Copyright Warning & Restrictions

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

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

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

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

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

NJIT New Jersey's Science & Technology University

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

ABSTRACT

TYPE I ERROR RATE CONTROLLING PROCEDURES FOR MULTIPLE HYPOTHESES TESTING

by
Beibei Li

This dissertation addresses several different but related topics arising in the field of multiple testing, including weighted procedures and graphical approaches for controlling the familywise error rate (FWER), and stepwise procedures with control of the false discovery rate (FDR) for discrete data. It consists of three major parts.

The first part investigates weighted procedures for controlling the FWER. In many statistical applications, hypotheses may be differentially weighted according to their different importance. Many weighted multiple testing procedures (wMTPs) have been developed for controlling the FWER. Among these procedures, two weighted Holm procedures are commonly used in practice: one is based on ordered weighted p-values and is called WHP; the alternative weighted Holm procedure that is based on ordered raw p-values is named WAP. This part of dissertation studies statistical properties of these two weighted procedures and make recommendation for their applications. First, the corresponding closed testing procedures (CTPs) of both weighted procedures are obtained and the WHP is proved to be uniformly more powerful than the WAP. Following this, in order to provide an intuitive and clear way to communicate with non-statisticians, two procedures are visualized with graphical approaches through a common initial graph and their corresponding updating strategies. Next, the adjusted p-values are derived for these two procedures. Finally, the optimality of these two procedures is discussed and it is shown that the WHP is an optimal procedure in the sense that the procedure cannot be improved by increasing even one of its critical values without losing control over the FWER. Simulations were conducted to provide numerical evidence of superior performance of the WHP in terms of the FWER control and average power.

In the second part of the dissertation, two graphical approaches are investigated. One is the original graphical approach which is introduced in Bretz et al. (2009) and widely used in clinical trials studies, and the other one is the default graphical approach, proposed in Burman et al. (2009). These two graphical approaches are commonly considered to be equivalent in the literature. However, this study shows that their equivalence can only be achieved under certain conditions or in the case of three hypotheses. When the conditions are satisfied, a general method is developed for deriving the equivalent graph. The nonuniqueness property of the original graphical approach is also discussed. Moreover, a simple and direct proof is offered for showing the FWER control of the original graphical approach. This is helpful for understanding the original graphical approach thoroughly and provides some guideline to develop new graphical approaches.

In the third part of the dissertation, a new generalized step-up FDR controlling procedure is developed for discrete data. Most existing FDR controlling procedures are developed for continuous data, which are often conservative when analyzing discrete data. Lynch and Guo (2016) introduced a generalized stepwise procedure which generalizes the usual stepwise procedure to the case where each hypothesis is tested with a different set of critical constants. Under the framework of the generalized step-up approach, by taking the discreteness and heterogeneity properties of discrete data into account and fully utilizing known marginal distributions of true null p-values, a powerful generalized step-up procedure is proposed for discrete case. Theoretically, it is shown that the proposed procedure strongly controls the FDR under independence and is more powerful than the popular BH procedure. Nevertheless, some theoretical as well as simulation issues still remain to be fully addressed.

TYPE I ERROR RATE CONTROLLING PROCEDURES FOR MULTIPLE HYPOTHESES TESTING

by
Beibei Li

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
Department of Mathematics and Computer Science, Rutgers-Newark

May 2022

Copyright © 2022 by Beibei Li

ALL RIGHTS RESERVED

APPROVAL PAGE

TYPE I ERROR RATE CONTROLLING PROCEDURES FOR MULTIPLE HYPOTHESES TESTING

Beibei Li

Dr. Wenge Guo, Dissertation Advisor — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Ji Meng Loh, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Zuofeng Shang, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Antai Wang, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Zhi Wei, Committee Member — Date
Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Beibei Li

Degree: Doctor of Philosophy

Date: May 2022

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2022
- Master of Science in Applied Statistics,
  New Jersey Institute of Technology, Newark, NJ, 2016
- Master of Science in Obstetrics and Gynecology,
  China Medical University, Shengyang, China, 2012
- Bachelor of Science in Medicine,
  Zhengzhou University, Zhengzhou, China, 2009

Major: Mathematical Sciences

Presentations and Publications:

B. Li and W. Guo. On weighted Holm procedures, Joint Statistical Meeting (JSM2021), Virtual conference, August 8-12, 2021.

B. Li and W. Guo. On weighted Holm procedures, Manuscript, 2022.

Y. Zhu, B. Li and W. Guo. A selective inference-based two-stage procedure for clinical safety studies, Manuscript, 2022.

To my husband and son, Ke Geng and Ethan Geng, who always help me go through the most difficult times with their love.

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Wenge Guo, for his guidance and support throughout the years. He has taught me, with great patience and dedication, how to conduct research project wisely and efficiently. At those moments when I felt very low and nothing seems to work, he always guided me to find where the hope is and encouraged me to pick up my confidence. As someone who does not have really strong and solid prior background in statistics or math, I feel very fortunate to have the opportunity to work with Dr. Guo. More importantly, in heavy teaching and research workload, Dr. Guo has always had an opened door for every student, offering them his valuable time and suggestions for the best of their interests, not only for their school work problems, but also for their intern/job pre-interview anxieties. I am very grateful to have such an outstanding mentor during my PhD study.

Besides, I would like to thank all my dissertation committee members Dr. Ji Meng Loh, Dr. Zuofeng Shang, Dr. Antai Wang, and Dr. Zhi Wei, for being an important part of my scientific journey. I truly appreciate for their valuable time and suggestions for my dissertation research.

I would also like to thank Department of Mathematical Sciences for supporting me financially through parts of my doctoral program, and thank Ms. Clarisa Gonzalez-Lenahan and Dr. Sotirios Ziavras who helped me to revise this dissertation.

I would also like to thank my friends who create such a warm and supporting environment and always help me without hesitation. Thank you Dr. Atefeh Javidialsaadi, Dr. Gan Luan, Hewei Zhang, Subhrasish Chakraborty, Chhavi Tyagi. I sincerely wish all of you have a great future and life, hope we all cherish our golden times at NJIT.

Finally, I would like to thank my parents for their unconditional support. To my son, Ethan, thank you for the happiness you bring to us every day and your sweetest love in this world. I also deeply appreciate the love and support from my husband, Ke Geng, as he

always understands and holds my hands through the good and bad times of these years. I cannot accomplish any of this without any one of you.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter Page

TABLE OF CONTENTS
(Continued)

Chapter Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Introduction

In many statistical applications, multiple testing problems are very common in clinical trial studies and genome-wide association studies (GWAS), such as estimating the effects of a new treatment, finding the optimal dose of a new drug, detecting adverse events and testing gene sets or individual genes for differential expression. Consequently, multiplicity problem, which we called type I error rate inflation, arises and the magnitude of the error rate increases as the number of hypotheses becomes larger. For this kind of problem, the difficulty is how to control the type I error rate at a pre-specified significant level α and meanwhile achieve a high power. Unlike the single hypothesis testing, there are various measures of type I error rate, the most common two are the familywise error rate (FWER), the probability of making at least one false rejections, and the false discovery rate (FDR) which is the expected proportion of false rejections among the rejected hypotheses. Many multiple testing procedures (MTPs) have been developed to address the multiplicity problem with the control of proper error rates. For small scale multiple testing problems, the procedures controlling FWER are needed, especially in clinical studies, while for large scale testing the procedures controlling FDR are more suitable. The Bonferroni (1936) procedure, a single-step procedure, is the most basic method to control FWER under arbitrary dependence and it is usually used to develop some advanced multiple testing procedures, for example, Holm (1979) procedure which is a step-down version of the Bonferroni procedure. Another popular procedure based on the Simes global test is Hochberg (1988) procedure which is a step-up procedure and uniformly more powerful than Holm procedure, but it controls the FWER under independence or positive regression dependence. However, controlling FWER will be conservative when the number

of hypotheses is large and in such a situation one can often tolerate a small number or proportion of type I errors. Benjamini and Hochberg (1995) developed a classical step-up procedure, BH procedure, which strongly controls the FDR under independence that allows to make a pre-specified proportion of false rejections and is suitable for testing a large number of hypotheses. Benjamini and Yekutieli (2001) showed that the FDR control of BH procedure can be extended to the cases of positive regression dependence and proposed a modified BH procedure controlling the FDR under arbitrary dependence.

Moreover, it often happens that some hypotheses are more important than the others, which suggests us to assign different weights to different hypotheses according to their importance. In the existing literature, some progress has been made to develop weighted procedures. Rosenthal and Rubin (1983) proposed a weighted Bonferroni procedure which permits greater power for the important hypotheses. Holm (1979) and Benjamini and Hochberg (1997) developed two different weighted Holm procedures. One is based on ordered weighted p-values that we called WHP, see Holm (1979); the alternative weighted Holm procedure that is based on ordered original p-values is named WAP, refer to Benjamini and Hochberg (1997). Wiens et al. (2013) pointed out that hypotheses ordering is more relevant for stepwise procedures with asymmetric rules for updating the hypotheses weights, including the fixed-sequence and fallback procedures; however, the similarities and differences of ordering based on raw or weighted p-values have not been studied. Tamhane and Liu (2008) constructed weighted Hochberg-type step-up multiple test procedures including two closed procedures based on weighted Simes tests: one based on ordered raw p-values and the other one based on ordered weighted p-values. However, both two weighted procedures lack simple stepwise structure, therefore it is hard to be compared with each other and not easy to explain it to practitioners. And also these procedures were developed under independence structure. Thus, to study weighted procedures controlling the FWER with some simple and exact stepwise short-cuts is highly needed. It happens to be that weighted Holm procedures, the WHP and WAP are such

procedures. Therefore, it will be of interest and necessary to study these two weighted Holm procedures by investigating the similarities and differences between them and make recommendation for their use. In addition, it is also important to study the procedures' optimality property that the procedures cannot be improved without losing the control of error rates. Sarkar, Fu and Guo (2016) improved the Holm procedure by using pairwise dependencies, but without information of dependence it was shown that Holm procedure is an optimal procedure, see Gordon and Salzman (2008) and Gordon (2011). However, there are few studies to explore the optimal property of weighted procedures, in the first part of thesis proposal, we will study the optimal properties of both WHP and WAP. Mielke et al. (2021) and Guilbaud (2021) proposed a stepdown MTP of Holm procedure and its weighted version and a slight modification based on marginal p-values for rejecting at least k out of m null hypotheses.

Closure principle (Marcus et al., 1976) is a powerful tool for constructing multiple testing methods controlling the FWER. The multiple testing procedures constructed by using the closure principle are called closed testing procedures (CTPs) and closed testing is a flexible and easily explained approach to control the overall error rate that has been widely used in pharmaceutical research, particularly in clinical trials settings, see Dmitrienko et al. (2007), Brannath and Bretz (2010) and Henning and Westfall (2015). A hypothesis is rejected in the context of multiple testing if and only if all intersection hypotheses containing this hypothesis are rejected by the local tests in the context of single test; however, sometimes the number of hypotheses is large, then the number of intersection hypotheses increases rapidly and the CTPs are in general difficult to apply. In contrast, graphs are usually easier to communicate with clinical teams than long and abstract decision tables, which typically are not intuitive and can avoid unnecessary computer programming. In order to make the weighted procedures being clear, intuitive and simple to communicate with clinical teams, the graphical representations of both WHP and WAP will be provided. Moreover, one more important reason to find powerful weighted procedures

is that gatekeeping procedures and graphical approaches are commonly used in clinical studies and the weighted procedures are fundamental to develop gatekeeping procedures and graphical approaches, refer to Westfall and Krishen (2001), Ghulam, Wang and Xie (2016), Bretz et al. (2009) and Bretz et al. (2011).

Even though we won't study how to choose proper weights in our projects, its importance can not be ignored. Except for choosing weights based on a priori importance of the hypotheses or prior information, refer to Westfall et al. (2001), the researchers can also choose weights depending on the concurrent data set to improve power without compromising significance levels. There are many approaches using data dependent weights and yet maintain familywise or generalized familywise error control, see Finos and Salmaso (2007), Kang et al. (2009), Dalmasso et al. (2008), Westfall, Kropf and Finos (2004) and Wang (2019). Also some papers have studied weighted parametric procedures based on utilizing the joint distribution of test statistics (See Xie, 2011 and Xi et al., 2017). There are also many weighted procedures with the control over FDR or wFDR, see Benjamini and Hochberg (1997), Genovese, Roeder and Wasserman (2006), Benjamini and Heller (2007), Zhao and Zhang (2014), Benjamini and Cohen (2017), Ramdas et al. (2019) and so on.

For the second part of this dissertation, we focus on investigating two independently developed graphical approaches, the original graphical approach in Bretz et al. (2009) and the default graphical approach in Burman et al. (2009), and providing a direct proof of the FWER control for the original graph. As aforementioned, the original graphical approach is a simple, flexible and clear graphical visualization approach. And both of the original graphical approach and the default graph are used to visualize Bonferroni-based sequentially rejective procedures, so they are usually considered to be equivalent, refer to Bretz et al. (2011), Robertson, Wason and Bretz (2020), etc. In fact, the study finds that two graphical approaches are different, especially, when the number of hypotheses is larger than 3. Moreover, given either one graph of two graphical approaches for testing multiple

hypotheses, we want to see if we can find the other corresponding graphical approach. In addition, the original graphical approach is commonly treated as a unique approach, we want to study if it has this property or not. So in this part, we aim to study both approaches thoroughly and provide theoretical results reference for future study, such as developing a more general graphical approach.

For the third part of this dissertation, we focus on developing a new generalized step-up FDR controlling procedure for discrete data. In many applications such as clinical safety analysis, genome-wide association studies (GWAS) and next generation sequencing data (NGS), the data of the experiments usually are represented by frequency counts. Also many of such experiments often involve a large number of hypotheses to test. In the analysis of such data, researchers often face the problem of multiple testing based on discrete test statistics, aimed at controlling false discovery rate (FDR). Most existing FDR controlling procedures are developed for continuous data, which are often conservative when analyzing discrete data. The reason for this phenomenon is that the distribution of the p-values for discrete data is stochastically larger than uniform (0, 1). Consequently, to develop a procedure, taking the properties of discreteness and heterogeneity of discrete data into account, becomes necessary and inevitable.

In the literature, some FWER controlling procedures for discrete data were developed by considering the special property of discrete data. Tarone (1990) proposed a modified Bonferroni procedure for discrete data, which improves the power by reducing the number of tested hypotheses through eliminating those hypotheses with relatively large minimal attainable p-values. There are also many other FWER controlling procedures, see Hommel and Krummenauer (1998), Roth (1999), Leon and Heo (2005), He and Heyse (2019) and Zhu and Guo (2019). However, only several studies are related to FDR control for discrete data. Gilbert (2005) developed a modified FDR procedure for discrete data, which is a simple two-step combination of the Tarone and BH procedures. Same as Tarone procedure, Gilbert procedure only used the information of minimal attainable p-values. Heyse (2011)

proposed a BH-type procedure relying on the averaged cumulative distribution functions (CDFs) of the p-values of hypotheses when they are true. Heyse procedure is powerful for discrete data; however, it can not be shown to control the FDR at a pre-specified level α. Döhler, Durand and Roquain (2018) provided new FDR upper bounds to help to construct BH-type procedures that incorporate the discrete and heterogeneous structure of the data and provably control the FDR for any fixed number of null hypotheses under independence. As modified versions of Heyse procedure, their procedures are not proved to be more powerful than BH procedure. Döhler (2018) modified the Benjamini-Yekutieli procedure by considering the special properties of discrete data. Recently, some adaptive and weighted procedures were developed to control the FDR for discrete data, refer to Chen, Doerge and Heyse (2018) and Chen , Doerge and Sarkar (2020). Lynch and Guo (2016) presented a generalized step-wise procedure which generalizes the usual step-wise procedure to the case where each hypothesis is tested with a different set of critical constants. But this generalized procedure is only developed for continuous case. In our project, under the same framework of the aforementioned generalized step-up approach, by fully utilizing known marginal distributions of true null p-values, we plan to develop a powerful generalized step-up procedure for discrete case.

1.2 Basic Concepts in Multiple Hypotheses Testing

Consider simultaneously testing m null hypotheses $H_1, \cdots, H_m$ for which m_0 of them are true and m_1 of them are false. Let V, S and R denote the numbers of false rejections, correct rejections and total rejections, respectively. Here, m, m_0 and m_1 are fixed numbers while m is known and m_0 and m_1 are unknown, R can be observable while V and S are unobservable.

1.2.1 Type I error rates and power

Unlike the single hypothesis testing, there are various measures of type I error rate and power. It is important to choose a suitable error rate before developing any MTP. The commonly used ones are described as follows.

(a) Comparisonwise error rate (CWER) is defined as the proportion of incorrectly rejected hypotheses among all tested hypotheses, which is given by:

$$\text{CWER} = E\left(\frac{V}{m}\right) = \frac{E\left(V\right)}{m}.$$

The simple rule of p-value $\leq \alpha$ can control the CWER well, so controlling the CWER does not impose any multiplicity adjustment.

(b) Perfamily error rate (PFER): the expected number of false rejections,

$$\text{PFER} = E\left(V\right).$$

This is the strongest type I error rate in the literature.

(c) Familywise error rate (FWER): the probability of making at least one type I errors, which is given by:

$$\text{FWER} = Pr\left(V \geq 1\right).$$

Controlling the FWER is suitable for the small scale test. And Strong control of the FWER for the primary objections is mandated by FDA in all confirmatory clinical trials.

A MTP strongly controls type I error rate means that it can control type I error rate under any combination of true and false null hypotheses. While weak control means controlling type I error rate only when all null hypotheses are true. Generally, in our proposal thesis, we will focus on strong control of type 1 error rate, since it is hard to know which combination of true and false hypotheses the actual setting is.

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

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

It is introduced by Benjamini and Hochberg (1995) and suitable for the large scale testing problems.

The order of four errors is CWER $\leq$ FDR $\leq$ FWER $\leq$ PFER, when all the null hypotheses are true, we have FDR $=$ FWER.

In addition to the type I error rates control, the performance of a MTP is usually to be evaluated by a suitable overall power. The definitions of two commonly used power are described as follows.

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

$$\text{Minimal power} = Pr\left(S \geq 1\right).$$

The average power is the expected proportion of rejected false null hypotheses among all false null hypotheses,

$$\text{Average power} = E\left(\frac{S}{m_1}\right) = \frac{E\left(S\right)}{m_1}.$$

The average power is commonly used in practice for large scale multiple testing.

1.2.2 Closure principle

Closure principle (Marcus et al., 1976) is a powerful tool and fundamental principle for constructing multiple testing methods controlling the FWER. The MTPs constructed by using the closure principle are called CTPs and closed testing is a flexible and easily explained approach to control the overall error rate that has been widely used in pharmaceutical research, particularly in clinical trials settings. A hypothesis is rejected in the context of multiple testing if and only if all intersection hypotheses containing this

hypothesis are rejected by the local tests in the context of single test. Given a closed test procedure with m null hypotheses $H_i, i \in \{1, \ldots, m\}$, for each hypothesis H_i, we test all the intersection hypotheses

$$H_I = \bigcap_{j \in I} H_j,$$

by valid local tests at a suitable level α , where I is an non-empty index set, such that $I \subseteq \{1, \ldots, m\}$ and $i \in I$. H_i can be rejected provided that all H_I are rejected. And any closed testing procedure strongly controls the FWER at level α.

1.2.3 Adjusted p-values

Westfall and Young (1993) provided a definition that adjusted p-value is the smallest significant level at which one can reject the hypothesis using the given multiple test procedure. And the adjusted p-values incorporate the structure of the underlying decision rule that can be quite complex; thus, they can be compared with global significance level α directly and give the same results as applying the multiple test procedures to the p-values.

1.2.4 Assumptions on p-values

Consider simultaneously testing m hypotheses H_i, $i = 1, ..., m$, with the associated p-values $P_1, ..., P_m$. Under true null hypotheses, the distribution of marginal p-values is assumed to be distributed as follows,

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

where I_0 is the index set of true null hypotheses.

In MTPs, we usually consider several types of joint dependence structure: arbitrary dependence, independence, and positive regression dependence on subset (PRDS). Under arbitrary dependence, we do not know any specific dependence structure of the p-values.

Definition 1.2.1 *A set D is called increasing if $x \in D$ and $y \geq x$ implies $y \in D$. Here, the inequality between vectors x and y is interpreted coordinatewise.*

Assumption 1.2.1 (PRDS) *The random variables $(P_1, ..., P_m)$ are PRDS on I_0 if for any increasing D and each $i \in I_0$,*

$$Pr\left((P_1, ..., P_m) \in D \mid P_i = p\right)$$

is non-decreasing in p.

First and second parts of this dissertation is discussed under arbitrary dependence, while the third part, independent structure of the p-values is assumed.

1.3 Multiple Testing Procedures

Consider we simultaneously test m hypotheses H_1, ..., H_m with p-values P_1, ..., P_m and positive weights w_1, ..., w_m, correspondingly. Then, the weighted p-values are $\tilde{P}_i = \frac{P_i}{w_i}$, $i = 1, ..., m$. Let $P_{(1)} \leq ... \leq P_{(m)}$ be the ordered version of the p-values P_1, ..., P_m with corresponding hypotheses $H_{(1)}$, ..., $H_{(m)}$ and corresponding weights $w_{(1)}, ..., w_{(m)}$. Let $\tilde{P}_{(1)} \leq ... \leq \tilde{P}_{(m)}$ be the ordered version of the weighted p-values $\tilde{P}_1$, ..., $\tilde{P}_m$ with corresponding hypotheses $H^*_{(1)}$, ..., $H^*_{(m)}$ and corresponding weights $w^*_{(1)}, ..., w^*_{(m)}$. Suppose there are m_0 true null hypotheses and m_1 false null hypotheses and let I_0 denote the indices of true nulls. Let $\alpha_1 \leq \alpha_2 \leq ... \leq \alpha_m$ be a sequence of increasing critical constants and let $\alpha_0 = 0$.

There are three main types MTPs: p-value based MTPs, parametric MTPs and resampling based MTPs. In this proposal thesis, we only focus on p-value based MTPs. Usually they are stepwise procedures.

1.3.1 Several classical stepwise FWER controlling procedures

Single-step procedure: A stepwise procedure with the same critical constant c, reject any hypothesis H_i if and only if $P_i \leq c$.

Bonferroni Procedure(1936): the most classical single-step procedure, reject H_i if $P_i \leq \alpha/m$, $i = 1, ..., m$.

When the number of hypotheses m is large, eg, m is tens of thousands, Bonferroni procedure will be very conservative.

Step-down Procedure: A step-down procedure begins with the most significant hypothesis $H_{(1)}$, gradually steps down to the least significant hypothesis $H_{(m)}$. Reject $H_{(1)}, ..., H_{(R)}$, where $R = max\left\{0 \leq i \leq m\colon P_{(j)} \leq \alpha_j,\ \forall j \in \{0, ..., i\}\right\}$.

Holm procedure (1979): is a typical step-down version of the Bonferroni procedure, reject $H_{(i)}$ when

$$P_{(j)} \leq \frac{\alpha}{m - j + 1}, \quad j = 1, ..., i. \tag{1.1}$$

Step-up Procedure: A step-up procedure begins with the least significant hypothesis $H_{(m)}$, gradually steps up to the most significant hypothesis $H_{(1)}$. Reject $H_{(1)}, ..., H_{(R)}$, where $R = max\left\{0 \leq i \leq m\colon P_{(i)} \leq \alpha_i\right\}$.

Hochberg Procedure (1988): a popular step-up procedure, reject $H_{(1)}$, $H_{(2)}$,..., $H_{(j)}$, when

$$P_{(j)} \leq \frac{\alpha}{m - j + 1}, \quad j = m, ..., 1. \tag{1.2}$$

The adjusted p-value P_i^{adj} for a hypothesis H_i is the required smallest FWER level at which one would reject the hypothesis using the given multiple testing procedure for controlling the FWER. The hypothesis will be rejected if the corresponding adjusted p-value is less than or equal to α. The following are adjusted p-values for Bonferroni procedure, Holm procedure and Hochberg procedure.

Bonferroni procedure:

$$P_i^{adj} = \min\left(1,\ mP_i\right), \quad i = 1, ..., m.$$

Holm procedure:

$$P_{(i)}^{adj} = \begin{cases} \min\left(1\ ,\ mP_{(i)}\right), & \text{if i = 1,} \\ \max\left(P_{(i-1)}^{adj}\ ,\ (m-i+1)P_{(i)}\right), & \text{if i = 2,...,m.} \end{cases}$$

Hochberg procedure:

$$P_{(i)}^{adj} = \begin{cases} P_{(i)}, & \text{if i = m,} \\ \min\left(P_{(i+1)}^{adj}\ ,\ (m-i+1)P_{(i)}\right), & \text{if i = m-1,...,1.} \end{cases}$$

1.3.2 Weighted procedures with control over FWER

The most popular weighted procedures are Rubin and Rosenthal's weighted Bonferroni procedure and Holm's weighted Holm procedure (WHP). Another popular weighted procedure is Benjamini and Hochberg's weighted alternative Holm procedure (WAP). Tamhane and Liu (2008)'s weighted Hochberg procedures lack a simple stepwise structure and were developed under independence or PRDS structure. The weighted procedures are very fundamental to develop popular gatekeeping procedures and graphical approaches, which are commonly used in clinical studies (see Dmitrienko, 2003, Dmitrienko et al., 2007, Bretz et al., 2009, Bretz et al., 2011, etc).

Procedure 1.3.1 (Rubin and Rosenthal, 1983) *Weighted Bonferroni procedure: Reject* H_i *if* $P_i \leq \frac{w_i}{\sum_{k=1}^{m} w_k}\alpha$, $i = 1, ..., m$.

Weighted Bonferroni procedure is a basic and typical procedure, but again, if the testing hypotheses' number is large, it might be conservative.

Procedure 1.3.2 (Holm, 1979) *Weighted Holm Procedure (WHP): Reject* $H_{(i)}^{*}$ *when*

$$\tilde{P}_{(j)} \leq \frac{\alpha}{\sum_{k=j}^{m} w_{(k)}^{*}}, \quad j = 1, ..., i. \tag{1.3}$$

Procedure 1.3.3 (Benjamini and Hochberg, 1997) *An alternative weighted procedure (WAP): Reject $H_{(i)}$ when*

$$P_{(j)} \leq \frac{w_{(j)}}{\sum_{k=j}^{m} w_{(k)}}\alpha, \quad j = 1, ..., i. \tag{1.4}$$

We can see that the testing order of the WHP is based on weighted p-values while the WAP is based on original p-values. However, when we calculate weighted p-values ($\tilde{P}_i = \frac{P_i}{w_i}$) if the weights do not change the order of p-values, which means $\tilde{P}_{(i)} = \frac{P_{(i)}}{w_{(i)}}$, $i = 1, ..., m$, the WHP and WAP are equivalent. For example, suppose simultaneously test $m = 3$ hypotheses $\{H_1, H_2, H_3\}$ with the corresponding weights $w_i = i$, $i = 1, 2, 3$, assume that p-values $p_1 = 0.01$, $p_2 = 0.03$ and $p_3 = 0.09$ were observed; thus, the observed weighted p-values $\tilde{p}_1 = 0.01$, $\tilde{p}_2 = 0.015$, and $\tilde{p}_3 = 0.03$. We can see that the weights do not change the testing order, and both of WHP and WAP compare sequentially p_1, p_2 and p_3 with $\alpha/6$, $2\alpha/5$ and α.

1.3.3 Graphical approach

The decision tables of the aforementioned multiple testing procedures might 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 Bonferroni-type sequentially rejective procedures. The graphical approaches are easy to communicate with other clinical team members and can avoid unnecessary computer programming.

In a graph, each null hypothesis is located at a vertex and the overall critical value α is allocated to each vertex initially and the relationships between null hypotheses are expressed by transition coefficients. Define initial significance levels $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)$ with $\sum_{i=1}^{m} \alpha_i = \alpha \in (0, 1)$ if based on original p-values. Let $M = \{1, ..., m\}$ and $G = (g_{ij})_{m\times m}$, where g_{ij} is the fraction of the level of H_i that is propagated to H_j with $0 \leq g_{ij} \leq 1$, $g_{ii} = 0$, and $\sum_{j=1}^{m} g_{ij} \leq 1, \forall i = 1, ..., m$. $(\boldsymbol{G}, \boldsymbol{\alpha})$ determines a graph with an associated

multiple test. And all local critical values and transition coefficients will be updated as well based on the following rules:

Bretz et al. (2009) Algorithm I:

0. Set $I = M$.

1. Let $j = \arg min_{i \in I} p_i/\alpha_i$

2. If $p_j \leq \alpha_j$, reject H_j; otherwise stop.

3. Update the graph:

$$I \to I \setminus \{j\}$$

$$\alpha_l \to \begin{cases} \alpha_l + \alpha_j g_{jl}, & l \in I \\ 0 & \text{otherwise} \end{cases}$$

$$g_{lk} = \begin{cases} \frac{g_{lk} + g_{lj} g_{jk}}{1 - g_{lj} g_{jl}}, & l,\, k \in I, l \neq k \\ 0 & \text{otherwise} \end{cases}$$

4. If $|I| \geq 1$, go to step 1; otherwise stop.

Bretz et al. (2009) shows that the graphical approach strongly controls the FWER at level α if the constraint conditions $\sum_{i=1}^{m} \alpha_i \leq \alpha \in (0, 1)$, $\sum_{j=1}^{m} g_{ij} \leq 1, \forall i = 1, ..., m.$ and $0 \leq g_{ij} \leq 1$, $g_{ii} = 0$ are satisfied.

1.3.4 FDR controlling procedures

There are many existing FDR controlling procedures, for example, the most popular ones are BH procedure and BY procedure. However, most of them are developed for continuous data, without considering the discreteness and heterogeneity, those procedures might be

much conservative when applying to discrete data. BH procedure and BY procedure are described as follows.

BH procedure: Benjamini and Hochberg (1995) first introduced the FDR and developed a step-up procedure controlling FDR with the critical constants : $\alpha_i = \frac{i\alpha}{m}$, $i = 1, ..., m$. BH procedure strongly controls the FDR at level α under independence.

BY procedure: Benjamini and Yekutieli (2001) showed BH procedure can also control the FDR under positive regression dependence on subset *PRDS* and developed a step-up procedure controlling FDR with the critical constants: $\alpha_i = \frac{i\alpha}{mC_m}$, where $C_m = \sum_{j=1}^{m} 1/j$, $i = 1, ...m$, under arbitrary dependence.

1.4 Motivation and Outline

In many statistical applications, such as clinical studies, hypotheses might be assigned different weights according to their different importance. Among the existing weighted procedures, the weighted Holm procedures are the most popular and easy to be applied without any assumption of dependence structure. There are two common weighted Holm procedures: the WHP whose testing order depends on weighted p-values; while the WAP depends on raw p-values. But the relation between them is still not clear yet. In addition, the weighted Bonferroni procedure might be too conservative when the number of hypotheses is large and weighted Hochberg procedures lack a simple stepwise short-cuts; thus, it's interesting to study the statistical properties of the weighted Holm procedures and recommend their application.

Closure principle is a powerful tool for constructing multiple testing methods controlling the FWER. Through finding the underlying CTPs of the procedures we will understand the weighted procedures clearly and thoroughly. Besides, the graphical approaches are usually easier to communicate with other clinical team members intuitively than long and abstract decision tables and can avoid unnecessary computer programming. Moreover, the weighted procedures are fundamental to develop gatekeepings and graphical

procedures. Therefore, it will be of interest and necessary to study these two weighted Holm procedures by investigating the similarities and differences between them based on comparing their corresponding closed testing procedures, graphical representations and adjusted p-values, and make recommendation for their use.

In addition, it is also important to study the MTPs' optimality property that the procedures cannot be improved without losing the control of error rates. To the best of our knowledge, there is no research studying the optimality of the WHP and WAP. Thus, in the first part of this dissertation, we also investigate if the WHP and WAP are optimal procedures.

In the second part of this dissertation, we mainly focus on studying the similarities and differences between two graphical approaches which are independently developed and usually considered to be same. One is called the original graphical approach which is popular and widely used in clinical trials. The other one is named the default graphical approach, a combination of fixed sequences of hypotheses. As there is no transition coefficient involved, the recycled of critical values of the rejected hypotheses is straightforward in the default graph. Moreover, in Bretz et al. (2009), it provided an indirectly proof of the FWER control by showing the equivalence between the original graphical approach and short-cut of CTPs. In this part, in order to understand the original graphical approach thoroughly and provide theoretical reference, we will give an elegant theoretical result about the FWER control directly by the original graphical approach.

For the third part of this dissertation, we focus on developing a generalized step-up FDR controlling procedure for discrete data. Most existing MTPs procedures are developed for continuous data, which are often conservative when analyzing discrete data because the true null distributions of discrete data are stochastically larger than uniform (0, 1). Also, many procedures, which were developed for discrete data, are based on the minimal attainable p-values. In fact, if we know the minimal attainable p-values, the CDFs under true nulls are also known. Moreover, the CDFs under different true null hypotheses might

be different which implies using a unique function of rejection rule might cause power loss. Consequently, to develop a procedure, taking the properties of discreteness and heterogeneity of discrete data into account, becomes necessary and inevitable.

In the literature, many FWER controlling procedures for discrete data were developed by considering the special property of discrete data, refer to Tarone (1990), Hommel and Krummenauer (1998), Roth (1999) and Zhu and Guo (2019). However, there are only a few literature discussing FDR control for discrete data, see Gilbert (2005), Heyse (2011) and Döhler, Durand and Roquain (2018). Usually, for large scale hypotheses testing problem, the procedures with FDR control are highly needed. And none of the aforementioned FDR controlling procedures controls the FDR and be more powerful than BH procedure theoretically at the same time. According to Lynch and Guo's (2016) generalized step-up procedure, we can see that each hypothesis can have its own critical function; therefore, this framework is especially suitable for developing powerful FDR controlling procedures for discrete data.

The rest of this dissertation is outlined as follows: in Chapter 2, we study the similarities and differences between the WHP and WAP and show that the WHP is more powerful than the WAP by constructing and comparing their corresponding closed testing procedures, graphical representations, and adjusted p-values. Also, we provide a proof that the WHP is an optimal procedure in the sense that the procedure cannot be improved by increasing even one of its critical values without losing control over the FWER. Simulations were conducted to provide numerical evidence of superior performance of the WHP in terms of the FWER controlling and average power. In Chapter 3, we investigate the similarities and differences between the original graphical approach and the default graphical approach. Also, we provide an elegant theoretical result that is a direct proof of the FWER control for the original graphical approach. In Chapter 4, we develop a new generalized step-up procedure for discrete data with proven of FDR control by fully

utilizing the distributions of marginal p-values. In Chapter 5, we have summarized the main results of this dissertation and discussed the future works.

CHAPTER 2

ON WEIGHTED HOLM PROCEDURES

2.1 Introduction

In this chapter, we aim to study the similarities and differences of two common weighted Holm procedures and make recommendation for their use. In many clinical applications, it often happens that some hypotheses are more important than the others, which suggests us to assign different weights to different hypotheses according to their importance. In the existing literature, some weighted procedures based on Bonferroni test, Simes test and resampling-based tests were developed, see Rosenthal and Rubin (1983), Holm (1979), Benjamini and Hochberg (1997), Tamhane and Liu (2008), Westfall and Young (1993) and so on. Rosenthal and Rubin (1983) proposed a weighted Bonferroni procedure which permits greater power for the important hypotheses. There are two common weighted Holm procedures. One is based on ordered weighted p-values that we called WHP, see Holm (1979); the alternative weighted Holm procedure that is based on ordered raw p-values is named WAP, refer to Benjamini and Hochberg (1997). The questions arise that what the similarities and differences are between these two procedures and which one performs better in terms of power and the FWER control? Apparently, the later procedure is more objective than the former one since the testing order, based on raw p-values, does not depend on the weights; however, it also loses the monotone property because of this reason. For stepwise procedures with asymmetric rules for updating the hypothesis weights, the ordering of the testing hypotheses is much relevant and utmost importance, including the fixed-sequence (Maurer et al., 1995) and fallback procedures (Wiens, 2003). Wiens et al. (2013) examined the behavior of different classes of multiple testing procedures (MTPs) in problems with unequally weighted hypotheses and a prior ordered hypotheses and provided practical guidelines for the choice of hypothesis weights and hypothesis ordering. However, the similarities and differences of ordering based on original or weighted p-values have

not been studied. Tamhane and Liu (2008) constructed weighted Hochberg-type step-up multiple test procedures including two closed procedures based on weighted Simes tests: one based on ordered raw p-values and the other one based on ordered weighted p-values. However, both two weighted procedures lack simple stepwise structure, therefore it is hard to be compared with each other and not easy to explain it to practitioners. And also these procedures were developed under independence structure. Thus, it's interesting to study the statistical properties of the weighted Holm procedures and recommend their application.

Moreover, the Bonferroni-type weighted procedures can often be described as some specific closed testing procedures (CTPs) and can be visualized by some specific graphics. A CTP constructed by using the closure principle (Marcus et al., 1976) is a flexible and easily explained approach to control the FWER, particularly in clinical trials settings, see Brannath and Bretz (2010) and Henning and Westfall (2015). But when simultaneously testing a large number of hypotheses, the number of intersection hypotheses increases rapidly and the CTPs are in general difficult to apply. In contrast, graphs can visualize many MTPs and gatekeeping procedures; thus, graphs are usually easier to communicate with clinical teams than long and unintuitive decision tables of the CTPs and they also can avoid unnecessary computer programming. In the literature, graphical approaches are commonly used in clinical studies, see Bretz et al.(2009) and Bretz et al. (2011). The weighted multiple testing procedures (wMTPs) are also very fundamental to develop popular gatekeeping procedures which are commonly used in clinical studies, such as serial gatekeeping see Maurer et al. (1995) and Westfall and Krishen (2001), parallel gatekeeping, refer to Dmitrienko et al. (2003) and Dmitrienko et al. (2007) and mixture procedures for gatekeeping, see Dmitrienko and Tamhane (2013). Therefore, to find simple and powerful weighted procedures will be especially helpful to develop useful and powerful graphical approaches and gatekeeping procedures. In addition, it is also important to study the optimality property of weighted procedures that the procedures cannot be improved without

losing the control of error rates. To the best of our knowledge, there is no research studying the optimality of the WHP and WAP.

In summary, we have studied these two weighted Holm procedures by investigating the similarities and differences between them according to their corresponding closed testing procedures, graphical representations and adjusted p-values and make recommendation for their use. Theoretically we show that the WHP is uniformly more powerful than the WAP. In addition, we have an interesting finding that the WAP is not monotone in the sense that a procedure rejects a hypothesis with a smaller p-value whenever it rejects another hypothesis with a larger p-value and also it does not satisfy the monotonicity condition in the sense of the weighted functions being monotone in terms of the index subsets of the corresponding CTP; however, it is a consonant procedure which usually leads to a shortcut of a procedure. Moreover, we have found that the graphical approach corresponding to the WAP is beyond the domain of the original graphical approach (Bretz et al., 2009). Then, we also discuss optimality of these two procedures and show that the WHP is an optimal procedure in the sense that the procedure cannot be improved by increasing even one of its critical values without losing control over the FWER, while the WAP is an optimal procedure under some condition that is the proportions of smallest weight to other weights are not less than the global significance level α. Through some clinical examples and simulations, we also give some numerical results that the WHP is always more powerful than the WAP and properly choosing weights will improve the performance of both procedures in terms of the average power, especially the WAP. The rest of the chapter is organized as follows: Section 2.2 introduces some basic notations and concepts used in this paper. In Section 2.3, we find the corresponding closed testing procedures. And we also discuss two different monotone properties of both procedures. The graphical representations are provided in Section 2.4. We derive the adjusted p-values and adjusted weighted p-values for both procedures in Section 2.5. In Section 2.6, we have discussed the optimality properties of the WHP and the WAP. Simulation studies are conducted in

Section 2.7. Two real data examples are used to show the performance of the WHP and WAP in Section 2.8. In Section 2.9, some discussions are briefly provided.

2.2 Preliminaries

In this section, we introduce some general notations and two weighted Holm procedures that will be studied in this paper.

2.2.1 Basic notations

Consider simultaneously testing m hypotheses H_1, ..., H_m which are associated with p-values P_1, ..., P_m and positive weights w_1, ..., w_m, correspondingly. Then, the weighted p-values are $\tilde{P}_i = \frac{P_i}{w_i}$, $i = 1, ..., m$. Let $P_{(1)} \leq ... \leq P_{(m)}$ be the ordered version of the p-values P_1, ..., P_m with corresponding hypotheses $H_{(1)}$, ..., $H_{(m)}$ and corresponding weights $w_{(1)}, ..., w_{(m)}$. Let $\tilde{P}_{(1)} \leq ... \leq \tilde{P}_{(m)}$ be the ordered version of the weighted p-values $\tilde{P}_1$, ..., $\tilde{P}_m$ with corresponding hypotheses $H^*_{(1)}$, ..., $H^*_{(m)}$ and corresponding weights $w^*_{(1)}, ..., w^*_{(m)}$. Suppose there are m_0 true null hypotheses and m_1 false null hypotheses and let I_0 denote the indices of true nulls. Let V be the number of true null hypotheses among the R rejected null hypotheses in a multiple testing procedure. The the familywise error rate (FWER) is defined as the probability of making at least one type I error rate.

2.2.2 Two weighted Holm procedures

Two most popular and widely used weighted Holm procedures that we will study are described as follows.

Procedure 2.2.1 *(Holm, 1979)*

The weighted Holm procedure (WHP): Reject $H^*_{(i)}$ when

$$\tilde{P}_{(j)} \leq \frac{\alpha}{\sum_{k=j}^{m} w^*_{(k)}}, \quad j = 1, ..., i. \tag{2.1}$$

The testing order of the WHP is based on the order of weighted p-values, and this procedure was developed based on the weighted Bonferroni test.

Procedure 2.2.2 *(Benjamini and Hochberg, 1997)*

An alternative weighted Holm procedure (WAP): Reject $H_{(i)}$ when

$$P_{(j)} \leq \frac{w_{(j)}}{\sum_{k=j}^{m} w_{(k)}}\alpha, \quad j = 1, ..., i. \tag{2.2}$$

To explore the similarities and differences between the WHP and WAP, we will study from three aspects: the underlying closed testing procedures, the graphical representations and adjusted p-values. First, we study the underlying CTPs of both procedures and their monotonicity properties and consonant property. Then we also use the graphical approaches to visualize both procedures and give two clear and intuitive examples. Finally, we will give the adjusted p-values and adjusted weighted p-values for the WAP and the WHP, respectively, since the adjusted p-values incorporate the structure of the underlying decision rules that can be quite complex; thus, they can be compared with global significance level α directly and give the same results as applying the multiple test procedures to the p-values.

2.3 The Underlying CTPs of the WAP and WHP and Some Theoretical Results

In order to study the statistical properties of the WHP and the WAP, we investigate the underlying CTPs for both procedures, which will help us to understand them well and conduct a comparison easily. Also, the consonance property of a MTP has not received much attention yet, especially for the WAP. While studying the local tests of both procedures will bring us with new understanding of consonance property.

First, we will study the CTP of the WAP and provide a proof of equivalence between the WAP and its corresponding CTP. The WAP can be expressed as a closed testing procedure where each intersection hypothesis in the closure of the family is tested with a modified weighted Bonferroni test.

For any intersection hypothesis $H_I = \cap_{i\in I} H_i$, where $I \subseteq \{1, 2, ..., m\}$, let $P^I_{(1)} = \min_{i\in I}\{P_i\}$ with corresponding weight $w^I_{(1)}$. Here, $P^I_{(1)}$ and $w^I_{(1)}$ denote the smallest p-value in $\{P_i, i \in I\}$ and the corresponding weight, respectively. Then we use local test

$$P^I_{(1)} \leq \frac{w^I_{(1)}}{\sum_{i\in I} w_i}\alpha \tag{2.3}$$

to reject H_I; otherwise, retain it.

The above local test is valid, the proof is given as follows.

The type 1 error rate of this local test is:

$$\begin{aligned} Pr\left(P^I_{(1)} \leq \frac{w^I_{(1)}}{\sum_{i\in I} w_i}\alpha\right) &\leq Pr\left(\cup_{i\in I} P_i \leq \frac{w_i}{\sum_{i\in I} w_i}\alpha\right) \\ &\leq \sum_{i\in I} Pr\left(P_i \leq \frac{w_i}{\sum_{i\in I} w_i}\alpha\right) \\ &\leq \alpha \end{aligned} \tag{2.4}$$

Thus, by closure principle, the FWER is controlled at level α.

Proposition 2.3.1 *The closed testing procedure with the local tests defined in (2.3) is equivalent to Benjamini and Hochberg(1997)'s weighted alternative Holm procedure(WAP).*

Second, in order to find the similarity and difference between the WHP and the WAP, we will study the corresponding CTP of the WHP. The WHP can be expressed as a closed testing procedure (CTP) where each intersection hypothesis in the closure of the family is tested with a weighted Bonferroni test, refer to Westfall and Krishen (2001).

For testing any intersection hypothesis $H_I = \cap_{i\in I} H_i$, where $I \subseteq \{1, 2, ..., m\}$, the local test for each intersection hypothesis is

$$P_i \leq \frac{w_i}{\sum_{i\in I} w_i}\alpha, \tag{2.5}$$

for some $i \in I$ to reject H_I; otherwise, retain it.

The local test , weighted Bonferroni test, is a valid test. The reason is given as follows. The type 1 error rate of this local test is:

$$\begin{aligned} Pr\left(\cup_{i\in I}P_i \leq \frac{w_i}{\sum_{i\in I} w_i}\alpha\right) &\leq \sum_{i\in I} Pr\left(P_i \leq \frac{w_i}{\sum_{i\in I} w_i}\alpha\right) \\ &\leq \alpha \end{aligned} \tag{2.6}$$

Proposition 2.3.2 (Westfall and Krishen, 2001) *The closed testing procedure with the local tests defined in (2.5): the weighted Bonferroni test, is equivalent to Holm (1979)'s weighted Holm procedure(WHP).*

By looking into the local tests of both WHP and WAP, we have the following result:

Theorem 1 *For any hypothesis H_i, $i = 1, ..., m$, if it is rejected by Alternative Weighted Holm procedure (WAP) (Benjamini and Hochberg, 1997), it will be rejected by Weighted Holm procedure (WHP) (Holm, 1979). In other words, the WHP is always more powerful than the WAP under arbitrary dependence.*

Proof for Theorem 1:

Note that the event $\left\{P^I_{(1)} \leq \frac{w^I_{(1)}}{\sum_{i\in I} w_i}\alpha\right\}$ implies $\left\{\cup_{i\in I}P_i \leq \frac{w_i}{\sum_{i\in I} w_i}\alpha\right\}$ for any intersection hypothesis H_I, $I \subseteq \{1, ..., m\}$. Therefore, if the intersection hypothesis H_I is rejected by WAP's local test, it must be rejected by WHP's local test.

Thus, by closure principle, the hypothesis H_i, $i = 1, ..., m$, which is rejected by the WAP, will be always rejected by the WHP. Thus, the WHP performs better than the WAP in terms of power. □

Remark 1 *We know that critical values for the WAP do not have the monotone property in the following Definition 2.3.1 (see Benjamini and Hochberg, 1997) which will cause power loss. And we can see that when we calculate weighted p-values ($\tilde{P}_i = \frac{P_i}{w_i}$) if the weights do not change the order of p-values, which means $\tilde{P}_{(i)} = \frac{P_{(i)}}{w_{(i)}}$, $i = 1, ..., m$, the WHP and the WAP are equivalent. The reason is described as follows,*

$$\tilde{P}_{(i)} \leq \frac{\alpha}{\sum_{k=i}^{m} w^*_{(k)}} \iff P_{(i)} \leq \frac{w_{(i)}}{\sum_{k=i}^{m} w_{(k)}}\alpha.$$

Next, we will discuss the monotone and consonant properties of the WHP and WAP. Two different monotone definitions are described as follows. Dimitrienko et al.(2009) introduced a monotone property which is defined for the general MTPs in terms of p-values. While the monotonicity condition is defined for the CTPs with weighted Bonferroni-type local tests in the sense of the weighted functions being monotone in terms of the index subsets, see Romano and Wolf (2005), Hommel, Bretz and Maurer (2007) and Bretz et al. (2009). For such CTPs, they are always monotone in terms of the p-values.

Definition 2.3.1 *(Monotone property)*
A multiple testing procedure is called to be p-value monotone if some p-values become smaller, then at least the same or even more hypotheses would be rejected by this procedure.

Definition 2.3.2 *(Monotonicity condition)*
Let $\alpha_i(I)$, $i \in I \subseteq M = \{1, ..., m\}$ denote the local significance levels for an intersection hypothesis such that $\sum_{i \in I} \alpha_i(I) \leq \alpha$, then the monotonicity condition is $\alpha_i(I) \leq \alpha_i(J)$ for all i, I, J with $i \in J$ and $J \subset I \subseteq M$.

Definition 2.3.3 *(Consonance, Gabriel, 1969)*
A multiple testing procedure is termed consonant if the rejection of an intersection hypothesis H_I with $I \subseteq \{1, ..., m\}$ and $|I| > 1$ always leads to the rejection of at least one H_J implied by H_I, i.e., H_J with $J \subset I$.

The WHP is monotone in the sense that the WHP rejects a hypothesis with a smaller p-value whenever it rejects another hypothesis with a larger p-value while the WAP does not have this property, see Benjamini and Hochberg (1997). It is easy to see the WHP is monotone by looking into its local test for each intersection hypothesis of the corresponding CTP of WHP, for example, for testing any intersection hypothesis $H_I = \cap_{i \in I} H_i$, $I \subseteq M = \{1, ..., m\}$, if some $P_i \leq \frac{w_i \alpha}{\sum_{i \in I} w_i}$ for all I such that $i \in I$, then by closure principle, H_i will be rejected. Obviously, if making such P_i smaller, all the H_I including H_i will have

a larger chance to be claimed significant, then by closure principle, the monotone property in Definition 2.3.1 is achieved.

In addition, for the WHP, through its corresponding Bonferroni-type local tests, we can know that its weighting function is monotone; thus, the CTP of WHP satisfy the monotonicity condition (See Hommel et al., 2007 and Bretz et al., 2009) which leads to a consonant closed test. By looking into the corresponding local tests of the WAP, which is $P^I_{(1)} \leq \frac{w^I_{(1)}}{\sum_{i\in I} w_i}\alpha$, the local significance level is a constant for all the component hypotheses in I; thus, the WAP does not satisfy the monotonicity condition. For example, consider simultaneously testing 3 hypotheses H_1, H_2 and H_3, with $P_1 \leq P_2 \leq P_3$ and weights $w_1 = 4$, $w_2 = w_3 = 1$, respectively. For testing the intersection hypothesis $H_1 \cap H_2 \cap H_3$, the local critical value for H_2 is $2\alpha/3$; while when testing the intersection hypothesis $H_2 \cap H_3$ the local critical value for H_2 is $\alpha/2$. However, its corresponding closed test is consonant, and we have the following result.

Proposition 2.3.3 *The corresponding CTP of the WAP is a consonant procedure, although it does not satisfy the monotonicity condition in Definition 2.3.2 and it is also not monotone in Definition 2.3.1.*

The proof of Proposition 2.3.3 is provided in the Appendix.

The following example 2.3.1 explains if the orders of the original p-values and the weighted p-values are same, the WHP and the WAP are equivalent when applying them to the data, respectively. While if the orders are different, the WHP performs better than the WAP in terms of power, see example 2.3.2.

Example 2.3.1 *Suppose simultaneously test $m = 3$ hypotheses $\{H_1, H_2, H_3\}$ with the corresponding weights $w_i = i$, $i = 1, 2, 3$, assume that p-values $p_1 = 0.01$, $p_2 = 0.03$ and $p_3 = 0.09$ were observed; thus, the observed weighted p-values $\tilde{p}_1 = 0.01$, $\tilde{p}_2 = 0.015$, and $\tilde{p}_3 = 0.03$. We can see that the weights do not change the testing order, and both of the WHP and WAP compare sequentially p_1, p_2 and p_3 with $\alpha/6$, $2\alpha/5$ and α.*

Example 2.3.2 *Suppose* $m = 3$*,* $w_i = i$*,* $i = 1, 2, 3$*, and* $\alpha = 0.05$*, assume that* p*-values* $p_1 = 0.01$*,* $p_2 = 0.014$ *and* $p_3 = 0.3$ *were observed; thus, the observed weighted* p*-values* $\tilde{p}_1 = 0.01$*,* $\tilde{p}_2 = 0.007$*, and* $\tilde{p}_3 = 0.1$*; thus, we can see the weights do change the testing order of hypotheses. The WAP compares sequentially* p_1*,* p_2 *and* p_3 *with* $\alpha/6 = 0.0083$*,* $2\alpha/5 = 0.02$*, and* 0.05 *while the WHP compares sequentially* p_2*,* p_1 *and* p_3 *with* $2\alpha/6 = 0.0167$*,* $\alpha/4 = 0.0125$ *and 0.05. Then the WAP does not reject any hypothesis but the WHP rejects* H_1 *and* H_2*.*

2.4 The Graphical Representations of the WHP and WAP

In a graph, each null hypothesis is located at a node and the overall critical value α is allocated to each node initially and the relationships between null hypotheses are expressed by transition coefficients. Define initial significance levels $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)$ with $\sum_{i=1}^{m} \alpha_i \leq \alpha \in (0, 1)$. Let $M = \{1, ..., m\}$ and $G = (g_{ij})_{m \times m}$, where g_{ij} is the fraction of the level of H_i that is propagated to H_j when H_i is rejected, with $0 \leq g_{ij} \leq 1$, $g_{ii} = 0$, and $\sum_{j=1}^{m} g_{ij} \leq 1$, $\forall i = 1, ..., m$. Then the initial graph $(\boldsymbol{G}, \boldsymbol{\alpha})$ and a proper updating algorithm determine graphs with an associated multiple test.

Let the initial transition coefficient matrix is

$$\boldsymbol{G} = \begin{bmatrix} 0 & \frac{w_2}{\sum_{j=2}^{m} w_j} & \cdots & \frac{w_m}{\sum_{j=2}^{m} w_j} \\ \frac{w_1}{\sum_{j=1, j\neq 2}^{m} w_j} & 0 & \cdots & \frac{w_m}{\sum_{j=1, j\neq 2}^{m} w_j} \\ & \cdots & & \\ \frac{w_1}{\sum_{j=1}^{m-1} w_j} & \frac{w_2}{\sum_{j=1}^{m-1} w_j} & \cdots & 0 \end{bmatrix}.$$

Let the initial local significance levels $\alpha_i = \frac{w_i}{\sum_{i=1}^{m} w_i} \alpha$, $i = 1, ..., m$; then, the initial levels satisfy $\sum_{i=1}^{m} \alpha_i \leq \alpha \in (0, 1)$.

Graphical algorithm corresponding to the WHP according to Bretz et al. (2009):

0. Set $I = M$.

1. Let $j = \arg min_{i \in I} \tilde{p}_i$.

2. If $p_j \leq \alpha_j$, reject H_j; otherwise stop.

3. Update the graph:

$$I \to I \setminus \{j\}$$

$$\alpha_l \to \begin{cases} \alpha_l + \alpha_j g_{jl}, & l \in I \\ 0 & \text{otherwise} \end{cases}$$

$$g_{lk} = \begin{cases} \frac{g_{lk} + g_{lj} g_{jk}}{1 - g_{lj} g_{jl}}, & l,\, k \in I, l \neq k \\ 0 & \text{otherwise} \end{cases}$$

4. If $|I| \geq 1$, go to step 1; otherwise stop.

Formal graphical algorithm corresponding to WAP according to Bretz et al. (2009):

0. Set $I = M$.

1. Let $j = \arg min_{i \in I} p_i$.

2. If $p_j \leq \alpha_j$, reject H_j; otherwise stop.

3. Update the graph:

$$I \to I \setminus \{j\}$$

$$\alpha_l \to \begin{cases} \alpha_l + \alpha_j g_{jl}, & l \in I \\ 0 & \text{otherwise} \end{cases}$$

$$g_{lk} = \begin{cases} \frac{g_{lk} + g_{lj} g_{jk}}{1 - g_{lj} g_{jl}}, & l,\, k \in I, l \neq k \\ 0 & \text{otherwise} \end{cases}$$

4. If $|I| \geq 1$, go to step 1; otherwise stop.

The graphical approach of the WHP was mentioned in Alosh et al. (2014) and Guilbaud (2018), but only restricted to some specific number (e.g., $m = 3$) of hypotheses examples; however, we have provided a general graphical representation for the WHP.

Remark 2 *Moreover, we can note that the graphical approach corresponding to the WAP is beyond the domain of the original graphical approach (Bretz et al., 2009) since its local tests are not weighted Bonferroni tests and also do not satisfy the monotonicity conditions. So it will be highly needed to develop a more general graphical approach.*

Given the initial graphs and algorithms, we have the following result.

Proposition 2.4.1 *The WHP and the WAP are equivalent to their corresponding graphical representations that found above, respectively.*

The following example 2.4.1 is used to illustrate different performance of the corresponding graphical representations of WHP and WAP.

Example 2.4.1 *Here, we use same setting of example 2.3.2. We can see from Figure 2.1 and Figure 2.2, the results are same as example 2.3.2, the WAP rejects none while the WHP rejects H_1 and H_2.*

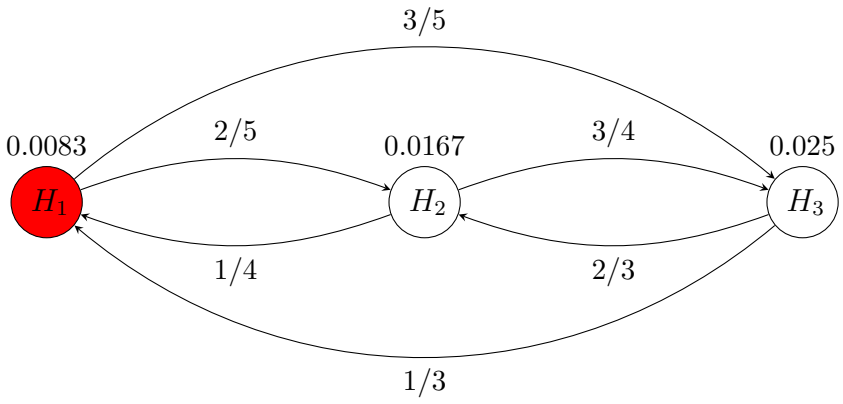

Figure 2.1 The graphical approach for the WAP: set $\alpha = 0.05$, and given $P_1 = 0.01$, $P_2 = 0.014$ and $P_3 = 0.3$; $w_i = i$, $i = 1, 2, 3$. Initial allocation $\boldsymbol{\alpha} = \{\alpha/6, \alpha/3, \alpha/2\}$. Yellow color means rejection; red color means acceptance.

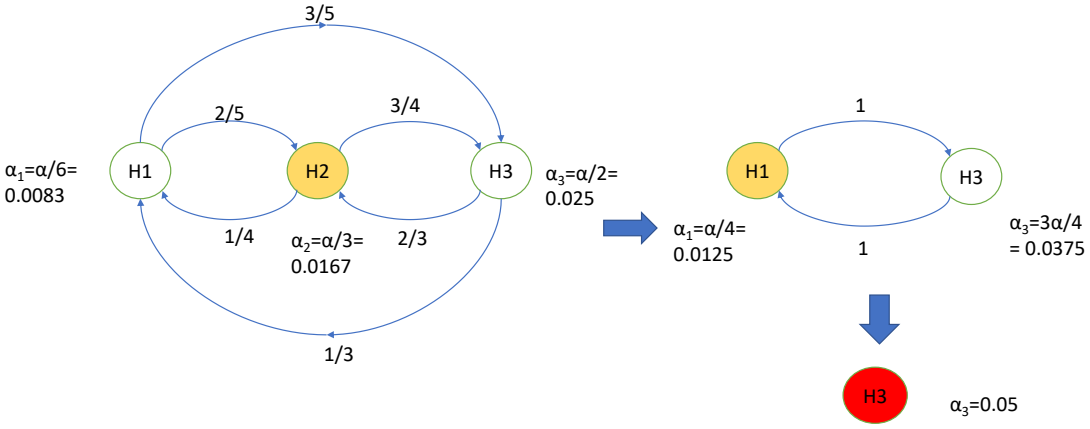

Figure 2.2 The graphical approach for the WHP: set $\alpha = 0.05$, and given $P_1 = 0.01$, $P_2 = 0.014$ and $P_3 = 0.3$; $w_i = i$, $i = 1, 2, 3$. Then, $\tilde{P}_1 = 0.01$, $\tilde{P}_2 = 0.007$ and $\tilde{P}_3 = 0.1$. Yellow color means rejection; red color means acceptance.

2.5 Adjusted p-values

In this section, we find adjusted p-values for both the WAP and the WHP which can be compared with global significance level α directly then give the same results as applying the multiple test procedures to the p-values.

For the WHP, the adjusted p-values $\tilde{P}_{(i)}^{adj}$ for corresponding $H_{(i)}^{*}$ can be calculated as follows.

$$\tilde{P}_{(i)}^{adj} = \begin{cases} \min\left\{\tilde{P}_{(1)} \sum_{k=1}^{m} w_{(k)}^{*}\,,\, 1\right\}, & \text{i=1} \\ \max\left\{\tilde{P}_{(i)} \sum_{k=i}^{m} w_{(k)}^{*}\,,\, \tilde{P}_{(i-1)}^{adj}\right\}, & \text{i=2,...,m.} \end{cases}$$

For the WAP, the adjusted p-values $P_{(i)}^{adj}$ for corresponding $H_{(i)}$ can be calculated as follows.

$$P_{(i)}^{adj} = \begin{cases} \min\left\{ \frac{P_{(1)}}{w_{(1)}} \sum_{k=1}^{m} w_{(k)} \, , 1 \right\}, & \text{i=1} \\ \max\left\{ \frac{P_{(i)}}{w_{(i)}} \sum_{k=i}^{m} w_{(k)} \, , P_{(i-1)}^{adj} \right\}, & \text{i=2,...,m.} \end{cases}$$

Proposition 2.5.1 *For each hypothesis H_i, $i = 1, ..., m$, the adjusted p-value of the WHP is always less than or equal to its adjusted p-value of the WAP. Thus, the hypothesis H_i, which is rejected by WAP, can always be rejected by WHP.*

We can see that the result of Proposition 2.5.1 is consistent with Theorem 1.

Example 2.5.1 *Here, we use same setting of example 2.3.2 to find both adjusted p-values and adjusted weighted p-values. Based on formulas $\tilde{P}_{(i)}^{adj}$ and $P_{(i)}^{adj}$, we find $P_1^{adj} = P_2^{adj} = 0.06$, $P_3^{adj} = 0.3$ and $\tilde{P}_1^{adj} = \tilde{P}_2^{adj} = 0.042$, $\tilde{P}_3^{adj} = 0.3$; then, compared with $\alpha = 0.05$, respectively, WHP rejects H_1 and H_2 while WAP rejects none, which is consistent with the result of example 2.3.2.*

2.6 The Optimality of the WHP and WAP

It is important to study the optimality property that the procedures cannot be improved without losing the control of error rates. To the best of our knowledge, there is no research studying the optimality of the WHP and WAP. Thus, in this paper, we will discuss the optimality property of both procedures.

2.6.1 The optimality property of the WHP

By finding a specific joint distribution for the p-values $\{P_1, ..., P_m\}$ under arbitrary dependence, we show that the WHP is an optimal procedure if the equality of $FWER \leq \alpha$ can be attained.

Theorem 2 *The WHP is an optimal procedure in the sense that the procedure cannot be improved by increasing even one of its critical values without losing control over the FWER.*

The optimality property of Theorem 2 is in a general sense which can be applied to either monotone procedures such as the WHP or non-monotone procedures such as the WAP. The following Proposition 2.6.1 is another result of the optimality of the WHP by narrowing the family of the weighted MTPs to the weighted monotone step-down MTPs. Some definitions and notations are introduced as follows.

Definition 2.6.1 *(Weighted monotone step-down procedure)*
Reject $H^*_{(i)}$ *when*

$$\tilde{P}_{(j)} \leq \alpha_j, \quad j = 1, ..., i, \tag{2.7}$$

where $\tilde{P}_{(1)} \leq ... \leq \tilde{P}_{(m)}$ *is the ordered version of weighted p-values* $\tilde{P}_i = P_i/w_i$*, associated with hypotheses* $H^*_{(1)}, ..., H^*_{(m)}$*. And* $\alpha_1 \leq \alpha_2 \leq ... \leq \alpha_m$ *be a sequence of increasing critical values.*

Proposition 2.6.1 *Let* $\mathcal{M}^w$ *be a weighted monotone step-down multiple testing procedure with*

$$FWER \leq \alpha < 1.$$

Then, for any hypothesis, if it is rejected by the procedure $\mathcal{M}^w$*, it will be rejected by the WHP.*

We can see the Proposition 2.6.1 is only restricted to the weighted monotone step-down procedure of Definition 2.6.1.

2.6.2 The optimality property of the WAP

Next, we will study the optimality property of the WAP.

Proposition 2.6.2 *The weighted alternative Holm procedure (WAP) (Benjamini and Hochberg, 1997) is an optimal procedure when the proportions of smallest weight to other weights are not less than the global significance level* α*, that is,* $\frac{w_j}{w_i} \geq \alpha$*, where* $w_j = \min\{w_1, ..., w_m\}$*,* $1 \leq i \leq m$ *and* $i \neq j$*.*

Remark 3 *Note: The constraint for Proposition 2.6.2 can be satisfied in many applications, eg, let* $\alpha = 0.05$*, then we have* $w_j \leq w_i \leq 20w_j$*, where* $w_j = \min\{w_1, ..., w_m\}$.

Remark 4 *The question is that if* $\frac{w_j}{w_i} < \alpha$*,* $w_j = \min\{w_1, ..., w_m\}$*, can we find a joint distribution of p-values* P_i*,* $i = 1, ..., m$ *under some arbitrary dependence, such that* $FWER = \alpha$*. Suppose* $m = 2$*, both* H_1 *and* H_2 *are true null hypotheses, and let* w_1 *be close to 1,* $w_2 > 0$ *but is close to 0, eg,* $w_1 = 0.999$*,* $w_2 = 0.001$*, and let* $f(P_1, P_2)$ *denote the joint pdf of* P_1 *and* P_2*, therefore we have*

$$\begin{aligned} FWER_{WAP} &= Pr\left(P_1 \leq 0.999\alpha, P_1 \leq P_2\right) + Pr\left(P_2 \leq 0.001\alpha, P_1 \geq P_2\right) \\ &= \int_0^{0.999\alpha} \int_{P_1}^{1} f(P_1, P_2) dP_2 dP_1 + \int_0^{0.001\alpha} \int_{P_2}^{1} f(P_1, P_2) dP_1 dP_2 \\ &\leq \int_0^{0.999\alpha} \int_{0}^{1} f(P_1, P_2) dP_2 dP_1 + \int_0^{0.001\alpha} \int_{0}^{1} f(P_1, P_2) dP_1 dP_2 \\ &= \alpha \end{aligned} \tag{2.8}$$

We can see, for the inequality of Equation (2.8), the equality holds only if (1) $\int_{P_1}^{1} f(P_1, P_2) dP_2 = \int_0^1 f(P_1, P_2) dP_2$ *which means* $\int_0^{P_1} f(P_1, P_2) dP_2 = 0 \iff P_1 \leq P_2$*; and (2) similarly,* $\int_{P_2}^{1} f(P_1, P_2) dP_1 = \int_0^1 f(P_1, P_2) dP_1$ *which means* $\int_0^{P_2} f(P_1, P_2) dP_1 = 0 \iff P_2 \leq P_1$*. Thus, for continuous data, there is no such* $f(P_1, P_2)$ *existing to make* $FWER_{WAP} = \alpha$*.*

So, according to Proposition 2.6.2 and Remark 4, the weights are α dependent, since only when $\frac{w_j}{w_i} \geq \alpha$, $w_j = \min\{w_1, ..., w_m\}$, $1 \leq i \leq m$ and $i \neq j$, the WAP is an optimal procedure.

2.7 Simulation Studies

In this section, simulation studies were conducted to compare the performances of the WHP and WAP in terms of average power and FWER control under dependent structure. The considered dependent structure is equal correlation dependence where off-diagonal components of covariance matrix are equal to ρ and the diagonal components are equal to 1.

In each simulation, n independent m dimensional random normal vectors with covariance matrix Σ and components $Z_i \sim N(\mu_i, 1)$, $i = 1, ..., m$ were generated. The p-value for testing $H_i : \mu_i = 0$ vs. $H_i^{'} : \mu_i > 0$ was calculated using a one-sided, one-sample t-test for each i.

Four different weight settings were used in this research: (1) The weights w_i's were generated from uniform distribution $U(1, 2)$ for true nulls and $U(2, 10)$ for false nulls to see if properly choosing the weights and the distance is large, how the power is affected. (2) The weights w_i's were generated from uniform distribution $U(1, 2)$ for true nulls and $U(2, 6)$ for false nulls to see if the distance gets smaller, how the power is affected. (3) The weights w_i's were generated from uniform distribution $U(1, 2)$ for true nulls and $U(6, 10)$ for false nulls to see if distance is same as (2) but weights are larger, how the power is affected. (4) The weights were generated from $U(1, 6)$ for both true and false nulls in order to see if the weights were poorly selected, how the performance of WHP, WAP and Holm procedure, respectively.

We set $\alpha = 0.05$, the number of hypotheses $m = \{5, 10, 15\}$, the true null proportion $\pi_0 = \{0.2, 0.4, 0.6, 0.8\}$, the correlation coefficients $\rho = \{0, 0.1 ..., 0.9\}$ and the whole sample size $n = 15$. Among m hypotheses, $m\pi_0$ of m μ_i's are equal to 0, and the remaining μ_i are equal to 0.7. Our simulation runs 5000 times.

Figures 2.3 to 2.7 have showed the simulated FWER levels and average powers of three procedures: the WHP, the WAP and Holm procedures under one-sided one-sample t test setting. From these simulation results we can observe:

All the three procedures always control the FWER at the pre-specified level α. But the WHP always has higher FWER level and greater power than the WAP no matter how to choose weights. If poorly choosing weights, see Figure 2.7, Holm procedure performs better than the WHP and WAP, and the average power decreases as the correlation increases.

If choosing weights properly, see Figures 2.4, 2.5 and 2.6, both the WHP and the WAP perform better than Holm in terms of power. The average powers for both WHP and

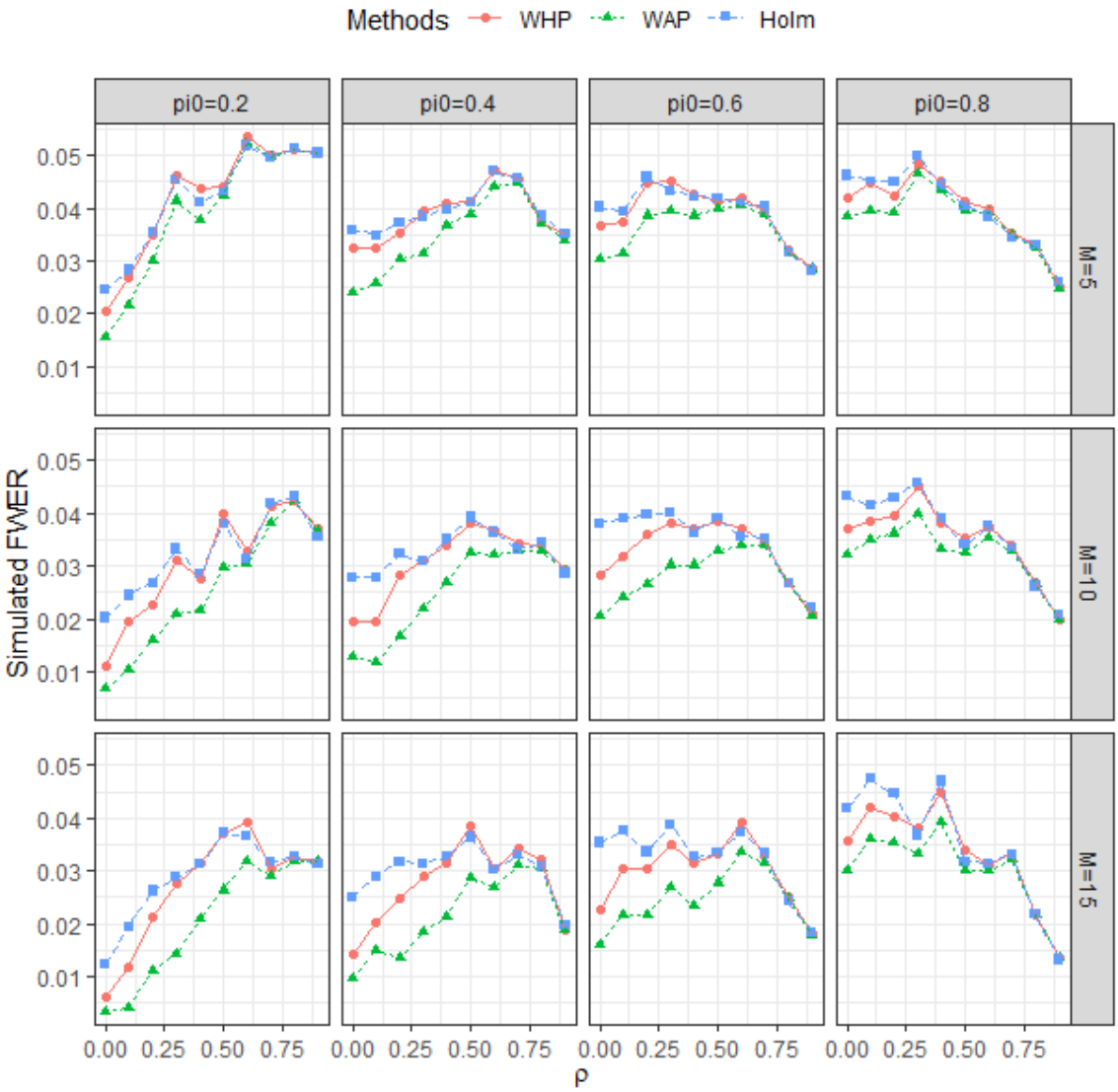

Figure 2.3 Simulated FWER comparisons for weighted Holm procedures based on arbitrary dependent one-sided t-test, including WHP and WAP, and the weights were generated from $U(1,\ 2)$ for true nulls and $U(2,\ 10)$ for false nulls.

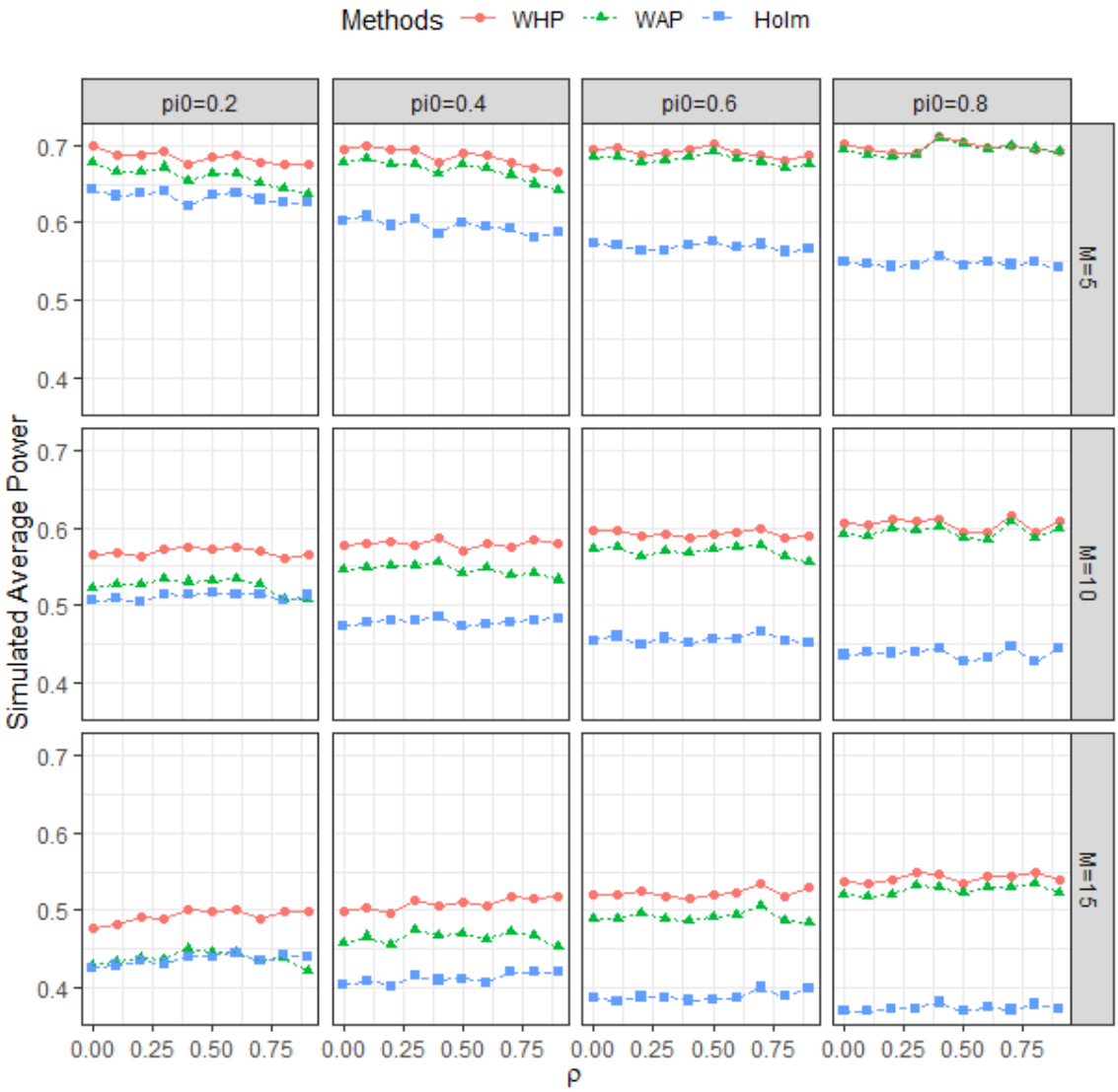

Figure 2.4 Simulated average power comparisons for weighted Holm procedures based on arbitrary dependent one-sided t-test, including WHP and WAP, and the weights were generated from $U(1,\ 2)$ for true nulls and $U(2,\ 10)$ for false nulls.

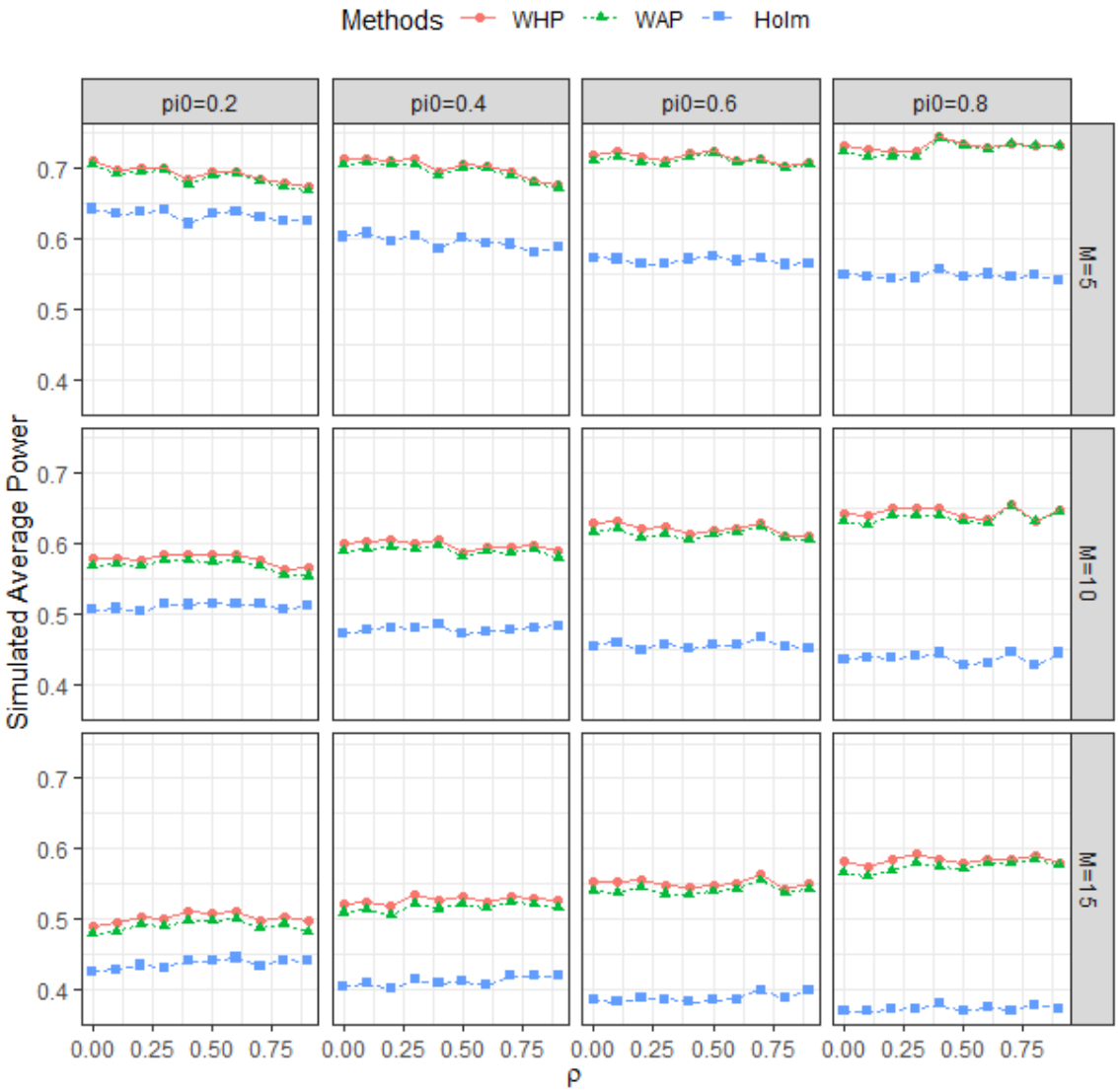

Figure 2.5 Simulated average power comparisons for weighted Holm procedures based on arbitrary dependent one-sided t-test, including WHP and WAP, and the weights were generated from $U(1,\ 2)$ for true nulls and $U(6,\ 10)$ for false nulls.

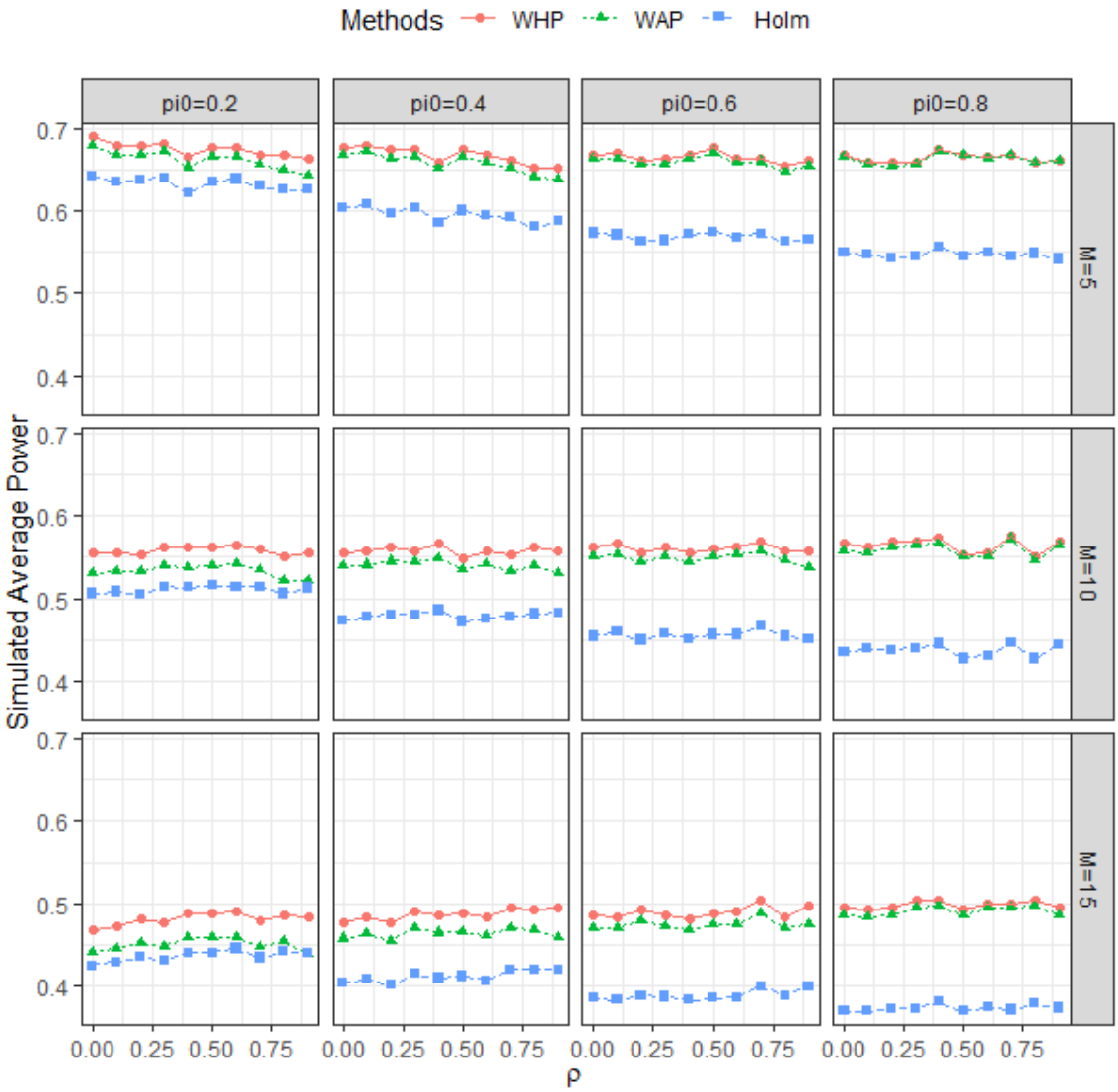

Figure 2.6 Simulated average power comparisons for weighted Holm procedures based on arbitrary dependent one-sided t-test, including WHP and WAP, and the weights w_i's were generated from uniform distribution $U(1,\ 2)$ for true nulls and $U(2,\ 6)$ for false nulls.

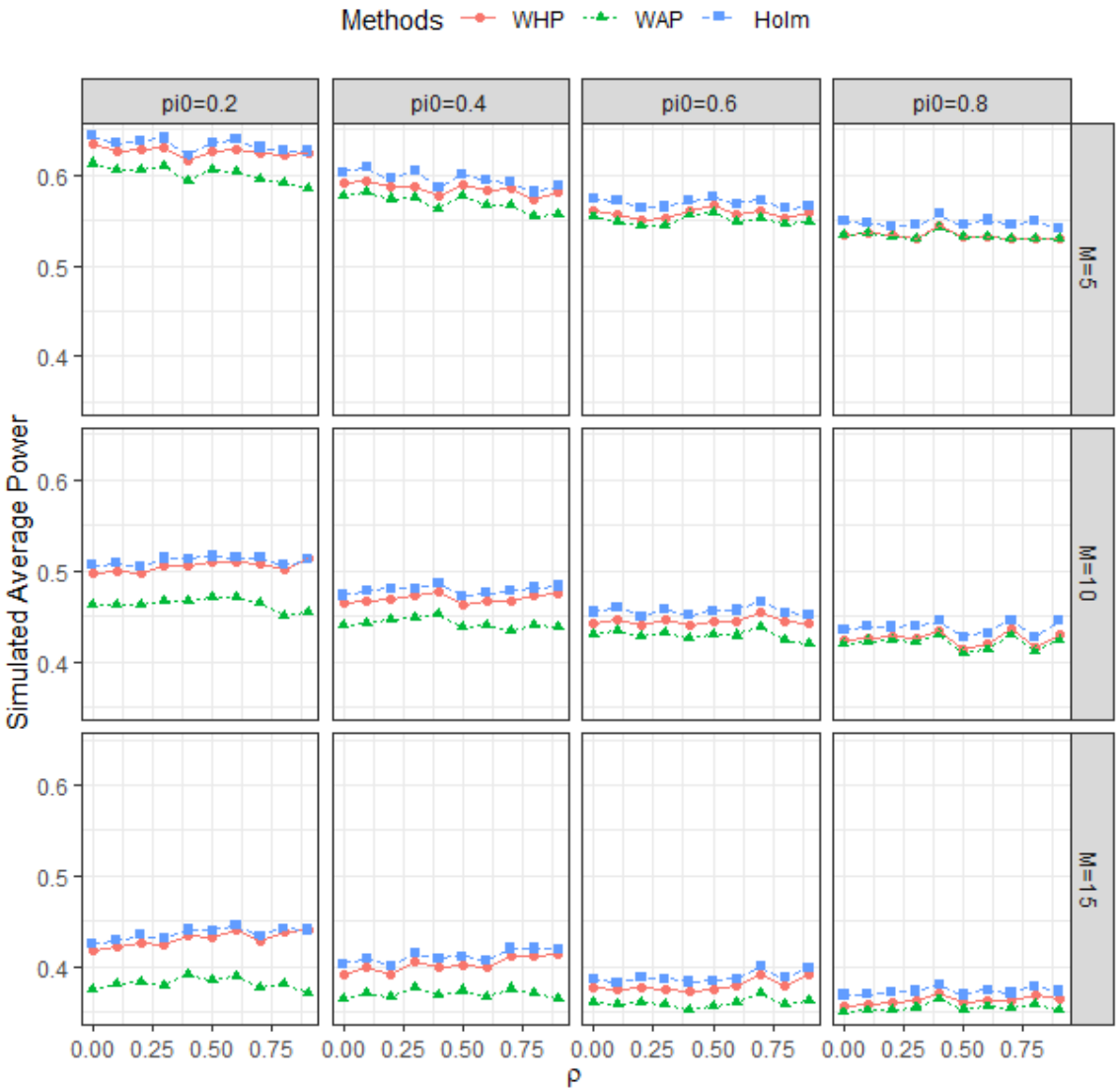

Figure 2.7 Simulated average power comparisons for weighted Holm procedures based on arbitrary dependent one-sided t-test, including WHP and WAP, and the weights w_i's were generated from uniform distribution $U(1,\ 6)$ for both true and false nulls.

WAP procedures are decreasing as the correlation ρ increases when hypotheses number m is small, eg, $m = 5$. More importantly, the WHP is always powerful than WAP no matter how the correlation ρ changes. And the power advantage for the WHP is larger for larger proportion of false null hypotheses and larger distance between weights.

From Figures 2.4, 2.5 and 2.6, we can find that the power will be improved for both procedures if the weights are chosen properly, especially for smaller proportion of false null hypotheses. The finding is interesting, because in many applications, the proportion of false nulls is not large and the power usually increases as the proportion of false null hypotheses increases, for example, Holm procedure.

2.8 Real Data Analysis: Clinical Examples

Example 2.8.1 *Consider a clinical trial in patients with acute respiratory distress syndrome (ARDS), refer to ARDS Network (2000) and Dmitrienko, Offen and Westfall (2003). The trial is conducted to compare one dose of a new drug to placebo. The therapeutic benefits of experimental treatments in ARDS trials are commonly measured using the number of days alive and off mechanical ventilation during a 28-day study period and 28-day all-cause mortality rate. There are two primary endpoints, denoted by H_1 and H_2, the null hypotheses of no treatment effect with respect to the number of ventilator-free days and 28-day all-cause mortality, respectively. Denote the secondary hypotheses H_3 and H_4 associated with the drug effects on the number of days the patients were out of the intensive care unit (ICU-free days) and general quality of life in the product label, respectively.*

Suppose the weights for the primary hypotheses are given by $w_1 = 2$ and $w_2 = 1.5$, and the secondary hypotheses are equally weighted, that is, $w_3 = w_4 = 0.25$. The observed raw p-values are $p_1 = 0.024$, $p_2 = 0.003$, $p_3 = 0.026$ and $p_4 = 0.002$. Then the adjusted p-values for both procedures are presented in the following table 2.1. The table shows that all of 4 hypotheses will be rejected by both procedures; however, the adjusted p-values,

given by the WHP, are always smaller than its corresponding adjusted p-values, given by the WAP.

Table 2.1 Comparison Between the WHP and the WAP Using Adjusted p-values for ARDS Data

Hypothesis	Weights	Raw p-values	$\tilde{P}_{(i)}^{adj}$	$P_{(i)}^{adj}$
H_1	2	0.024	0.027	0.032
H_2	1.5	0.003	0.008	0.032
H_3	0.25	0.026	0.027	0.032
H_4	0.25	0.002	0.020	0.032

Example 2.8.2 *A clinical trial was conducted to compare a new formulation of an insulin therapy (Formulation A) to a standard formulation (Formulation B) in patients with Type 2 diabetes. Patients were allocated to three treatment groups (A, B and A + B) and the efficacy analysis was based on the mean change in hemoglobin A1c from baseline to a 6-month endpoint, see Dmitrienko et al. (2007). To use the raw p-values from Dmitrienko et al. (2007) Table* IV*, then let $w_i = \{6, 6, 5, 4, 2, 1\}$. The results are shown in the following table 2.2.*

Table 2.2 Comparison Between the WHP and the WAP Using Adjusted p-values for the Formulation Clinical Trial

Hypothesis	Weights	Raw p-values	Adjusted weighted p-value	Adjusted p-value
H_1	6	0.011	0.0348	0.0348
H_2	6	0.023	0.0498	0.0585
H_3	5	0.006	0.0288	0.0288
H_4	4	0.018	0.0498	0.0585
H_5	2	0.042	0.063	0.063
H_6	1	0.088	0.088	0.088

From above table 2.2, we can see the WHP rejects $\{H_1,\ H_2,\ H_3,\ H_4\}$ *while the WAP rejects* $\{H_1,\ H_2\}$.

2.9 Conclusion

Although several weighted procedures were developed and commonly used to control the FWER, such as the weighted Bonferroni procedure, the weighted Holm procedure and the weighted Hochberg procedure, there is no clear conclusion that which procedures will be preferred and easy to be applied in real applications, especially in clinical trial studies. In this paper, we have investigated the similarities and differences between the WHP and the WAP from three aspects: exploring their corresponding underlying closed testing procedures, visualizing both procedures by proper graphical approaches and finding the adjusted p-values and adjusted weighted p-values. And we have provided a theoretical result that the WHP is more powerful than the WAP under arbitrary dependence. Then, we have studied the optimality property of both procedures and have showed the WHP is an optimal procedure and can dominate all the other monotone step-down weighted procedures controlling the FWER and the WAP is an optimal procedure when the proportions of smallest weight to other weights are not less than the global significance level α. We also give some numerical results of the power and FWER performance through our simulation studies.

Moreover, the WHP is monotone in the sense that the WHP always rejects a hypothesis with the smaller p-value whenever it rejects another hypothesis with a larger p-value and this procedure also satisfies the monotonicity condition from Definition 2.3.2. While we find that the WAP has neither the monotone property in terms of p-values nor satisfied the monotonicity condition in the sense of the weighted functions being monotone in terms of the index subsets. Bretz et al. (2009) pointed out that monotonicity condition leads to consonant closed tests and shortcut procedures. Even though the WAP does not satisfy the monotonicity condition, it is a consonant procedure which usually leads to a

shortcut of the procedure. Thus, an interesting observation is that the WAP does not belong to the class of CTPs with weighted Bonferroni-type local tests and it's also not monotone in terms of the p-values. In the future, we can develop a new and more general weighted graphical approach, which is a natural extension of the usual graphical approach and the WHP and WAP both can be expressed as special cases of the new graphical approach.

CHAPTER 3

ON GRAPHICAL APPROACHES

3.1 Introduction

In this chapter, we aim to study the similarities and differences of the original graphical approach in Bretz et al. (2009) and the default graph in Burman et al. (2009), and provide a direct proof of the FWER control for the original graph. As aforementioned, there are existing many FWER controlling procedures to address multiplicity issues when we simultaneously test multiple hypotheses, such as Bonferroni-based sequentially rejective procedures, Simes-based procedures. However, it is too complicated and not intuitive when using long and non-visualized decision tables, especially when the number of hypotheses is large. Bretz et al. (2009) proposed an original graphical approach which can visualize the Bonferroni-type based sequentially rejective MTPs with intuitive graphs and relationships between two hypotheses via directed edges. The original graphical approach is flexible and powerful to develop Bonferroni-type procedures according to various objectives of clinical trial studies. This graphical approach strongly controls the FWER and has been widely used in clinical trials and further extended in many literatures, including Bretz et al. (2011), Maurer and Bretz (2014), Sugitani, Bretz and Maurer (2016), Robertson, Wason and Bretz (2020), Zhan et al. (2022), etc. One limitation of the original graphical approach is that at most one rejection is allowed in each step; thus, the computation can be inefficient as the allocated critical values and transition coefficients need to be updated at each step, especially when test large number of hypotheses.

The default graph is an independently developed graphical approach in Burman et al. (2009), it is also based on weighted Bonferroni test and usually considered to be equivalent to the original graphical approach. That is also one of the reasons, the default graph is not that popular as the previous approach. The default graph consists of a group of fixed sequences of pre-ordered testing hypotheses, with allocated critical values to each

sequence. The testing order for each sequence is prespecified, once the corresponding hypothesis is rejected, the allocated critical value is fully passed to next hypothesis in the current sequence. An example of visualizations of both graphical approaches for a parallel gatekeeping problem is given in Figure 3.1 and 3.2. From Figure 3.2, we can see the default graph consists of four fixed sequences of hypotheses, which can be denoted as $\frac{\alpha}{4}(H_1 \to H_3 \to H_4)$, $\frac{\alpha}{4}(H_1 \to H_4 \to H_3)$, $\frac{\alpha}{4}(H_2 \to H_3 \to H_4)$ and $\frac{\alpha}{4}(H_2 \to H_4 \to H_3)$, respectively. At first step, both H_1 and H_2 can be test at level $\alpha/2$. Suppose H_1 is rejected, then for the first sequence from left side, the allocated critical value $\alpha/4$ is fully passed to H_3, and for the second sequence, $\alpha/4$ is fully passed to H_4. Comparing with the original graphical approach, more than one hypothesis can be rejected at each step, and the transition of allocated critical values is more straightforward if the some hypotheses are rejected. However, there is no clear updating algorithm provided in Burman et al. (2009).

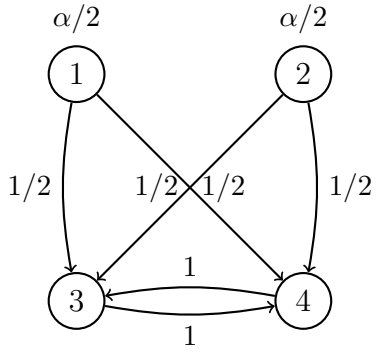

Figure 3.1 The original graph for a parallel gatekeeping problem.

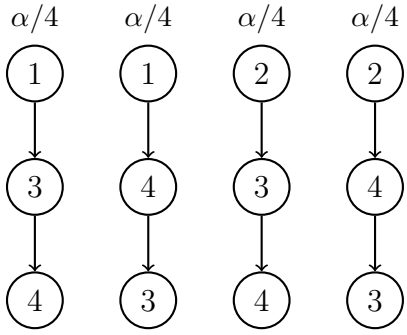

Figure 3.2 The corresponding default graph.

Both of graphical approaches are Bonferroni-based procedures, they are usually considered to be similar, refer to Bretz et al. (2011), Robertson, Wason and Bretz (2020), etc. Actually, based on our investigation, two approaches are different, especially, when the number of hypotheses is larger than 3. Initially, we can use the degree of freedom to explain the reason of difference. Let's take Holm procedure as an example. Consider simultaneously testing m hypotheses, and there are $m!$ sequences. Let df_1 and df_2 denote the degrees of freedom of local critical values for the original and default graphical approaches, respectively. For the default graph, $df_1 = m! - 1$; however, for the original approach, we have $df_2 = (m-1) + m(m-2) = m^2 - m - 1$, since (1) we have m hypotheses and $\sum_{i=1}^{m} \alpha_i = \alpha$, and (2) for each of m hypotheses, we have $m-1$ flexible transition coefficients and the summation of them equals 1, which produces m(m-2) degrees of freedom. Thus, we have $df_1 = df_2$ when $m \leq 3$ (when $m = 2$, $df_1 = df_2 = 1$), and $df_1 > df_2$ when $m \geq 4$. Therefore, there are less degree of freedoms for the original graph, which means we have more restrictions for this method and the default graph is more flexible.

In summary, we have studied the similarity and difference between these two approaches and the conditions that should be satisfied. We also have provided the method to find the corresponding equivalent graphical approach when the other approach is given for testing three hypotheses. Moreover, Bretz et al. (2009) provided an indirect proof of FWER control for the original graphical approach by showing the equivalence between the graphs with updating algorithm and a short-cut of the corresponding closed testing procedures (CTPs). In our study, we will give an elegant proof to show the FWER control directly, which can help to understand the original graph approach thoroughly and provide theoretical reference to develop new graphical approaches. The rest of the chapter is organized as follows: Section 3.2 introduces some basic notations and concepts used in this paper. Two algorithms of the default graphical approach are provided in Section 3.3. In Section 3.4, we have provided theoretical results of the similarity and difference between

two graphical approaches, and an elegant and direct proof of the FWER control for the original graphical approach. Some clinical examples are given in Section 3.5. In Section 3.6, a brief summary and future plan are given.

3.2 Preliminaries

In this section, some general notations and definitions are introduced in this chapter. Moreover, as the original graphical approach is well explained in Chapter 2, we will introduce the default graph in this section.

Consider simultaneously testing m null hypotheses in the framework of the original graphical approach. Let $\mathcal{H}$ and M denote the collection of m hypotheses and its corresponding index set, respectively. Let $\mathcal{T}$ and $\mathcal{F} = \mathcal{H} \setminus \mathcal{T}$ denote the set of all true nulls and the set of all false null hypotheses, respectively. The initial local critical values for m hypotheses H_k: $\boldsymbol{\alpha} = (\alpha_1, \ ..., \ \alpha_m)$. The initial transition coefficients matrix $G = (g_{lk})_{m \times m}$, where g_{lk} is the fraction of the critical value of H_l that is passed to H_k when H_l is rejected.

Given the following regularity conditions on the initial critical values and transition coefficients:

1. $\sum_{l=1}^{m} \alpha_l \leq \alpha \in (0, 1)$
2. $0 \leq g_{lk} \leq 1$, $g_{ll} = 0, l, k = 1, ..., m.$
3. $\sum_{j=1}^{m} g_{lk} \leq 1, \forall l = 1, ..., m.$

Let $\mathcal{H} = (H_{j_1}, \ H_{j_2}, ..., H_{j_m})$, with associated p-values $\mathcal{P} = (P_{j_1}, ..., P_{j_m})$, denote a sequence of ordered testing hypotheses, such that H_{j_k} will be tested at step k by the original graphical approach given that all the previous hypotheses H_{j_1},..., $H_{j_{k-1}}$ are rejected at step $i = 1, ..., k-1$, correspondingly. Let R denote the rejection number, where

$$R = \max \left\{ 1 \leq k \leq m : P_{j_i} \leq \alpha_{j_i}(\mathcal{R}_{i-1}), \forall i \in \{1, ..., k\} \right\},$$

where $\mathcal{R}_{i-1} = \left(H_{j_1}, ..., H_{j_{i-1}}\right)$ denotes the rejection sequence at first $i-1$ steps. Therefore, we have the final rejection sequence $\mathcal{R} = \{H_{j_i} \in \mathcal{H} : P_{j_i} \leq \alpha_{j_i}(\mathcal{R}_{i-1}), i = 1, ..., R\}$. And let I_i denote the index set of non-rejected hypotheses before testing at step i, suppose $\mathcal{R}_0 = \emptyset$ and $I_1 = M$. Let $g_{lk}(\mathcal{R}_{i-1})$ denote transition coefficients of the remaining hypotheses after rejecting all the hypotheses in $\mathcal{R}_{i-1}$, which is a function of $\mathcal{R}_{i-1}$, $l, k \in I_i$. Let m_0 denote the number of true null hypotheses with the index set I_0.

For the default graphical approach, we know that the default graph consists of a group of fixed sequences of pre-ordered hypotheses, each sequence includes at least one hypothesis which belongs to $\mathcal{H}$. Let $S_1, ..., S_n$ denote the sequences of testing hypotheses of the default graph, where n is the total number of sequences. Let $\tilde{\alpha}_k$, $k = 1, ..., n$, be the allocated critical values for each sequence. Then, we have $S_k = \tilde{\alpha}_k(H_{j_1} \to H_{j_2} \to ... \to H_{j_i})$, $j_1, ..., j_i \in \{1, ..., m\}$, $j_l \neq j_t$ if $l \neq t$. The following are definitions of the default graph, complete default graph, complete original graph which are basic settings that we studied with.

Definition 3.2.1 (Burman et al., 2009) *A default graph splits the nominal critical value into several parts (not necessarily of equal size), each allocated to a fixed-sequence of hypotheses.*

Definition 3.2.2 *A complete default graph consists of $m!$ sequences when it is used to test m hypotheses, and each sequence consists of all m hypotheses. For the sequences critical values, $\sum_{k=1}^{m!} \tilde{\alpha}_k = \alpha$, and $0 < \tilde{\alpha}_k < \alpha$.*

Definition 3.2.3 *A complete original graph is with initial critical values $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_m\}$, $0 < \alpha_i < \alpha$ and $\sum_{i \in M} \alpha_i = \alpha$, $M = \{1, ..., m\}$. And the transition coefficient matrix is denoted by $\boldsymbol{G} = (g_{ij})_{m \times m}$, where $0 < g_{ij} < 1$, $i \neq j$, $g_{ii} = 0$ and $\sum_{j \in M} g_{ij} = 1$.*

3.3 Algorithms of the Default Graphical Approach

In this section, first we will provide an updating algorithm for the default graph. Let α_i denote the initial critical value of hypothesis H_i, then we have $\sum_{i=1}^{m} \alpha_i \leq \alpha$ and $\alpha_i =$

$\sum_{k=1}^{n} \tilde{\alpha}_k I(S_k(1) = i)$, $i = 1, ..., m$, where $S_k(1)$ denotes the index of initial hypothesis (tail hypothesis) for the sequence S_k at first stage, suppose $S_k(1) = S_k^1(1)$.

Algorithm 1 *1. Initialize $I = \{1, ..., m\}$, $t = 1$, $n^t = n$.*

2. *Let $\alpha_i = \sum_{k=1}^{n^t} \tilde{\alpha}_k I(S_k^t(1) = i)$ for $i \in I$, where $S_k^t(1)$ denotes the index of initial hypothesis for the sequence S_k at stage t, for simplicity.*

3. *If $P_i \leq \alpha_i$, reject the corresponding hypotheses H_i. Let $\mathcal{A}_t$ and I_t denote the set and corresponding index set of the rejected hypotheses; otherwise stop.*

4. *Update the default graph:*

   (a) $I \to I \setminus I_t$;

   (b) Remove all the hypotheses $H_i \in \mathcal{A}_t$ from the sequences, if all the elements in some sequences are removed, the corresponding critical values won't be recycled; otherwise, the critical value of each remaining sequence is unchanged;

   (c) Combine the same remaining sequences by adding the corresponding sequence critical values.

$$\tilde{\alpha}_k \longrightarrow \tilde{\alpha}_k + \sum_{j \in I} \tilde{\alpha}_j I(S_j = S_k), \quad k \in I, k \neq j.$$

   Only one updated combined $\tilde{\alpha}_k$ for the same sequences is kept. Increase t by 1, and let n^t denote the updated number of sequences, and the updated $k = 1, ..., n^t$.

5. *If $|I| \geq 1$, go to step 2; otherwise stop.*

Remark 5 *We take a fallback procedure example to explain the statements in step 4(ii) and (iii). Suppose we simultaneously test three hypotheses H_1, H_2, H_3 with corresponding weights w_1, w_2, w_3. If the testing order is H_1, H_2 and H_3, correspondingly. Then applying fallback procedure, we have three sequences, $S_1 = w_1\alpha(H_1 \to H_2 \to H_3)$, $S_2 = w_2\alpha(H_2 \to H_3)$ and $S_3 = w_3\alpha(H_3)$. (1) Suppose only H_3 is rejected at stage 1 as $P_3 \leq w_3\alpha$, $P_1 > w_1\alpha$ and $P_2 > w_2\alpha$. Then, sequence S_3 is removed and its assigned critical value won't be recycled. (2) If only H_2 is rejected at stage 1 when $P_1 > w_1\alpha$, $P_2 \leq w_2\alpha$ and $P_3 > w_3\alpha$. Consequently, H_2 is removed from sequences S_1 and S_2; therefore the updated sequences are $S_1 = w_1\alpha(H_1 \to H_3)$, $S_2 = w_2\alpha(H_3)$ and $S_3 = w_3\alpha(H_3)$. After combining same sequences we have updated sequences $S_1 = w_1\alpha(H_1 \to H_3)$, $S_2 = (w_2 + w_3)\alpha(H_3)$.*

The following two examples are used to illustrate Algorithm 1 of the default graph.

Example 3.3.1 *Consider a fallback procedure example. Given three hypotheses H_1, H_2 and H_3, with associated p-values P_1, P_2 and P_3, respectively. Suppose $P_1 = 0.5\alpha$, $P_2 = 0.8\alpha$ and $P_3 > \alpha$. The yellow nodes mean that the hypotheses are rejected, red means that the procedure stops.*

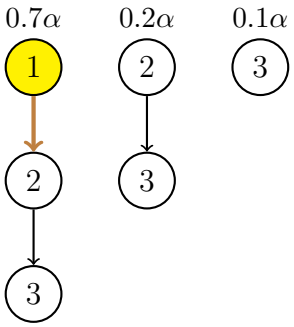

Figure 3.3 The initial default graph.

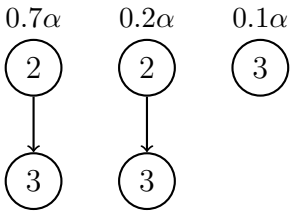

Figure 3.4 The updated default graph after rejecting H_1 at stage 1 as $P_1 < 0.7\alpha$.

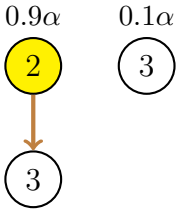

Figure 3.5 Combining the remaining same sequences.

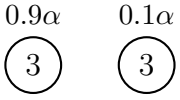

Figure 3.6 The updated default graph after rejecting H_2 at stage 2, as $P_2 < 0.9\alpha$.

From this example, we can see that at first stage, only H_1 is rejected as $P_1 < 0.7\alpha$, $P_2 > 0.2\alpha$ and $P_3 > 0.1\alpha$. After removing the node of H_1 in the graph, the allocated

Figure 3.7 Combining same sequences.

critical value 0.7α *is passed to* H_2 *in the first sequence from left side. Then, we combine first two same sequences* $S_1 = 0.7\alpha(H_2 \to H_3)$ *and* $S_2 = 0.2\alpha(H_2 \to H_3)$ *by adding* 0.7α *and* 0.2α*, which is* $0.9\alpha(H_2 \to H_3)$*. The following steps are updated by the same way.*

Example 3.3.2 *Consider a Holm procedure example. Given three hypotheses* H_1*,* H_2 *and* H_3*, with associated* p*-values* P_1*,* P_2 *and* P_3*, respectively. Suppose* $P_1 = 0.5\alpha$*,* $P_2 = 0.3\alpha$ *and* $P_3 > \alpha$*.*

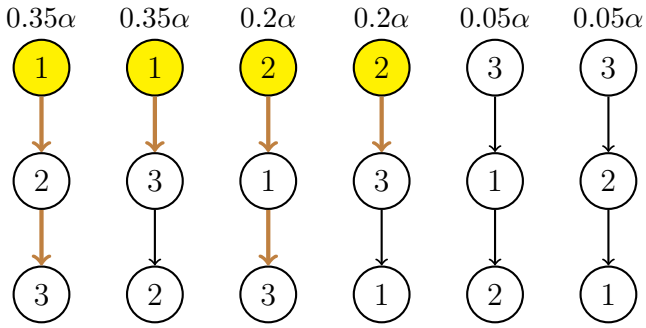

Figure 3.8 The initial default graph.

Figure 3.9 The updated default graph as $P_1 < 0.7\alpha$ and $P_2 < 0.4\alpha$.

From this example, we can see both H_1 *and* H_2 *are rejected at stage 1 which is different with the original graphical approach, and the allocated critical values of these two hypotheses will be fully passed to* H_3*.*

We also provide an algorithm to find the corresponding original graphical approach when a complete default graph is given. From the definition of complete default graph, we know that it consists of $m!$ sequences, each sequence is one of permutations of m hypotheses. Let $S_k^t(l)$ denote the lth hypothesis in the sequence S_k at stage t.

Algorithm 2 *1. Initialize* $I = \{1, ..., m\}$, $t = 1$, $n^t = m!$ *which is the number of sequences for the initial graph.*

2. *First, according to the complete default graph, we find the initial transition coefficient matrix.*

   (a) *For* $i \in I$, *define summation of critical values of the sequences with same first two hypotheses* H_i *and* H_j,

$$\alpha_j^i = \sum_{k=1}^{n^t} \tilde{\alpha}_k I(S_k^t(1) = i,\, S_k^t(2) = j),\, j \in I \setminus \{i\}\,.$$

   (b) *Therefore, we have initial transition coefficients*

$$g_{ij} = \frac{\alpha_j^i}{\sum_{k=1}^{n^t} \tilde{\alpha}_k I(S_k(1) = i)}, \forall i \in I,\, j \in I \setminus \{i\}\,. \tag{3.1}$$

3. *If some hypotheses are rejected at stage* t, *update the graph according to Algorithm 1. Let* I_t *denote the index set of the rejected hypotheses at stage* t, *and* $I \to I \setminus I_t$. *otherwise stop.*

4. *Increase* t *by 1, and let* n^t *denote the updated number of sequences, and the updated* $k = 1, ..., n^t$.

5. *If* $|I| \geq 3$, *go to step 2(a); otherwise stop.*

3.4 Main Theoretical Results

In this section, first, we will provide the theoretical results of similarity and difference between two graphical approaches. Second, we will give elegant theoretical results to prove the FWER control directly for the original graphical approach.

3.4.1 The similarity and difference between two graphical approaches

First, in the following Theorem 3, we show that when simultaneously test three hypotheses, if the complete original graph is given, we can always find a corresponding equivalent complete default graph.

Theorem 3 *When simultaneously test three hypotheses, the original graph according to Bretz et al. (2009) is given, with* $\sum_{i=1}^{3} \alpha_i = \alpha$, $g_{ii} = 0$, $0 < g_{ij} < 1$, *and* $\sum_{j \in M} g_{ij} = 1$, $i \neq j$, $i, j \in M$ *and* $M = \{1, 2, 3\}$. *The default graph is obtained based on finding all*

the possible routes for each sequence, and adding all the allocated critical values to the routes together. Then the performances of the original graphical approach and the default graphical approach are equivalent.

The proof of Theorem 3 is given as follows.

Suppose three hypotheses H_1, H_2 and H_3 are simultaneously tested. The initial critical values for the hypotheses and transition coefficients are given in Figure 3.10. Suppose $\sum_{i=1}^{3} \alpha_i = \alpha$, $g_{ii} = 0$ and $0 < g_{ij} < 1$, $\sum_{j \in M} g_{ij} = 1$, $i \neq j$, $i, j \in M$ and $M = \{1, 2, 3\}$.

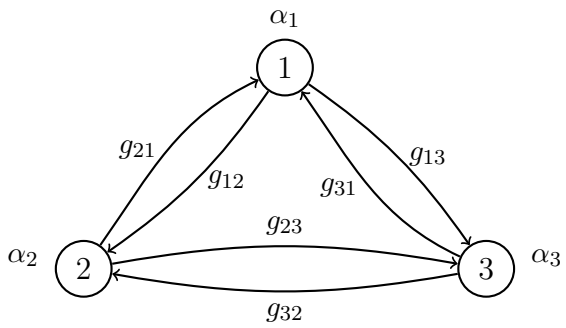

Figure 3.10 The complete original graph with $0 < g_{ij} < 1$, $\forall\, i \neq j$, $i, j \in M$ and $M = \{1, 2, 3\}$. Yellow node means the tail hypothesis in the sequence.

According to the given original graph, we list all the possible direct routes as a default graph in Figure 3.11. The key point is to find the critical values for each sequence $\tilde{\alpha}_k$, $k = 1, ..., 6$.

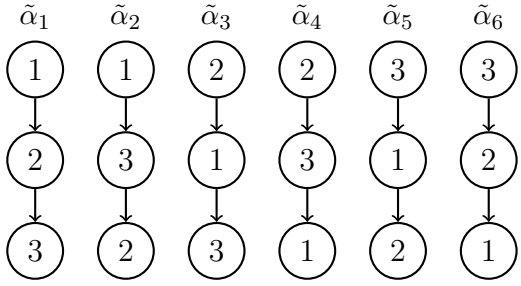

Figure 3.11 The corresponding default graph of the original graph in Figure 3.10.

Since the methods to find the critical values for 6 sequences are similar, we only show the process for the first sequence: $H_1 \rightarrow H_2 \rightarrow H_3$, which consists of two types of routes:

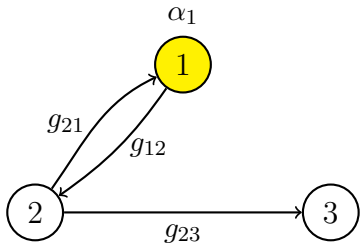

Figure 3.12 First type of routes for the sequence $H_1 \to H_2 \to H_3$ when all $0 < g_{ij} < 1$, $i \neq j$.

(1) First type of routes for the sequence $H_1 \to H_2 \to H_3$ is as follows.

We can express this type of routes as $(H_1 \to H_2)^t \to H_3$, where $(H_1 \to H_2)^t$ means the loop between H_1 and H_2 repeats $t-1$ times and the allocated critical values are $\alpha_1 (g_{12}g_{21})^{t-1} g_{12}g_{23}$, $t \in \mathbb{N}$, correspondingly. For example, when $t = 1$ means there is no loop between H_1 and H_2, then the route is the simplest one: $H_1 \to H_2 \to H_3$. When $t = 2$, we have one loop: $H_1 \to H_2 \to H_1 \to H_2 \to H_3$.

Thus, for this type of routes, we have assigned the following critical value in total

$$\sum_{t=1}^{\infty} \alpha_1 (g_{12}g_{21})^{t-1} g_{12}g_{23} = \alpha_1 g_{12} \frac{g_{23}}{1 - g_{12}g_{21}}, \tag{3.2}$$

this is due to the fact that $\sum_{t=1}^{\infty} x^{t-1} = \frac{1}{1-x}$, when $0 < x < 1$ and $0 < g_{ij} < 1$, $\forall\, i, j \in M$, $i \neq j$ and $M = \{1, 2, 3\}$.

(2) The graph of second type of routes is attached as below.

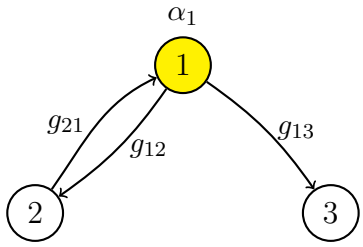

Figure 3.13 Second type of routes for the sequence $H_1 \to H_2 \to H_3$ when all $0 < g_{ij} < 1$, $i \neq j$.

This type of routes can be expressed as $(H_1 \to H_2)^t \to H_1 \to H_3$ and the allocated critical values to these routes are $\alpha_1 (g_{12}g_{21})^t g_{13}$, $t \in \mathbb{N}$, correspondingly.

Thus, for the second type of routes, we have assigned the following critical value in total

$$\sum_{t=1}^{\infty} \alpha_1 \left(g_{12}g_{21}\right)^t g_{13} = \alpha_1 g_{12}g_{21}g_{13}(1 + \sum_{t=1}^{\infty} \left(g_{12}g_{21}\right)^t) = \alpha_1 g_{12}\frac{g_{21}g_{13}}{1 - g_{12}g_{21}}, \tag{3.3}$$

this is also because $1 + \sum_{t=1}^{\infty} x^t = \frac{1}{1-x}$, when $0 < x < 1$.

Therefore, for the given sequence $H_1 \to H_2 \to H_3$, the critical value can be obtained by adding Equation (3.2) and (3.3) together, which is,

$$\alpha_1 g_{12}\frac{g_{23}}{1 - g_{12}g_{21}} + \alpha_1 g_{12}\frac{g_{21}g_{13}}{1 - g_{12}g_{21}} = \alpha_1 g_{12}\left(\frac{g_{23} + g_{21}g_{13}}{1 - g_{12}g_{21}}\right) = \alpha_1 g_{12}g_{23}(H_1) = \alpha_1 g_{12},$$

where $g_{23}(H_1)$ denotes the updated transition coefficient from H_2 to H_3 after removing hypothesis $\{H_1\}$ at first stage, and $g_{23}(H_1) = \frac{g_{23}+g_{21}g_{13}}{1-g_{12}g_{21}} = \frac{1-g_{21}+g_{21}(1-g_{12})}{1-g_{12}g_{21}} = 1$.

Thus, by similar method, finally we have $\tilde{\alpha}_1 = \alpha_1 g_{12}$, $\tilde{\alpha}_2 = \alpha_1 g_{13}$, $\tilde{\alpha}_3 = \alpha_2 g_{21}$, $\tilde{\alpha}_4 = \alpha_2 g_{23}$, $\tilde{\alpha}_5 = \alpha_5 g_{31}$, and $\tilde{\alpha}_6 = \alpha_6 g_{32}$. Consequently, we have $\tilde{\alpha}_1 + \tilde{\alpha}_2 = \alpha_1$, $\tilde{\alpha}_3 + \tilde{\alpha}_4 = \alpha_2$ and $\tilde{\alpha}_5 + \tilde{\alpha}_6 = \alpha_3$. Therefore, at first stage, both approaches can test hypotheses at exactly same critical values.

Suppose H_i, $i \in M$, is rejected at first stage, (1) for the original graphical approach, we have $\alpha_l(H_i) = \alpha_l + \alpha_i g_{il}$, $l \in M \setminus \{i\}$. And $g_{lk}(H_i) = 1$, $l, k \in M \setminus \{i\}$, $l \neq k$. (2) For the default graph, we have $\alpha_l(H_i) = \alpha_l + \alpha_i g_{il} g_{lk}(H_i) = \alpha_l + \alpha_i g_{il}$, $l \in M \setminus \{i\}$. Thus, at stage 2, two approaches can also test the remaining hypotheses at same critical values. And it is easy to show after removing second hypothesis at stage 2, both approaches can test last hypothesis at $\sum_{i=1}^{3} \alpha_i$ level.

Thus, for testing 3 hypotheses, given an original graph, we can find the corresponding default graph and prove it they are equivalent when $0 < g_{ij} < 1$, $i \neq j$, $i, j \in M$. □

Second, in Theorem 4, we show that when simultaneously test three hypotheses, if the complete default graph is given, we can also find a corresponding equivalent original default graph.

Theorem 4 *When simultaneously test three hypotheses, the default graph according to Burman et al. (2009) is given, with* $\sum_{k=1}^{6} \tilde{\alpha}_k = \alpha$, *and* $0 < \tilde{\alpha}_k < \alpha$. *The corresponding original graphical approach according to Algorithm 2 is equivalent to the default graphical approach.*

The proof of Theorem 4 is given as follows.

Suppose we simultaneously test 3 hypotheses H_1, H_2 and H_3, with associated p-values P_1, P_2 and P_3. The default graph is given in Figure 3.14, then we find the corresponding initial original graph according to Algorithm 2 (refer to Figure 3.15), and prove that these two approaches are equivalent when they are performed based on their own algorithms. Suppose $\sum_{k=1}^{6} \tilde{\alpha}_k = \alpha$, and $0 < \tilde{\alpha}_k < \alpha$.

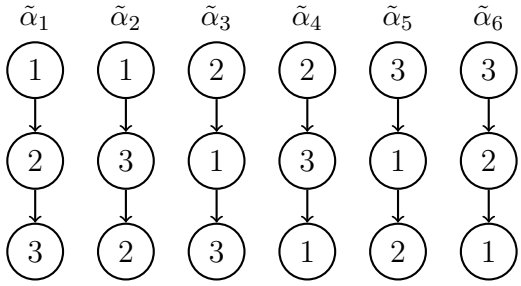

Figure 3.14 The complete default graph for testing 3 hypotheses with $\sum_{k=1}^{6} \tilde{\alpha}_k = \alpha$, and $0 < \tilde{\alpha}_k < \alpha$.

According to the above graph, we can find $\alpha_1 = \sum_{k=1}^{6} \tilde{\alpha}_k I(S_k(1) = 1) = \tilde{\alpha}_1 + \tilde{\alpha}_2$. Similarly, $\alpha_2 = \tilde{\alpha}_3 + \tilde{\alpha}_4$, and $\alpha_3 = \tilde{\alpha}_5 + \tilde{\alpha}_6$.

The initial transition coefficients can be found by $g_{ij} = \frac{\sum_{k=1}^{n} \tilde{\alpha}_k I(S_k(1)=i, S_k(2)=j)}{\alpha_i}$, $j \in M \setminus \{i\}$, where $S_k(1) = i$, $S_k(2) = j$ denote the first and second hypothesis indices of sequence S_k. For instance, $g_{12} = \tilde{\alpha}_1/\alpha_1$.

Without loss of generality, we suppose H_1 is rejected first by both approaches, since at stage 1, both methods can test three hypotheses at same levels, correspondingly. The updated graphs are attached as below for both approaches.

(1) For the default graphical approach, the updated default graph according to Algorithm 1 is in Figure 3.16.

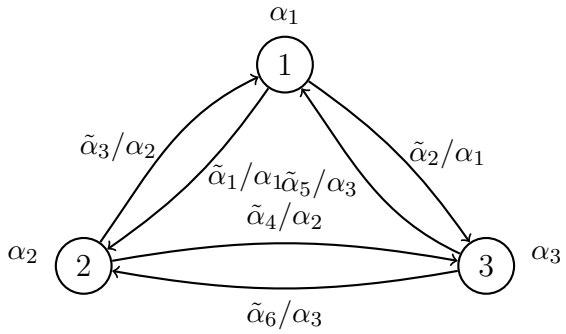

Figure 3.15 The corresponding initial original graph of the default graph in Figure 3.14 according to Algorithm 2.

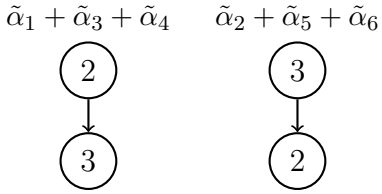

Figure 3.16 The updated default graph of Figure 3.14 after rejecting H_1 according to Algorithm 1.

Based on Algorithm 2, the updated original graph (refer to Figure 3.17) is corresponding to Figure 3.16.

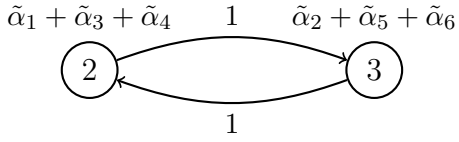

Figure 3.17 The updated original graph of Figure 3.16 according to Algorithm 2.

(2) According to the updating algorithm in Bretz et al. (2009), $\alpha_2(H_1) = \alpha_2 + \alpha_1 g_{12} = \tilde{\alpha}_1 + \tilde{\alpha}_3 + \tilde{\alpha}_4$ and $\alpha_3(H_1) = \alpha_3 + \alpha_1 g_{13} = \tilde{\alpha}_2 + \tilde{\alpha}_5 + \tilde{\alpha}_6$. And $g_{23}(H_1) = g_{32}(H_1) = 1$.

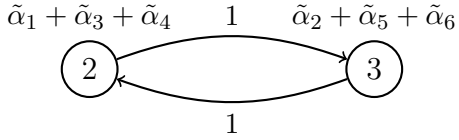

Figure 3.18 The updated original graph of Figure 3.15 according to the algorithm in Bretz et al. (2009).

Therefore the updated original graphs based on two different algorithms are same. Suppose H_2 is removed at stage 2. It is easy to show both approaches can test H_3 at level

$\sum_{k=1}^{6} \tilde{\alpha}_k$. Thus, when only three hypotheses are simultaneously tested, two approaches are equivalent when all the $0 < \tilde{\alpha}_k < \alpha$. □

We also provide an algebraic method to find the corresponding equivalent graphical approach when the other approach is given to test three hypotheses.

Suppose we simultaneously test three hypotheses H_1, H_2 and H_3. Then we applied the default graphical approach and the original graphical approach to test these hypotheses, respectively. Only complete graphs are considered. See Figure 3.10 and 3.11.

By solving the following equations, we want to show if two approaches are equivalent when there are only three hypotheses. Equations (3.4) to (3.6) mean the critical values are equivalent correspondingly at first stage for both the original graphical approach and the default graphical approach; Equations (3.7) to (3.12) mean after removing one hypothesis, the updated critical values are equivalent for two approaches; Equation (3.13) means after removing two hypotheses, the critical values for the remaining hypothesis are equivalent. As both approaches are sequentially rejective procedures, if a unique group of solutions exists, then, two approaches are equivalent.

$$
\begin{cases}
\tilde{\alpha}_1 + \tilde{\alpha}_2 = \alpha_1 & (3.4)\\
\tilde{\alpha}_3 + \tilde{\alpha}_4 = \alpha_2 & (3.5)\\
\tilde{\alpha}_5 + \tilde{\alpha}_6 = \alpha_3 & (3.6)\\
\tilde{\alpha}_1 + \tilde{\alpha}_3 + \tilde{\alpha}_4 = \alpha_2 + \alpha_1 g_{12} & (3.7)\\
\tilde{\alpha}_2 + \tilde{\alpha}_5 + \tilde{\alpha}_6 = \alpha_3 + \alpha_1 g_{13} & (3.8)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_3 = \alpha_1 + \alpha_2 g_{21} & (3.9)\\
\tilde{\alpha}_4 + \tilde{\alpha}_5 + \tilde{\alpha}_6 = \alpha_3 + \alpha_2 g_{23} & (3.10)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_5 = \alpha_1 + \alpha_3 g_{31} & (3.11)\\
\tilde{\alpha}_3 + \tilde{\alpha}_4 + \tilde{\alpha}_6 = \alpha_2 + \alpha_3 g_{32} & (3.12)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_3 + \tilde{\alpha}_4 + \tilde{\alpha}_5 + \tilde{\alpha}_6 = \alpha_1 + \alpha_2 + \alpha_3 & (3.13)
\end{cases}
$$

Case 1: When we test three hypotheses simultaneously, suppose the original graph is given, we find the critical values for sequences of the default graph based on the above equations. On the right side of equations, the critical values are corresponding to the initial critical values and updated critical values which can be obtained from initial critical values and transition coefficients.

From Equations (3.5) and (3.7), we can find $\tilde{\alpha}_1 = \alpha_1 g_{12}$; from Equations (3.4) and (3.9), we can find $\tilde{\alpha}_3 = \alpha_2 g_{21}$. Similarly, we can find $\tilde{\alpha}_2 = \alpha_1 g_{13}$, $\tilde{\alpha}_4 = \alpha_2 g_{23}$, $\tilde{\alpha}_5 = \alpha_3 g_{31}$ and $\tilde{\alpha}_6 = \alpha_3 g_{32}$. Since Equation (3.13) can be obtained from (3.4) to (3.6), then for the exhaustive procedure, the solutions don't depend on last stage. Therefore, when an original graph is given, we can find a unique default graph which is equivalent to the original one for three hypotheses.

Case 2: Suppose a default graph is given, we want to find a corresponding initial graph to see if it is unique. From first three equations we can find three initial critical values in terms of sequence critical values. According to Equations (3.4), (3.5) and (3.7), we can have $g_{12} = \frac{\tilde{\alpha}_1}{\tilde{\alpha}_1+\tilde{\alpha}_2}$; from Equations (3.4), (3.6) and (3.8), we can have $g_{13} = \frac{\tilde{\alpha}_2}{\tilde{\alpha}_1+\tilde{\alpha}_2}$. Similarly, we can find $g_{21} = \frac{\tilde{\alpha}_3}{\tilde{\alpha}_3+\tilde{\alpha}_4}$, $g_{23} = \frac{\tilde{\alpha}_4}{\tilde{\alpha}_3+\tilde{\alpha}_4}$, $g_{31} = \frac{\tilde{\alpha}_5}{\tilde{\alpha}_5+\tilde{\alpha}_6}$, and $g_{32} = \frac{\tilde{\alpha}_6}{\tilde{\alpha}_5+\tilde{\alpha}_6}$. The solutions are also unique, which means both approaches are equivalent.

However, when a given default graph is not complete, it is not necessary to find a corresponding original graph. See the following example 3.4.1.

Example 3.4.1 *A default graph is given in Figure 3.19, which consists of three sequences, with the sequence critical values $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$, respectively. Suppose $\sum_{k=1}^{3} \tilde{\alpha}_k = \alpha$.*

To find the corresponding original graph, a group of equations are found by setting equal critical values for the remaining hypotheses at each stage for both approaches. Then we will see if we can find the parameters for the original graph in terms of given information.

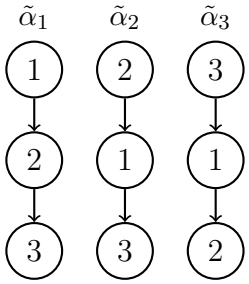

Figure 3.19 An incomplete default graph for three hypotheses, where the number of sequences is three.

$$
\begin{cases}
\tilde{\alpha}_1 = \alpha_1 & (3.14)\\
\tilde{\alpha}_2 = \alpha_2 & (3.15)\\
\tilde{\alpha}_3 = \alpha_3 & (3.16)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 = \alpha_2 + \alpha_1 g_{12} = \alpha_2(H_1) \quad \textit{Remove } H_1 & (3.17)\\
\tilde{\alpha}_3 = \alpha_3 + \alpha_1 g_{13} = \alpha_3(H_1) \quad \textit{Remove } H_1 & (3.18)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 = \alpha_1 + \alpha_2 g_{21} = \alpha_1(H_2) \quad \textit{Remove } H_2 & (3.19)\\
\tilde{\alpha}_3 = \alpha_3 + \alpha_2 g_{23} = \alpha_3(H_2) \quad \textit{Remove } H_2 & (3.20)\\
\tilde{\alpha}_1 + \tilde{\alpha}_3 = \alpha_1 + \alpha_3 g_{31} = \alpha_1(H_3) \quad \textit{Remove } H_3 & (3.21)\\
\tilde{\alpha}_2 = \alpha_2 + \alpha_3 g_{32} = \alpha_2(H_3) \quad \textit{Remove } H_3 & (3.22)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_3 = \alpha_3(H_1) + \alpha_2(H_1) g_{23}(H_1) \quad \textit{Remove } H_1 \textit{ and } H_2 & (3.23)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_3 = \alpha_2(H_1) + \alpha_3(H_1) g_{32}(H_1) \quad \textit{Remove } H_1 \textit{ and } H_3 & (3.24)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_3 = \alpha_1(H_2) + \alpha_3(H_2) g_{31}(H_2) \quad \textit{Remove } H_2 \textit{ and } H_3 & (3.25)
\end{cases}
$$

Therefore, we can find the initial transition coefficients $g_{12} = g_{21} = g_{31} = 1$ *and* $g_{13} = g_{23} = g_{32} = 0$ *according to Equations (3.14) to (3.22). According to the initial critical values and transition coefficients that we found, the corresponding initial original graph is attached in Figure 3.20.*

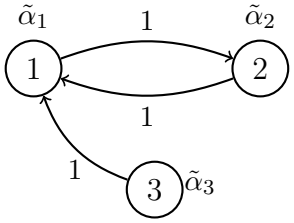

Figure 3.20 The corresponding initial incomplete original graph for three hypotheses in Figure 3.19.

Then, substitute these values to Equations (3.23) to (3.25), we can have the updated transition coefficients $g_{23}(H_1) = g_{32}(H_1) = g_{31}(H_2) = 1$. *However, according to the algorithm in Bretz et al. (2009), we have* $g_{23}(H_1) = 0$. *Thus for this given default graph, there is no corresponding updated original graph existing. From this example, we can note that* H_3 *have a chance to be tested at* α *level by using a default graph; however, in the corresponding initial original graph,* H_3 *can only be tested at level* $\tilde{\alpha}_3$.

If an original graph is incomplete but with $\sum_{j \in M} g_{ij} = 1$, $M = \{1, 2, 3\}$, we can find an equivalent default graph. See Example 3.4.2. However, if an original graph is incomplete and with $\sum_{j \in M} g_{ij} < 1$, $M = \{1, 2, 3\}$, it is not necessarily to find an equivalent default graph. See Remark 6.

Example 3.4.2 *When an original graph is given, with* $\sum_{i=1}^{3} \alpha_i = \alpha$, $g_{ii} = 0$ *and* $\sum_{j \in M} g_{ij} = 1$, $i, j \in M$ *and* $M = \{1, 2, 3\}$. *But at least one of* g_{ij}*'s are equal to 0 or 1.*

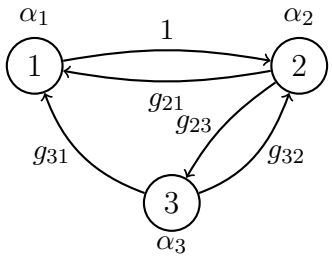

Figure 3.21 $g_{12} = 1$ and other $0 < g_{ij} < 1$.

For special cases from Figure 3.21 to 3.24, once we have $\sum_{j \in M} g_{ij} = 1$, *there must be existing an unique corresponding default graph. Take Figure 3.24 as an example, the corresponding default graph is given in Figure 3.25.*

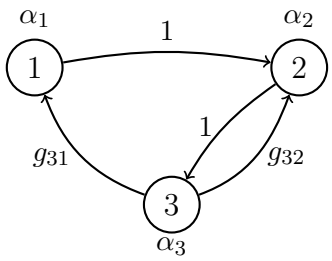

Figure 3.22 $g_{12} = g_{23} = 1$ and other $0 < g_{ij} < 1$.

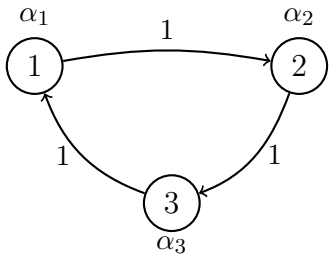

Figure 3.23 One example, $g_{12} = g_{23} = g_{31} = 1$.

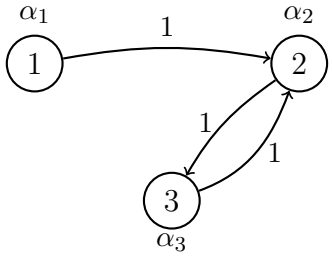

Figure 3.24 Another example, $g_{12} = g_{23} = g_{32} = 1$.

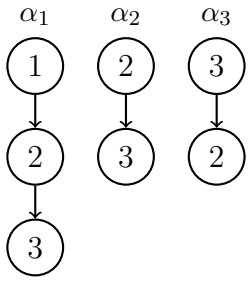

Figure 3.25 The corresponding default graph to original graph in Figure 3.24.

Remark 6 *However, when* $\sum_{j \in M} g_{ij} < 1$*, it is not necessarily to find a corresponding default graph, refer to Figure 3.26.*

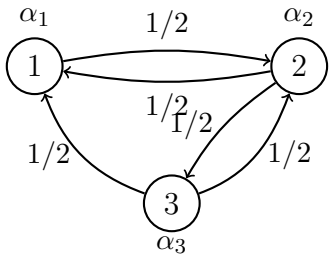

Figure 3.26 $g_{12} = 1/2$, $g_{13} = 0$.

According to the algebraic method, we can find the initial default graph for Figure 3.26.

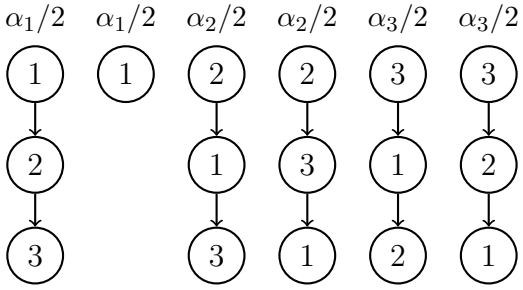

Figure 3.27 The corresponding default graph of the Figure 3.26.

The following graphs in Figures 3.28 and 3.29 are the updated graphs for two approaches, respectively. It is easy to see the updated graphs are not equivalent after removing either H_2 *or* H_3 *of the remaining two hypotheses.*

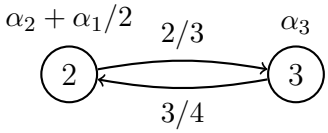

Figure 3.28 The updated original graph in Figure 3.26 after removing H_1 according to the algorithm in Bretz at el. (2009).

Third, the uncertainty of equivalency between two graphical approaches for more than three hypotheses are also discussed. We have already discussed the equivalence of two complete graphical approaches and also non-equivalence between two incomplete graphs

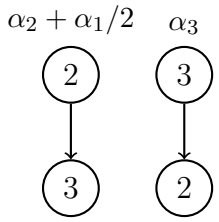

Figure 3.29 The updated default graph in Figure 3.27 after removing H_1 according to Algorithm 1.

for testing three hypotheses. In the following section, at beginning, we use an example to illustrate the complete original graph is corresponding to multiple default graphs when we simultaneously test four hypotheses. Next, a theorem of nonuniqueness is given. Then, we will discuss the non-equivalent property when there are more than three hypotheses, even when the given default graphs are complete.

Example 3.4.3 (Four hypotheses) *Suppose a complete original graph is given to simultaneously test 4 hypotheses H_1, H_2, H_3 and H_4. Suppose $\sum_{i=1}^{4} \alpha_i = \alpha$, $g_{ii} = 0$, $0 < g_{ij} < 1$, $i \neq j$, and $\sum_{j \in M} g_{ij} = 1$, $i, j \in M$ and $M = \{1, ..., 4\}$. See the following original graph in Figure 3.30.*

Then the corresponding default graph with 4!=24 sequences is attached in Figures 3.31 and 3.32. We will use the algebraic method to find the critical value for each sequence.

In the following group of equations, only a part of equations is listed, which is needed for finding $\tilde{\alpha}_1$, $\tilde{\alpha}_2$, $\tilde{\alpha}_7$, and $\tilde{\alpha}_8$. Equations (3.26) to (3.29) are defined to find equivalent initial critical values for both approaches. Equations (3.30) to (3.32) are defined to find equivalent updated critical values after removing H_1 for two approaches. Equations (3.33) to (3.35) are defined to find equivalent updated critical values after removing H_2 for two approaches. Equations (3.36) and (3.37) are defined to find equivalent updated critical values for H_3 and H_4 after removing H_1 and H_2, no matter how the removing order is. Allocated critical values for other sequences are found by the similar way.

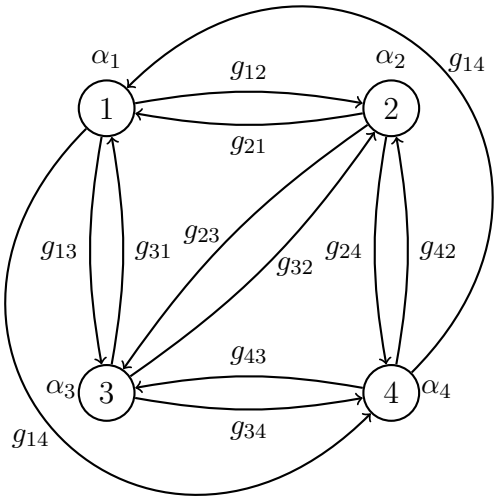

Figure 3.30 A complete original graph for testing 4 hypotheses.

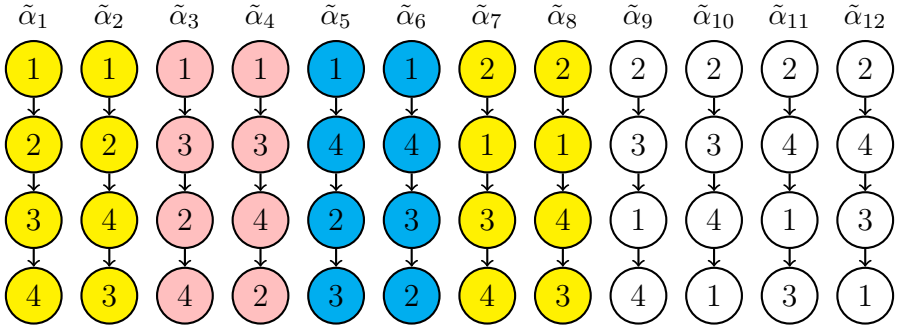

Figure 3.31 First 12 sequences in the complete default graph for testing 4 hypotheses.

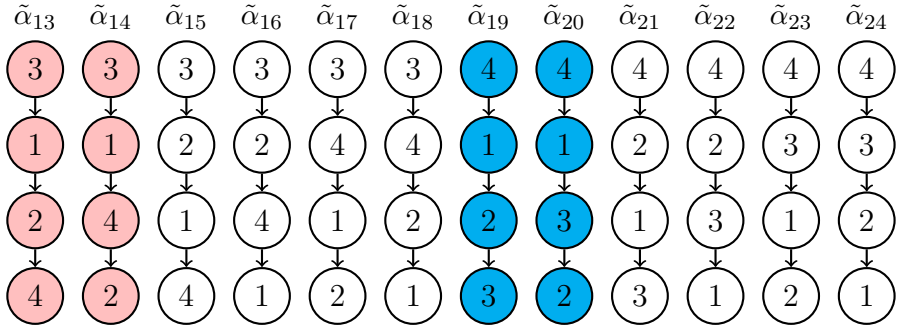

Figure 3.32 First 12 sequences in the complete default graph for testing 4 hypotheses.

$$
\begin{cases}
\sum_{k=1}^{6} \tilde{\alpha}_k = \alpha_1 & (3.26)\\
\sum_{k=7}^{12} \tilde{\alpha}_k = \alpha_2 & (3.27)\\
\sum_{k=13}^{18} \tilde{\alpha}_k = \alpha_3 & (3.28)\\
\sum_{k=19}^{24} \tilde{\alpha}_k = \alpha_4 & (3.29)\\
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \sum_{k=7}^{12} \tilde{\alpha}_k = \alpha_2 + \alpha_1 g_{12} & (3.30)\\
\tilde{\alpha}_3 + \tilde{\alpha}_4 + \sum_{k=13}^{18} \tilde{\alpha}_k = \alpha_3 + \alpha_1 g_{13} & (3.31)\\
\tilde{\alpha}_5 + \tilde{\alpha}_6 + \sum_{k=19}^{24} \tilde{\alpha}_k = \alpha_4 + \alpha_1 g_{14} & (3.32)\\
\tilde{\alpha}_7 + \tilde{\alpha}_8 + \sum_{k=1}^{6} \tilde{\alpha}_k = \alpha_1 + \alpha_2 g_{21} & (3.33)\\
\tilde{\alpha}_9 + \tilde{\alpha}_{10} + \sum_{k=13}^{18} \tilde{\alpha}_k = \alpha_3 + \alpha_2 g_{23} & (3.34)\\
\tilde{\alpha}_{11} + \tilde{\alpha}_{12} + \sum_{k=19}^{24} \tilde{\alpha}_k = \alpha_4 + \alpha_2 g_{24}, & (3.35)
\end{cases}
$$

and

$$\tilde{\alpha}_1 + \tilde{\alpha}_3 + \tilde{\alpha}_4 + \tilde{\alpha}_7 + \tilde{\alpha}_9 + \tilde{\alpha}_{10} + \sum_{k=13}^{18} \tilde{\alpha}_k = \alpha_3 + \alpha_1 g_{13} + (\alpha_2 + \alpha_1 g_{12}) g_{23}(H_1), \quad (3.36)$$

$$\tilde{\alpha}_2 + \tilde{\alpha}_5 + \tilde{\alpha}_6 + \tilde{\alpha}_8 + \tilde{\alpha}_{11} + \tilde{\alpha}_{12} + \sum_{k=19}^{24} \tilde{\alpha}_k = \alpha_4 + \alpha_1 g_{14} + (\alpha_2 + \alpha_1 g_{12}) g_{24}(H_1) \quad (3.37)$$

After simplifying above group of equations we get the following four equations:

$$\begin{cases} \tilde{\alpha}_1 + \tilde{\alpha}_2 = \alpha_1 g_{12} \\ \tilde{\alpha}_7 + \tilde{\alpha}_8 = \alpha_2 g_{21} \\ \tilde{\alpha}_1 + \tilde{\alpha}_7 = \alpha_1 g_{12} g_{23}(H_1) + \alpha_2 g_{21} g_{13}(H_2) \\ \tilde{\alpha}_2 + \tilde{\alpha}_8 = \alpha_1 g_{12} g_{24}(H_1) + \alpha_2 g_{21} g_{14}(H_2). \end{cases}$$

From Equations (1) to (4*), we can find*

$$\begin{cases} \tilde{\alpha}_1 = \alpha_1 g_{12} g_{23}(H_1) + \delta_1 \\ \tilde{\alpha}_2 = \alpha_1 g_{12} g_{24}(H_1) - \delta_1 \\ \tilde{\alpha}_7 = \alpha_2 g_{21} g_{13}(H_2) - \delta_1 \\ \tilde{\alpha}_8 = \alpha_2 g_{21} g_{14}(H_2) + \delta_1, \end{cases}$$

Then by similar method, we can get

$$\begin{cases} \tilde{\alpha}_3 = \alpha_1 g_{13} g_{32}(H_1) + \delta_2 \\ \tilde{\alpha}_4 = \alpha_1 g_{13} g_{34}(H_1) - \delta_2 \\ \tilde{\alpha}_{13} = \alpha_3 g_{31} g_{12}(H_3) - \delta_2 \\ \tilde{\alpha}_{14} = \alpha_3 g_{31} g_{14}(H_3) + \delta_2, \end{cases}$$

$$\begin{cases} \tilde{\alpha}_5 = \alpha_1 g_{14} g_{42}(H_1) + \delta_3 \\ \tilde{\alpha}_6 = \alpha_1 g_{14} g_{43}(H_1) - \delta_3 \\ \tilde{\alpha}_{19} = \alpha_4 g_{41} g_{12}(H_4) - \delta_3 \\ \tilde{\alpha}_{20} = \alpha_4 g_{41} g_{13}(H_4) + \delta_3, \end{cases}$$

$$\begin{cases} \tilde{\alpha}_9 = \alpha_2 g_{23} g_{31}(H_2) + \delta_4 \\ \tilde{\alpha}_{10} = \alpha_2 g_{23} g_{34}(H_2) - \delta_4 \\ \tilde{\alpha}_{15} = \alpha_3 g_{32} g_{21}(H_3) - \delta_4 \\ \tilde{\alpha}_{16} = \alpha_3 g_{32} g_{24}(H_3) + \delta_4, \end{cases}$$

$$\begin{cases} \tilde{\alpha}_{11} = \alpha_2 g_{24} g_{41}(H_2) + \delta_5 \\ \tilde{\alpha}_{12} = \alpha_2 g_{24} g_{43}(H_2) - \delta_5 \\ \tilde{\alpha}_{21} = \alpha_4 g_{42} g_{21}(H_4) - \delta_5 \\ \tilde{\alpha}_{22} = \alpha_4 g_{42} g_{23}(H_4) + \delta_5, \end{cases}$$

and

$$\begin{cases} \tilde{\alpha}_{17} = \alpha_3 g_{34} g_{41}(H_3) + \delta_6 \\ \tilde{\alpha}_{18} = \alpha_3 g_{34} g_{42}(H_3) - \delta_6 \\ \tilde{\alpha}_{23} = \alpha_4 g_{43} g_{31}(H_4) - \delta_6 \\ \tilde{\alpha}_{24} = \alpha_4 g_{43} g_{32}(H_4) + \delta_6. \end{cases}$$

We find that if δ-values change over some specific intervals, the default graphs are different, but equivalent. And these default graphs are corresponding to an unique original graph with initial critical values $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ and initial transition coefficients $\boldsymbol{G} = \{g_{ij}\}_{4\times 4}$.

For example, even though the solutions of sequence critical values are not unique; however, we have unique $\sum_{k=1}^{6} \tilde{\alpha}_k$, $\sum_{k=7}^{12} \tilde{\alpha}_k$, $\sum_{k=13}^{18} \tilde{\alpha}_k$, and $\sum_{k=19}^{24} \tilde{\alpha}_k$. And after removing any hypotheses, the summations of the corresponding sequence critical values for each of remaining hypotheses are also unique, for example, after removing H_1, we have unique

$\tilde{\alpha}_1 + \tilde{\alpha}_2 + \sum_{k=7}^{12} \tilde{\alpha}_k$, $\tilde{\alpha}_3 + \tilde{\alpha}_4 + \sum_{k=13}^{18} \tilde{\alpha}_k$ *and* $\tilde{\alpha}_5 + \tilde{\alpha}_6 + \sum_{k=19}^{24} \tilde{\alpha}_k$. *This is also true for removing two or three hypotheses. Therefore, all the corresponding default graphs are equivalent.*

Therefore, when an original graph is given for testing four hypotheses, even though the solutions for default graphs are not unique, but they are equivalent in the sense that when testing hypotheses are same, we will have the same rejection sets.

Consequently, we have the following result. And the proof of Theorem 5 is deferred to Appendix B.

Theorem 5 *A complete original graph, described as Definition 3.2.3, is used to test* m *hypotheses. Then, there are existing multiple corresponding default graphs, and they are equivalent.*

Next, we will introduce a necessary condition for two different original graphs are equivalent, which is , $\sum_{l \in M} g_{kl} = \sum_{l \in M} g'_{kl} = 1$, $M = \{1, ..., m\}$.

Consider a problem simultaneously test m hypotheses, two different original graphs are used to test the given hypotheses and they have same results. Suppose for one graph, the assigned initial critical values are $\alpha_1, ..., \alpha_m$, and we have $\alpha_i \geq 0$, $\sum_{i=1}^m \alpha_i = \alpha$. The initial transition coefficients are g_{ij}, and we have $0 \leq g_{ij} \leq 1$, $\sum_{j=1}^m g_{ij} \leq 1$, where $i, j \in M = \{1, ..., m\}$. For the other graph, the initial critical values are α'_i, and suppose $\alpha_i = \alpha'_i$, $\forall\, i \in M$. Let g'_{ij} denote the transition coefficients for this graph, where $0 \leq g'_{ij} \leq 1$, $\sum_{j=1}^m g'_{ij} \leq 1$, where $i, j \in M$.

Let $\alpha_l^{(t)}$ and $\alpha_l^{(t)\prime}$ denote the updated critical values for hypothesis H_l for both graphs at stage $1 \leq t \leq m$. Let $g_{jl}^{(t)}$ and $g_{jl}^{(t)\prime}$ denote the updated transition coefficients from H_j to H_l at stage $1 \leq t \leq m-1$. Since two original graphs are equivalent for testing m given hypotheses, they must have same initial and updated critical values at each stage if the rejection orders are same. And according to the Algorithm 1 in Bretz et al. (2009), if hypothesis H_j is rejected at stage t, we have following equations for two graphs, respectively.

$$\alpha_l^{(t+1)} = \alpha_l^{(t)} + \alpha_j^{(t)} g_{jl}^{(t)},$$

and

$$\alpha_l^{(t+1)\prime} = \alpha_l^{(t)\prime} + \alpha_j^{(t)\prime} g_{jl}^{(t)\prime},$$

where $l \in I_{t+1}$, I_{t+1} denotes the index set of remaining hypotheses. $\alpha_l^{(1)} = \alpha_l$ and $g_{jl}^{(1)} = g_{jl}$, $\alpha_l^{(1)\prime} = \alpha_l'$ and $g_{jl}^{(1)} = g_{jl}'$.

Moreover, since two original graphs are equivalent, and because of the uniqueness property of the rejection set, we can assume that they have same rejection orders of hypotheses. And also due to the uniqueness properties of critical values, we will have

$$\alpha_l^{(t)} = \alpha_l^{(t)\prime}, \quad t = 1, ..., m,$$

and

$$g_{jl}^{(t)} = g_{jl}^{(t)\prime}, \quad \text{if } \alpha_j^{(t)} = \alpha_j^{(t)\prime} \neq 0. \tag{3.38}$$

That means if $\alpha_j^{(t)} = \alpha_j^{(t)\prime} = 0$, at stage t, the transition coefficients from H_j to other hypotheses might be different for two different graphs. However, if the initial or updated critical values to H_j for two graphs are equal to 0, the hypothesis H_j does not have a chance to be rejected at the corresponding stages. Thus, this situation does not necessarily affect the result.

Once the assigned critical value for some hypothesis H_k is updated to a positive number from 0, the updated transition coefficients from H_k to other hypotheses will be same for two graphs. For example, we consider H_j is rejected at first stage, and $\alpha_k = \alpha_k' = 0$, $\alpha_k^{(2)} = \alpha_j g_{jk}$, and $\alpha_k^{(2)\prime} = \alpha_j' g_{jk}'$, $j \in M$, $k \in M \setminus \{j\}$. Then according to Equation (3.38), we have $g_{jk} = g_{jk}'$, and suppose they are non-zeros; therefore, $\alpha_k^{(2)} = \alpha_k^{(2)\prime} \neq 0$.

The updated transition coefficients from H_k to other hypotheses H_l, $l \in M \setminus \{j,k\}$ are as follows,

$$g_{kl}^{(2)} = \frac{g_{kl} + g_{kj}g_{jl}}{1 - g_{kj}g_{jk}},$$

and

$$g_{kl}^{(2)\prime} = \frac{g'_{kl} + g'_{kj}g'_{jl}}{1 - g'_{kj}g'_{jk}} = \frac{g'_{kl} + g'_{kj}g_{jl}}{1 - g'_{kj}g_{jk}}.$$

Then, according to equation (3.38), we have

$$\frac{g_{kl} + g_{kj}g_{jl}}{1 - g_{kj}g_{jk}} = \frac{g'_{kl} + g'_{kj}g_{jl}}{1 - g'_{kj}g_{jk}}.$$

Next, based on above equation, we want to find the condition that we need to satisfy the equality.

Consequently, we will have

$$\begin{aligned} \sum_{l\in M\setminus\{j,k\}} \frac{g_{kl} + g_{kj}g_{jl}}{1 - g_{kj}g_{jk}} &= \sum_{l\in M\setminus\{j,k\}} \frac{g'_{kl} + g'_{kj}g_{jl}}{1 - g'_{kj}g_{jk}} \\ \frac{\sum\limits_{l\in M} g_{kl} - g_{kj} + \sum\limits_{l\in M\setminus\{j,k\}} g_{kj}g_{jl}}{1 - g_{kj}g_{jk}} &= \frac{\sum\limits_{l\in M} g'_{kl} - g'_{kj} + \sum\limits_{l\in M\setminus\{j,k\}} g'_{kj}g_{jl}}{1 - g'_{kj}g_{jk}}, \end{aligned} \tag{3.39}$$

which is

$$\begin{aligned} &\sum_{l\in M} g_{kl} - g'_{kj}g_{jk}\sum_{l\in M} g_{kl} - g_{kj} + \sum_{l\in M\setminus\{j,k\}} g_{kj}g_{jl} \\ &= \sum_{l\in M} g'_{kl} - g_{kj}g_{jk}\sum_{l\in M} g'_{kl} - g'_{kj} + \sum_{l\in M\setminus\{j,k\}} g'_{kj}g_{jl} \\ &\sum_{l\in M} g_{kl} - g'_{kj}g_{jk}(\sum_{l\in M} g_{kl} - 1) + g_{kj}(\sum_{l\in M} g_{jl} - 1) \\ &= \sum_{l\in M} g'_{kl} - g_{kj}g_{jk}(\sum_{l\in M} g'_{kl} - 1) + g'_{kj}(\sum_{l\in M} g_{jl} - 1). \end{aligned} \tag{3.40}$$

From above equations, we can see when $\sum\limits_{l\in M} g_{kl} = \sum\limits_{l\in M} g'_{kl} = \sum\limits_{l\in M} g_{jl} = 1$, two sides are equal.

The above statement is also applied to an arbitrary stage $t+1$. Suppose H_j is rejected at stage t, and for some hypothesis H_k, $\alpha_k^{(t)} = \alpha_k^{(t)\prime} = 0$. Then after rejecting H_j, assume we have $\alpha_k^{(t+1)} = \alpha_k^{(t+1)\prime} > 0$; therefore, according to Equation (3.38), we have $g_{kl}^{(t+1)} = g_{kl}^{(t+1)\prime}$, $l \in I_{t+1}$, which is

$$\begin{aligned} \frac{g_{kl}^{(t)} + g_{kj}^{(t)} g_{jl}^{(t)}}{1 - g_{kj}^{(t)} g_{jk}^{(t)}} &= \frac{g_{kl}^{(t)\prime} + g_{kj}^{(t)\prime} g_{jl}^{(t)}}{1 - g_{kj}^{(t)\prime} g_{jk}^{(t)}} \\ \sum_{l \in I_t \setminus \{j,k\}} \frac{g_{kl}^{(t)} + g_{kj}^{(t)} g_{jl}^{(t)}}{1 - g_{kj}^{(t)} g_{jk}^{(t)}} &= \sum_{l \in I_t \setminus \{j,k\}} \frac{g_{kl}^{(t)\prime} + g_{kj}^{(t)\prime} g_{jl}^{(t)}}{1 - g_{kj}^{(t)\prime} g_{jk}^{(t)}}. \end{aligned} \tag{3.41}$$

Similarly, we will have the condition $\sum_{l \in I_t} g_{kl}^{(t)} = \sum_{l \in I_t} g_{kl}^{(t)\prime} = \sum_{l \in I_t} g_{jl}^{(t)} = 1$, two sides are equal.

Remark: If for all hypotheses H_k, $k \in M$, the assigned initial critical values $\alpha_k \neq 0$, then according to Equation (3.38) we have $g_{kl} = g'_{kl}$, $\forall\, k, l \in M$ if two graphs are equivalent. We can also say that for this case, the graph is unique.

However, the above condition is only a necessary condition which means when two different original graphs are equivalent, if some hypothesis H_k with zero initial critical value, the transition coefficients from H_k to other hypotheses for two graphs can be different, but the summation of them of one graph must be equal to 1. However, $\sum_{l \in M} g_{kl} = \sum_{l \in M} g'_{kl} = 1$ is not a sufficient condition.

We also provide an example 3.5.1 in Section 3.5 to illustrate that even if the default graph is complete when it is used to test more than three hypotheses, there is not necessarily an equivalent original graph existing. Also, we give an example to show that for some incomplete default graphs, by adding some edges between two hypotheses with associated infinitesimally small transition coefficients, we can find corresponding equivalent original graphs.

3.4.2 A direct proof of FWER control for the original graphical approach

In this section, we have shown that the three regularity conditions are achieved at each testing step, and monotonicity property of critical values and transition coefficient holds; thus, a direct proof of FWER control is obtained, consequently. Moreover, we also showed the final rejection set is unique and the original approach graphical approach is monotone in terms of p-values directly. These results offered a complete theoretical results for the original graphical approach and it will help us to understand it thoroughly and develop new graphical approaches.

The proofs of Proposition 3.4.1, Proposition 3.4.2, Proposition 3.4.3, Theorem 6 and Theorem 7 are deferred to Appendix B.

Proposition 3.4.1 *The critical value function $\alpha_k(\mathcal{R})$ and transition coefficients function $g_{lk}(\mathcal{R})$ are unique in terms of rejection set no mater what the rejection order is, where $\mathcal{R}$ is an arbitrary rejection set and l, k are the indices of hypotheses $H_l, H_k \in \mathcal{H} \setminus \mathcal{R}$.*

Proposition 3.4.2 *Three regularity conditions holds for rejection sets at any step,*

$$\sum_{k \in I_i} g_{lk}(\mathcal{R}_{i-1}) \leq 1, \tag{3.42}$$

$$0 \leq g_{lk}(\mathcal{R}_{i-1}) \leq 1, \quad g_{ll}(\mathcal{R}_{i-1}) = 0, \quad \forall l, k \in I_i,\ i = 1, ..., m, \tag{3.43}$$

$$\sum_{k \in I_i} \alpha_k(\mathcal{R}_{i-1}) \leq \alpha, \quad i = 1, ..., m. \tag{3.44}$$

Proposition 3.4.3 *Given two rejection sets $\mathcal{R}$ and $\mathcal{S}$, such that $\mathcal{R} \subset \mathcal{S}$, then we have shown $\alpha_k(\mathcal{R}) \leq \alpha_k(\mathcal{S})$ and $g_{lk}(\mathcal{R}) \leq g_{lk}(\mathcal{S})$, l, k are the indices of hypotheses $H_l, H_k \in \mathcal{H} \setminus \mathcal{S}$. The rejection set $\mathcal{R}$ consists of an arbitrary number of rejected hypotheses from 1 to $m-1$.*

Example 3.4.4 *If simultaneously test 4 hypotheses $\mathcal{H} = \{H_1,\ H_2,\ H_3,\ H_4\}$, suppose $\mathcal{R} = \{H_2\}$ and $\mathcal{S} = \{H_2,\ H_4\}$, then according to Proposition 3.4.3, we have $\alpha_1(\mathcal{R}) \leq \alpha_1(\mathcal{S})$, $\alpha_3(\mathcal{R}) \leq \alpha_3(\mathcal{S})$, $g_{13}(\mathcal{R}) \leq g_{13}(\mathcal{S})$ and $g_{31}(\mathcal{R}) \leq g_{31}(\mathcal{S})$.*

Theorem 6 (Uniqueness of rejection set) *When apply the original graphical approach to test a given number of hypotheses, the final rejection set $\mathcal{R}$ is unique no matter what the rejection order is.*

Theorem 7 (Monotone property in terms of p-values) *The original graphical approach is monotone in terms of p-values, in the sense that the original graphical approach always rejects a hypothesis with the smaller p-value whenever it rejects another hypothesis with a larger p-value.*

A direct proof of the FWER control for the original graphical approach is achieved. Suppose the first rejection of true null hypothesis happens in step i, in other words, at first $i-1$ steps, the graphical approach does not reject any true null hypotheses. Therefore, we have $\mathcal{R}_{i-1} \subset \mathcal{F}$; thus, $\alpha_l(\mathcal{R}_{i-1}) \leq \alpha_l(\mathcal{F})$, $l \in I_0$, which is due to the monotonicity of the critical value function. Correspondingly,

$$\begin{aligned} FWER &= Pr(V \geq 1) \\ &= Pr\left(\cup_{l\in I_0}\left\{P_l \leq \alpha_l(\mathcal{R}_{i-1})\right\}\right) \\ &\leq Pr\left(\cup_{l\in I_0}\left\{P_l \leq \alpha_l(\mathcal{F})\right\}\right) \\ &\leq \sum_{l\in I_0} Pr\left(P_l \leq \alpha_l(\mathcal{F})\right) \\ &\leq \sum_{l\in I_0} \alpha_l(\mathcal{F}) \\ &\leq \alpha. \end{aligned} \tag{3.45}$$

The first inequality of Equation (3.45) is due to the monotonicity of the critical value function. The second and third inequalities are based on Bonferroni inequality and the distribution of true nulls. Thus, the original graphical approach strong controls the FWER at α under arbitrary dependence.

Moreover, from the first inequality of Equation (3.45), $Pr\left(\cup_{l\in I_0}\left\{P_l \leq \alpha_l(\mathcal{R}_{i-1})\right\}\right) \leq Pr\left(\cup_{l\in I_0}\left\{P_l \leq \alpha_l(\mathcal{F})\right\}\right)$ we can see only all of false null hypotheses are rejected we may

achieve the largest type I error α. Therefore, if all the p-values of false nulls are equal to 0, and true nulls follow the uniform distribution from 0 to 1, the FWER could be equal to α based on the monotone property and uniqueness property of $\alpha(\mathcal{R})$ and rejection set.

3.5 Clinical Examples

Example 3.5.1 is used to explain for some default graphs when they are used to test more than three hypotheses, it is not necessarily to find a corresponding original graph even if the default graph is complete. As aforementioned, for testing more than three hypotheses, the degree of freedom of a complete default graph is more than that of an complete original graph; thus, a complete default graph is more flexible.

Example 3.5.1 *See the default graph in Example 3.4.3. Here, for simplicity I only attach part of the default graph where only the sequences with the tail hypotheses H_1 or H_2 are included, see Figure 3.33. Also, we have $\tilde{\alpha}_{13} = \tilde{\alpha}_{14} = \tilde{\alpha}_{17} = \tilde{\alpha}_{18} = 0.001$, $\tilde{\alpha}_{15} = \tilde{\alpha}_{16} = \tilde{\alpha}_{19} = \tilde{\alpha}_{20} = \tilde{\alpha}_{23} = \tilde{\alpha}_{24} = 0.002$, and $\tilde{\alpha}_{21} = \tilde{\alpha}_{22} = 0.003$. Suppose the associated p-values $P_1 = 0.01$, $P_2 = 0.015$, $P_3 = 0.04$ and $P_4 = 0.025$.*

Then, according to Algorithm 1, we have $\alpha_1 = \sum_{k=1}^{6} \tilde{\alpha}_k = 0.016$, $\alpha_2 = \sum_{k=7}^{12} \tilde{\alpha}_k = 0.012$, $\alpha_3 = \sum_{k=13}^{18} \tilde{\alpha}_k = 0.008$ and $\alpha_4 = \sum_{k=19}^{24} \tilde{\alpha}_k = 0.014$.

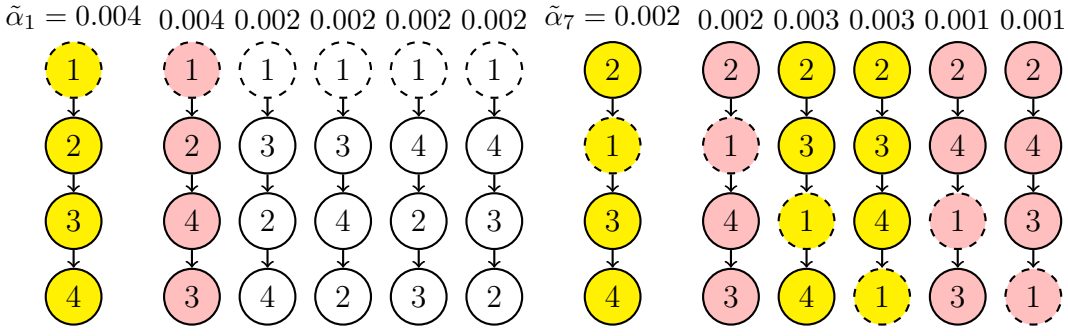

Figure 3.33 A part of the complete default graph for testing four hypotheses for which the sequences has H_1, H_2 as tail hypotheses.

According to Algorithm 2, we can find the initial transition coefficients, eg,

$$g_{12} = \frac{\sum_{k=1}^{24} \tilde{\alpha}_k I(S_k(1) = 1,\ S_k(2) = 2)}{\alpha_1} = \frac{0.004 + 0.004}{0.016} = \frac{1}{2}, \tag{3.46}$$

or

$$g_{21} = \frac{\sum_{k=1}^{24} \tilde{\alpha}_k I(S_k(1) = 2,\ S_k(2) = 1)}{\alpha_2} = \frac{0.002 + 0.002}{0.012} = \frac{1}{3}. \tag{3.47}$$

From above two equations and default graph, we can see the numerator of g_{12} are determined by the portions of critical value α_1 which are allocated to first and second sequences. If g_{12} is fixed, then there can be any combination of first two sequences critical values as long as the summation equals 0.008. So is g_{21}. Therefore, if we don't have additional restriction for the setting of initial critical values of sequences, after removing a hypothesis at first stage, then if we find the corresponding original graph according to the updated default graph, the updated transition coefficients $g_{23}(H_1)$ and $g_{24}(H_2)$ will vary depending on how we set $\tilde{\alpha}_1$, $\tilde{\alpha}_2$, $\tilde{\alpha}_7$ and $\tilde{\alpha}_8$. Only when we set $\tilde{\alpha}_1 = 0.0056$, $\tilde{\alpha}_2 = 0.0024$, $\tilde{\alpha}_7 = 0.0024$, and $\tilde{\alpha}_8 = 0.0016$, then $g_{23}(H_1)$ and $g_{24}(H_2)$, found according to the original graphical approach can be equal to the ones found by the default graphical approach, and $g_{23}(H_1) = 7/10$, $g_{24}(H_1) = 3/10$. This is because for this setting we also consider how to pass the updated critical value of H_2 to H_3 or H_4 after removing H_1.

Then, by similar method, we get the initial transition coefficient matrix is as follows.

$$\boldsymbol{G} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{2}{7} & \frac{3}{7} & \frac{2}{7} & 0 \end{bmatrix}.$$

Return to this example, since $P_1 < \alpha_1$, $P_2 > \alpha_2$, $P_3 > \alpha_3$ and $P_4 > \alpha_4$, H_1 is rejected at the first stage. Then, according to Algorithm 1, the updated graph is attached in Figure 3.34. Take the sequences with tail hypothesis H_2 as an example, we have two different sequences: (1) $H_2 \to H_3 \to H_4$ (yellow color labeled), and (2) $H_2 \to H_4 \to H_3$ (pink color labeled).

Also, according to the algorithm in Bretz et al. (2009), after rejecting H_1, the updated original graph is attached in Figure 3.35.

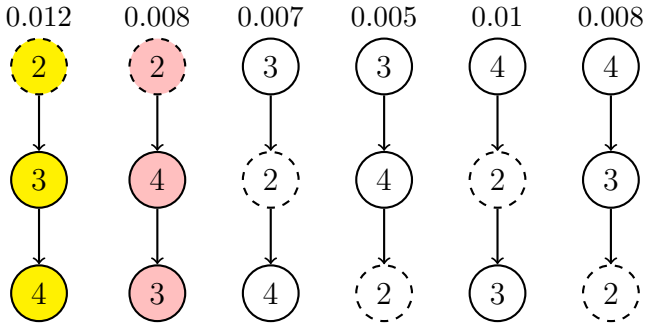

Figure 3.34 The updated default graph after removing H_1 according to Algorithm 1.

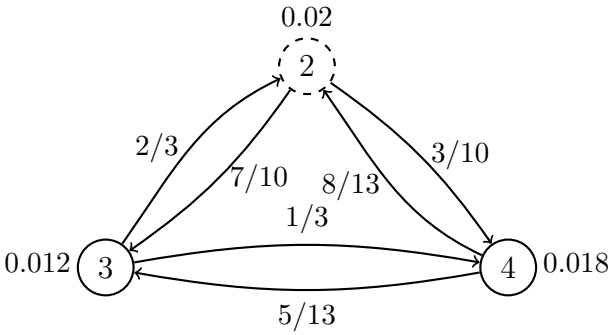

Figure 3.35 The updated original graph after removing H_1 according to the algorithm in Bretz et al. (2009).

We can see, based on both approaches, H_2 is rejected at second stage since $P_2 < \alpha_2(H_1)$, $P_3 > \alpha_3(H_1)$ and $P_4 > \alpha_4(H_1)$. The updated graphs in Figure 3.36 and 3.37 are corresponding to two updating algorithms of the default graph and original graph, respectively.

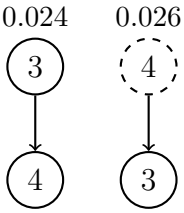

Figure 3.36 The updated default graph at stage 3 after rejecting H_1 and H_2 according to Algorithm 1.

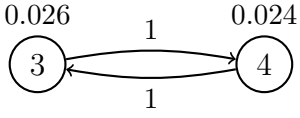

Figure 3.37 The updated original graph at stage 3 after rejecting H_1 and H_2 according to the algorithm in Bretz at el. (2009).

Based on the updated default graph, we can see H_4 is rejected as $P_4 < \alpha_4(H_1, H_2)$. However, the original graphical approach stops testing since both of remaining p-values are larger than corresponding critical values. This means for the given default graph, there is no corresponding original graph existing.

Example 3.5.2 *In this example, we want to show for some incomplete default graph, we may find a corresponding original graph by adding some sequences with infinitesimally small allocated critical values. Consider Example 3.4.1 and Figure 3.19 again.*

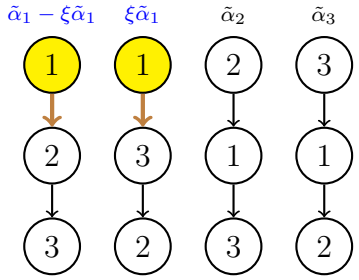

Figure 3.38 Add a sequence $\xi\tilde{\alpha}_1 : H_1 \to H_3 \to H_2$.

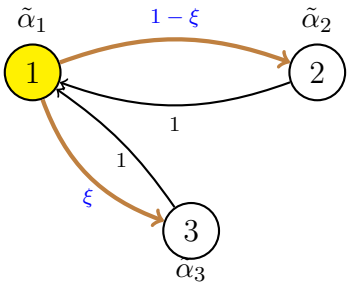

Figure 3.39 Add an edge from H_1 to H_3 with an infinitesimally small number $g_{13} = \xi$.

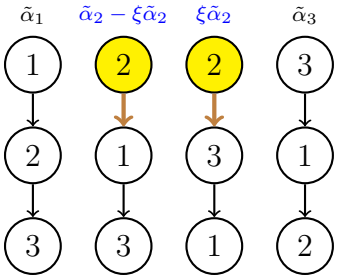

Figure 3.40 Add a sequence $\xi\tilde{\alpha}_2 : H_2 \to H_3 \to H_1$.

From Example 3.4.1, we know there is no corresponding original graph to Figure 3.19. However, for the method 1, by adding a sequence $\xi\tilde{\alpha}_1 : H_1 \to H_3 \to H_2$ where ξ is

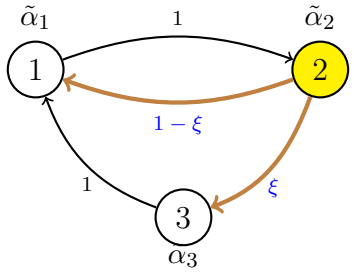

Figure 3.41 Add an edge from H_2 to H_3 with an infinitesimally small number $g_{23} = \xi$.

an infinitesimally small number to Figure 3.38 and adding an edge with an infinitesimally small transition coefficient ξ from H_1 to H_3 to Figure 3.39, we can find the equivalent corresponding graphs, especially when ξ goes to 0. And for method 2, by adding a sequence $\xi\tilde{\alpha}_2 : H_2 \to H_3 \to H_1$ where ξ is an infinitesimally small number to Figure 3.40 and adding an edge with an infinitesimally small transition coefficient ξ from H_2 to H_3 to Figure 3.41, we also can find the equivalent corresponding graphs, especially when ξ goes to 0.

3.6 Conclusion

In this chapter, we mainly investigated the similarities and differences between the original graphical approach and the default graph. Currently, these two approaches are usually considered to be equivalent since both of them are used to visualize Bonferroni-type sequentially rejective procedures. Actually, when the graphs are not complete and the number of hypotheses are more than three, their performance may be different. We have shown the equivalence between two graphical approaches when we simultaneously test three hypotheses and both of graphs are complete. For testing three hypotheses, if a default graph is incomplete, it is not necessarily to find an equivalent original graph; however, even if an original graph is incomplete but with $\sum_{j=1}^{3} g_{ij} = 1$, $i = 1, 2, 3$, we can find an equivalent default graph. We also find if an original graph is incomplete and with $\sum_{j=1}^{3} g_{ij} < 1$, it is not necessarily to find an equivalent default graph. For testing more than three hypotheses, we have proved that if an original graph is complete, there will be corresponding to multiple equivalent default graphs. We also provide a proof of a

necessary condition of non-uniqueness of original graph which is $\sum_{l \in M} g_{kl} = \sum_{l \in M} g'_{kl} = 1$, $M = \{1, ..., m\}$. However, given an complete default graph for testing more than three hypotheses, it is not necessarily to find an equivalent original graph. One of the reasons is the degree of freedom of a complete default graph is more than that of a complete original graph. In this chapter, we also give an elegant theoretical result: a direct proof of the FWER control for the original graphical approach. Moreover, for the families of hypotheses testing problems, for example when we use the serial gatekeeping procedure, it is convenient to use default graph since there are existing problems to use original graph without adding edges with infinitesimally small transition coefficients between two families. This work, adding infinitesimally small transition coefficients ξ for application purpose, is not complete yet. We are working on providing general guidelines for how to add edges. Specially, between which two hypotheses should the edges be placed and what values of transition coefficients should be associated with the added edges?

CHAPTER 4
A GENERALIZED STEP-UP FDR CONTROLLING PROCEDURE FOR DISCRETE DATA

4.1 Introduction

In this chapter, we focus on developing a new generalized step-up FDR controlling procedure for discrete data. In many applications, such as clinical safety analysis, genome-wide association studies (GWAS) and next generation sequencing data (NGS), the experiment data often takes the forms of counts and the number of hypotheses for such studies is usually large. Thus, the procedures that considering the discrete properties of the data and controlling the FDR are required for this kind of problems. Most existing FDR controlling procedures are developed for continuous data (Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)), which are often conservative when analyzing discrete data because the distributions of p-values under true nulls for discrete case are usually stochastically larger than U(0, 1). In addition, for discrete data, under true null hypotheses, different p-values often have different distributions. To overcome this conservatism, we aim to develop a FDR controlling procedure, taking the properties of discreteness and heterogeneity of discrete data into account.

Several FDR controlling procedures were developed for analyzing discrete data by considering various information of the discrete data: the minimal attainable p-values, see Gilbert (2005), the mid$P$-values, see Heller and Gur (2011) and the averaged cumulative distribution functions of the p-values, refer to Heyse (2011) and Dohler, Durand and Roquain (2018). However, none of these procedures can be proved to control the FDR and be more powerful than BH procedure theoretically at same time. Jiang et al. (2017) introduced a discrete FDR method which coincides with the permutation-based FDR estimation procedure of Li and Tibshirani (2013) by using the nonparametric method to exploit the discreteness of the test statistics, but it is not proved to be more powerful than BH procedure. Moreover, most existing multiple testing procedures (MTPs) are stepwise

methods with a unique critical value function which is suitable for testing the continuous data, since under the hypotheses are true, the distributions of p-values are uniformly distributed over an interval (0, 1). However, because of the heterogeneity property of discrete data the existing rejection rule may not be proper any more and consequently, it will be promising to find a specific sequence of critical values for each hypothesis by fully utilizing its underlying distribution, which will be known if the minimal attainable p-values are known. Lynch and Guo (2016)introduced a new concept of generalized stepwise procedure which generalizes the usual stepwise procedure to the case where each hypothesis is tested with a sequence of its own critical constants. In this research project, under the same framework of the aforementioned generalized step-up approach, by fully utilizing known marginal distributions of true null p-values, we develop a powerful generalized step-up procedure for discrete case under independence structure with proven control of the FDR.

The rest of this chapter is organized as follows: in Section 4.2, we introduce basic notations used in this chapter. In section 4.3, we discuss some existing FDR controlling procedures for discrete data and their limitations. Section 4.4 introduces a generalized step-up procedure. The FDR control of the proposed procedure is proved in Section 4.5. In Section 4.6, we demonstrate clinical trial examples to illustrate the proposed procedure. In Section 4.7, a summary and future plan are given.

4.2 Preliminary

The FDR is defined as $FDR = E(\frac{V}{R})$, where we use the convention that $\frac{V}{R} \equiv 0$ when $R = 0$. Here, R and V are number of rejected hypotheses and falsely rejected hypotheses among the m hypotheses. Consider we simultaneously test m hypotheses $H_1, ..., H_m$, suppose there are m_0 true null hypotheses and m_1 false null hypotheses, assuming the testing statistics are discrete, with p-values $P_1, ..., P_m$ correspondingly. Let F_i denote the

CDF of P_i under H_i is true and let $\mathbb{P}_i$ denotes the full set of all attainable values for P_i, $i = 1, ..., m$.

In our project,we make use of the following assumption regarding marginal p-values: under the hypotheses are true,

$$F_i(u) \leq u \text{ for any } 0 \leq u \leq 1,$$

here if $u \in \mathbb{P}_i$, the above inequality becomes equality; otherwise, $F_i(u) < u$.

4.3 The Existing FDR Controlling Procedures

Lynch and Guo (2016) presented a generalized stepwise procedure where each hypothesis is tested with a different set of critical constants.

For the generalized step-up procedure, if we simultaneously test m hypotheses $H_1, ..., H_m$ with p-values $P_1, ..., P_m$ correspondingly, then given m non-decreasing critical functions $\alpha_r^{(i)}$, $i = 1, ..., m$, this procedure rejects H_i if $P_i \leq \alpha_R^{(i)}$ for each $i = 1, ..., m$ where R is determined as follows : $R = max\left\{0 \leq r \leq m : r \leq \sum_{i=1}^{m} I\left\{P_i \leq \alpha_r^{(i)}\right\}\right\}$.

Remark 7 *From this generalized step-up procedure, we can see that each hypothesis can have its own critical function; therefore, this framework is especially suitable for discrete data. That is because under true null hypotheses, p-values for different hypotheses have different distributions, which is not like continuous case that all p-values under true nulls will follow an unique uniform distribution from 0 to 1. Thus, it will be promising to find a specific sequence of critical values for each hypothesis by fully utilizing its underlying distribution.*

Gilbert procedure (2005): a simple two-step combination of the Tarone procedure (1990) and BH procedure, which is described below.

(1) Tarone's procedure (1990): is a modified Bonferroni procedure for discrete data. Suppose that p_i^* is the smallest attainable p-value for H_i, i=1,...,m. Let

$$m(k) = \sum_{i=1}^{m} I\left\{p_i^* \leq \frac{\alpha}{k}\right\} \quad and \quad K = min\left\{1 \leq k \leq m : m(k) \leq k\right\}.$$

Then, reject H_i if $P_i \leq \frac{\alpha}{K}$. Here, m(k) is the number of hypotheses with $p_i^* < \alpha/k$, R_K is the set of indices satisfying $p_i^* < \alpha/K$, which contains m(K) indices.

(2) The newly proposed modified FDR procedure: first, compute the integer K and the subset of m(K) indices R_K among the m hypotheses; second, perform BH procedure at level α on the subset of hypotheses R_K.

We can see this modified procedure carries out BH procedure using cut-off levels $\frac{i\alpha}{m(K)}$, $i = 1, ..., m(K)$. Because K, $m(K)$ and R_K are calculated on the basis of marginal information only, it follows that the FDR procedure conducted on the subset of indices in R_K controls the FDR at level less than or equal to $\{m_0(K)/m(K)\}\,\alpha \leq \alpha$.

Gilbert (2005) also showed, under arbitrary dependence, the FDR can be controlled at level α if the cut-off critical values are $\frac{i\alpha}{m(K)\sum_{j=1}^{m(K)}(1/j)}$, $i = 1, ..., m(K)$.

Remark 8 *When Gilbert procedure is applied to continuous data, $R_k = m$ since $p_i^* = 0$. Thus, it is reduced to BH procedure. However, this procedure cannot be theoretically shown more powerful than BH procedure since it heavily depends on the minimal attainable p-values.*

Heyse procedure (2011): Heyse's procedure is relying on the following averaged cumulative distribution functions (CDFs):

$$\bar{F}(t) = \frac{1}{m}\sum_{i=1}^{m} F_i(t), \quad t \in [0, 1],$$

where each F_i corresponds to the CDF of the ith p-value under the true null hypothesis.

The critical values of the Heyse procedure can be obtained by inverting $\bar{F}$ at the values $\alpha k/m$, $k = 1, ..., m$. Thus, the smaller the $\bar{F}$-values, the larger the critical values.

Remark 9 *Heyse procedure is a BH-type procedure and it is more powerful than BH procedure when the heterogeneity and discreteness both exist; however, it cannot be theoretically shown to control the FDR.*

The adjusted p-values for Heyse procedure are

$$P_{(j)}^{adj} = \begin{cases} P_{(j)}, & \text{if j = m,} \\ \min\left(P_{(j+1)}^{adj},\ \left[\sum_{i=1}^{m} F_i(P_{(j)})\right]/j\right) & \text{if j = m-1,...,1.} \end{cases}$$

Hypotheses with $P_{(j)}^{adj} \leq \alpha$ are declared significant.

Döhler, Durand and Roquain (2018) heterogeneous step-up procedure [HSU]: The step-up procedure $\mathbf{SU}_{(\tau)}$ using the critical values defined in the following way:

$$\tau_m = max\left\{t \in \mathbb{P} : \frac{1}{m}\sum_{i=1}^{m}\frac{F_i(t)}{1-F_i(t)} \leq \alpha\right\},$$

$$\tau_k = max\left\{t \in \mathbb{P} : t \leq \tau_m, \frac{1}{m}\sum_{i=1}^{m}\frac{F_i(t)}{1-F_i(\tau_m)} \leq \alpha k/m\right\}, \quad 1 \leq k \leq m-1,$$

where $\mathbb{P} = \bigcup_{i=1}^{m}\mathbb{P}_i$ denotes the overall p-value support and $\mathbb{P}_i$ denotes the attainable values for p-value P_i, $i = 1, ..., m$, under the null hypotheses.

Then the upper bound of FDR when applying $\mathbf{SU}_{(\tau)}$ is given by

$$FDR(SU_{(\tau)}) \leq min\left(\sum_{i=1}^{m}\max_{1\leq k\leq m}\frac{F_i(\tau_k)}{k}, \quad \max_{1\leq k\leq m}\max_{\substack{A\subset\{1,...,m\}\\|A|=m-k+1}}\left(\frac{1}{k}\sum_{i\in A}\frac{F_i(\tau_k)}{1-F_i(\tau_m)}\right)\right).$$

Remark 10 *In order to overcome the problem that Heyse procedure cannot control the FDR, $\boldsymbol{SU}_{(\tau)}$ has a little bit conservative modification of Heyse procedure. We can see that if $\boldsymbol{SU}_{(\tau)}$ is applied to continuous data, then for any $t \in (0,1)$, we have $\frac{1}{m}\sum_{i=1}^{m}\frac{F_i(t)}{1-F_i(t)} = \frac{t}{1-t}$. Then $\tau_m = \frac{\alpha}{1+\alpha}$, consequently, $\tau_k = \frac{\alpha}{1+\alpha}\frac{k}{m}$, $k = 1, ..., m-1$. Thus, when applying $\boldsymbol{SU}_{(\tau)}$ to continuous case, it is same as applying BH procedure at $\frac{\alpha}{1+\alpha}$ level, which is a little bit more conservative than BH procedure.*

4.4 The Proposed Generalized Step-up Procedure

In this section, we introduce a new generalized step-up procedure, different from the traditional ones, refer to Gilbert (2005), Heyse (2011) and Dohler and Roquain (2017), which generalizes the usual step-wise procedure to the case where each hypothesis is tested with a different set of critical constants. To the best of our knowledge, the proposed procedure is the first procedure which can be theoretical proved to be more powerful than BH procedure with proven FDR control under independence.

Suppose we simultaneously test m hypotheses $H_1, ..., H_m$ with the corresponding discrete testing statistics $T_1, ..., T_m$. Let P_i denote the p-value for testing H_i and $\mathbb{P}_i = \left\{p_1^{(i)}, ..., p_n^{(i)}\right\}$ denotes the full set of all attainable p-values for H_i such that $P_i \in \mathbb{P}_i$, where $p_n^{(i)} = 1$ and let $p_0^{(i)} = 0$ for all i. Suppose F_i denote the CDF of P_i under H_i is true, such that $F_i(u) = u$, for any $u \in \mathbb{P}_i$; otherwise, $F_i(u) < u$.

Then we find the critical values sequences for m hypotheses, respectively. For hypothesis H_i, we find a non-decreasing $\alpha_r^{(i)}$ sequence by using the following method:

$$\alpha_r^{(i)} = \max_{1 \le j \le m} \left\{ \frac{j\alpha}{m} : F_i(\frac{j\alpha}{m}) \le \frac{r\alpha}{m} \right\}, \tag{4.1}$$

where $r = 1, ..., m$.

In order to make it easy to be applicable and provable, we introduce some notations and detailed steps to find critical value sequences as follows. Define

$$\psi^{(i)}(k) = max \left\{ 0 \le r \le m : \frac{r\alpha}{m} < p_k^{(i)} \right\}, \tag{4.2}$$

which is $\psi^{(i)}(k) = \lfloor \frac{mp_k^{(i)}}{\alpha} \rfloor$, for $k = 1, ..., n$, if the maximum does not exist, let $\psi^{(i)}(k) = 0$.

The function $\psi^{(i)}(k)$ is non-decreasing in k, since $p_k^{(i)}$ is increasing in k.

Let

$$S^{(i)} = \left\{\text{non-zero jump points of } \psi^{(i)}(k),\ k = 1, ..., n\right\} \text{ and } K^{(i)} = \mid S^{(i)} \mid;$$

then let

$$S^{(i)} = \left\{ j_1^{(i)},\, j_2^{(i)}, ...,\, j_{K^{(i)}}^{(i)} \right\},$$

where $j_{K^{(i)}}^{(i)} = m$ and let $j_0^{(i)} = 0$ for all i.

Define

$$\tilde{D}(u) = \sum_{i=1}^{m} \left[F_i(\frac{j_1^{(i)}}{m}u) + \sum_{s=1}^{K^{(i)}-1} \left[\frac{F_i(\frac{j_{s+1}^{(i)}}{m}u)}{j_s^{(i)}+1} - \frac{F_i(\frac{j_s^{(i)}}{m}u)}{j_{s-1}^{(i)}+1} \right] I\left\{ \frac{F_i(\frac{j_{s+1}^{(i)}}{m}u)}{j_s^{(i)}+1} > \frac{F_i(\frac{j_s^{(i)}}{m}u)}{j_{s-1}^{(i)}+1} \right\} \right],$$

for any given $u \in (0, 1)$.

Let $D(u) = \max_\beta \left\{ \tilde{D}(\beta) : \beta \le u \right\}$ which is monotonically non-decreasing in u. Then we find $\tilde{\alpha} = max\left\{ u : D(u) \le \alpha \right\}$. For simplicity, let $\tilde{\alpha} = D^{-1}(\alpha)$. Thus, $\alpha_r^{(i)} = \frac{j_s^{(i)}}{m}\tilde{\alpha}, \quad j_{s-1}^{(i)} < r \le j_s^{(i)}, s = 1, ..., K^{(i)}$.

4.4.1 A generalized step-up procedure

Suppose simultaneously testing m hypotheses $H_1, ..., H_m$ where the testing statistics are discrete, with p-values $P_1, ..., P_m$ correspondingly.

Procedure 4.4.1 *Given m non-decreasing sequences of critical constants $\alpha_r^{(i)}$, $i = 1, ..., m$, the hypotheses are tested as follows.*

1. *Determine R first, let $R = max\left\{ r \in \{0,\, 1,\, ...,\, m\} : r \le \sum_{i=1}^{m} I\left\{ P_i \le \alpha_r^{(i)} \right\} \right\}$, where $\alpha_0^{(i)} = 0$.*

   (a) *Step 1. If $m > \sum_{i=1}^{m} I\left\{ P_i \le \alpha_m^{(i)} \right\}$, go to the next step. Otherwise reject all hypotheses and stop.*

   (b) *Step $j = 2, ..., m-1$. If $m - j + 1 > \sum_{i=1}^{m} I\left\{ P_i \le \alpha_{m-j+1}^{(i)} \right\}$, go to the next step. Otherwise reject the corresponding $m - j + 1$ H_i's with $P_i \le \alpha_{m-j+1}^{(i)}$ and stop.*

   (c) *Step m. If $1 > \sum_{i=1}^{m} I\left\{ P_i \le \alpha_1^{(i)} \right\}$, retain all the hypotheses. Otherwise reject the corresponding H_i with $P_i \le \alpha_1^{(i)}$.*

2. *Test the hypotheses H_1, ..., H_m at level $\alpha_R^{(1)}$, ..., $\alpha_R^{(m)}$, respectively. Then reject H_i if and only if $P_i \le \alpha_R^{(i)}$.*

Remark 11 *For example, if $\frac{\alpha}{m}$ is located in an interval $[p_k^{(i)}, p_{k+1}^{(i)})$,*

$$\psi^{(i)}(j) = max\left\{0 \le r \le m : \frac{r\alpha}{m} < p_j^{(i)}\right\} = 0, \quad for\, j = 1, ..., k,$$

and

$$\psi^{(i)}(k+1) = max\left\{0 \le r \le m : \frac{r\alpha}{m} < p_{k+1}^{(i)}\right\} = j_1^{(i)},$$

then, we can check $F_i(\frac{j_1^{(i)}}{m}\alpha) = p_k^{(i)} \le \frac{\alpha}{m}$ and also $F_i(\frac{j_1^{(i)}+1}{m}\alpha) = p_{k+1}^{(i)} > \frac{\alpha}{m}$ since $j_1^{(i)}$ is the largest r which satisfies $\frac{r\alpha}{m} < p_{k+1}^{(i)}$. Next, we continue to find $\psi_{(i)}(k+j)$, $j = 2, ..., n-k$ until we once get a larger value than $j_1^{(i)}$, which will be denoted to $j_2^{(i)}$. Then, by similar way, we can obtain $j_s^{(i)}$, $s = 3, ..., K^{(i)}$, respectively. It's easy to show the critical value sequences that we find by the above method are same as those found by equation (4.1).

4.4.2 A simple algorithm to find $j_s^{(i)}$ and $K^{(i)}$

A simple algorithm to find $j_s^{(i)}$ and $K^{(i)}$ is described bellow.

Algorithm 3 *For simplicity, we will omit the index i for the attainable p-values of P_i, $j_s^{(i)}$ and $K^{(i)}$.*

1. *Initialize $j_0 = 0$, $t = 1$.*
2. *For any given $\alpha \in (0, 1)$, let $a_t = F_i(\frac{j_{t-1}+1}{m}\alpha)$, find $k \in \{0, 1, ..., n\}$ such that $p_k = a_t$, let $b_t = p_{k+1}$. Then, find the largest l which satisfy $\frac{l\alpha}{m} \in [a_t, b_t)$, l is a positive integer from 1 to m.*
3. *(a) If $l = m$, set $j_K = m$, $K = t$ and stop.*
   (b) If $l < m$, set $j_t = l$ and increase t by 1. Then repeat step 2 and 3.

4.5 The FDR Control

According to the Procedure 4.4.1, the method (4.1), and the algorithm to find $j_s^{(i)}$ and $K^{(i)}$, we have the following result.

Theorem 8 *Under independence structure of P_i, $i = 1, ..., m$, the generalized step-up procedure 4.4.1 with m critical constant sequences $\alpha_r^{(i)} = \frac{j_s^{(i)}}{m}\tilde{\alpha}$ that we find by the aforementioned method, strongly controls the FDR at level α.*

Therefore, we provide a theoretical result of the FDR control for the proposed method. Then proof of Theorem 8 is deferred to Appendix C. Combined with the following result, we provide a procedure which is theoretically more powerful than BH procedure and meanwhile, it is also proved to control the FDR under independence structure.

Proposition 4.5.1 *The generalized step-up procedure 4.4.1 is uniformly more powerful than BH procedure, that is, procedure 4.4.1 will reject any hypotheses that are rejected by BH procedure.*

The result just simply follows Equation (4.1), that is, each critical value sequence is obtained by inverting the CDF of P_i at $r\alpha/m$, $r = 1, ..., m$, and the distribution is stochastically larger than U(0, 1). Thus, all m critical value sequences are not less than those for BH procedures in coordinatewise.

Remark 12 *The generalized step-up procedure 4.4.1 will be reduced to BH procedure when it is applied to continuous data. Moreover, for discrete case, when the distributions of P_i, i = 1, ..., m are homogeneous. There is no space to improve the procedure. One special case is given as follows.*

Suppose that p_i^*'s are the smallest attainable p-values for H_i's, $i = 1, ..., m$. (1) For any given α, if $p_i^* > \alpha$, $D(\alpha) = 0$ since for any $j \leq k$, $\frac{i_j}{m}\alpha \leq \alpha < p_i^*$. (2) If $p_i^* = \alpha$, in the summation terms of $D(\alpha)$, only when j = k - 1, $F_i(\frac{i_{j+1}}{m}\alpha) = F_i(\frac{i_k}{m}\alpha) = \alpha = p_i^*$, others equals 0; thus, $D(\alpha) = \sum_{i=1}^{m} \frac{p_i^*}{i_{k-1}+1} \geq p_i^*$.

4.6 Clinical Examples

In this section, we will use a sample of simultaneous testing $m = 10$ hypotheses, which consists of two different distributions under true null hypotheses, to demonstrate that the proposed generalized procedure has a great potential to perform better than the Döhler, Durand and Roquain (2018) procedure and Gilbert (2005) procedure.

Consider simultaneously testing m independent hypotheses, suppose for each hypothesis we conduct a two sample test with two binomial responses X_{i1} and X_{i2}, observed for each

of n individuals in each group, such as $X_{i1} \sim b(n, p_{i1})$, $X_{i2} \sim b(n, p_{i2})$, $i = 1, ..., m$. Then, to test two-sided hypothesis $H_i : p_{i1} = p_{i2}$ vs $H_i^{'} : p_{i1} \neq p_{i2}$ where p_{i1} and p_{i2} are the success probabilities for ith treatment group and control group, respectively. Let $X_i = X_{i1} + X_{i2}$. Let T_i denote the number of success belongs to ith treatment group in the sample. Thus, the test statistics $T_i \sim Hypergeometric(2n,\ n,\ X_i)$.

Example 4.6.1 *Consider simultaneously testing 10 independent hypotheses with two distributions under true nulls, and the sample size for each group is* $n = 5$*. Suppose* $i = 1, ..., 5$*, we have* $X_{i1} = 4$ *and* $X_{i2} = 0$*, when* $i = 6, ..., 10$*, we have* $X_{i1} = 4$ *and* $X_{i2} = 1$*. Then, we have 5 test statistics* $T_i \sim Hypergeometric\,(10,\ 5, 4)$*,* $i = 1, ..., 5$ *and the other 5 test statistics* $T_i \sim Hypergeometric\,(10,\ 5,\ 5)$*,* $i = 6, ..., 10$*.*

Therefore, for the first 5 hypotheses, $P_i = 0.0476$*, the full set of attainable* p*-values* $\mathbb{P}_i = \{0.0476,\ 0.524,\ 1\}$*, while for the remaining 5 hypotheses,* $P_i = 0.206$ *and the corresponding attainable* p*-values set* $\mathbb{P}_i = \{0.0079,\ 0.206,\ 1\}$*. The table 4.1 shows the critical values for the proposed generalized step-up procedure and Döhler, Durand and Roquain (2018) procedure.*

In table 4.1, The column Distribution 1 represents the critical value sequence of first 5 hypotheses where $X_i \sim Hypergeometric\,(10,\ 5,\ 4)$*, while the column Distribution 2 for the hypotheses where* $X_i \sim Hypergeometric\,(10,\ 5,\ 5)$*; Döhler, Durand and Roquain (2018) procedure is denoted by* $\boldsymbol{SU}_{(\tau)}$*. From table 4.1, we can see, except for* $\alpha_1^{(i)}$*,* $i = 6, ..., 10$*, all the other critical values are larger than those of Döhler, Durand and Roquain (2018) procedure, which means our procedure has a great potential more powerful than it.*

Example 4.6.2 *Under the same setting of example 4.6.1, By Gilbert procedure (2005),* $K(\alpha) = min\left\{1 \leq k \leq 10 : \sum_{k=1}^{10} I\left\{p_i^* \leq \frac{\alpha}{k}\right\} \leq k\right\}$ *gives* $K(\alpha) = 5$*. Consequently,* $R_k = \left\{H_i : p_i^* \leq \frac{\alpha}{5} = 0.01\right\}$ *returns* $R_k = \{H_6, ..., H_{10}\}$ *with* P*-value: 0.206 from which we can see the hypotheses with smaller* P*-value 0.046 don't even have the chance to be tested.*

Table 4.1 Comparison Between the Generalized Step-up Procedure and Döhler, Durand and Roquain (2018) Procedure

α	Distribution 1	Distribution 2	$\mathbf{SU}_{(\tau)}$
α_1	0.0476	0.0068	0.0079
α_2	0.0476	0.068	0.0079
α_3	0.0476	0.068	0.0079
α_4	0.0476	0.068	0.0079
α_5	0.0476	0.068	0.0079
α_6	0.0476	0.068	0.0476
α_7	0.0476	0.068	0.0476
α_8	0.068	0.068	0.0476
α_9	0.068	0.068	0.0476
α_{10}	0.068	0.068	0.0476

4.7 Discussion and Future Work

In this chapter, we have introduced a new generalized step-up procedure by taking the special properties of discrete data which are discreteness and heterogeneity into account. The proposed procedure is theoretically more powerful than BH procedure and proved to control the FDR under independence structure. Compared to the existing FDR controlling procedures for discrete data, our procedure fully utilizes the information of marginal distribution of p-values that each hypothesis is tested with a different set of critical constants. Such procedure is especially suitable for discrete data since under the true null hypotheses, the discrete testing statistics will have their specific distributions. In the future work, we aim to develop the corresponding generalized step-up procedures under various dependent structures of the p-values under true null hypotheses. Also the simulation studies should be given as future work.

The hypotheses with a known graphical structure often arise in many large scale multiple testing applications, such as clinical safety analysis and gene ontology in gene expression data. Exploiting this graphical structure in multiple testing procedures can improve power as well as aid in interpretation. Many MTPs were developed for the hierarchical structured hypotheses with continuous data, see Yekutieli (2008), Barber and Ramdas (2016), Guo, Lynch and Romano (2018) and Bogomolov et al. (2018). And few studies are related to structured hypotheses with discrete data, for example, adverse events in clinical safety study, see Mehrotra and Heyse (2004), Berry and Berry (2004) and Mehrotra and Adewale (2012). Thus, a hierarchical procedure based on the generalized step-up procedure for discrete data will be meaningful to be developed in the future work.

CHAPTER 5

SUMMARY AND FUTURE WORK

In this dissertation, first, we mainly focus on investigating two weighted Holm procedures, the WHP and the WAP, in Chapter 2. We have studied the basic statistical properties and the optimality property of both procedures, such as finding the underlying CTPs, visualizing both procedures by graphs, providing the adjusted p-values. We have shown the WHP is uniformly more powerful than the WAP by theoretical and numerical methods. Also, we have an interesting observation which is the WAP does not belong to the class of CTPs with weighted Bonferroni-type local tests and it's also not monotone in terms of the p-values. However, the WAP is a consonant procedure although it does not satisfy the monotonicity condition. And this finding can provide an inspiration to develop a more general graphical approach. In this chapter, we also discussed how the weights affect the performance of both procedures by simulation method.

Next, in Chapter 3, we have studied similarities and differences between two independently developed graphical approaches. Both of two graphical approaches are used to visualize Bonferroni-type sequentially rejective procedures. We have theoretically showed that for testing three hypotheses, two graphical approaches are equivalent when the graphs are complete. However, if the graphs are incomplete, it is not necessarily to find an equivalent corresponding one. Moreover, when we test more than three hypotheses, they are not equivalent, this is due to the degree of freedom of the default graph is larger than the original graph which means the default graph is more flexible. We also provide a direct proof of the FWER control for the original graphical approach. For the families of hypotheses testing problems, it is convenient to use default graph; however, we should be able to solve this problem by adding edges with infinitesimally small transition coefficients

between two families or modify the updating algorithm when using original approach. We are working on providing general guidelines on this topic.

Finally, in Chapter 4, we have introduced a new generalized step-up procedure for discrete data by taking discreteness and heterogeneity properties into account. To the best of our knowledge, the proposed procedure is the first procedure for discrete data that is theoretically more powerful than BH procedure and proved to control the FDR under independence structure. The proposed procedure provides a specific sequence of critical values for each hypothesis, which is especially suitable for discrete data since the discrete testing statistics may have different distributions under true null hypotheses.

Some future works for the chapters are summarized as follows. For Chapters 2 and 3, we plan to develop a new general graphical approach for clinical trial studies with familywise error rate control, including the original graphical approach, the WHP, and the WAP as special cases. Moreover, we are planning to provide general guidelines on modifying the graph or the updating algorithm of the original graphical approach in order to make it to be a more general approach which can be applied to more multiple testing problems, such as serial gatekeeping problems. For Chapter 4, we want to develop the corresponding generalized step-up procedures under various dependent structures of the p-values under true null hypotheses. Also the simulation studies should be given as future work. A hierarchical procedure based on the generalized step-up procedure for discrete data will be meaningful to be developed in the future work.

APPENDIX A

ON WEIGHTED HOLM PROCEDURES

This appendix shows the proofs of the theorems and propositions for the statistical properties and optimal property of two weighted Holm procedures.

A.1 Proof of Proposition 2.3.1

Given an index set I for which $P_{(i)} = \min\{P_l,\, l \in I\}$ where i denote the index of the ith smallest p-value. Let $I^+_{(i)}$ denote the index set corresponding to the $n-i+1$ largest p-values $\{P_{(i)},\, P_{(i+1)},\, ...,\, P_{(m)}\}$. Thus, $I \subseteq I^+_{(i)}$ for any I including $P_{(i)}$ as the smallest p-value in the index set.

Define ϕ is an indicate function of the local test, eg, for testing any intersection hypothesis H_I, if reject, $\phi_I = 1$.

First, to show $\phi_{I^+_{(i)}} = 1$ implies $\phi_I = 1$. The proof is provided as follows.

If $\phi_{I^+_{(i)}} = 1$, then based on local test of CTP for WAP, we have

$$P_{(i)} \leq \frac{w_{(i)}}{\sum_{k=i}^{m} w_{(k)}}\alpha.$$

In addition, as $P_{(i)}$ is the smallest p-value with index in the index set I, we have $\sum_{k\in I} w_{(k)} \leq \sum_{k=i}^{m} w_{(k)}$.

Thus, $P_{(i)} = \min\{P_l,\, l \in I\} \leq \frac{w_{(i)}}{\sum_{k\in I} w_{(k)}}\alpha$, so $\phi_{I^+_{(i)}} = 1$ implies $\phi_I = 1$.

Second, given the first observation, we can know that $H_{(i)}$ is rejected if and only if

$$\phi_{I^+_{(j)}} = 1, \quad j = 1, ..., i,$$

which is equivalent to

$$P_{(j)} \leq \frac{w_{(j)}}{\sum_{k=i}^{m} w_{(k)}}\alpha, \quad j = 1, ..., i.$$

The last inequality is just the decision rule of the WAP for rejecting $H_{(i)}$, $i = 1, ..., m$; thus, the WAP and its corresponding CTP are equivalent. □

The above arguments can also be used to prove the equivalence between the WHP and its corresponding CTP, which is different with the proof provided by Westfall and Krishen (2001).

A.2 Proof of Proposition 2.3.3

Let the index set $M = \{1, ..., m\}$. According to Definition 2.3.3, we need to show that if an intersection hypothesis H_I, $I \subseteq M$ is rejected in the multiple setting, at least one H_i implied by H_I will be rejected, i.e., H_i with $i \in I$.

(i) If $I = M$ and H_I is rejected, consequently, the global intersection hypothesis $\cap_{i \in M} H_i$ is rejected by its local test $P_{(1)} \leq \frac{w_{(1)}}{\sum_{i \in M} w_i}\alpha$. Then, according to the local tests of the WAP, all the intersection hypotheses H_J, $J \subset M$ including $H_{(1)}$ will be rejected. Therefore, by closure principle, $H_{(1)}$ will be rejected.

(ii) If $I \subset M$, and H_I is rejected, then we have

$$P^I_{(1)} \leq \frac{w^I_{(1)}}{\sum_{i \in I} w_i}\alpha \quad and \quad P^S_{(1)} \leq \frac{w^S_{(1)}}{\sum_{i \in S} w_i}\alpha \quad \forall S,\, I \subset S \subseteq M.$$

where $P^I_{(1)}$ and $w^I_{(1)}$ are the smallest p-value in I and its corresponding weight, let H_j, P_j and w_j denote the associated hypothesis, its p-value and weight, respectively. And let $P^S_{(1)}$ and $w^S_{(1)}$ are the smallest p-value in S and its corresponding weight. Then, $\forall\, J$, where $j \in J \subset I \subseteq M$, we have

$$P^J_{(1)} \leq \frac{w^J_{(1)}}{\sum_{i \in J} w_i}\alpha \quad \forall J \subset I \subseteq M.$$

Next, we need to show that $\forall S, J$, where $j \in J \not\subset I, J \subseteq S$, the intersection hyptheses H_J will be rejected.

Let $I^c = M \backslash I$ and $P^{I^c}_{(1)} \leq ... \leq P^{I^c}_{(|I^c|)}$ be the ordered version of the p-values in the index set I^c with the corresponding weights $w^{I^c}_{(1)}, ..., w^{I^c}_{(|I^c|)}$ and hypotheses $H^{I^c}_{(1)}, ..., H^{I^c}_{(|I^c|)}$, where $|I^c|$ is the cardinality of I^c. Consider all intersection hypotheses H_S, including both

H_j and $H_{(1)}^{I^c}$ first, where $I \subset S$, since such intersection hypotheses are claimed significant, we have

$$P_j \leq \frac{w_j}{\sum_{i \in S} w_i}\alpha \text{ or } P_{(1)}^{I^c} \leq \frac{w_{(1)}^{I^c}}{\sum_{i \in S} w_i}\alpha.$$

Therefore, by the similar method as above, all the intersection hypotheses H_J including both H_j and $H_{(1)}^{I^c}$, $J \subseteq S$, are significant according to the following facts,

$$P_j \leq \frac{w_j}{\sum_{i \in J} w_i}\alpha \text{ or } P_{(1)}^{I^c} \leq \frac{w_{(1)}^{I^c}}{\sum_{i \in J} w_i}\alpha.$$

Consequently, we can continue testing the intersection hypotheses H_J including both H_j and $H_{(i)}^{I^c}$, $i = 2, ..., |I^c|$ gradually by the similar method as above.

Therefore, all the intersection hypotheses including H_j and itself are rejected, then by closure principle, H_j is rejected. Thus, the WAP is a consonant procedure. □

A.3 Proof of Proposition 2.4.1

In the following part, we will show that the above graphical representations are equivalent to the corresponding procedures, the WHP or the WAP, based on the initial graph and algorithms. Since the only difference between the graphical representations of the WHP and the WAP is the ordering rule: taking the argument of the minimum of weighted p-values or original p-values in step 1 of algorithm, so we only need to discuss one proof of them here, for example, to show equivalence between WHP and its graphical representation.

Let $I = \{1, ..., m\}$ denote the whole index set. Let j_i, $i = 1, ..., m$ denote the indices of hypotheses H_{j_i} that are sequentially tested based on the algorithm, then let $I_i = I \backslash \{j_1, ..., j_i\}$ denote the index set of the remaining hypotheses after rejecting ith hypothesis. Suppose $I_0 = I$.

(i) First, for each step of graphical approach, to show a weighted Bonferroni procedure will be applied to the subset I_i, $i = 0, 1, ..., m-1$. Eg, at $i+1$ th step, $\alpha_j = \frac{w_j}{\sum_{i \in I_i} w_i}\alpha$, $j \in I_i$, $i = 0, 1, ..., m-1$. Here, the induction method will be used.

At step 1 when $i = 0$, we start from the minimum weighted p-value P_{j_1}, if $P_{j_1} \leq \alpha_{j_1} = \frac{w_{j_1}}{\sum_{i\in I} w_i}\alpha$ based on the selection rule at step 1 of algorithm and initial local significance level, reject H_{j_1}; otherwise stop.

At step 2 when $i = 1$, H_{j_1} is rejected, then $I_1 = I\backslash\{j_1\}$. According to the initial graph and updating rule, for any $l, k \in I_1$, $l \neq k$,

$$\begin{aligned}\alpha_l &\leftarrow \alpha_l + \alpha_{j_1} g_{j_1 l} \\ &= \frac{w_l \alpha}{\sum_{i\in I} w_i} + \frac{w_{j_1}\alpha}{\sum_{i\in I} w_i}\frac{w_l}{\sum_{i\in I\backslash\{j_1\}} w_i} \\ &= \frac{w_l}{\sum_{i\in I\backslash\{j_1\}} w_i}\alpha = \frac{w_l}{\sum_{i\in I_1} w_i}\alpha\end{aligned} \tag{A.1}$$

$$\begin{aligned}g_{lk} &\leftarrow \frac{g_{lk} + g_{lj_1} g_{j_1 k}}{1 - g_{lj_1} g_{j_1 l}} \\ &= \frac{\frac{w_k}{\sum_{i\in I\backslash\{l\}} w_i} + \frac{w_{j_1}}{\sum_{i\in I\backslash\{l\}} w_i}\frac{w_k}{\sum_{i\in I\backslash\{j_1\}} w_i}}{1 - \frac{w_{j_1}}{\sum_{i\in I\backslash\{l\}} w_i}\frac{w_l}{\sum_{i\in I\backslash\{j_1\}} w_i}} \\ &= \frac{w_k}{\sum_{i\in I_1\backslash\{l\}} w_i}\end{aligned} \tag{A.2}$$

Based on the above updated local significance levels, we can see, at step 2, a weighted Bonferroni procedure will be applied to the subset of hypotheses with index set $I_1 = I\backslash\{j_1\}$.

Next, by induction method, assume at step i, after rejecting $\{H_{j_1}, ..., H_{j_{i-1}}\}$ the results hold, which means for any $l, k \in I_{i-1}$, $l \neq k$, $\alpha_l = \frac{w_l}{\sum_{i\in I_{j-1}} w_i}\alpha$ and $g_{lk} = \frac{w_k}{\sum_{i\in I_{i-1}\backslash\{l\}} w_i}$. Then we only need to show, at step $i+1$, the results still hold.

Based on (6) and (7) and the assumption of ith step, it is easy to find, for any $l, k \in I_i$, $l \neq k$,

$$\alpha_l \leftarrow \alpha_l + \alpha_{j_i} g_{j_i l} = \frac{w_l}{\sum_{i\in I_{i-1}\backslash\{j_i\}} w_i}\alpha = \frac{w_l}{\sum_{i\in I_i} w_i}\alpha,$$

$$g_{lk} \leftarrow \frac{g_{lk} + g_{lj_i} g_{j_i k}}{1 - g_{lj_i} g_{j_i l}} = \frac{w_k}{\sum_{i\in\{I_{i-1}\backslash\{j_i\}\}\backslash\{l\}} w_i} = \frac{w_k}{\sum_{i\in I_i\backslash\{l\}} w_i}.$$

Besides, considering the selection rule at step 1 of algorithm, we have $\frac{P_{j_1}}{w_{j_1}} \leq \frac{P_{j_2}}{w_{j_2}} \leq \cdots \leq \frac{P_{j_m}}{w_{j_m}}$.

Thus, we can find the graphical approach will reject H_{j_l} if

$$P_{j_i} \leq \alpha_{j_i} = \frac{w_{j_i}}{\sum_{k \in I_{i-1}} w_k} \alpha \Rightarrow \frac{P_{j_i}}{w_{j_i}} \leq \frac{\alpha}{\sum_{k \in I_{i-1}} w_k}, \quad i = 1, ..., l,$$

which is equivalent to

$$\tilde{P}_{(i)} \leq \frac{\alpha}{\sum_{k=i}^{m} w^*_{(k)}}, \quad i = 1, ..., l.$$

Thus, the WHP is equivalent to its graphical representation.

The above arguments can also be used for the graphical representation of the WAP. Compared to the graphical representation of the WHP, for the WAP, the only difference is the selection rule, accordingly, $P_{j_i} = P_{(i)}$, $i = 1, ..., m$.

Then, to reject H_{j_l} which is $H_{(l)}$ if

$$P_{j_i} \leq \alpha_{j_i} = \frac{w_{j_i}}{\sum_{k \in I_{i-1}} w_k} \alpha \quad i = 1, ..., l,$$

which is equivalent to

$$P_{(i)} \leq \frac{w_{(i)}}{\sum_{k=i}^{m} w_{(k)}} \alpha, \quad i = 1, ..., l.$$

Thus, the result follows. □

A.4 Proof of Proposition 2.5.1

For any $i \in \{1, ..., m\}$, we can find a j, such that $\tilde{P}_{(j)} = \frac{P_{(i)}}{w_{(i)}}$, then we have two possible results: (1) $j \geq i$ and (2) $j < i$. Under both cases (1) and (2) we need to show $\tilde{P}^{adj}_{(j)} \leq P^{adj}_{(i)}$, which means we need to show $\tilde{P}_{(r)} \sum_{k=r}^{m} w^*_{(k)} \leq P^{adj}_{(i)}, \quad r = 1, ..., j$.

If $\tilde{P}^{adj}_{(1)} = 1$, then because $\frac{P_{(1)}}{w_{(1)}} \sum_{k=1}^{m} w_{(k)} \geq \tilde{P}_{(1)} \sum_{k=1}^{m} w^*_{(k)}$ we have $P^{adj}_{(1)} = 1$. In addition, by the monotonicity increasing of adjusted p-values and adjusted weighted p-values, all of the values are equal to 1; thus, the result follows.

If $\tilde{P}^{adj}_{(1)} < 1$, the proof is shown as follows.

(1) We start from first case $j \geq i$.

For any $r \in \{1, ..., j\}$, we can find a s_r which satisfies

$$s_r = \min \left\{ l : \tilde{P}_{(r)} \leq \frac{P_{(l)}}{w_{(l)}} \leq \tilde{P}_{(m)} \right\}. \tag{A.3}$$

From condition (A.3), we can see $\tilde{P}_{(r)} \leq \tilde{P}_{(j)} = \frac{P_{(i)}}{w_{(i)}} \leq \tilde{P}_{(m)}$ and s_r takes the minimum index of $\frac{P_{(l)}}{w_{(l)}}$, therefore, we have $s_r \leq i$.

Then, if there are $m - r + 1$ of $l's$ that satisfy the condition (A.3), and all such l's satisfy $l \geq r$; thus, $s_r = r$ and the two hypotheses subsets, $\left\{ H^*_{(r)}, ..., H^*_{(m)} \right\}$ and $\left\{ H_{(s_r)}, ..., H_{(m)} \right\}$, have the same components, therefore we have $\sum_{k=r}^{m} w^*_{(k)} = \sum_{k=s_r}^{m} w_{(k)}$, $r = 1, ..., j$.

If some of the l's, such that $l < r$, considering the condition (A.3), we have $\left\{ H^*_{(r)}, ..., H^*_{(m)} \right\} \subset \left\{ H_{(s_r)}, ..., H_{(m)} \right\}$, therefore $\sum_{k=r}^{m} w^*_{(k)} < \sum_{k=s_r}^{m} w_{(k)}$, $r = 1, ..., j$.

Then we have

$$\tilde{P}_{(r)} \leq \frac{P_{(s_r)}}{w_{(s_r)}} \quad and \quad \sum_{k=r}^{m} w^*_{(k)} \leq \sum_{k=s_r}^{m} w_{(k)},\ r = 1, ..., j;$$

thus,

$$\tilde{P}_{(r)} \sum_{k=r}^{m} w^*_{(k)} \leq \frac{P_{(s_r)}}{w_{(s_r)}} \sum_{k=s_r}^{m} w_{(k)},$$

and

$$P^{adj}_{(s_r)} \leq P^{adj}_{(i)} \Rightarrow \tilde{P}_{(r)} \sum_{k=r}^{m} w^*_{(k)} \leq P^{adj}_{(i)}, \quad r = 1, ..., j.$$

Finally, we have $\tilde{P}^{adj}_{(j)} \leq P^{adj}_{(i)}$ when $j \geq i$.

(2) Second case when $j < i$, still for any $r \in \{1, ..., j\}$, we can find a s_r which satisfies $s_r = \min \left\{ l : \tilde{P}_{(r)} \leq \frac{P_{(l)}}{w_{(l)}} \leq \tilde{P}_{(m)} \right\}$, here, $1 \leq s_r \leq j < i$.

Then we have the same proof as first case, the only minor difference is

$$P^{adj}_{(s_r)} \leq P^{adj}_{(j)} \Rightarrow \tilde{P}_{(r)} \sum_{k=r}^{m} w^*_{(k)} \leq P^{adj}_{(j)}, \quad r = 1, ..., j.$$

Since $j < i$, consequently, $\tilde{P}_{(j)}^{adj} \leq P_{(j)}^{adj} \leq P_{(i)}^{adj}$.

Thus, the result follows. □

A.5 Proof of Theorem 2

To show WHP is an optimal procedure, we need to show the equality of $FWER \leq \alpha$ can be attained by finding a joint distribution for the p-values $\{P_1, ..., P_m\}$.

We will prove Theorem 2 under the least favorable configuration, where the false null p-values are all 0 with probability 1, and the true null p-values are the same with each following $U(0,1)$ distribution. This is because, under arbitrary dependence, the least favorable configuration leads to the largest error rate (Finner and Roters, 2001). Let m_0 denote the number of true null hypotheses, and for simplicity, let first m_0 hypotheses H_i, $i = 1, ..., m_0$ are true null hypotheses.

Consider the following joint distribution. Choose exactly one true null hypothesis H_i, $i \in \{1, ..., m_0\}$ with probability $\frac{w_i}{\sum_{k=1}^{m_0} w_k}$, and let $P_i = w_i U_1$, where U_1 is uniform on $\left(0, \frac{1}{\sum_{k=1}^{m_0} w_k}\right)$, then $P_i \sim U\left(0, \frac{w_i}{\sum_{k=1}^{m_0} w_k}\right)$ and $\tilde{P}_i = U_1$. Given H_i is not being selected, let $P_i = w_i U_2^i$, where U_2^i is independent of U_1 and $U_2^i \sim U\left(\frac{1}{\sum_{k=1}^{m_0} w_k}, \frac{1}{w_i}\right)$, then $P_i \sim U\left(\frac{w_i}{\sum_{k=1}^{m_0} w_k}, 1\right)$ and $\tilde{P}_i = U_2^i$. Then, unconditionally,

$$P_i \sim \frac{w_i}{\sum_{k=1}^{m_0} w_k} U\left(0, \frac{w_i}{\sum_{k=1}^{m_0} w_k}\right) + \left(1 - \frac{w_i}{\sum_{k=1}^{m_0} w_k}\right) U\left(\frac{w_i}{\sum_{k=1}^{m_0} w_k}, 1\right).$$

Indeed, if $u \leq \frac{w_i}{\sum_{k=1}^{m_0} w_k}$,

$$Pr\left(P_i \leq u\right) = \frac{w_i}{\sum_{k=1}^{m_0} w_k} \frac{u}{w_i / \sum_{k=1}^{m_0} w_k} = u$$

and if $u > \frac{w_i}{\sum_{k=1}^{m_0} w_k}$,

$$Pr(P_i \leq u) = \frac{w_i}{\sum_{k=1}^{m_0} w_k} \cdot 1 + \left(1 - \frac{w_i}{\sum_{k=1}^{m_0} w_k}\right) \frac{u - w_i / \sum_{k=1}^{m_0} w_k}{1 - w_i / \sum_{k=1}^{m_0} w_k} = u.$$

Thus, by construction, all the true null p-values follow uniform distribution from 0 to 1 and exactly one of constructed weighted p-values $\tilde{P}_i$ is less than or equal to $\frac{1}{\sum_{k=1}^{m_0} w_k}$. Therefore,

$$\begin{aligned} FWER &= Pr\left(V \geq 1\right) \\ &= Pr\left(\min_{i \in I_0} \tilde{P}_i \leq \frac{\alpha}{\sum_{k=1}^{m_0} w_k}\right) \\ &= Pr\left(U_1 \leq \frac{\alpha}{\sum_{k=1}^{m_0} w_k}\right) \\ &= \frac{\alpha / \sum_{k=1}^{m_0} w_k}{1 / \sum_{k=1}^{m_0} w_k} = \alpha, \end{aligned} \tag{A.4}$$

where U_1 is distributed as $U\left(0, \frac{1}{\sum_{k=1}^{m_0} w_k}\right)$ and the second equality is because of least favorable configuration. Thus, the WHP under arbitrary dependency is shown to be unimprovable without losing control of FWER. □

A.6 Proof of Proposition 2.6.1

Given p-values $\mathbf{q} = \{q_1, ..., q_m\}$ associated with m hypotheses $H_1, ..., H_m$ and pre-specified positive weights $w_1, ..., w_m$, respectively. A m-dimensional simplex:

$$Simp^m = \{\mathbf{t} = (t_1, ..., t_m) \in \mathbf{R}^m : 0 \leq t_1 \leq \cdots \leq t_m \leq 1\}. \tag{A.5}$$

Let $\tilde{q}_i = q_i / w_i$, $i = 1, ..., m$. Then, according to equation (A.5), we can always find a vector $\mathbf{t} = \{t_1, ..., t_m\} \in Simp^m$, such that the vector of the ordered version of $\tilde{q}_i$: $\tilde{\mathbf{q}} = \left\{\tilde{q}_{(1)}, ..., \tilde{q}_{(m)}\right\}$ is equal to $\mathbf{t}$, where $t_i = \tilde{q}_{(i)}$, $i = 1, ..., m$. Let $w^*_{(i)}$ and $H^*_{(i)}$ correspond to $\tilde{q}_{(i)}$ (also t_i).

Let $\mathcal{M}^w$ be a weighted monotone step-down procedure with $FWER \leq \alpha < 1$. Suppose the weighted procedure $\mathcal{M}^w$ finds the rth of $\mathbf{t}$, t_r, significant, then what is the upper bound of t_r?

Let $l = \sum_{k=r}^{m} w^*_{(k)}$, $\tau = \min\{t_r,\ 1/l\}$ and $\beta = l\tau$, so that $\beta \leq 1$.

Next, we define $m - r + 1$ random variables, corresponding to weighted p-values,

$$\tilde{P}_r,\ \tilde{P}_{r+1},\ ...,\ \tilde{P}_m \tag{A.6}$$

as follows. First, randomly choose an integer j equal to one of the numbers r, $r+1$, ..., m, 0 with probabilities $w^*_{(r)}\tau$, $w^*_{(r+1)}\tau$, ..., $w^*_{(m)}\tau$, $1-\beta$, respectively. Then generate, independently of each other (given j), the random numbers (A.6), if $j \neq 0$, $\tilde{P}_j$ is uniformly distributed in the interval $[0, \tau]$ while the other variables are distributed as $U\left[\tau, \frac{1}{w^*_{(i)}}\right]$, $i \in \{r, r+1, ..., m\} \setminus \{j\}$, where $w^*_{(i)}$ associated with the corresponding random variables $\tilde{P}_i$. If $j = 0$, all $\tilde{P}_i$ will follow $U[\tau, \frac{1}{w^*_{(i)}}]$, $i \in \{r, r+1, ..., m\}$.

Consequently, the original p-value corresponding to the weighted p-value $\tilde{P}_j$ will follow $U\left[0, w^*_{(j)}\tau\right]$, $j \in \{r, ..., m\}$, while all the other original p-values, corresponding to other $\tilde{P}_i$, $i \in \{r, ..., m\} \setminus \{j\}$, will follow $U[w^*_{(i)}\tau, 1]$. It follows the construction that the distribution of each of the original p-values corresponding to the random variables (A.6) are uniformly distributed in the interval [0, 1]. Therefore, these corresponding $m - r + 1$ hypotheses are true nulls.

Also, by construction, if $j \neq 0$, we have

$$\min_{r \leq i \leq m} \tilde{P}_i \leq \tau. \tag{A.7}$$

Then, let $\tilde{P}_1$, $\tilde{P}_2$, ..., $\tilde{P}_{r-1}$ be arbitrary random variables, such that

$$0 \leq \tilde{P}_i \leq t_i, \quad 1 \leq i \leq r-1. \tag{A.8}$$

Denote $\tilde{\mathbf{P}} = \left\{\tilde{P}_{(1)}, ..., \tilde{P}_{(m)}\right\}$ the nondecreasing rearrangement of the random vector $\left\{\tilde{P}_1, ..., \tilde{P}_m\right\}$. Then, inequalities (A.7), $\tau = \min\{t_r, 1/l\} \leq t_r$ and (A.8) imply that, with probability at least β,

$$\tilde{P}_{(i)} \leq t_i, \quad i = 1, 2, .., r. \tag{A.9}$$

Since the weighted procedure $\mathcal{M}^w$ finds the first r terms t_i of $\mathbf{t}$ significant, being a step-down procedure, $\mathcal{M}^w$ will still find them significant if the remaining terms $t_{r+1}, ..., t_m$ are replaced by 1's. Then, because $\mathcal{M}^w$ is monotone, the first r terms of the $\tilde{\mathbf{P}}$ will also be found significant, if inequalities (A.9) hold. However, based on construction, there are at

most $r-1$ hypotheses are false, so $\mathcal{M}^w$ makes at least one type I error with probability at least β. Therefore, we have $FWER \geq \beta$. By assumption, we have $FWER \leq \alpha < 1$, consequently, $\beta \leq \alpha$, which is $l\tau \leq \alpha$.

Thus, in particular, $\beta < 1$, we have $\tau = t_r$ and

$$t_r \leq \frac{\alpha}{\sum_{k=r}^{m} w^*_{(k)}}. \tag{A.10}$$

Because, given the p-values $\mathbf{q}$, the weighted procedure $\mathcal{M}^w$ finds all weighted p-values t_j, $j = 1, ..., r$, significant, it follows that inequalities

$$t_j \leq \frac{\alpha}{\sum_{k=j}^{m} w^*_{(k)}}, \quad j = 1, ..., r$$

all hold, which means that the hypothesis with rth weighted p-value t_r can be rejected by the WHP.

Thus, given p-values $\mathbf{q}$, if the rth weighted p-value is rejected by the weighted monotone step-down procedure $\mathcal{M}^w$, it is also rejected by the WHP. □

A.7 Proof of Proposition 2.6.2

Still under least favorable configuration, the same setting and joint distribution as proof of Theorem 2.

Suppose $w_j = \min\{w_1, ..., w_{m_0}\}$, which is the smallest weight among true null hypotheses. And we know $P_i \sim U\left(0, \frac{w_i}{\sum_{k=1}^{m_0} w_k}\right)$ once H_i is selected with probability $\frac{w_i}{\sum_{k=1}^{m_0} w_k}$, $i \in \{1, ..., m_0\}$ when we choose exactly one hypothesis. For other non-selected hypotheses, $P_i \sim U\left(\frac{w_i}{\sum_{k=1}^{m_0} w_k}, 1\right)$, $i \in I_0 \backslash \{j\}$. Let $\mathbb{P}_0 = \{P_1, ..., P_{m_0}\}$.

Obviously, as w_j is the smallest weights, so when H_j is selected with probability $\frac{w_j}{\sum_{k=1}^{m_0} w_k}$, we have

$$Pr\left(P_j \leq \frac{w_j}{\sum_{k=1}^{m_0} w_k}\alpha,\ P_j \leq P_i \in \mathbb{P}_0 \backslash P_j\right) = \frac{w_j}{\sum_{k=1}^{m_0} w_k}\alpha.$$

When H_i, $i \in I_0 \backslash \{j\}$ is selected,

$$Pr\left(P_i \leq \frac{w_i}{\sum_{k=1}^{m_0} w_k}\alpha,\ P_i \leq P_l \in \mathbb{P}_0 \backslash P_i\right) = \frac{w_i}{\sum_{k=1}^{m_0} w_k}\alpha,$$

where we only need $\frac{w_i}{\sum_{k=1}^{m_0} w_k}\alpha \leq \frac{w_j}{\sum_{k=1}^{m_0} w_k} \iff \frac{w_j}{w_i} \geq \alpha$, $i \in I_0 \backslash \{j\}$, which means the proportions of the smallest weight to any other weights of true null hypotheses are not less than α. Actually this constraint is easy to be satisfied for many applications.

Therefore, when $\frac{w_j}{w_i} \geq \alpha$, we have

$$FWER = \sum_{i=1}^{m_0} Pr\left(P_i \leq \frac{w_i}{\sum_{k=1}^{m_0} w_k}\alpha,\ P_i \leq P_l \in \mathbb{P}_0 \backslash P_i\right) = \alpha.$$

Thus, we can say if the weights of true null hypotheses satisfy the constraint $\frac{w_j}{w_i} \geq \alpha$, $w_j = \min\{w_1, ..., w_{m_0}\}$ and $i \in I_0 \backslash \{j\}$, the WAP is an optimal procedure. Thus, the result follows. □

APPENDIX B

ON GRAPHICAL APPROACHES

This appendix shows the proofs of the theorems and the propositions that we do not provide in Chapter 3.

B.1 Proof of Theorem 5

Suppose we simultaneously test m hypotheses, $m \geq 4$, the number of sub-default graphs is $n_{sub-default} = \binom{m}{2}$, and for each sub-default graph, we will have $(m-2)!2!$ sequences. Suppose the last two hypotheses of sequences are permutations of H_s and H_t, then any sequences, consisting of permutations of remaining hypotheses H_i, $i \in M \setminus \{s,t\}$ followed by permutations of H_s and H_t, will compose a sub-default graph. We can consider use δ_i for each sub-default graph, where $i = 1, ..., \binom{m}{2}$. Then we can let two arbitrary sequences of this sub-default graph be $H_{j_1} \to H_{j_2} \to ... \to H_{j_{m-2}} \to H_s \to H_t$ and $H_{j_1} \to H_{j_2} \to ... \to H_{j_{m-2}} \to H_t \to H_s$, respectively. Suppose $\mathcal{R}_0 = \emptyset$. The sequence critical values are denoted by

$$\alpha_{j_1} \prod_{k=1}^{m-2} g_{j_k j_{k+1}}(\mathcal{R}_{k-1}) \pm \delta$$

, where $\mathcal{R}_{k-1} = \left\{H_{j_1}, ..., H_{j_{k-1}}\right\}$.

For two sequences, if the order of first $m-2$ hypotheses are same, when $j_{m-1} = t$, $j_m = s$, $\alpha_{j_1} \prod_{k=1}^{m-2} g_{j_k j_{k+1}}(\mathcal{R}_{k-1}) + \delta$ is assigned, then when $j_{m-1} = s$, $j_m = t$, $\alpha_{j_1} \prod_{k=1}^{m-2} g_{j_k j_{k+1}}(\mathcal{R}_{k-1}) - \delta$ will be assigned. Moreover, in this sub-default graph, if the sequences with same order of last two hypotheses H_s and H_t, for example, $j_{m-1} = t$, we should also make sure they have same numbers of δ and $-\delta$, for each one, the number is $(m-2)!/2$. Then by adding corresponding sequences together when some hypotheses are rejected, positive or negative δ-values can be canceled. □

B.2 Proof of Proposition 3.4.1

In the following, we will use the mathematical induction to prove the uniqueness of the critical value function $\alpha_l(\mathcal{R})$ and transition coefficients function $g_{lk}(\mathcal{R})$ no mater how the rejection order is, where $\mathcal{R}$ is an arbitrary rejection set and l, k are the indices of hypotheses $H_l, H_k \in \mathcal{H} \setminus \mathcal{R}$.

When $i = 2$, $\mathcal{R}_1$ consists of one rejected hypothesis H_{j_1}, then according to Algorithm 1 of Bretz et al. (2009) we have

$$g_{lk}(\mathcal{R}_1) = \frac{g_{lk} + g_{lj_1} g_{j_1 k}}{1 - g_{lj_1} g_{j_1 l}}, \quad l, k \in I_2,\, l \neq k,$$

$$\alpha_k(\mathcal{R}_1) = \alpha_k + g_{j_1 k} \alpha_{j_1}, \quad k \in I_2.$$

Therefore, the transition coefficient function and critical value function are unique when the rejection set consists only one component.

Assume these two functions are unique on the rejection set consisting of $s - 1$ hypotheses, e.g., $i = s$, no mater how the rejection order of hypotheses in $\mathcal{R}_{s-1}$ is. That is, we have that the following equations are unique for any j, where j is the index of hypothesis $H_j \in \mathcal{R}_{s-1}$.

$$g_{lk}(\mathcal{R}_{s-1}) = \frac{g_{lk}(\mathcal{R}_{s-2}) + g_{lj}(\mathcal{R}_{s-2}) g_{jk}(\mathcal{R}_{s-2})}{1 - g_{lj}(\mathcal{R}_{s-2}) g_{jl}(\mathcal{R}_{s-2})}, \quad l, k \in I_s,\, l \neq k, \tag{B.1}$$

$$\alpha_k(\mathcal{R}_{s-1}) = \alpha_k(\mathcal{R}_{s-2}) + g_{jk}(\mathcal{R}_{s-2}) \alpha_j(\mathcal{R}_{s-2}), \quad k \in I_s, \tag{B.2}$$

where $\mathcal{R}_{s-2} = \mathcal{R}_{s-1} \setminus \{H_j\}$.

We need to prove the transition coefficient function and critical value function are still unique for $\mathcal{R}_s$ no mater how the rejection order is. Suppose H_t is rejected at step $i = s$, such that $\mathcal{R}_s = \mathcal{R}_{s-1} \cup \{H_t\}$, based on the updating rule we have

$$g_{lk}(\mathcal{R}_s) = \frac{g_{lk}(\mathcal{R}_{s-1}) + g_{lt}(\mathcal{R}_{s-1})g_{tk}(\mathcal{R}_{s-1})}{1 - g_{lt}(\mathcal{R}_{s-1})g_{tl}(\mathcal{R}_{s-1})}, \quad l, k \in I_{s+1}, l \neq k, \tag{B.3}$$

$$\alpha_k(\mathcal{R}_s) = \alpha_k(\mathcal{R}_{s-1}) + g_{tk}(\mathcal{R}_{s-1})\alpha_t(\mathcal{R}_{s-1}), \quad k \in I_{s+1}. \tag{B.4}$$

Then we need to show if the rejection order of H_t is exchanged with any hypothesis H_j that belongs to $\mathcal{R}_{s-1}$, the Equations (B.3) and (B.4) are still unique.

According to the assumption on $\mathcal{R}_{s-1}$, and let $\mathcal{R}'_{s-1} = \mathcal{R}_s \setminus \{H_j\}$, where $H_j \in \mathcal{R}_{s-1}$, the Equation (B.3) is equal to

$$\begin{aligned}
&\frac{\frac{g_{lk}(\mathcal{R}_{s-2})+g_{lj}(\mathcal{R}_{s-2})g_{jk}(\mathcal{R}_{s-2})}{1-g_{lj}(\mathcal{R}_{s-2})g_{jl}(\mathcal{R}_{s-2})} + \frac{g_{lt}(\mathcal{R}_{s-2})+g_{lj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}{1-g_{lj}(\mathcal{R}_{s-2})g_{jl}(\mathcal{R}_{s-2})} \frac{g_{tk}(\mathcal{R}_{s-2})+g_{tj}(\mathcal{R}_{s-2})g_{jk}(\mathcal{R}_{s-2})}{1-g_{tj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}}{1 - \frac{g_{lt}(\mathcal{R}_{s-2})+g_{lj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}{1-g_{lj}(\mathcal{R}_{s-2})g_{jl}(\mathcal{R}_{s-2})} \frac{g_{tl}(\mathcal{R}_{s-2})+g_{tj}(\mathcal{R}_{s-2})g_{jl}(\mathcal{R}_{s-2})}{1-g_{tj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}} \\
&= \frac{\frac{g_{lk}(\mathcal{R}_{s-2})+g_{lj}(\mathcal{R}_{s-2})g_{jk}(\mathcal{R}_{s-2})}{1-g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} + \frac{g_{lt}(\mathcal{R}_{s-2})+g_{lj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}{1-g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} \frac{g_{tk}(\mathcal{R}_{s-2})+g_{tj}(\mathcal{R}_{s-2})g_{jk}(\mathcal{R}_{s-2})}{1-g_{tj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}}{\frac{1-g_{lj}(\mathcal{R}_{s-2})g_{jl}(\mathcal{R}_{s-2})}{1-g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} - \frac{g_{lt}(\mathcal{R}_{s-2})+g_{lj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}{1-g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} \frac{g_{tl}(\mathcal{R}_{s-2})+g_{tj}(\mathcal{R}_{s-2})g_{jl}(\mathcal{R}_{s-2})}{1-g_{tj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}}.
\end{aligned} \tag{B.5}$$

The numerator of Equation (B.5) is equal to

$$\begin{aligned}
&\frac{g_{lk}(\mathcal{R}_{s-2}) + g_{lt}(\mathcal{R}_{s-2})g_{tk}(\mathcal{R}_{s-2})}{1 - g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} \\
&\quad + \frac{g_{lj}(\mathcal{R}_{s-2}) + g_{lt}(\mathcal{R}_{s-2})g_{tj}(\mathcal{R}_{s-2})}{1 - g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} \frac{g_{jk}(\mathcal{R}_{s-2}) + g_{jt}(\mathcal{R}_{s-2})g_{tk}(\mathcal{R}_{s-2})}{1 - g_{tj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})} \\
&= g_{lk}(\mathcal{R}'_{s-1}) + g_{lj}(\mathcal{R}'_{s-1})g_{jk}(\mathcal{R}'_{s-1}).
\end{aligned} \tag{B.6}$$

From above equation, based on assumption of the uniqueness of transition coefficient function on $\mathcal{R}_{s-1}$, we can see that if t in the indices of three fractional expressions on the left hand side of Equation (B.6) changes to be r which is any index of hypothesis that belongs to $\mathcal{R}'_{s-1} = \mathcal{R}_s \setminus \{H_j\}$, and the corresponding $\mathcal{R}_{s-2} = \mathcal{R}'_{s-1} \setminus \{H_r\}$, the value of left hand side of the equation stays same.

Similar as above, according to the assumption, the denominator of Equation (B.5) is equal to

$$\begin{aligned} & 1 - \frac{g_{lj}(\mathcal{R}_{s-2}) + g_{lt}(\mathcal{R}_{s-2})g_{tj}(\mathcal{R}_{s-2})}{1 - g_{lt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})} \frac{g_{jl}(\mathcal{R}_{s-2}) + g_{jt}(\mathcal{R}_{s-2})g_{tl}(\mathcal{R}_{s-2})}{1 - g_{jt}(\mathcal{R}_{s-2})g_{tj}(\mathcal{R}_{s-2})} \\ & = 1 - g_{lj}(\mathcal{R}'_{s-1})g_{jl}(\mathcal{R}'_{s-1}). \end{aligned} \tag{B.7}$$

Therefore, we have

$$\begin{aligned} g_{lk}(\mathcal{R}_s) &= \frac{g_{lk}(\mathcal{R}_{s-1}) + g_{lt}(\mathcal{R}_{s-1})g_{tk}(\mathcal{R}_{s-1})}{1 - g_{lt}(\mathcal{R}_{s-1})g_{tl}(\mathcal{R}_{s-1})}, \quad l, k \in I_{s+1}, l \neq k, \\ &= \frac{g_{lk}(\mathcal{R}'_{s-1}) + g_{lj}(\mathcal{R}'_{s-1})g_{jk}(\mathcal{R}'_{s-1})}{1 - g_{lj}(\mathcal{R}'_{s-1})g_{jl}(\mathcal{R}'_{s-1})}, \end{aligned} \tag{B.8}$$

where $\mathcal{R}'_{s-1} = \mathcal{R}_s \setminus \{H_j\}$, for any $H_j \in \mathcal{R}_{s-1}$. Thus, the transition coefficient function is unique for the arbitrary rejection set.

Based on the assumption of uniqueness of critical value function on $\mathcal{R}_{s-1}$, for any index j of hypothesis $H_j \in \mathcal{R}_{s-1}$, the Equation (B.4) is equal to

$$\begin{aligned} & \alpha_k(\mathcal{R}_s) = \alpha_k(\mathcal{R}_{s-1}) + g_{tk}(\mathcal{R}_{s-1})\alpha_t(\mathcal{R}_{s-1}), \quad k \in I_{s+1} \\ & = \alpha_k(\mathcal{R}_{s-2}) + g_{jk}(\mathcal{R}_{s-2})\alpha_j(\mathcal{R}_{s-2}) + \frac{g_{tk}(\mathcal{R}_{s-2}) + g_{tj}(\mathcal{R}_{s-2})g_{jk}(\mathcal{R}_{s-2})}{1 - g_{tj}(\mathcal{R}_{s-2})g_{jt}(\mathcal{R}_{s-2})}(\alpha_t(\mathcal{R}_{s-2}) \\ & \quad + g_{jt}(\mathcal{R}_{s-2})\alpha_j(\mathcal{R}_{s-2})) \\ & = \alpha_k(\mathcal{R}_{s-2}) + g_{tk}(\mathcal{R}_{s-2})\alpha_t(\mathcal{R}_{s-2}) + \frac{g_{jk}(\mathcal{R}_{s-2}) + g_{jt}(\mathcal{R}_{s-2})g_{tk}(\mathcal{R}_{s-2})}{1 - g_{jt}(\mathcal{R}_{s-2})g_{tj}(\mathcal{R}_{s-2})}(\alpha_j(\mathcal{R}_{s-2}) \\ & \quad + g_{tj}(\mathcal{R}_{s-2})\alpha_t(\mathcal{R}_{s-2})) \\ & = \alpha_k(\mathcal{R}'_{s-1}) + g_{jk}(\mathcal{R}'_{s-1})\alpha_j(\mathcal{R}'_{s-1}), \end{aligned} \tag{B.9}$$

where $\mathcal{R}'_{s-1} = \mathcal{R}_s \setminus \{H_j\}$, for any $H_j \in \mathcal{R}_{s-1}$. In the third equality of Equation (B.9), according to the assumption, if t is exchanged with any index r which is index of an hypothesis that belongs to $\mathcal{R}'_{s-1} = \mathcal{R}_s \setminus \{H_j\}$, and the corresponding $\mathcal{R}_{s-2} = \mathcal{R}'_{s-1} \setminus \{H_r\}$,

the equality still holds. Thus, the critical value function is unique for the arbitrary rejection set. □

B.3 Proof of Proposition 3.4.2

We have shown the uniqueness of the transition coefficient function and critical value function on any rejection set. Next, by mathematical induction, we want to show three regularity conditions holds for any rejection sets,

$$\sum_{k\in I_i} g_{lk}(\mathcal{R}_{i-1}) \leq 1 \tag{B.10}$$

and

$$0 \leq g_{lk}(\mathcal{R}_{i-1}) \leq 1, \quad g_{ll}(\mathcal{R}_{i-1}) = 0, \quad \forall l, k \in I_i,\ i = 1, ..., m, \tag{B.11}$$

$$\sum_{k\in I_i} \alpha_k(\mathcal{R}_{i-1}) \leq \alpha, \quad i = 1, ..., m. \tag{B.12}$$

When $i = 1$, by the assumption and initial construction, it is easy to have

$$\sum_{k\in I_1} g_{lk}(\mathcal{R}_0) = \sum_{k\in M} g_{lk} \leq 1,$$

$0 \leq g_{lk}(\mathcal{R}_0) = g_{lk} \leq 1$ and $g_{ll}(\mathcal{R}_0) = g_{ll} = 0$, $l, k \in I_1$, and $\sum_{k\in I_1} \alpha_l(\mathcal{R}_0) = \sum_{k\in M} \alpha_k \leq \alpha$.

Assume at step $i = s$, we have $\sum_{k\in I_s} g_{lk}(\mathcal{R}_{s-1}) \leq 1$, for all $l \in I_s$. Consequently, the conditions $0 \leq g_{lk}(\mathcal{R}_{s-1}) \leq 1$, $g_{ll}(\mathcal{R}_{s-1}) = 0$ hold for all $l, k \in I_s$. Moreover, assume $\sum_{k\in I_s} \alpha_k(\mathcal{R}_{s-1}) \leq \alpha$. And Then we need to show the conditions are also satisfied at step $i = s + 1$, which are

$$\sum_{k\in I_{s+1}} g_{lk}(\mathcal{R}_s) \leq 1, \quad \text{for all } l \in I_{s+1}, \tag{B.13}$$

$$0 \leq g_{lk}(\mathcal{R}_s) \leq 1, \quad g_{ll}(\mathcal{R}_s) = 0,\ l, k \in I_{s+1}, \tag{B.14}$$

$$\sum_{k \in I_{s+1}} \alpha_k(\mathcal{R}_s) \leq \alpha. \tag{B.15}$$

In the following proof, we will not consider the case $k = l$, because according to the updating rule, $g_{lk}(\mathcal{R}_{s-1}) = 0$ for all $s \in M$ in such case. Suppose H_{j_s} is rejected at step s, then we have

$$\begin{aligned}
\sum_{k \in I_{s+1}} g_{lk}(\mathcal{R}_s) &= \sum_{k \in I_{s+1}} \frac{g_{lk}(\mathcal{R}_{s-1}) + g_{lj_s}(\mathcal{R}_{s-1}) g_{j_s k}(R_{s-1})}{1 - g_{lj_s}(\mathcal{R}_{s-1}) g_{j_s l}(\mathcal{R}_{s-1})}, \quad l, k \in I_{s+1}, k \neq l \\
&\leq \frac{\left(1 - g_{lj_s}(\mathcal{R}_{s-1})\right) + g_{lj_s}(\mathcal{R}_{s-1}) \left(1 - g_{j_s l}(\mathcal{R}_{s-1})\right)}{1 - g_{lj_s}(\mathcal{R}_{s-1}) g_{j_s l}(\mathcal{R}_{s-1})} \\
&= 1.
\end{aligned} \tag{B.16}$$

The inequality of Equation (B.16) is due to three following conditions:

(1) According to the assumption $\sum_{k \in I_s} g_{lk}(\mathcal{R}_{s-1}) \leq 1$, for all $l \in I_s$ and H_{j_s} is rejected at step s, we have $\sum_{k \in I_{s+1},\, k \neq l} g_{lk}(\mathcal{R}_{s-1}) + g_{lj_s}(\mathcal{R}_{s-1}) \leq 1$, where $g_{ll}(\mathcal{R}_{s-1}) = 0$, $l \in I_s$.

(2) Given $l, k \in I_{s+1}, k \neq l$, we have $g_{lk}(\mathcal{R}_s) = \frac{g_{lk}(\mathcal{R}_{s-1}) + g_{lj_s}(\mathcal{R}_{s-1}) g_{j_s k}(R_{s-1})}{1 - g_{lj_s}(\mathcal{R}_{s-1}) g_{j_s l}(\mathcal{R}_{s-1})}$; otherwise, $g_{lk}(\mathcal{R}_s) = 0$. Therefore,

$$\sum_{k \in I_{s+1},\, k \neq l} g_{j_s k}(\mathcal{R}_{s-1}) + g_{j_s l}(\mathcal{R}_{s-1}) = \sum_{k \in I_s} g_{j_s k}(\mathcal{R}_{s-1}) \leq 1,$$

where $g_{j_s j_s}(\mathcal{R}_{s-1}) = 0$, $j_s \in I_s$.

(3) Based on assumption $0 \leq g_{lk}(\mathcal{R}_{s-1}) \leq 1$, $l, k \in I_s$ at step $i = s$, we can know that $1 - g_{j_s l}(\mathcal{R}_{s-1}) > 0$, $j_s, l \in I_s$.

Therefore, we have $\sum_{k\in I_{s+1},\, k\neq l} g_{lk}(\mathcal{R}_s) \leq 1$ and $g_{lk}(\mathcal{R}_s) = 0$, when $k = l$; consequently, we have

$$\sum_{k\in I_{s+1}} g_{lk}(\mathcal{R}_s) \leq 1$$

for all $l \in I_{s+1}$. Thus, we proved the regularity condition (B.10).

Then, the regularity condition (B.11) is achieved by proving the satisfaction of the regularity condition (B.10) and the updating rule for the case when $l = k$.

Finally, we want to prove the regularity condition (B.12),

$$\begin{aligned}\sum_{k\in I_{s+1}} \alpha_k(\mathcal{R}_s) &= \sum_{k\in I_{s+1}} \alpha_k(\mathcal{R}_{s-1}) + \sum_{k\in I_{s+1}} \alpha_{j_s}(\mathcal{R}_{s-1}) g_{j_s k}(\mathcal{R}_{s-1}) \\ &\leq \alpha - \alpha_{j_s}(\mathcal{R}_{s-1}) + \alpha_{j_s}(\mathcal{R}_{s-1}) \\ &= \alpha.\end{aligned} \tag{B.17}$$

The inequality of Equation (B.17) is due to the following two reasons: (1) Based on the rejection of hypothesis H_{j_s} at step $i = s$ and the assumption $\sum_{k\in I_s} \alpha_k(\mathcal{R}_{s-1}) \leq \alpha$, we have $\sum_{k\in I_{s+1}} \alpha_k(\mathcal{R}_{s-1}) + \alpha_{j_s}(\mathcal{R}_{s-1}) = \sum_{k\in I_s} \alpha_k(\mathcal{R}_{s-1}) \leq \alpha$. Consequently, $\sum_{k\in I_{s+1}} \alpha_k(\mathcal{R}_{s-1}) \leq \alpha - \alpha_{j_s}(\mathcal{R}_{s-1})$. (2) $\sum_{k\in I_{s+1}} g_{j_s k}(\mathcal{R}_{s-1}) = \sum_{k\in I_s} g_{j_s k}(\mathcal{R}_{s-1}) - g_{j_s j_s}(\mathcal{R}_{s-1}) \leq 1$, as $\sum_{k\in I_s} g_{j_s k}(\mathcal{R}_{s-1}) \leq 1$ and $g_{j_s j_s}(\mathcal{R}_{s-1}) = 0$ which are shown above.

Therefore, the regularity condition (B.12) is proved. □

B.4 Proof of Proposition 3.4.3

Given two rejection sets $\mathcal{R}$ and $\mathcal{S}$, such that $\mathcal{R} \subset \mathcal{S}$, we need to show $\alpha_k(\mathcal{R}) \leq \alpha_k(\mathcal{S})$ and $g_{lk}(\mathcal{R}) \leq g_{lk}(\mathcal{S})$, l, k are the indices of hypotheses $H_l, H_k \in \mathcal{H} \setminus \mathcal{S}$. The rejection set $\mathcal{R}$ consists of an arbitrary number of rejected hypotheses from 1 to $m - 1$.

Let $\mathcal{A}_i$, consisting of i rejected hypotheses, denote a mutually exclusive rejection set with the rejection set $\mathcal{R}$, and $\mathcal{S}_i = \mathcal{R} \cup \mathcal{A}_i$. Therefore, we have $\mathcal{R} \subset \mathcal{S}_i$ and there are i more rejected hypotheses in $\mathcal{S}_i$ more than the number of hypotheses in $\mathcal{R}$. Let J_i denote the index set of $\mathcal{H} \setminus \mathcal{S}_i$.

We use the mathematical induction to prove the monotone property of transition coefficient function and critical value function. When $i = 1$, $\mathcal{A}_1$ consists of one hypothesis, say H_j, where $H_j \in \mathcal{H} \setminus \mathcal{R}$. Then we have

$$\alpha_k(\mathcal{S}_1) = \alpha_k(\mathcal{R}) + g_{jk}(\mathcal{R})\alpha_j(\mathcal{R}) \geq \alpha_k(\mathcal{R}), \quad k \in J_1,$$

$$g_{lk}(\mathcal{S}_1) = \frac{g_{lk}(\mathcal{R}) + g_{lj}(\mathcal{R})g_{jk}(\mathcal{R})}{1 - g_{lj}(\mathcal{R})g_{jl}(\mathcal{R})} \geq g_{lk}(\mathcal{R}) + g_{lj}(\mathcal{R})g_{jk}(\mathcal{R}) \geq g_{lk}(\mathcal{R}), \quad k, l \in J_1,$$

the above equations are due to the regularity condition (B.11), the property of the critical values and the uniqueness of these two functions.

Assume when $i = r - 1$, $\mathcal{S}_{r-1} = \mathcal{R} \cup \mathcal{A}_{r-1}$, the following relations are satisfied,

$$\alpha_k(\mathcal{S}_{r-1}) \geq \alpha_k(\mathcal{R}), \quad k \in J_{r-1},$$

and

$$g_{lk}(\mathcal{S}_{r-1}) \geq g_{lk}(\mathcal{R}), \quad l, k \in J_{r-1}.$$

Then we need to show when $i = r$, the relations still hold for the rejection sets $\mathcal{S}_r$ and $\mathcal{R}$, where $\mathcal{R} \subset \mathcal{S}_r$. Suppose H_j, $j \in J_{r-1}$ is rejected. Therefore, we have

$$\begin{aligned} \alpha_k(\mathcal{S}_r) &= \alpha_k(\mathcal{S}_{r-1}) + g_{jk}(\mathcal{S}_{r-1})\alpha_j(\mathcal{S}_{r-1}), \quad k \in J_r \\ &\geq \alpha_k(\mathcal{R}) + g_{jk}(\mathcal{R})\alpha_j(\mathcal{R}) \\ &\geq \alpha_k(\mathcal{R}). \end{aligned} \tag{B.18}$$

The first inequality of Equation (B.18) is due to the assumption on S_{r-1}.

Moreover, we have

$$
\begin{aligned}
g_{lk}(\mathcal{S}_r) &= \frac{g_{lk}(\mathcal{S}_{r-1}) + g_{lj}(\mathcal{S}_{r-1})g_{jk}(\mathcal{S}_{r-1})}{1 - g_{lj}(\mathcal{S}_{r-1})g_{jl}(\mathcal{S}_{r-1})}, \quad , l, k \in J_r \\
&\geq g_{lk}(\mathcal{S}_{r-1}) + g_{lj}(\mathcal{S}_{r-1})g_{jk}(\mathcal{S}_{r-1}) \\
&\geq g_{lk}(\mathcal{R}),
\end{aligned} \tag{B.19}
$$

the first inequality of Equation (B.19) is due to the regularity condition (B.11), the second inequality is due to the assumption on $\mathcal{S}_{r-1}$ and the regularity condition (B.11).

Therefore, we have shown $\alpha_k(\mathcal{R}) \leq \alpha_k(\mathcal{S})$ and $g_{lk}(\mathcal{R}) \leq g_{lk}(\mathcal{S})$, l, k are the indices of hypotheses $H_l, H_k \in \mathcal{H} \setminus \mathcal{S}$, given $\mathcal{R} \subset \mathcal{S}$. □

B.5 Proof of Theorem 6

We use the contradiction method to prove that the rejection set by the original graphical approach is unique.

Suppose there exists two different final rejection sets $\mathcal{R}$ and $\mathcal{R}^{'}$, eg, $\mathcal{R} \neq \mathcal{R}^{'}$, such that we can find one hypothesis $H_{j_r} \in \mathcal{R}$, but $H_{j_r} \notin \mathcal{R}^{'}$, and suppose H_{j_r} is the first different rejected hypothesis at some step $i = r \in M = \{1, ..., m\}$, which means for both rejection sets, they have same rejected components at first $r-1$ steps, that is, $\mathcal{R}_{r-1} = \mathcal{R}^{'}_{r-1}$, let $\mathcal{R}_0 = \mathcal{R}^{'}_0 = \emptyset$. According to the uniqueness of the critical value function on any rejection set, we have

$$
\alpha_{j_r}(\mathcal{R}_{r-1}) = \alpha_{j_r}(\mathcal{R}^{'}_{r-1});
$$

thus, the event $P_{j_r} \leq \alpha_{j_r}(\mathcal{R}_{r-1})$ as $H_{j_r} \in \mathcal{R}$ is conflict with the event $P_{j_r} > \alpha_{j_r}(\mathcal{R}^{'}_{r-1})$ as $H_{j_r} \notin \mathcal{R}^{'}$. Therefore, the rejection set is unique. □

B.6 Proof of Theorem 7

From section 3.2, we have the ordered testing hypotheses $\mathcal{H} = (H_{j_1}, H_{j_2}, ..., H_{j_m})$ with associated p-values $\mathcal{P} = (P_{j_1}, ..., P_{j_m})$, and the final rejection sequence

$$\mathcal{R} = \{H_{j_i} \in \mathcal{H} : P_{j_i} \leq \alpha_{j_i}(\mathcal{R}_{i-1}), i = 1, ..., R\}, \tag{B.20}$$

where $R = \max\{1 \leq k \leq m : P_{j_i} \leq \alpha_{j_i}(\mathcal{R}_{i-1}), \forall i \in \{1, ..., k\}\}$.

Suppose some of p-values become smaller, others remains unchanged, then the corresponding rejection sequence is $\mathcal{S}$, and let $\mathcal{S}_i$ denote the rejection sequence at first i steps, $i = 1, ..., m$. Suppose $\mathcal{S}_0 = \emptyset$.

To show the original graphical approach is monotone in terms of p-values, we need to show $\mathcal{R} \subseteq \mathcal{S}$.

From Equation (B.20), it is easy to see H_{j_1} can always be rejected since $\alpha_{j_1}(\mathcal{S}_0) = \alpha_{j_1}$, then we can have $\mathcal{S}_1 = \mathcal{R}_1 = (H_{j_1})$. According to the uniqueness property of $\alpha_{j_i}(\mathcal{S}_{i-1})$, $i = 2, ..., R$, then H_{j_i} can be rejected sequentially no mater if the corresponding p-value becomes smaller or unchanged. Therefore, we have $\mathcal{R} \subseteq \mathcal{S}$. The result follows. □

APPENDIX C

A GENERALIZED STEP-UP FDR CONTROLLING PROCEDURE FOR DISCRETE DATA

This appendix shows the proof of the FDR control for the generalized step-up procedure for discrete data. The FDR is defined as follows.

$$\begin{aligned} FDR &= E\left[\frac{V}{R \vee 1}\right] \\ &= \sum_{i \in I_0} \sum_{r=1}^{m} \frac{1}{r} E\left(I\left\{P_i \leq \alpha_r^{(i)}\right\} I\left\{R = r\right\}\right) \\ &= \sum_{i \in I_0} \sum_{r=1}^{m} \frac{1}{r} Pr\left(R = r, P_i \leq \alpha_r^{(i)}\right) \\ &= \sum_{i \in I_0} \sum_{r=1}^{m} \frac{1}{r} Pr\left(R = r \mid P_i \leq \alpha_r^{(i)}\right) Pr\left(P_i \leq \alpha_r^{(i)}\right) \\ &= \sum_{i \in I_0} \sum_{r=1}^{m} \frac{F_i\left(\alpha_r^{(i)}\right)}{r} Pr\left(R = r \mid P_i \leq \alpha_r^{(i)}\right), \end{aligned} \tag{C.1}$$

where I_0 is the indices of true nulls and F_i is the distribution of the null p-values P_i.

Note that under positive dependence, we have

$$\begin{aligned} Pr\left(R = r \mid P_i \leq \alpha_r^{(i)}\right) &= Pr\left(R \geq r \mid P_i \leq \alpha_r^{(i)}\right) - Pr\left(R \geq r+1 \mid P_i \leq \alpha_r^{(i)}\right) \\ &\leq Pr\left(R \geq r \mid P_i \leq \alpha_r^{(i)}\right) - Pr\left(R \geq r+1 \mid P_i \leq \alpha_{r+1}^{(i)}\right). \end{aligned} \tag{C.2}$$

Then,

$$FDR \leq \sum_{i \in I_0} \sum_{r=1}^{m} \frac{F_i\left(\alpha_r^{(i)}\right)}{r} \left(Pr\left(R \geq r \mid P_i \leq \alpha_r^{(i)}\right) - Pr\left(R \geq r+1 \mid P_i \leq \alpha_{r+1}^{(i)}\right)\right),$$

(C.3)

which is equivalent to,

$$\sum_{i\in I_0}\sum_{r=1}^{m}\frac{F_i\left(\alpha_r^{(i)}\right)}{r}Pr\left(R\geq r\mid P_i\leq\alpha_r^{(i)}\right)-\sum_{i\in I_0}\sum_{r=2}^{m}\frac{F_i\left(\alpha_{r-1}^{(i)}\right)}{r-1}Pr\left(R\geq r\mid P_i\leq\alpha_r^{(i)}\right)$$
$$=\sum_{i\in I_0}\left[F_i\left(\alpha_1^{(i)}\right)+\sum_{r=2}^{m}\left[\frac{F_i\left(\alpha_r^{(i)}\right)}{r}-\frac{F_i\left(\alpha_{r-1}^{(i)}\right)}{r-1}\right]Pr\left(R\geq r\mid P_i\leq\alpha_r^{(i)}\right)\right]$$
$$=\sum_{i\in I_0}\left[F_i(\alpha_1^{(i)})+\sum_{r=2}^{m}\left[\frac{F_i(\alpha_r^{(i)})}{r}-\frac{F_i(\alpha_{r-1}^{(i)})}{r-1}\right]Pr(R^{(-i)}\geq r-1)\right]$$
$$\leq\sum_{i\in I_0}[F_i(\frac{j_1^{(i)}}{m}\tilde{\alpha})+\sum_{s=0}^{K^{(i)}-1}\sum_{r=j_s^{(i)}+2}^{j_{s+1}^{(i)}}\left[\frac{F_i(\alpha_r^{(i)})}{r}-\frac{F_i(\alpha_{r-1}^{(i)})}{r-1}\right]Pr(R^{(-i)}\geq j_{s+1}^{(i)})$$
$$+\sum_{s=1}^{K^{(i)}-1}\left[\frac{F_i(\frac{j_{s+1}^{(i)}}{m}\tilde{\alpha})}{j_s^{(i)}+1}-\frac{F_i(\frac{j_s^{(i)}}{m}\tilde{\alpha})}{j_s^{(i)}}\right]Pr(R^{(-i)}\geq j_s^{(i)})],$$

(C.4)

which is less than or equal to

$$\sum_{i\in I_0}\left[F_i(\frac{j_1^{(i)}}{m}\tilde{\alpha})+\sum_{s=1}^{K^{(i)}-1}\left[\frac{F_i(\frac{j_{s+1}^{(i)}}{m}\tilde{\alpha})}{j_s^{(i)}+1}-\frac{F_i(\frac{j_s^{(i)}}{m}\tilde{\alpha})}{j_{s-1}^{(i)}+1}\right]Pr(R^{(-i)}\geq j_s^{(i)})\right]$$
$$\leq\sum_{i=1}^{m}\left[F_i(\frac{j_1^{(i)}}{m}\tilde{\alpha})+\sum_{s=1}^{K^{(i)}-1}\left[\frac{F_i(\frac{j_{s+1}^{(i)}}{m}\tilde{\alpha})}{j_s^{(i)}+1}-\frac{F_i(\frac{j_s^{(i)}}{m}\tilde{\alpha})}{j_{s-1}^{(i)}+1}\right]I\left\{\frac{F_i(\frac{j_{s+1}^{(i)}}{m}\tilde{\alpha})}{j_s^{(i)}+1}>\frac{F_i(\frac{j_s^{(i)}}{m}\tilde{\alpha})}{j_{s-1}^{(i)}+1}\right\}\right]$$
$$=\tilde{D}(\tilde{\alpha})\leq D(\tilde{\alpha})\leq\alpha.$$

(C.5)

where $j_0^{(i)}=0$ and $j_{K^{(i)}}^{(i)}=m$.

(1)The second equality of Equation (C.4) is because $R^{(-i)}$ is independent of P_i under the independence assumption of p-values.

(2) The inequality of Equation (C.4) holds because of the following reasons: (i)Since when $j_s^{(i)}+1 \leq r \leq j_{s+1}^{(i)}$, $s=0,...,K^{(i)}-1$, $\left[\frac{F_i(\alpha_r^{(i)})}{r}-\frac{F_i(\alpha_{r-1}^{(i)})}{r-1}\right]<0$. (ii)$Pr(R^{(-i)} \geq r-1)$ is decreasing in r.

(3)In the above upper bound of the FDR, there are $K^{(i)}-1$ terms $\left[\frac{F_i(\frac{j_{s+1}^{(i)}}{m}\tilde{\alpha})}{j_s^{(i)}+1}-\frac{F_i(\frac{j_s^{(i)}}{m}\tilde{\alpha})}{j_{s-1}^{(i)}+1}\right]$, we only keep the positive terms and replace the following probability $Pr(R^{(-i)} \geq j_s^{(i)})$ with 1; if some terms are negative, just remove them.

(4) Finally, we can see that the last two inequalities hold is because of $D(u) = \max_\beta\left\{\tilde{D}(\beta):\beta \leq u\right\}\tilde{\alpha} = max\left\{u : D(u) \leq \alpha\right\}$. □

REFERENCES

[1] M. Alosh, F. Bretz and M. Huque. Advanced multiplicity adjustment methods in clinical trials. *Statistics in Medicine*, 33:693 - 713, 2014.

[2] ARDS Network. Ventilation with lower tidal volumes for acute lung injury and the acute respiratory distress syndrome. *New England Journal of Medicine*, 342:1301 - 1308, 2000.

[3] R. F. Barber and A. Ramdas. The p-filter: multi-layer FDR control for grouped hypotheses. *arXiv preprint arXiv:1512.03397v3*, 2016.

[4] S. M. Berry and D. A. Berry. Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model. *Biometrics*, 60:418 - 426, 2004.

[5] W. Brannath and F. Bretz. Shortcuts for locally consonant closed test procedures. *Journal of the American Statistical Association*, 105:660-669, 2010.

[6] Y. Benjamini and R. Cohen. Weighted false discovery rate controlling procedures for clinical trials. *Biostatistics*, 18:91 - 104, 2017.

[7] Y. Benjamini and R. Heller. False discovery rates for spatial signals. *Journal of the American Statistical Association*, 102:1272 - 1281, 2007.

[8] 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.

[9] Y. Benjamini and Y. Hochberg. Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, 24: 407418, 1997.

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

[11] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29: 1165 - 1188, 2001.

[12] 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.

[13] 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, 2011.

[14] F. Bretz, M. Posch, E. 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, 2011.

[15] X. Chen, R. W. Doerge and J. F. Heyse. Multiple testing with discrete data: Proportion of true null hypotheses and two adaptive FDR procedures. *Biometrical Journal*, 60:761 - 779, 2018.

[16] X. Chen, R. W. Doerge and S. K. Sarkar. A weighted FDR procedure under discrete and heterogeneous null distributions. *Biometrical Journal*, 1 - 20, 2020.

[17] C. Dalmasso, E. Génin and D. A. Trégouet. A weighted-Holm procedure accounting for allele frequencies in genomewide association studies. *Genetics Society of America*, 180:697-702, 2008.

[18] 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.

[19] A. Dmitrienko and A. C. Tamhane. Gatekeeping procedures with clinical trial applications. *Pharmaceutical Statistics*, 6:171 - 180, 2007.

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

[21] A. Dmitrienko, A. C. Tamhane and F. Bretz. Multiple testing problems in Pharmaceutical Statistics. *Boca Raton, FL: CRC Press*, 2009.

[22] S. Döhler. A discrete modification of the Benjamini-Yekutieli procedure. *Econometrics and Statistics*, 5:137 - 147, 2018.

[23] S. Döhler, G. Durand and E. Roquain. New FDR bounds for discrete and heterogeneous tests. *Electronic Journal of Statistics*, 12:1867 - 1900, 2018.

[24] C. W. Dunnett and A. C. Tamhane. Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine*, 10:939 - 947, 1991.

[25] J. A. Ferreira. The Benjamini-Hochberg method in the case of discrete test statistics. *The International Journal of Biostatistics*, 3:article 11, 2007.

[26] H. Finner and M. Roters. On the false discovery rate and expected type I errors. *Biometrical Journal*, 43:985 - 1005, 2001.

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

[28] K. R. Gabriel. Simultaneous test procedures-Some theory of multiple comparisons. *Annals of Mathematical Statistics*, 40:224 - 250, 1969.

[29] E. Ghulam, K. Wang and C. Xie. The Mixture Gatekeeping Procedure Based on Weighted Multiple Testing Correction for Correlated Tests. *Advanced Statistical Methods in Data Science*, 3 - 11, 2016.

[30] P. B. Gilbert. A modified false discovery rate multiple-comparisons procedure for discrete data, applied to human immunodeficiency virus genetics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54:143 - 158, 2005.

[31] A. Y. Gordon. A new optimality property of the Holm step-down procedure. *Statistical Methodology*, 8:129 - 135, 2011.

[32] A. Y. Gordon and P. Salzman. Optimality of the Holm procedure among general step-down multiple testing procedures. *Statistics and Probability Letters*, 78:1878 - 1884, 2008.

[33] O. J. M. Guilbaud. Note on a stepwise procedure for rejecting at least k out of m hypotheses: A simple Holm-type formulation and proof. *Biometrical Journal*, 1 - 7, 2021.

[34] O. J. M. Guilbaud. Simultaneous confidence intervals compatible with sequentially rejective graphical procedures. *Statistics in Biopharmaceutical Research*, 10:220 - 232, 2018.

[35] W. Guo, G. Lynch and J. P. Romano. A new approach for large scale multiple testing with application to FDR control for graphically structured hypotheses. *arXiv preprint arXiv:1812.00258v1*, 2018.

[36] L. He and J. F. Heyse. Improved power of familywise error rate procedures for discrete data under dependency. *Biometrical Journal*, 61:101 - 114, 2019.

[37] R. Heller and H. Gur. False discovery rate controlling procedures for discrete tests. *arXiv preprint arXiv:1112.4627*, 2011.

[38] K. S. Henning and P. H. Westfall. Closed testing in pharmaceutical research: historical and recent developments. *Statistics in Biopharmaceutical Research*, 7:126 - 147, 2015.

[39] J. F. Heyse. A false discovery rate procedure for categorical data. *Resent Advances in Biostatistics: False Discovery Rates, Survival Analysis, and Related Topics*, 43 - 58, 2011.

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

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

[42] G. Hommel and F. Krummenauer. Improvements and Modifications of Tarone's multiple test procedure for discrete data. *Biometrics*, 54:673 - 681, 1998.

[43] L. Jiang, A. Amir, J. T. Morton, R. Heller, E. Arias-Castro and R. Knight. Discrete false-discovery rate improves identification of differentially abundant microbes. *American Society for Microbiology Journals*, 2:e00092-17, 2017.

[44] G. Kang, K. Ye, N. Liu, D. B. Allison and G. Gao. Weighted multiple hypothesis testing procedures. *Statistical Applications in Genetics and Molecular Biology*, 8:article 23, 2009.

[45] A. C. Leon and M. Heo. A comparison of multiplicity adjustment strategies for correlated binary endpoints. *Journal of Biopharmaceutical Statistics*, 15: 839 - 855, 2005.

[46] J. Li and R. Tibshirani. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical Methods in Medical Research*, 22: 519 - 536, 2013.

[47] K. Liang. False discovery rate estimation for large-scale homogeneous discrete p-values. *Biometrics*, 72: 639 - 648, 2016.

[48] W. Liu. Multiple tests of a non-hierarchical finite family of hypotheses. *Journal of the Royal Statistical Society. Series B.*, 58:455 - 461, 1996.

[49] G. Lynch and W. Guo. On Procedures Controlling the FDR for Testing Hierarchically Ordered Hypotheses. *arXiv preprint arXiv:1612.04467v1*, 2016.

[50] W. Maurer and F. Bretz. A note on testing families of hypotheses using graphical procedures. *Statistics in Medicine*, 33:5340 - 5346, 2014.

[51] 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-in-Pharmazeutischen Industrie*, 6:3 - 18, 1995.

[52] R. Marcus, E. Peritz and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63:655 - 660, 1976.

[53] D. V. Mehrotra and A. J. Adewale. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. *Statistics in Medicine*, 31:1918 - 1930, 2012.

[54] D. V. Mehrotra and J. F. Heyse. Use of the false discovery rate for evaluating clinical safety data. *Statistical Methods in Medical Research*, 13:227 - 238, 2004.

[55] J. Mielke, B. Jones, M. Posch and F. König. Testing procedures for claiming success on at least k out of m hypotheses with an application to biosimilar development. *Statistics in Biopharmaceutical Research*, 13:106 - 112, 2021.

[56] S. Nandi and S. K. Sarkar. Adapting BH to one- and two-way classified structures of hypotheses. *arXiv preprint arXiv:1812.06551v2*, 2019.

[57] A. K. Ramdas, R. F. Barber, M. J. Wainwright and M. I. Jordan. A unified treatment of multiple testing with prior knowledge using the p-filter. *The Annals of Statistics*, 47:2790 - 2821, 2019.

[58] K. Roeder and L. Wasserman. Genome-Wide significance levels and weighted hypothesis testing. *Statistical Science*, 24:398 - 413, 2009.

[59] R. Rosenthal and D. Rubin. Ensemble-adjusted p values. *Psychological Bulletin*, 94:540 - 541, 1983.

[60] A. J. Roth. Multiple comparison procedures for discrete test statistics. *Journal of Statistical Planning and Inference*, 82:101 - 117, 1999.

[61] D. S. Robertson, J. M. S. Wason and F. Bretz. Graphical approaches for the control of generalized error rates. *Statistics in Medicine*, 39:3135 - 3155, 2020.

[62] S. K. Sarkar, Y. Fu and W. Guo. Improving Holm's procedure using pairwise dependencies. *Biometrika*, 103:237 - 243, 2016.

[63] R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 63:655 - 660, 1986.

[64] 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.

[65] A. C. Tamhane and L. Liu. On weighted Hochberg procedures. *Biometrika*, 95:279 - 294, 2008.

[66] R. E. Tarone. A modified bonferroni method for discrete data. *Biometrics*, 46:515 - 522, 1990.

[67] L. Wang. Weighted multiple testing procedure for grouped hypotheses with k-FWER control. *Computational Statistics*, 34:885 - 909, 2019.

[68] 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.

[69] P. H. Westfall, S. Kropf and L. Finos. Weighted FWE-controlling methods in high-dimensional situations. *Institute of Mathematical Statistics*, 47: 143 - 154, 2004.

[70] P. H. Westfall and S. Young. Resampling-based multiple testing: Examples and methods for p-value adjustment. *New York, NY: John Wiley & Sons*, 1993.

[71] B. L. Wiens. A fixed sequence Bonferroni proceudre for testing multiple endpoints. *Pharmaceutical Statistics*, 2:211 - 215, 2003.

[72] B. L. Wiens, A. Dmitrienko and O. Marchenko. Selection of hypothesis weights and ordering when testing multiple hypotheses in clinical trials. *Journal of Biopharmaceutical Statistics*, 23:1403 - 1419, 2014.

[73] D. Xi, E. Glimm, W. Maurer and F. Bretz. A unified framework for weighted parametric multiple test procedures. *Biometrical Journal*, 00:1 - 14, 2017.

[74] C. Xie. Weighted multiple testing correction for correlated tests. *Statistics in Medicine*, 31:341 - 352, 2012.

[75] C. Xie. Weighted multiple testing correction for correlated binary endpoints. *Communications in Statistics-Simulation and Computation*, 42:1693 - 1702, 2014.

[76] D. Yekutieli. Hierarchical false discovery ratecontrolling methodology. *Journal of the American Statistical Association*, 103:309 - 316, 2008.

[77] Y. Zhu and W. Guo. Family-Wise Error Rate Controlling Procedures for Discrete Data. *Statistics in Biopharmaceutical Research*, 12:117 - 128, 2020.

[78] T. Zhan, A. Hartford, J. Kang and W. Offen. Optimizing graphical procedures for multiplicity control in a confirmatory clinical trial via deep learning. *Statistics in Biopharmaceutical Research*, 14:92 - 102, 2022.

[79] H. Zhao and J. Zhang. Weighted p-value procedures for controlling FDR of grouped hypotheses. *Journal of Statistical Planning and Inference*, 151-152:90 - 106, 2014.