FWER.

5.5.2 Block Dependence

In this simulation, under block dependent structure of testing statistics among all tested hypotheses, we aim to explore the difference between the proposed selective inference procedures with two-stage Bonferroni and conventional Bonferroni procedure by comparing the average power and FWER control vs. correlation coefficient (ρ).

Our simulation runs 1000 times, where $B = 1000$. And we assign different values to correlation coefficient ρ from 0 to 0.95. The whole sample size $n = 15$. For each fixed B and t, we will simulate a data as follows to test $m = 100$ two-sided hypotheses:

$$H_i : \mu = 0 \text{ vs. } H_i^{'} : \mu \neq 0.$$

Within each block, a sample is from a multivariated normal population, that is

$$X_i \sim MVN(\tilde{\mu}, \Sigma),$$

where the correlation $\rho_{ij} = \rho$ btween H_i and H_j for all $i, j = 1, \ldots, m$, and $i \neq j$. Among the 100 hypotheses, that is among $100\mu_i^{\prime}s$, $100\pi_0$ are equal to 0, and the remaining are equal to $\mu_i = 1$, where π_0 is the proportion of $\mu_i = 0$. And the population variance $\sigma^2 = 1$

In our simulation, we use an alternative way to simulate each block data which follows the same distribution mentioned above:

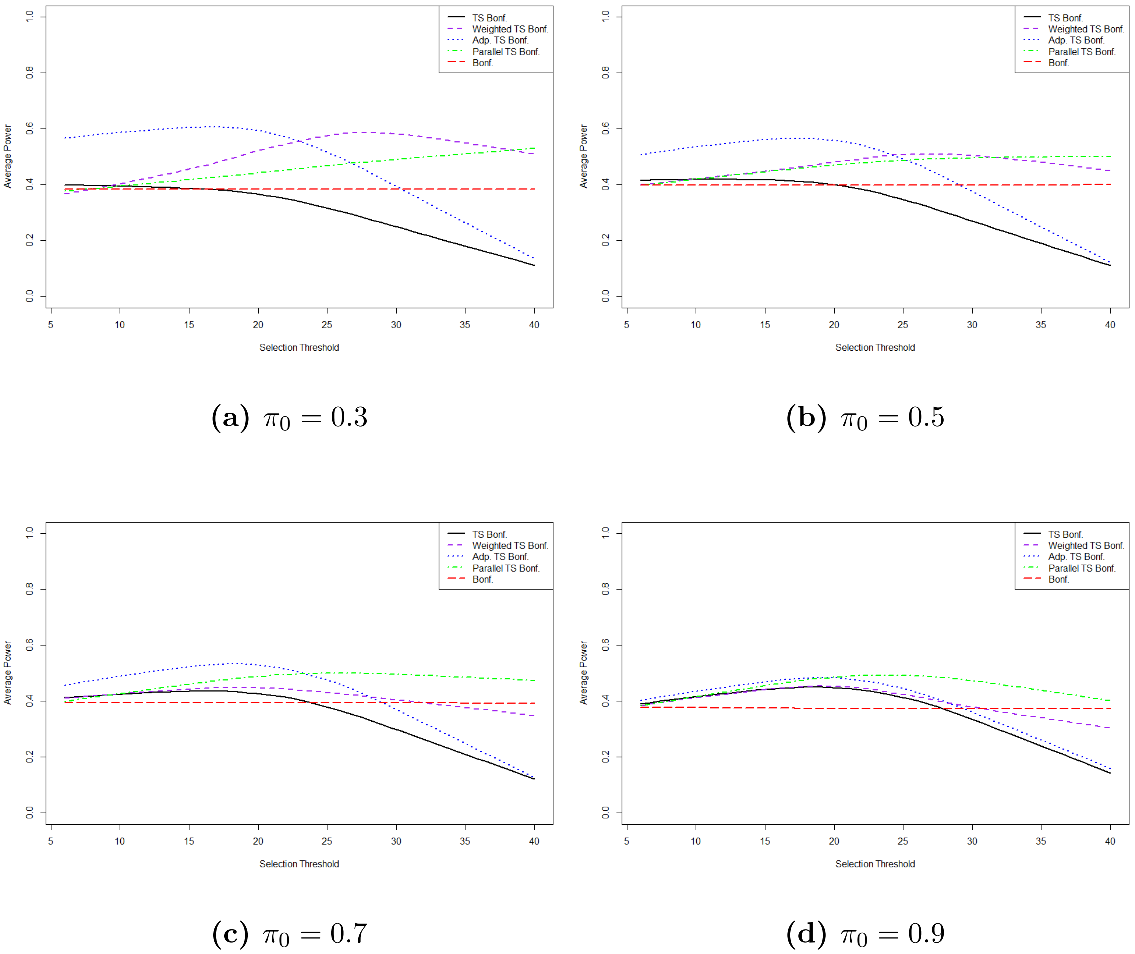
$$X_i = \sqrt{\rho} Z_0 + \sqrt{1-\rho} Z_i + \mu_i,$$

where $Z_0, Z_i \sim N(0, \sigma^2)$, for $i = 1, \ldots, m$. Overall, we have K blocks with size q, $K \times q = m$. Within each block, we have the same true null proportion π_0. Between blocks, the testing statistics are independent while within blocks, they are dependent with same correlation coefficient ρ. We use two block settings, that is from $K = 50$ to $K = 5$. With the decreaseing of K, the block dependence becomes stronger.

As seen from Figure 5.19, when the selection threshold $t = 15$ and $t = 20$, which guarantees approximately 50% and 20% of hypotheses are selected,respectively, the proposed three procedures almost perform better than the compared two procedures. While $t = 40$, the selective parallel gatekeeping procedure has the best power performance. The estimated FWER almost control at level α, though in some cases, it is a slightly greater than α. This may be due to the fact that our simulation only runs $1,000$.

5.6 Summary and Future Work

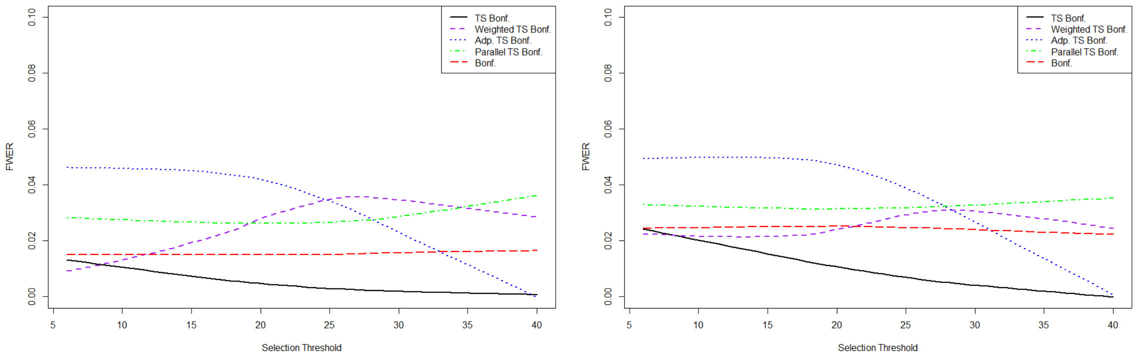
Through extensive simulation studies, we evaluate the proposed two-stage procedures by comparing the average power performance among all five procedure, and find out that there is no such procedure which is always more powerful than the others with proper control of FWER under independence. In different practical situations, one needs to choose a most suitable procedure. And our proposed two-stage procedures perform well under block dependence in terms of estimated FWER, while they may lose FWER control under equal correlation in some scenario. In the future, additional simulation studies will be conducted for the generalized three procedures: blockwise adaptive two-stage Bonferroni procedure, blockwise selective parallel gatekeeping procedure and blockwise data-driven weighted selective procedure.



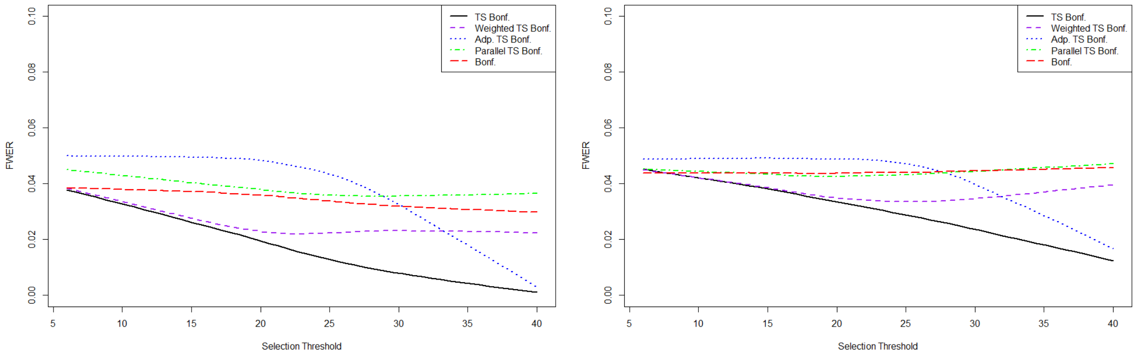
(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.8 Average power of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to 1. Here, the value of selection threshold t is from 6 to 40, $m = 100, n = 15, \alpha = 0.05$.

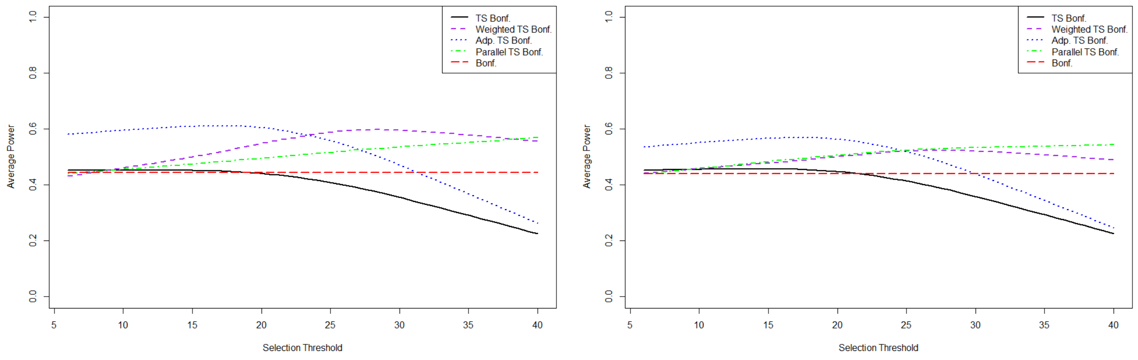


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

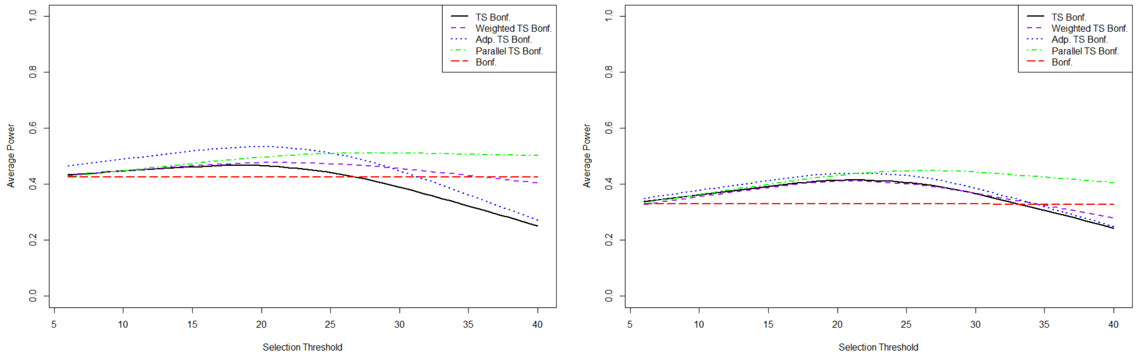


(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.9 Estimated FWER of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9, and variance $\sigma^2 \sim U(0.5, 1.5)$. For the mean values, $100\pi_0$ are equal to 0 and the rest are equal to 1. Here, the value of selection threshold t is from 6 to 40, $m = 100, n = 15, \alpha = 0.05$.

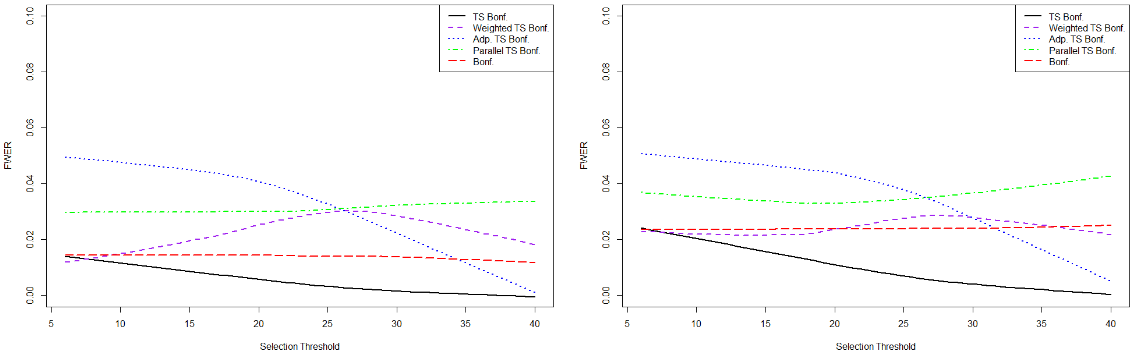


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

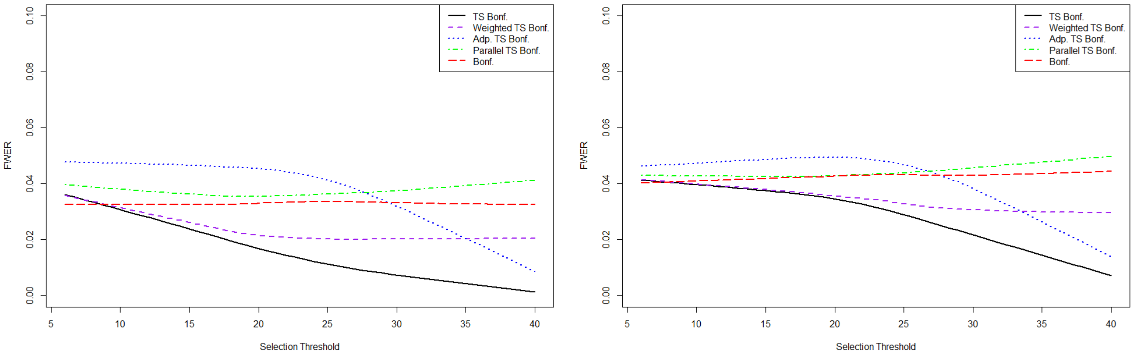


(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.10 Average power of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9. For false null hypothesis H_i, $\mu_i \sim U(0.5, 1.5)$ and variance $\sigma^2 \sim U(0.5, 1.5)$. Here, the value of selection threshold t is from 6 to 40, $m = 100, n = 15, \alpha = 0.05$.

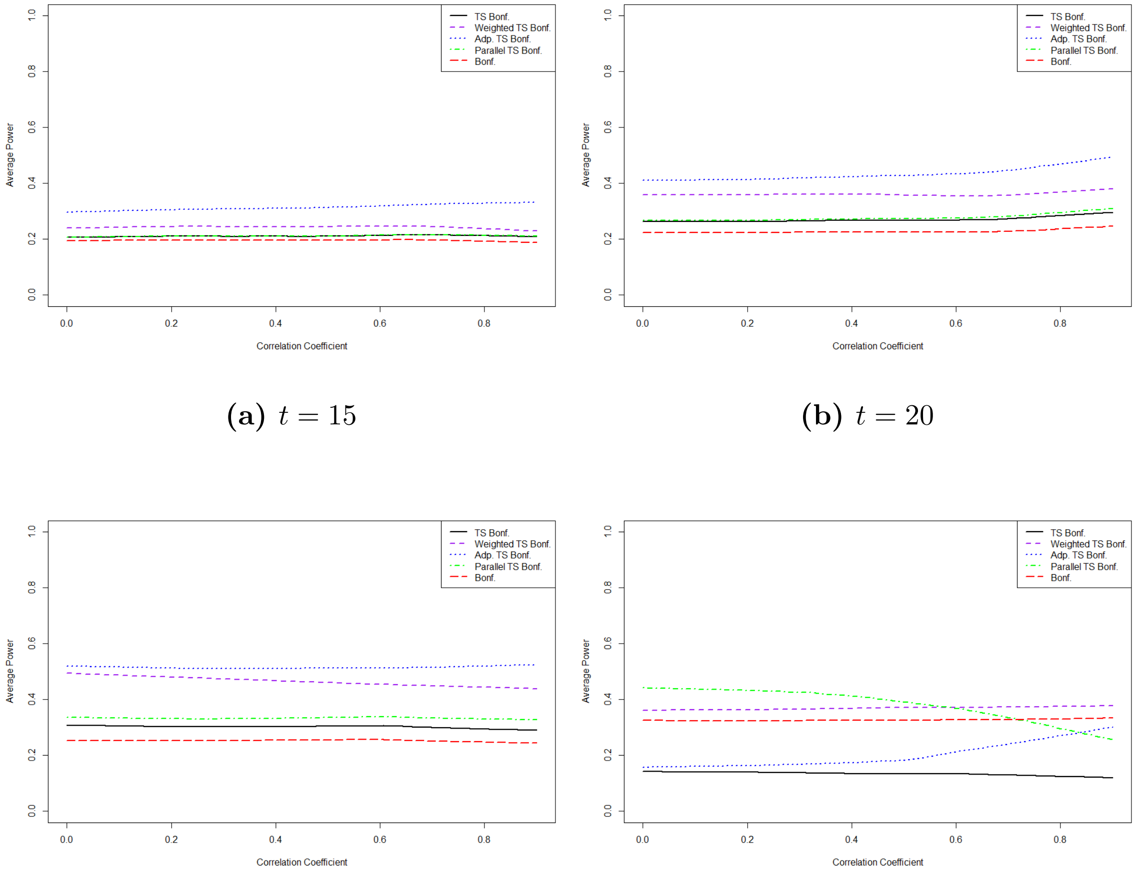


(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$



(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

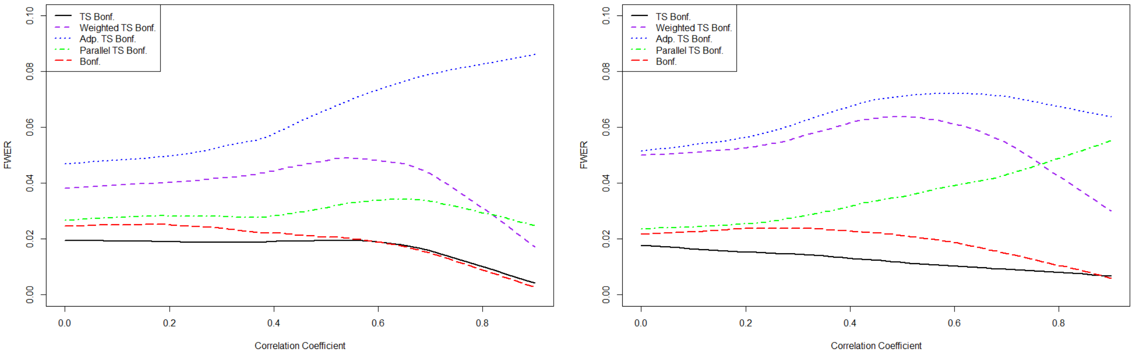
Figure 5.11 Estimated FWER of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion π_0 from 0.3 to 0.9. For false null hypothesis H_i, $\mu_i \sim U(0.5, 1.5)$ and variance $\sigma^2 \sim U(0.5, 1.5)$. Here, the value of selection threshold t is from 6 to 40, $m = 100, n = 15, \alpha = 0.05$.



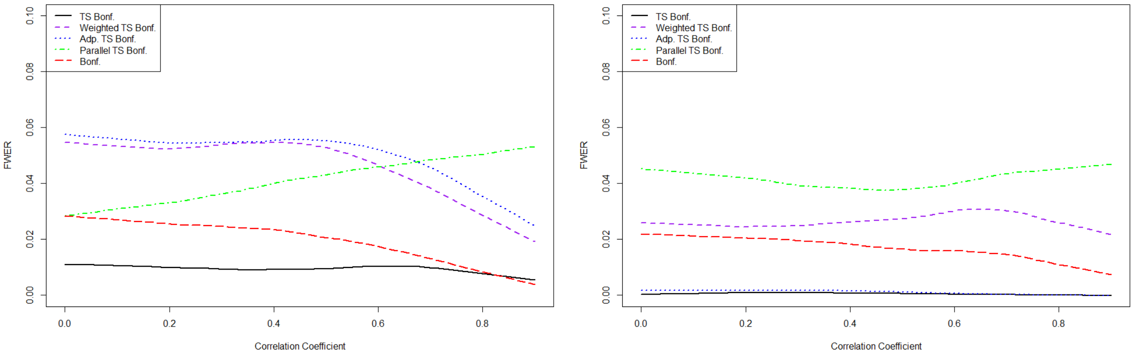
(a) $t = 15$ **(b)** $t = 20$

(c) $t = 25$ **(d)** $t = 40$

Figure 5.12 Under equal correlation, average power of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportions $\pi_0 = 0.5$, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i = 1$, and $\sigma_i^2 = \sigma^2 \sim U(0.5, 1.5)$. The correlation coefficient ρ takes values from 0 to 0.95.

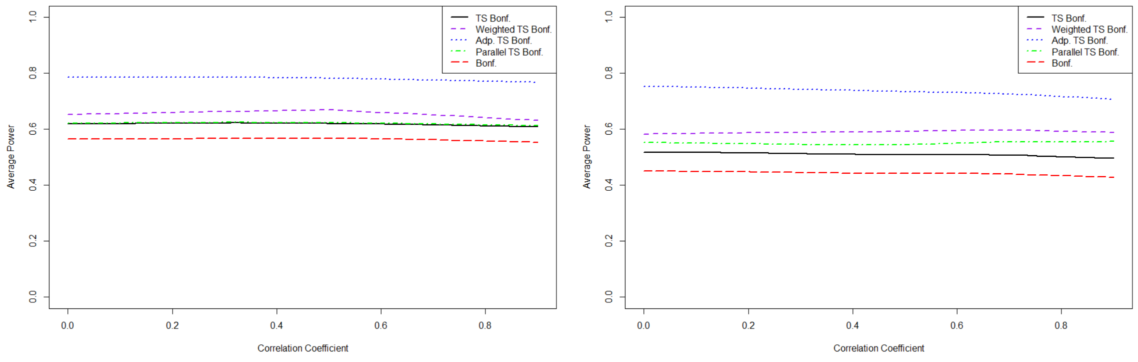


(a) $t = 15$ **(b)** $t = 20$

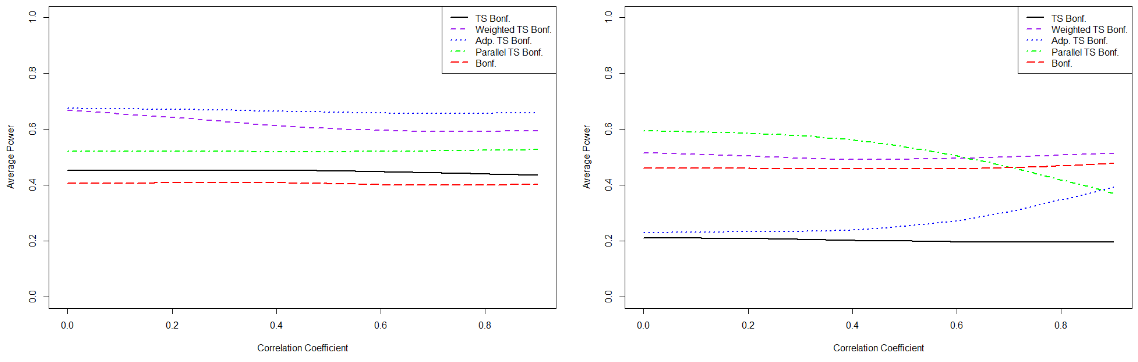


(c) $t = 25$ **(d)** $t = 40$

Figure 5.13 Under equal correlation, estimated FWER of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportions $\pi_0 = 0.5$, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i = 1$, and $\sigma_i^2 = \sigma^2 \sim U(0.5, 1.5)$. The correlation coefficient ρ takes values from 0 to 0.95.

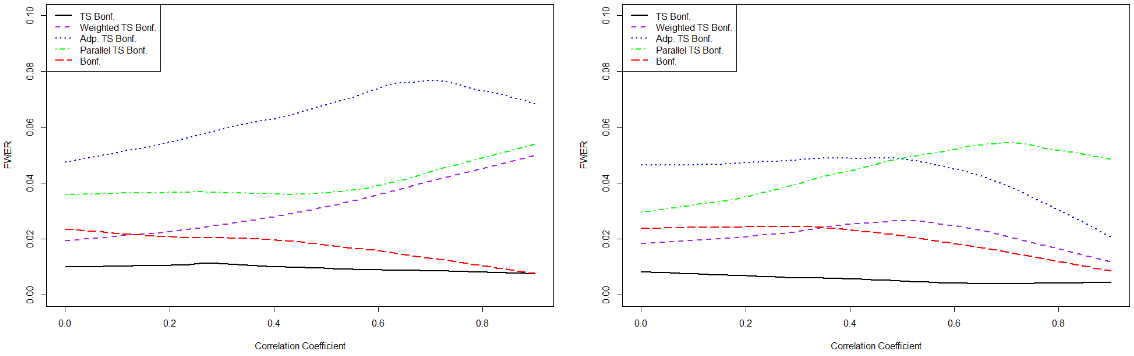


(a) $t = 15$ **(b)** $t = 20$

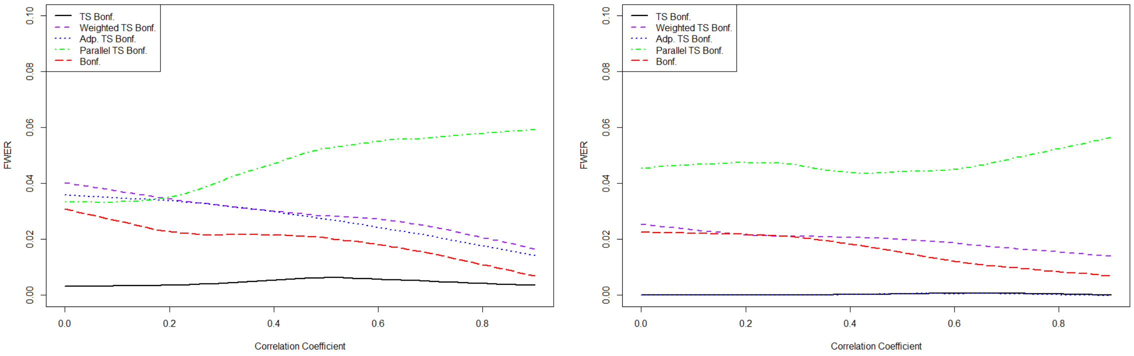


(c) $t = 25$ **(d)** $t = 40$

Figure 5.14 Under equal correlation, average power of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportions $\pi_0 = 0.5$, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i \sim U(0.5, 15)$, and $\sigma_i^2 = \sigma^2 \sim U(0.5, 1.5)$. The correlation coefficient ρ takes values from 0 to 0.95.

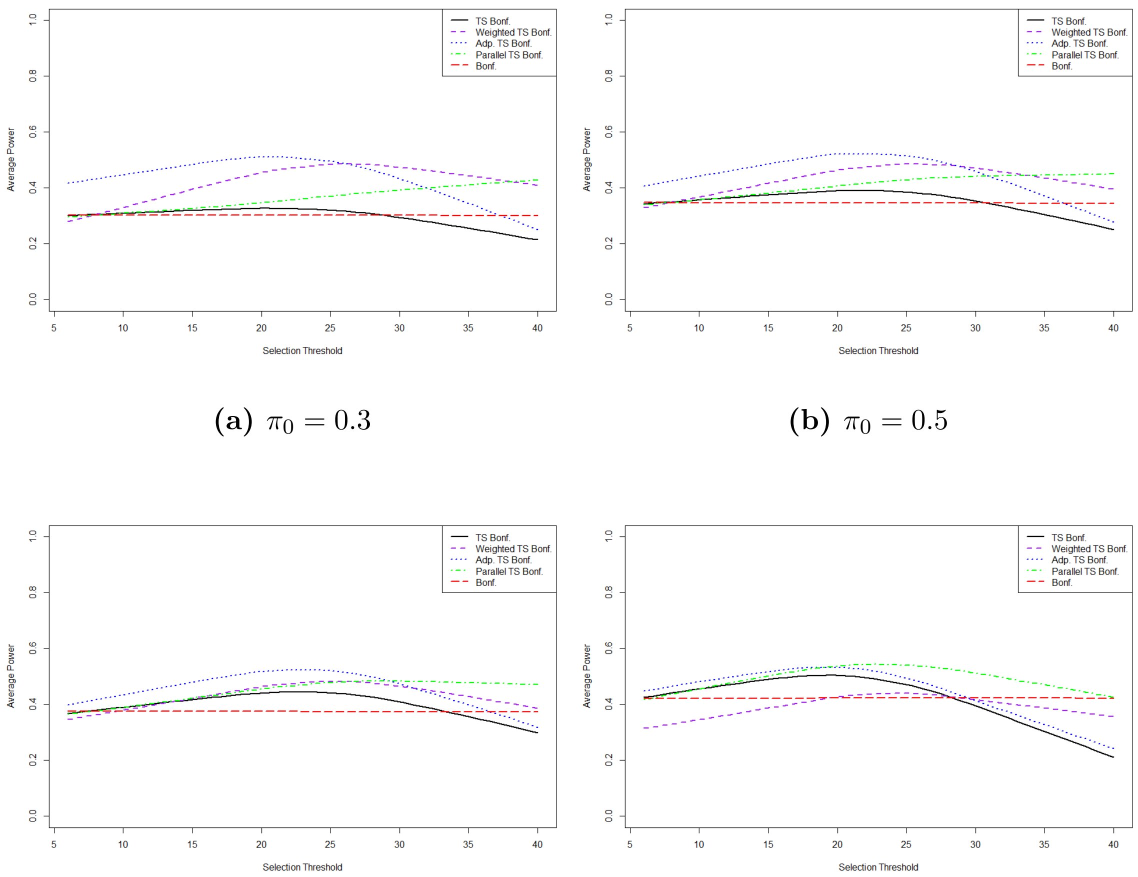


(a) $t = 15$ **(b)** $t = 20$



(c) $t = 25$ **(d)** $t = 40$

Figure 5.15 Under equal correlation, estimated FWER of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportions $\pi_0 = 0.5$, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i \sim U(0.5, 15)$, and $\sigma_i^2 = \sigma^2 \sim U(0.5, 1.5)$. The correlation coefficient ρ takes values from 0 to 0.95.



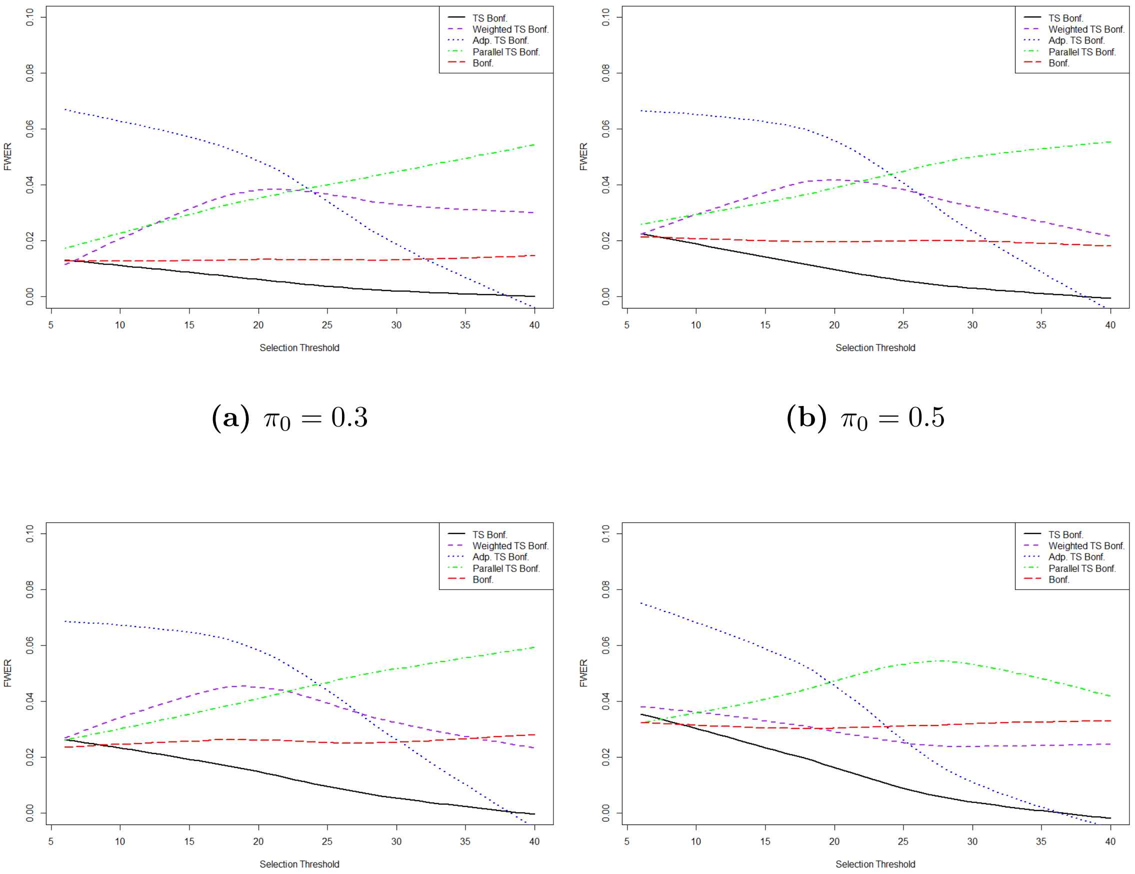
(a) $\pi_0 = 0.3$

(b) $\pi_0 = 0.5$

(c) $\pi_0 = 0.7$

(d) $\pi_0 = 0.9$

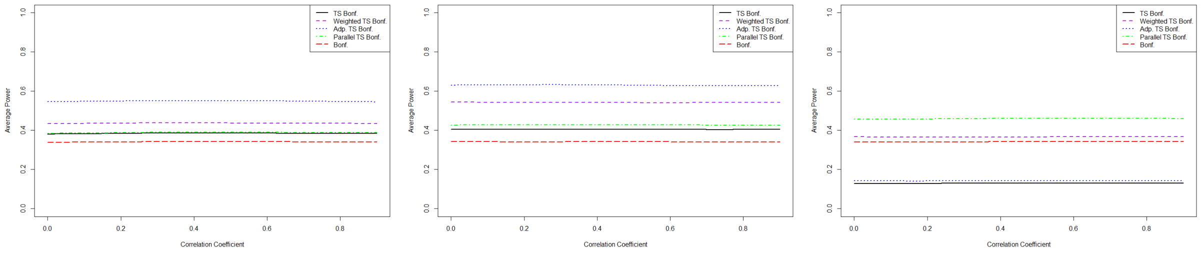
Figure 5.16 Under equal correlation with correlation coefficient $\rho = 0.5$, average power of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with different true null proportions π_0, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i \sim U(0.5, 15)$, and $\sigma_i^2 = \sigma^2 \sim U(0.5, 1.5)$.



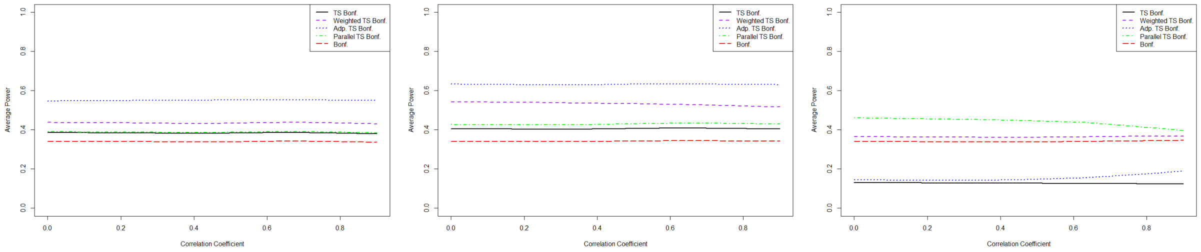
(a) $\pi_0 = 0.3$ **(b)** $\pi_0 = 0.5$

(c) $\pi_0 = 0.7$ **(d)** $\pi_0 = 0.9$

Figure 5.17 Under equal correlation with correlation coefficient $\rho = 0.5$, estimated FWER of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with different true null proportions π_0, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i \sim U(0.5, 15)$, and $\sigma_i^2 = \sigma^2 \sim U(0.5, 1.5)$.

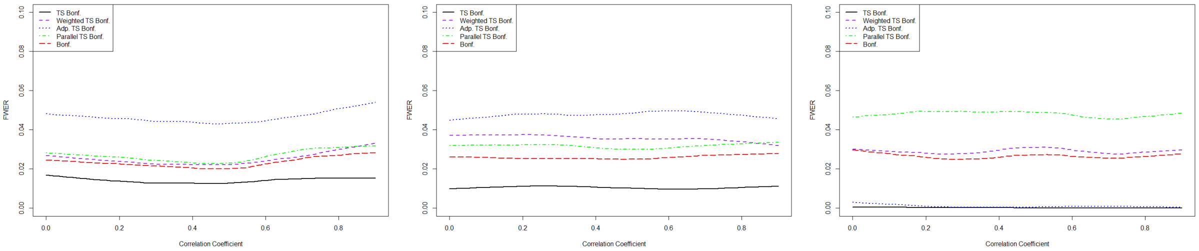


(a) $block = 50, t = 15$ **(b)** $block = 50, t = 20$ **(c)** $block = 50, t = 40$

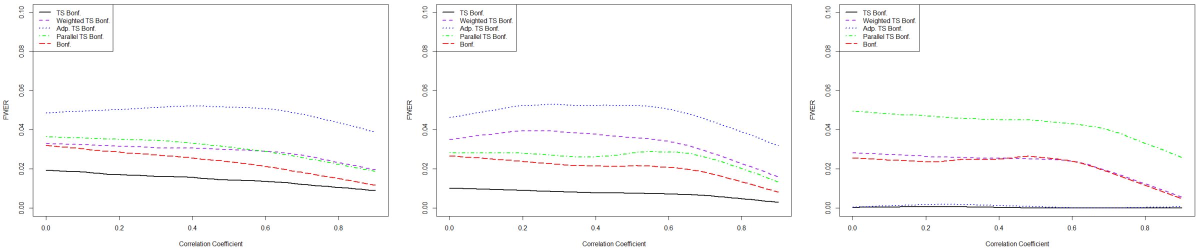


(d) $block = 5, t = 15$ **(e)** $block = 5, t = 20$ **(f)** $block = 5, t = 40$

Figure 5.18 Under block dependent structure, average power of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion $\pi_0 = 0.5$, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i = 1$, and $\sigma_i^2 = \sigma^2 = 1$. The correlation coefficient ρ takes values from 0 to 0.95.



(a) $block = 50, t = 15$ **(b)** $block = 50, t = 20$ **(c)** $block = 50, t = 40$



(d) $block = 5, t = 15$ **(e)** $block = 5, t = 20$ **(f)** $block = 5, t = 40$

Figure 5.19 Under block dependent structure, estimated FWER of the proposed data-driven weighted selective procedure (Weighted TS Bonf.), selective parallel gatekeeping procedure (Parallel TS Bonf.), adaptive two-stage Bonferroni procedure (Adp.TS Bonf.) along with two-stage Bonferroni procedure (TS Bonf.) and the conventional Bonferroni procedure (Bonf.), with true null proportion $\pi_0 = 0.5$, when $n = 15$, $m = 100$, for false null hypothesis H_i, $\mu_i = 1$, and $\sigma_i^2 = \sigma^2 = 1$. The correlation coefficient ρ takes values from 0 to 0.95.

CHAPTER 6

SUMMARY AND FUTURE WORK

In this dissertation, we first introduce a novel graphical approach, which allows one to reject more than one hypothesis at each step. This overcomes a main drawback of existing graphical approaches in which only one rejection is allowed at each step. Through clinical trail examples, we illustrate that the proposed approach is more flexible and computationally efficient than existing graphical approaches. As a by-product, we propose a generalized sequential rejection principle to show the FWER control of the proposed approach. Then, in Chapter 3 we propose a graphical approach for general logically restricted hypotheses testing. By re-assigning critical values between testable and non-testable (logically restricted) hypotheses, all critical values are made fully used. Through clinical trial examples, we demonstrate that the proposed graphical approach is more simple and flexible than few existing approaches for logically restricted hypotheses testing. Finally, we introduce three powerful two-stage selective inference procedures as well as their generalized procedures to deal with large scale data analysis in Chapter 4 and extensive simulation studies are provided in Chapter 5. The future works are discussed in the following.

For proposed graphical approaches in Chapter 2 and 3, a main work we plan to do is to implement them in R package. Thus users can easily and quickly conduct our approaches.

For proposed selective inference procedures, extensive simulation studies are to be done to evaluate and compare the performance of the proposed blockwise procedures with the proposed two-stage procedures and those of several existing methods in terms of FWER control and average power. We also plan to apply the constructed data-driven weights in Chapter 4 to the proposed weighted Bonferroni-based graphical approaches in Chapter 2 and 3 to develop newer and more powerful graphical approaches.

APPENDIX A

THE GENERALIZED GRAPHICAL APPROACH

A.1 Proof of (2.9) in Proposition 1

We use mathematical induction to show that (2.9) holds, that is

$$\sum_{k \in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1 \text{ for all } i = 1, ..., m.$$

For $i = 1$, $I_1 = \{1, ..., m\}$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \emptyset$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}$, by $\sum_{k=1}^{m} g_{lk} \leq 1$ in (2.4), we prove that $\sum_{k \in I_1} g_{lk}(\hat{R}_0) \leq 1$.

By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $\sum_{k \in I_{s-1}} g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq 1$ for every $l \in I_{s-1}$. Then for $i = s$, by the updating rule of transition coefficient function in Algorithm 1, for every fixed $l \in I_s$, we have

$$\begin{aligned}\sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-1}) &= \sum_{k \in I_s^{(l)}} \frac{g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2})}{1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})} \\ &= \frac{\sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right)}{1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})},\end{aligned} \tag{A.1}$$

where $I_s^{(l)} = I_s \backslash \{l\}$. Expanding the numerator of the right-side in (A.1), we have

$$\begin{aligned}
&\sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right) \\
=& \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{k \in I_s^{(l)}} \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \\
=& \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) \sum_{k \in I_s^{(l)}} g_{jk}(\hat{\mathbf{R}}_{s-2}) \\
\leq& \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) \left(1 - \sum_{q \in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}) - g_{jl}(\hat{\mathbf{R}}_{s-2}) \right) \\
=& \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) \sum_{q \in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}) \\
&- \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
=& \sum_{k \in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
\leq& \ 1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}),
\end{aligned}$$

where the first inequality holds since

$$\sum_{k \in I_{s-1}} g_{jk}(\hat{\mathbf{R}}_{s-2}) = \sum_{k \in I_s^{(l)}} g_{jk}(\hat{\mathbf{R}}_{s-2}) + g_{jl}(\hat{\mathbf{R}}_{s-2}) + \sum_{q \in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}) \leq 1,$$

thus,

$$\sum_{k \in I_s^{(l)}} g_{jk}(\hat{\mathbf{R}}_{s-2}) \leq 1 - g_{jl}(\hat{\mathbf{R}}_{s-2}) - \sum_{q \in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}).$$

The last equality holds since $I_s^{(l)} \cup J_{s-1} = I_{s-1}^{(l)}$, and $I_s^{(l)} \cap J_{s-1} = \emptyset$. And the second inequality holds since

$$\sum_{k \in I_{s-1}} g_{lk}(\hat{\mathbf{R}}_{s-2}) = \sum_{k \in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{l=k} g_{kk}(\hat{\mathbf{R}}_{s-2}) = \sum_{k \in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq 1.$$

Therefore, we have

$$\frac{\sum\limits_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right)}{1 - \sum\limits_{j \in J_{s-1}} \sum\limits_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})} \leq 1,$$

which implies $\sum\limits_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq 1$ for $i = s$. Moreover, we know that $g_{kk}(\hat{\mathbf{R}}_{s-1}) = 0$ is always true based on the updating rule of transition coefficient function in Algorithm 1. Thus $\sum\limits_{k \in I_s} g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq 1$ for $i = s$.

Summarizing the above arguments, $\sum\limits_{k \in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1$ holds for all $i = 1, ..., m$. ☐

A.2 Proof (2.10) in Proposition 1

We use mathematical induction to show (2.10) holds, that is

$$0 \leq g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1 \text{ for all } i = 1, ..., m.$$

For $i = 1$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \emptyset$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}$. By $0 \leq g_{lk} \leq 1$ in (2.3), we prove that $0 \leq g_{lk}(\hat{\mathcal{R}}_0) \leq 1$.

By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $0 \leq g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq 1$ for all $l, k \in I_{s-1}$. Then for $i = s$, by the updating rule of the transition coefficient function in Algorithm 1, for all $l, k \in I_s, l \neq k$, we have

$$g_{lk}(\hat{\mathbf{R}}_{s-1}) = \frac{g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2})}{1 - \sum\limits_{j \in J_{s-1}} \sum\limits_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})}, \quad \text{(A.2)}$$

where the numerator of the right-side in (A.2) is always non-negative due to the assumption that $0 \leq g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq 1$. Moreover, in A.1 we showed that

$$\sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right)$$

$$\leq \; 1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}).$$

And by the assumption again, we have

$$\sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right) \geq 0, \tag{A.3}$$

thus, $1 - \sum\limits_{j \in J_{s-1}} \sum\limits_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \geq 0$, which is the denominator of the right-side in (A.2). And when

$$1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) = 0,$$

we set $g_{lk}(\hat{\mathbf{R}}_{s-1}) = 0$ as presented in the article. Therefore, $g_{lk}(\hat{\mathbf{R}}_{s-1}) \geq 0$. Since we showed in A.1 that $\sum\limits_{k \in I_s} g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq 1$, thus $g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq 1$.

Therefore, $0 \leq g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1$ holds for all $i = 1, ..., m$. When $l = k$, it is always true that $g_{kk}(\hat{\mathbf{R}}_{i-1}) = 0$ for all $i = 1, ..., m$, based on the updating rule of transition coefficient function in Algorithm 1. □

A.3 Proof of (2.11) in Proposition 1

We use mathematical induction to show that (2.11) holds, that is

$$\sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha \text{ for all } i = 1, ..., m.$$

For $i = 1$, $I_1 = \{1, ..., m\}$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \emptyset$, and $\alpha_l(\hat{\mathcal{R}}_0) = \alpha_l$. By $\sum_{l=1}^{m} \alpha_l \leq \alpha$ in (2.2), we prove that

$$\sum_{l \in I_i} \alpha_l(\hat{\mathcal{R}}_0) \leq \alpha.$$

By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $\sum_{l \in I_{s-1}} \alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha$. Then for $i = s$, by the updating rule of the critical value function in Algorithm 1, we have

$$\begin{aligned}
\sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-1}) &= \sum_{l \in I_s} \left(\alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) \alpha_j(\hat{\mathbf{R}}_{s-2}) \right) \\
&= \sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{l \in I_s} \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) \alpha_j(\hat{\mathbf{R}}_{s-2}) \\
&= \sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} \alpha_j(\hat{\mathbf{R}}_{s-2}) \sum_{l \in I_s} g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
&\leq \sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} \alpha_j(\hat{\mathbf{R}}_{s-2}) = \sum_{l \in I_{s-1}} \alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha,
\end{aligned}$$

where the first inequality holds since for non-negative transition coefficients functions, and $I_s \subseteq I_{s-1}$, we have $\sum_{l \in I_s} g_{jl}(\hat{\mathbf{R}}_{s-2}) \leq \sum_{l \in I_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) \leq 1$. The last equality holds since $I_s \cup J_{s-1} = I_{s-1}$, and $I_s \cap J_{s-1} = \emptyset$. And the last inequality is due to the assumptation that $\sum_{l \in I_{s-1}} \alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha$.

Summarizing the above argument, $\sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha$ holds for all $i = 1, ..., m$. $\square$

A.4 Proof of monotonicity of transition coefficient function

We use mathematical induction to show that the monotonicity of transition coefficient function holds, that is

$$g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq g_{lk}(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m,$$

here, $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$, and $\mathbf{S}_{i-1} = (\mathcal{S}_1, ..., \mathcal{S}_{i-1})^T$.

For $i = 1$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \mathcal{S}_0 = \emptyset$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}(\mathcal{S}_0) = g_{lk}$. Thus, we prove that the monotonicity of trasnsition coefficient function holds when $i = 1$.

By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq g_{lk}(\mathbf{S}_{s-2})$ for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-2$, and $l, k \in I_{s-1}$. Then for $i = s$, when $l \neq k$, by the updating rule of transition coefficient function in Algorithm 1, we have

$$g_{lk}(\hat{\mathbf{R}}_{s-1}) = \frac{g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2})}{1 - \sum\limits_{j \in J_{s-1}} \sum\limits_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})} \tag{A.4}$$

and

$$g_{lk}(\mathbf{S}_{s-1}) = \frac{g_{lk}(\mathbf{S}_{s-2}) + \sum\limits_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jk}(\mathbf{S}_{s-2})}{1 - \sum\limits_{j \in T_{s-1}} \sum\limits_{q \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jq}(\mathbf{S}_{s-2}) - \sum\limits_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jl}(\mathbf{S}_{s-2})}, \tag{A.5}$$

where T_{s-1} is the index set of a collection of rejections in $\mathcal{S}_{s-1}$. For non-negative transition coefficient function (showed in A.2), we have

$$\begin{aligned}
& g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) && \text{(A.6)} \\
\leq \quad & g_{lk}(\mathbf{S}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jk}(\mathbf{S}_{s-2}) && \text{(A.7)} \\
\leq \quad & g_{lk}(\mathbf{S}_{s-2}) + \sum_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jk}(\mathbf{S}_{s-2}), && \text{(A.8)}
\end{aligned}$$

where (A.6) is the numerator of the right-side in (A.4), and the right-side in (A.8) is the numerator of the right-side in (A.5). The inequality in (A.7) holds due to the assumptation made at step $s-1$. And the inequality in (A.8) holds since $J_{s-1} \subseteq T_{s-1}$. Moreover, for both denominators of right-side in (A.4), (A.5), we have

$$\begin{aligned}
& 1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
\geq \quad & 1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jq}(\mathbf{S}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jl}(\mathbf{S}_{s-2}) && \text{(A.9)} \\
\geq \quad & 1 - \sum_{j \in T_{s-1}} \sum_{q \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jq}(\mathbf{S}_{s-2}) - \sum_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jl}(\mathbf{S}_{s-2}), && \text{(A.10)}
\end{aligned}$$

the inequality in (A.9) holds due to the assumptation made at step $s-1$. And the inequality in (A.10) holds since $J_{s-1} \subseteq T_{s-1}$. Therefore, by the above argument, we prove that when $l \neq k$, $g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq g_{lk}(\mathbf{S}_{s-2})$. When $l = k$, it is always true that $g_{lk}(\hat{\mathbf{R}}_{s-2}) = g_{lk}(\mathbf{S}_{s-2}) = 0$ by the updating rule of transition coefficient function in Algorithm 1. Therefore, $g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq g_{lk}(\mathbf{S}_{s-1})$, for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-1$.

Summarizing the above argument, we conclude that

$$g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq g_{lk}(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m. \ \square$$

A.5 Proof of (2.12) in Proposition 1

We use mathematical induction to show that (2.12) holds, that is

$$\alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m,$$

where, $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$, and $\mathbf{S}_{i-1} = (\mathcal{S}_1, ..., \mathcal{S}_{i-1})^T$.

For $i = 1$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \mathcal{S}_0 = \emptyset$, and $\alpha_l(\hat{\mathcal{R}}_0) = \alpha_l(\mathcal{S}_0) = \alpha_l$. Thus, we prove that (2.12) holds when $i = 1$.

By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $\alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha_l(\mathbf{S}_{s-2})$ for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-2$. Then for $i = s$, by the updating rule of critical value function in Algorithm 1, we have

$$\alpha_l(\hat{\mathbf{R}}_{s-1}) = \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) \alpha_j(\hat{\mathbf{R}}_{s-2}) \tag{A.11}$$

and

$$\alpha_l(\mathbf{S}_{s-1}) = \alpha_l(\mathbf{S}_{s-2}) + \sum_{j \in T_{s-1}} g_{jl}(\mathbf{S}_{s-2}) \alpha_j(\mathbf{S}_{s-2}), \tag{A.12}$$

where T_{s-1} is the index set of a collection of rejections in $\mathcal{S}_{s-1}$. The right-side in (A.11) has the following property by the monotonicity of transition coefficient function showed in A.4

$$
\begin{aligned}
& \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2})\alpha_j(\hat{\mathbf{R}}_{s-2}) && \text{(A.13)} \\
\leq \quad & \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\hat{\mathbf{R}}_{s-2}) && \text{(A.14)} \\
\leq \quad & \alpha_l(\mathbf{S}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\mathbf{S}_{s-2}) && \text{(A.15)} \\
\leq \quad & \alpha_l(\mathbf{S}_{s-2}) + \sum_{j \in T_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\mathbf{S}_{s-2}). && \text{(A.16)}
\end{aligned}
$$

The inequality in (A.15) holds due to the assumptation made at step $s-1$, i.e., $\alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha_l(\mathbf{S}_{s-2})$. And the inequality in (A.16) holds since $J_{s-1} \subseteq T_{s-1}$. Therefore, by the above argument, we prove that $\alpha_l(\hat{\mathbf{R}}_{s-1}) \leq \alpha_l(\mathbf{S}_{s-1})$, for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-1$.

Summarizing the above arguments, we conclude that

$$\alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m. \ \square$$

A.6 Proof of Theorem 3

We apply the extended sequention rejective principle in Theorem 2 to prove the FWER control of the proposed graphical approach. First, we show that the monotonicity condition of the successor function in Theorem 2 is satisfied. The successor function is defined as follows:

$$\mathcal{N}(\hat{\mathbf{R}}_{i-1}) = \hat{\mathcal{R}}_i = \{H_l \in \mathcal{H} \backslash \cup_{k=1}^{i-1} \hat{\mathcal{R}}_k : p_l \leq \alpha_l(\hat{\mathbf{R}}_{i-1})\}$$

and

$$\mathcal{N}(\mathbf{S}_{i-1}) = \mathcal{S}_i = \{H_l \in \mathcal{H} \backslash \cup_{k=1}^{i-1} \mathcal{S}_k : p_l \leq \alpha_l(\mathbf{S}_{i-1})\},$$

where $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$, and $\mathbf{S}_{i-1} = (\mathcal{S}_1, ..., \mathcal{S}_{i-1})^T$.

In A.5, we showed that

$$\alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, \text{ where } j = 1, ..., i-1, \text{ and for all } i = 1, ..., m.$$

Thus, for every $H_l \in \mathcal{N}(\hat{\mathbf{R}}_{i-1})$, we have $p_l \leq \alpha_l(\hat{\mathbf{R}}_{i-1})$, by the monotonicity of critical value function, we show that $p_l \leq \alpha_l(\mathbf{S}_{i-1})$, thus, $H_l \in \mathcal{N}(\mathbf{S}_{i-1})$. Therefore, $\mathcal{N}(\hat{\mathbf{R}}_{i-1}) \subseteq \mathcal{N}(\mathbf{S}_{i-1})$, which satisfies the monotonicity condition of successor function in Theorem 2.

Next, we show that the singel step FWER control in Theorem 2 is also satisfied for the proposed approach. Assume we reject all false null hypotheses in previous $i-1$ steps, then we have

$$\begin{aligned} & P(\text{reject at least one true null hypothesis at step } i) \\ = \; & P(\mathcal{N}(\mathbf{F}_{i-1}) \nsubseteq \mathcal{F}) \\ = \; & P\{\cup_{j \in \mathcal{T}} \{p_j \leq \alpha_j(\mathbf{F}_{i-1})\}\} \\ \leq \; & \sum_{j \in \mathcal{T}} P(p_j \leq \alpha_j(\mathbf{F}_{i-1})) \leq \sum_{j \in \mathcal{T}} \alpha_j(\mathbf{F}_{i-1}) \leq \alpha, \end{aligned}$$

where $\mathcal{T}$ and $\mathcal{F}$ are collection of true null hypotheses and false null hypotheses, respectively. And $\mathbf{F}_{i-1} = (\mathcal{F}_1, ..., \mathcal{F}_{i-1})^T$, is a vector, which includes all rejected false null hypotheses, and $\mathcal{F}_1, ..., \mathcal{F}_{i-1}$ are all mutually exclusive sets. The first inequality is due to Bonferroni Inequality. The second inequality is under the assumption that all unadjusted p-values follow uniform distribution. The third inequality is due to the property of critical value function (shown in A.3), that is:

$$\sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha \text{ for all } i = 1, ..., m.$$

Thus, the single step FWER control is also satisfied for the proposed approach. By the extended sequention rejective principle, we prove that the proposed graphical approach strongly controls FWER at level α. □

APPENDIX B

THE GRAPHICAL APPROACH FOR LOGICALLY RELATED HYPOTHESES TESTING

B.1 Proof of (3.8) in Proposition 2

We use mathematical induction to show that (3.8) holds, that is

$$\sum_{k\in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1}) = 1 \text{ for all } i = 1, ..., m.$$

For $i = 1$, $I_1 = \{1, ..., m\}$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \emptyset$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}$, by $\sum_{k=1}^{m} g_{lk}$ =1 in (3.3), we prove that $\sum_{k\in I_1} g_{lk}(\hat{R}_0)$ =1.

By induction, we assume that for $i = s-1 (s \leq m)$, the inequality holds, i.e., $\sum_{k\in I_{s-1}} g_{lk}(\hat{\mathbf{R}}_{s-2})$ =1 for every $l \in I_{s-1}$. Then for $i = s$, by the updating rule of transition coefficient function in Algorithm 3, for every fixed $l \in I_s$, we have

$$\begin{aligned}\sum_{k\in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-1}) &= \sum_{k\in I_s^{(l)}} \frac{g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jk}(\hat{\mathbf{R}}_{s-2})}{1 - \sum_{j\in J_{s-1}}\sum_{q\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jl}(\hat{\mathbf{R}}_{s-2})} \\ &= \frac{\sum_{k\in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jk}(\hat{\mathbf{R}}_{s-2}) \right)}{1 - \sum_{j\in J_{s-1}}\sum_{q\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jl}(\hat{\mathbf{R}}_{s-2})},\end{aligned} \tag{1}$$

where $I_s^{(l)} = I_s \backslash \{l\}$. Expanding the numerator of the right-side in (1), we have

$$
\begin{aligned}
& \sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right) \\
= \quad & \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{k \in I_s^{(l)}} \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \\
= \quad & \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) \sum_{k \in I_s^{(l)}} g_{jk}(\hat{\mathbf{R}}_{s-2}) \\
= \quad & \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) \left(1 - \sum_{q \in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}) - g_{jl}(\hat{\mathbf{R}}_{s-2}) \right) \\
= \quad & \sum_{k \in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) \sum_{q \in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}) \\
& - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
= \quad & \sum_{k \in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) \\
& - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
= \quad & \sum_{k \in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + g_{kk}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) \\
& - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
= \quad & 1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}),
\end{aligned}
$$

where the third equality holds since

$$\sum_{k\in I_{s-1}} g_{jk}(\hat{\mathbf{R}}_{s-2}) = \sum_{k\in I_s^{(l)}} g_{jk}(\hat{\mathbf{R}}_{s-2}) + g_{jl}(\hat{\mathbf{R}}_{s-2}) + \sum_{q\in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}) = 1,$$

thus,

$$\sum_{k\in I_s^{(l)}} g_{jk}(\hat{\mathbf{R}}_{s-2}) = 1 - g_{jl}(\hat{\mathbf{R}}_{s-2}) - \sum_{q\in J_{s-1}} g_{jq}(\hat{\mathbf{R}}_{s-2}).$$

The fifth equality holds since $I_s^{(l)} \cup J_{s-1} = I_{s-1}^{(l)}$, and $I_s^{(l)} \cap J_{s-1} = \emptyset$. And the last equality holds since

$$\sum_{k\in I_{s-1}} g_{lk}(\hat{\mathbf{R}}_{s-2}) = \sum_{k\in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{l=k} g_{kk}(\hat{\mathbf{R}}_{s-2}) = \sum_{k\in I_{s-1}^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-2}) = 1.$$

Therefore, we have

$$\frac{\sum\limits_{k\in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum\limits_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right)}{1 - \sum\limits_{j\in J_{s-1}} \sum\limits_{q\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum\limits_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})} = 1,$$

which implies $\sum_{k\in I_s^{(l)}} g_{lk}(\hat{\mathbf{R}}_{s-1}) = 1$ for $i = s$. Moreover, we know that $g_{kk}(\hat{\mathbf{R}}_{s-1}) = 0$ is always true based on the updating rule of transition coefficient function in Algorithm 3. Thus $\sum_{k\in I_s} g_{lk}(\hat{\mathbf{R}}_{s-1}) = 1$ for $i = s$.

Summarizing the above arguments, $\sum_{k\in I_i} g_{lk}(\hat{\mathbf{R}}_{i-1})$ =1 holds for all $i = 1, ..., m$. □

B.2 Proof of (3.9) in Proposition 2

We use mathematical induction to show that (3.9) holds, that is

$$0 \leq g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1 \text{ for all } i = 1, ..., m.$$

For $i = 1$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \emptyset$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}$. By $0 \leq g_{lk} \leq 1$ in (3.2), we prove that $0 \leq g_{lk}(\hat{\mathcal{R}}_0) \leq 1$. By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $0 \leq g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq 1$ for all $l, k \in I_{s-1}$. Then for $i = s$, by the updating rule of the transition coefficient function in Algorithm 3, for all $l, k \in I_s, l \neq k$, we have

$$g_{lk}(\hat{\mathbf{R}}_{s-1}) = \frac{g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2})}{1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})}, \qquad (2)$$

where the numerator of the right-side in (2) is always non-negative due to the assumption that $0 \leq g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq 1$. Moreover, in B.1 we showed that

$$\begin{aligned} & \sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right) \\ = \;& 1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}). \end{aligned}$$

And by the assumption again, we have

$$\sum_{k \in I_s^{(l)}} \left(g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) \right) \geq 0, \qquad (3)$$

thus,

$$1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) \geq 0,$$

which is the denominator of the left-side in (2). And when

$$1 - \sum_{j \in J_{s-1}} \sum_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2}) = 0,$$

we always set $g_{lk}(\hat{\mathbf{R}}_{s-1}) = 0$. Therefore, $g_{lk}(\hat{\mathbf{R}}_{s-1}) \geq 0$. Since we showed in B.1 that $\sum_{k \in I_s} g_{lk}(\hat{\mathbf{R}}_{s-1}) = 1$, thus $g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq 1$.

Therefore, $0 \leq g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq 1$ holds for all $i = 1, ..., m$. When $l = k$, it is always true that $g_{kk}(\hat{\mathbf{R}}_{i-1}) = 0$ for all $i = 1, ..., m$, based on the updating rule of transition coefficient function in Algorithm 3. □

B.3 Proof of (3.10) Proposition 2

We use mathematical induction to show that (3.10) holds, that is

$$\sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha \text{ for all } i = 1, ..., m.$$

For $i = 1$, $I_1 = \{1, ..., m\}$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \emptyset$, and $\alpha_l(\hat{\mathcal{R}}_0) = \alpha_l$. By $\sum\limits_{l=1}^{m} \alpha_l \leq \alpha$ in (**??**), we prove that

$$\sum_{l \in I_i} \alpha_l(\hat{\mathcal{R}}_0) \leq \alpha.$$

By induction, we assume for $i = s - 1 (s \leq m)$, the inequality holds, i.e., $\sum\limits_{l \in I_{s-1}} \alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha$. Then for $i = s$, by the updating rule of the critical value function in Algorithm 3, we have

$$\begin{aligned}
\sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-1}) &= \sum_{l \in I_s} \left(\alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) \alpha_j(\hat{\mathbf{R}}_{s-2}) \right) \\
&= \sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{l \in I_s} \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) \alpha_j(\hat{\mathbf{R}}_{s-2}) \\
&= \sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} \alpha_j(\hat{\mathbf{R}}_{s-2}) \sum_{l \in I_s} g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
&\leq \sum_{l \in I_s} \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} \alpha_j(\hat{\mathbf{R}}_{s-2}) = \sum_{l \in I_{s-1}} \alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha,
\end{aligned}$$

where the first inequality holds since for non-negative transition coefficient functions, and $I_s \subseteq I_{s-1}$, we have $\sum\limits_{l \in I_s} g_{jl}(\hat{\mathbf{R}}_{s-2}) \leq \sum\limits_{l \in I_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) = 1$. The equality holds only when for all $k \in J_{s-1}$, $g_{jk} = 0$, thus $\sum\limits_{l \in I_s} g_{jl}(\hat{\mathbf{R}}_{s-2}) = \sum\limits_{l \in I_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2}) = 1$. And the last inequality is due to the assumptation that $\sum\limits_{l \in I_{s-1}} \alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha$.

Summarizing the above arguments, we conclude that

$$\sum_{l \in I_i} \alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha \text{ holds for all } i = 1, ..., m. \square$$

B.4 Proof of monotonicity of transition coefficient function

We use mathematical induction to show that the monotonicity of transition coefficient function holds, that is

$$g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq g_{lk}(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m,$$

here, $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$, and $\mathbf{S}_{i-1} = (\mathcal{S}_1, ..., \mathcal{S}_{i-1})^T$.

For $i = 1$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \mathcal{S}_0 = \emptyset$, and $g_{lk}(\hat{\mathcal{R}}_0) = g_{lk}(\mathcal{S}_0) = g_{lk}$. Thus, we prove that the monotonicity of trasnsition coefficient function holds when $i = 1$.

By induction, we assume that for $i = s-1 (s \leq m)$, the inequality holds, i.e., $g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq g_{lk}(\mathbf{S}_{s-2})$ for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-2$, and $l, k \in I_{s-1}$. Then for $i = s$, when $l \neq k$, by the updating rule of transition coefficient function in Algorithm 3, we have

$$g_{lk}(\hat{\mathbf{R}}_{s-1}) = \frac{g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2})}{1 - \sum\limits_{j \in J_{s-1}} \sum\limits_{q \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jq}(\hat{\mathbf{R}}_{s-2}) - \sum\limits_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jl}(\hat{\mathbf{R}}_{s-2})}, \tag{4}$$

$$g_{lk}(\mathbf{S}_{s-1}) = \frac{g_{lk}(\mathbf{S}_{s-2}) + \sum\limits_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jk}(\mathbf{S}_{s-2})}{1 - \sum\limits_{j \in T_{s-1}} \sum\limits_{q \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jq}(\mathbf{S}_{s-2}) - \sum\limits_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jl}(\mathbf{S}_{s-2})}, \tag{5}$$

where T_{s-1} is the index set of a collection of rejections in $\mathcal{S}_{s-1}$. For non-negative transition coefficient function (showed in B.2), we have

$$\begin{aligned} g_{lk}(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2}) g_{jk}(\hat{\mathbf{R}}_{s-2}) &\leq g_{lk}(\mathbf{S}_{s-2}) + \sum_{j \in J_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jk}(\mathbf{S}_{s-2}) \qquad (6) \\ &\leq g_{lk}(\mathbf{S}_{s-2}) + \sum_{j \in T_{s-1}} g_{lj}(\mathbf{S}_{s-2}) g_{jk}(\mathbf{S}_{s-2}), \qquad (7) \end{aligned}$$

where the left-side in (6) is the numerator of the right-side in (4), and the right-side in (7) is the numerator of the right-side in (5). The inequality in (6) holds due to the assumptation made at step $s-1$. And the inequality in (7) holds since $J_{s-1} \subseteq T_{s-1}$.

Moreover, for both denominators of right-side in (4) and (5), we have

$$\begin{aligned}
& 1-\sum_{j\in J_{s-1}}\sum_{q\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jq}(\hat{\mathbf{R}}_{s-2})-\sum_{j\in J_{s-1}} g_{lj}(\hat{\mathbf{R}}_{s-2})g_{jl}(\hat{\mathbf{R}}_{s-2}) \\
\geq\;& 1-\sum_{j\in J_{s-1}}\sum_{q\in J_{s-1}} g_{lj}(\mathbf{S}_{s-2})g_{jq}(\mathbf{S}_{s-2})-\sum_{j\in J_{s-1}} g_{lj}(\mathbf{S}_{s-2})g_{jl}(\mathbf{S}_{s-2}) \qquad (8)\\
\geq\;& 1-\sum_{j\in T_{s-1}}\sum_{q\in T_{s-1}} g_{lj}(\mathbf{S}_{s-2})g_{jq}(\mathbf{S}_{s-2})-\sum_{j\in T_{s-1}} g_{lj}(\mathbf{S}_{s-2})g_{jl}(\mathbf{S}_{s-2}), \qquad (9)
\end{aligned}$$

the inequality in (8) holds due to the assumptation made at step $s-1$. And the inequality in (9) holds since $J_{s-1} \subseteq T_{s-1}$. Therefore, by the above argument, we prove that when $l \neq k$, $g_{lk}(\hat{\mathbf{R}}_{s-2}) \leq g_{lk}(\mathbf{S}_{s-2})$. When $l = k$, it is always true that $g_{lk}(\hat{\mathbf{R}}_{s-2}) = g_{lk}(\mathbf{S}_{s-2}) = 0$ by the updating rule of transition coefficient function in Algorithm 3. Therefore, $g_{lk}(\hat{\mathbf{R}}_{s-1}) \leq g_{lk}(\mathbf{S}_{s-1})$, for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-1$.

Summarizing the above arguments, we conclude that

$$g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq g_{lk}(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m. \;\square$$

B.5 Proof of (3.11) in Proposition 2

We use mathematical induction to show that (3.11) holds, that is

$$\alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m,$$

where, $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$, and $\mathbf{S}_{i-1} = (\mathcal{S}_1, ..., \mathcal{S}_{i-1})^T$.

For $i = 1$, we have $\hat{\mathcal{R}}_{i-1} = \hat{\mathcal{R}}_0 = \mathcal{S}_0 = \emptyset$, and $\alpha_l(\hat{\mathcal{R}}_0) = \alpha_l(\mathcal{S}_0) = \alpha_l$. Thus, we prove that (3.11) holds when $i = 1$.

By induction, we assume that for $i = s-1 (s \leq m)$, the inequality holds, i.e., $\alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha_l(\mathbf{S}_{s-2})$ for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-2$. Then for $i = s$, by the updating rule of critical value function in Algorithm 3, we have

$$\alpha_l(\hat{\mathbf{R}}_{s-1}) = \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j\in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2})\alpha_j(\hat{\mathbf{R}}_{s-2}) \qquad (10)$$

and

$$\alpha_l(\mathbf{S}_{s-1}) = \alpha_l(\mathbf{S}_{s-2}) + \sum_{j \in Q_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\mathbf{S}_{s-2}), \tag{11}$$

where Q_{s-1} is the index set of a collection of rejections in $\mathcal{S}_{s-1}$. The right-side in (10) has the following property by the monotonicity of transition coefficient function showed in B.4

$$\begin{aligned}
\alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\hat{\mathbf{R}}_{s-2})\alpha_j(\hat{\mathbf{R}}_{s-2}) \quad & \leq \alpha_l(\hat{\mathbf{R}}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\hat{\mathbf{R}}_{s-2}) && (12)\\
& \leq \alpha_l(\mathbf{S}_{s-2}) + \sum_{j \in J_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\mathbf{S}_{s-2}) && (13)\\
& \leq \alpha_l(\mathbf{S}_{s-2}) + \sum_{j \in Q_{s-1}} g_{jl}(\mathbf{S}_{s-2})\alpha_j(\mathbf{S}_{s-2}). && (14)
\end{aligned}$$

The inequality in (13) holds due to the assumptation made at step $s-1$, i.e., $\alpha_l(\hat{\mathbf{R}}_{s-2}) \leq \alpha_l(\mathbf{S}_{s-2})$. And the inequality in (14) holds since $J_{s-1} \subseteq Q_{s-1}$. Therefore, by the above argument, we prove that $\alpha_l(\hat{\mathbf{R}}_{s-1}) \leq \alpha_l(\mathbf{S}_{s-1})$, for every $\hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., s-1$.

Summarizing the above arguments, we conclude that

$$\alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m. \ \square$$

B.6 Proof of Theorem 4

We apply the extended sequention rejective principle in Theorem 2 to prove the FWER control of the proposed graphical approach. First, we show that the monotonicity condition of the successor function in Theorem 2 is satisfied. The successor function is defined as follows:

$$\mathcal{N}(\hat{\mathbf{R}}_{\mathbf{i-1}}) = \hat{\mathcal{R}}_i = \{H_j \in \mathcal{H} \setminus \cup_{k=1}^{i-1} \hat{\mathcal{R}}_k : p_j \leq \alpha_j(\hat{\mathbf{R}}_{\mathbf{i-1}}) + \sum_f b_{fj}(\hat{\mathbf{R}}_{i-1})\},$$

and

$$\mathcal{N}(\mathbf{S}_{i-1}) = \mathcal{S}_i = \{H_j \in \mathcal{H} \backslash \cup_{k=1}^{i-1} \mathcal{S}_k : p_j \leq \alpha_j(\mathbf{S}_{i-1}) + \sum_f b_{fj}(\mathbf{S}_{i-1})\},$$

where $\hat{\mathbf{R}}_{i-1} = (\hat{\mathcal{R}}_1, ..., \hat{\mathcal{R}}_{i-1})^T$, and $\mathbf{S}_{i-1} = (\mathcal{S}_1, ..., \mathcal{S}_{i-1})^T$.

In B.4, we showed that

$$g_{lk}(\hat{\mathbf{R}}_{i-1}) \leq g_{lk}(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, j = 1, ..., i-1, \text{and for all } i = 1, ..., m.$$

And in B.5, we showed that

$$\alpha_l(\hat{\mathbf{R}}_{i-1}) \leq \alpha_l(\mathbf{S}_{i-1}), \text{for every } \hat{\mathcal{R}}_j \subseteq \mathcal{S}_j, \text{ where } j = 1, ..., i-1, \text{ and for all } i = 1, ..., m.$$

The re-assigned critical value function $b_{fj}(\hat{\mathbf{R}}_{i-1})$ at any step i is a multiply of non-negative monotonic functions, thus it is also a monotonic function, as well as the the summation of such re-assigned critical value functions $\sum\limits_f b_{fj}(\hat{\mathbf{R}}_{i-1})$. So we have

$$\alpha_j(\hat{\mathbf{R}}_{\mathbf{i-1}}) + \sum_f b_{fj}(\hat{\mathbf{R}}_{i-1}) \leq \alpha_j(\mathbf{S}_{i-1}) + \sum_f b_{fj}(\mathbf{S}_{i-1}), \tag{15}$$

for every $\hat{\mathcal{R}}_k \subseteq \mathcal{S}_k$, where $k = 1, ..., i-1$, and for all $i = 1, ..., m$.

Thus, for every $H_j \in \mathcal{N}(\hat{\mathbf{R}}_{\mathbf{i-1}})$, we have $p_j \leq \alpha_j(\hat{\mathbf{R}}_{\mathbf{i-1}}) + \sum\limits_f b_{fj}(\hat{\mathbf{R}}_{i-1})$, by the monotonicity showed in (15), we show that $p_j \leq \alpha_j(\mathbf{S}_{i-1}) + \sum\limits_f b_{fj}(\mathbf{S}_{i-1})$, thus, $H_j \in \mathcal{N}(\mathbf{S}_{i-1})$. Therefore, $\mathcal{N}(\hat{\mathbf{R}}_{\mathbf{i-1}}) \subseteq \mathcal{N}(\mathbf{S}_{i-1})$, which satisfies the monotonicity condition of successor function in Theorem 2.

Next, we show that the singel step FWER control in Theorem 2 is also satisfied for the proposed approach. Assume we reject all false null hypotheses in previous

$i-1$ steps, then we have

$$\begin{aligned}
& P(\text{reject at least one true null hypothesis at step } i) \\
= \; & P(\mathcal{N}(\mathbf{F}_{i-1}) \nsubseteq \mathcal{F}) \\
= \; & P\left(\cup_{j \in \mathcal{T}} \{ p_j \leq \alpha_j(\mathbf{F}_{i-1}) + \sum_f b_{fj}(\mathbf{F}_{i-1}) \} \right) \\
\leq \; & \sum_{j \in \mathcal{T}} P\left(p_j \leq \alpha_j(\mathbf{F}_{i-1}) + \sum_f b_{fj}(\mathbf{F}_{i-1}) \right) \\
\leq \; & \sum_{j \in \mathcal{T}} \left(\alpha_j(\mathbf{F}_{i-1}) + \sum_f b_{fj}(\mathbf{F}_{i-1}) \right) \\
\leq \; & \sum_{j \in I_i} \left(\alpha_j(\mathbf{F}_{i-1}) + \sum_f b_{fj}(\mathbf{F}_{i-1}) \right) \\
= \; & \sum_{j \in T_i} \alpha_j(\mathbf{F}_{i-1}) + \sum_{f \in I_i \setminus T_i} \alpha_f(\mathbf{F}_{i-1}) \leq \alpha
\end{aligned}$$

where $\mathcal{T}$ and $\mathcal{F}$ are collection of true null hypotheses and false null hypotheses, respectively. And $\mathbf{F}_{i-1} = (\mathcal{F}_1, ..., \mathcal{F}_{i-1})^T$, is a vector, which includes all rejected false null hypotheses, and $\mathcal{F}_1, ..., \mathcal{F}_{i-1}$ are all mutually exclusive sets. The first inequality is due to Bonferroni inequality. The second inequality is under the assumption that all true null p-values follow uniform distribution. The third inequality is because of $\mathcal{T}_i \subseteq I_i$. The last inequality is due to the property of critical value function (showed in B.3), that is:

$$\sum_{l \in I_i} \alpha_l(\mathbf{\hat{R}_{i-1}}) \leq \alpha \text{ for all } i = 1, ..., m,$$

and the property of re-assignment that is the sum of critical values over all hypotheses at any step i remains the same before and after re-assignment.

Thus, the single step FWER control is also satisfied for the proposed approach. By the extended sequential rejection principle, we prove that the proposed graphical approach strongly controls the FWER at level α. □

BIBLIOGRAPHY

[1] R. F. Barber and E. J. Candes. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43: 2055-2085, 2015.

[2] R. F. Barber and E. J. Candes. The p-filter: multi-layer FDR control for grouped hypotheses. *arXiv:1512.03397v2*, 2016.

[3] J. Bartroff and J. Song. A rejection principle for sequential tests of multiple hypotheses controlling familywise error rates. *Scandinavian Journal of Statistics*, 43:3-19, 2016.

[4] P. Bauer, J. Rhmel J, W. Maurer and L. Hothorn. Testing strategies in multidose experiments including active control. *Statistics in Medicine*, 17:2133-2146, 1998.

[5] Y. Benjamini. Simultaneous and selective inference: current successes and future challenges. *Biometrical Journal*, 52:708-721, 2010.

[6] Y. Benjamini and M. Bogomolov. Selective inference on multiple families of hypotheses. *Journal of the Royal Statistical Society: Series B*, 76:297-318, 2014.

[7] 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.

[8] Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25:60-83, 2000.

[9] Y. Benjamini, K. Krieger and D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93:491-507, 2006.

[10] Y. Benjamini and D. Yekutieli. False discovery rateadjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100:71-81, 2005.

[11] R. Berk, L. Brown, A. Buja, K. Zhang and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, 41:802-837, 2013.

[12] G. Blanchard and E. Roquain. Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research*, 10:2837-2871, 2009.

[13] R. Bourgon, R. Gentleman and W. Huber. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences*, 107:9546-9551, 2010.

[14] 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.

[15] F. Bretz, W. Maurer and G. Hommel. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. *Statistics in Medicine*, 30:1489-1501, 2011a.

[16] F. Bretz, M. Posch, K. Glimm, F. Klinglmueller, W. Maurer and K. Rohmeyer. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biometrical Journal*, 53:894-913, 2011b.

[17] C. F. Burman, C. Sonesson and O. Guilbaud. A recycling framework for the construction of Bonferroni-based multiple tests. *Statistics in Medicine*, 28:739-761, 2009.

[18] J. Y. Dai, C. Kooperberg, M. Leblanc and R. L. Prentice. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*, 99:929-944, 2012.

[19] 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.

[20] A. Dmitrienko and A. C. Tamhane. Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine*, 30:1473-1488, 2011.

[21] A. Dmitrienko and A. C. Tamhane. General theory of mixture procedures for gatekeeping. *Biometrical Journal*, 55:402-419, 2013.

[22] A. Dmitrienko, A. C. Tamhane, X. Wang and X. Chen. Stepwise gatekeeping procedures in clinical trial applications. *Biometrical Journal*, 48:984-991, 2006.

[23] A. Dmitrienko, A. C. Tamhane and B. L. Wiens. General multistage gatekeeping procedures. *Biometrical Journal*, 50:667-677, 2008.

[24] L. Du and C. Zhang. Single-index modulated multiple testing. *The Annals of Statistics*, 42:1262-1311, 2014.

[25] L. Finos and L. Salmaso. FDR-and FWE-controlling methods using data-driven weights. *Journal of Statistical Planning and Inference*, 137:3859-3870, 2007.

[26] W. Fithian, D. Sun and J. Taylor. Optimal inference after model selection. *arXiv:1410.2597v2*, 2015.

[27] J. J. Goeman and A. Solari. The sequential rejection principle of familywise error control. *The Annals of Statistics*, 38:3782-3810, 2010.

[28] O. Guilbaud. Bonferroni parallel gatekeepingtransparent generalizations, adjusted p-values, and short direct proofs. *Biometrical Journal*, 49:917-927, 2007.

[29] O. Guilbaud. Note on simultaneous inferences about non-inferiority and superiority for a primary and a secondary endpoint. *Biometrical Journal*, 53:927-937, 2011.

[30] W. Guo and J. P. Romano. Analysis of error control in large scale two-stage multiple hypothesis testing. *arXiv:1703.06336*, 2017.

[31] R. Heller, N. Chatterjee, A. Krieger and J. Shi. Post-selection inference following aggregate level hypothesis testing in large scale genomic data. *Journal of the American Statistical Association*, 1-14, 2018.

[32] Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75:800-802, 1988.

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

[34] G. Hommel and F. Bretz. Aesthetics and power considerations in multiple testinga contradiction? *Biometrical Journal*, 50:657-666, 2008.

[35] G. Hommel, F. Bretz and W. Maurer. Powerful shortcuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine*, 26:4063-4073, 2007.

[36] H. Hung and S. Wang. Some controversial multiple testing problems in regulatory applications. *Journal of Biopharmaceutical Statistics*, 19:1-11, 2009.

[37] M. Huque, M. Alosh and R. Bhore. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. *Journal of Biopharmaceutical Statistics*, 21:610-634, 2011.

[38] N. Ignatiadis and W. Huber. Covariate-powered weighted multiple testing with false discovery rate control. *arXiv:1701.05179*, 2017.

[39] N. Ignatiadis, B. Klaus, J. B. Zaugg and W. Huber. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, 13:577-580, 2016.

[40] H. Kim, A. Entsuah and J. Shults. The union closure method for testing a fixed sequence of families of hypotheses. *Biometrika*, 98:391-401, 2011.

[41] S. Kim and P. Schliekelman. Prioritizing hypothesis tests for high throughput data. *Bioinformatics*, 32:850-858, 2016.

[42] F. Klinglmueller, M. Posch and F. Koenig. Adaptive graphbased multiple testing procedures. *Pharmaceutical Statistics*, 13:345-356, 2014.

[43] G. Kordzakhia and A. Dmitrienko. Superchain procedures in clinical trials with multiple objectives. *Statistics in Medicine*, 32:486-508, 2013.

[44] J. Lawrence. Testing non-inferiority and superiority for two endpoints for several treatments with a control. *Pharmaceutical Statistics*, 10:318-324, 2011.

[45] J. D. Lee, D. L. Sun, Y. Sun and J. E. Taylor. Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44:907-927, 2016.

[46] E. Lehmann and J. P. Romano. *Testing statistical hypotheses*, 3rd edition, Springer, New York, 2005.

[47] L. Lei and W. Fithian. AdaPT: an interactive procedure for multiple testing with side information. *arXiv:1609.06035*, 2016.

[48] A. Li and R. Barber. Multiple testing with the structure adaptive Benjamini-Hochberg algorithm. *arXiv:1606.07926*, 2016.

[49] K. Lu. Graphical approaches using a Bonferroni mixture of weighted Simes tests. *Statistics in Medicine*, 35:4041-4055, 2016.

[50] R. Marcus, P. Eric and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63:655-660, 1976.

[51] W. Maurer and F. Bretz. Memory and other properties of multiple test procedures generated by entangled graphs. *Statistics in Medicine*, 32:1739-1753, 2013.

[52] W. Maurer and F. Bretz. Multiple testing in group sequential trials using graphical approaches. *Statistics in Biopharmaceutical Research*, 5:311-320, 2013.

[53] W. Maurer and F. Bretz. A note on testing families of hypotheses using graphical procedures. *Statistics in Medicine*, 33:5340-5346, 2014.

[54] W. Maurer, E. Glimm and F. Bretz. Multiple and repeated testing of primary, co-primary and secondary hypotheses. *Statistics in Biopharmaceutical Research*, 3:336-352, 2011.

[55] W. Maurer, L. A. Hothorn and W. Lehmacher. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. *Biometrie in Der Chemisch-Pharmazeutischen Industrie*, 6:3-18, 1995.

[56] N. Meinshausen. Hierarchical testing of variable importance. *Biometrika*, 95:265-278, 2008.

[57] N. Meinshausen and J. Rice. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *The Annals of Statistics*, 34:373-393, 2006.

[58] B. Millen and A. Dmitrienko. A class of flexible closed testing procedures with clinical trial applications. *Journal of Biopharmaceutical Statistics*, 3:14-30, 2011.

[59] S. Panigrahi and J. Taylor. Scalable methods for Bayesian selective inference. *Electronic Journal of Statistics*, 12: 2355-2400, 2018.

[60] J. Peng, W. Liu, F. Bretz and Z. Shkedy. Multiple confidence intervals for selected parameters adjusted for the false coverage rate in monotone doseresponse microarray experiments. *Biometrical Journal*, 2016.

[61] L. Poisson, A. Sreekumar, A. Chinnaiyan and D. Ghosh. Pathway-directed weighted testing procedures for the integrative analysis of gene expression and metabolomic data. *Genomics*, 99:265-274, 2012.

[62] Z. Qiu, L. Yu and W. Guo. A family-based graphical approach for testing hierarchically ordered families of hypotheses. Working paper, 2018.

[63] G. Rauch and J. Beyersmann. Planning and evaluating clinical trials with composite time-to-first-event endpoints in a competing risk framework. *Statistics in Medicine*, 32:3595-3608, 2013.

[64] K. Roeder and L. Wasserman. Genome-wide significance levels and weighted hypothesis testing. *Statistical Science*, 24:398-413, 2009.

[65] D. Rugamer and S. Greven. Selective inference after likelihood-or test-based model selection in linear models. *Statistics Probability Letters*, 140: 7-12, 2018.

[66] S. K. Sarkar. On methods controlling the false discovery rate. *Sankhya, Series A*, 70:135-168, 2008.

[67] S. K. Sarkar, W. Guo and H. Finner. On adaptive procedures controlling the familywise error rate. *Journal of Statistical Planning and Inference*, 142:65-78, 2012.

[68] T. Schweder and E. Spjotvoll. Plots of p-values to evaluate many tests simultaneously. *Biometrika*, 69:493-502, 1982.

[69] A. Skol, L. Scott, G. Abecasis and M. Boehnke. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics*, 38:209-213, 2006.

[70] J. D. Storey, J. E. Taylor and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B*, 66:187-205, 2004.

[71] T. Sugitani, F. Bretz and W. Maurer. A simple and flexible graphical approach for adaptive group-sequential clinical trials. *Journal of Biopharmaceutical Statistics*, 26:202-216, 2016.

[72] T. Sugitani, T. Hamasaki and C. Hamada. Partition testing in confirmatory adaptive designs with structured objectives. *Biometrical Journal*, 55:341-359, 2013.

[73] J. Taylor and R. J. Tibshirani. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112:7629-7634, 2015.

[74] X. Tian, J. R. Loftus and J. Taylor. Selective inference with unknown variance via the square-root lasso. *Biometrika*, 105: 755-768, 2018.

[75] X. Tian and J. Taylor. Asymptotics of selective inference. *arXiv:1501.03588*, 2015b.

[76] X. Tian and J. E. Taylor. Selective inference with a randomized response. *The Annals of Statistics*, 46: 679-710, 2018.

[77] A. Weinstein, W. Fithian and Y. Benjamini. Selection adjusted confidence intervals with more power to determine the sign. *Journal of the American Statistical Association*, 108:165-176, 2013.

[78] A. Weinstein and D. Yekutieli. Selective sign-determining multiple confidence intervals with FCR control. *arXiv:1404.7403v2*, 2016.

[79] P. H. Westfall and A. Krishen. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *Journal of Statistical Planning and Inference*, 99:25-40, 2001.

[80] B. L. Wiens. A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics*, 2:211-215, 2003.

[81] B. L. Wiens and A. Dmitrienko. The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics*, 15:929-942, 2005.

[82] D. Xi, E. Glimm, W. Maurer and F. Bretz. On weighted parametric tests. *arXiv:1605.06397*, 2016.

[83] D. Xi and A. C. Tamhane. Allocating recycled significance levels in group sequential procedures for multiple endpoints. *Biometrical Journal*, 57:90-107, 2015.

[84] D. Yekutieli. Adjusted Bayesian inference for selected parameters. *Journal of the Royal Statistical Society: Series B*, 74:515-541, 2012.