Electronic Journal of Statistics
Vol. 11 (2017) 4649–4673
ISSN: 1935-7524
DOI: 10.1214/17-EJS1359

The control of the false discovery rate in fixed sequence multiple testing

Gavin Lynch

Catchpoint Systems, Inc., New York, NY 10003, USA
e-mail: gavlynch@yahoo.com

Wenge Guo*

Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, NJ 07102, USA
e-mail: wenge.guo@njit.edu

Sanat K. Sarkar†

Department of Statistical Science, Temple University
Philadelphia, PA 19122, USA
e-mail: sanat@temple.edu

and

Helmut Finner‡

Institute for Biometrics and Epidemiology
German Diabetes Center (DDZ)
Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf
Auf'm Hennekamp 65, 40225 Düsseldorf, Germany
e-mail: finner@ddz.uni-duesseldorf.de

Abstract: Controlling false discovery rate (FDR) is a powerful approach to multiple testing. In many applications, the tested hypotheses have an inherent hierarchical structure. In this paper, we focus on the fixed sequence structure where the testing order of the hypotheses has been strictly specified in advance. We are motivated to study such a structure, since it is the most basic of hierarchical structure, yet it is often seen in real applications such as statistical process control and streaming data analysis. We first consider a conventional fixed sequence method that stops testing once an acceptance occurs, and develop such a method controlling FDR under both arbitrary and negative dependencies. The method under arbitrary dependency is shown to be unimprovable without losing control of FDR and, unlike existing FDR methods; it cannot be improved even by restricting to the usual positive regression dependence on subset (PRDS) condition.

*The research of Wenge Guo was supported by NSF Grants DMS-1006021 and DMS-1309162.

†The research of Sanat Sarkar was supported by NSF Grant DMS-1006344 and DMS-1309273.

‡The research of Helmut Finner was supported by the Ministry of Science and Research of the State of North Rhine-Westphalia (MIWF NRW) and the German Federal Ministry of Health (BMG).

To account for any potential mistakes in the ordering of the tests, we extend the conventional fixed sequence method to one that allows more but a given number of acceptances. Simulation studies show that the proposed procedures can be powerful alternatives to existing FDR controlling procedures. The proposed procedures are illustrated through a real data set from a microarray experiment.

AMS 2000 subject classifications: Primary 62J15.

Keywords and phrases: Arbitrary dependence, false discovery rate, fixed sequence, multiple testing, negative association, PRDS property, p-values.

Received November 2016.

1. Introduction

In many applications of multiple testing, such as genomic research, clinical trials, and statistical process control, the hypotheses are so structured that they are to be tested in a particular sequence. This structure may be a natural one, as in [14], where Gene Ontology imposes a directed acyclic graph structure onto the tested hypotheses, or can be formed by using a data-driven approach for specifying the testing order of the hypotheses, as in [11, 19, 23, 24, 37], etc. In some applications, it is not even possible to use the conventional p-value based multiple testing methods, because of some inherent structure among the tested hypotheses. For example, in sequential change detection problems with stream data [33] where the hypotheses have a temporal structure, traditional stepwise procedures requiring the availability of all p-values cannot be used since the decision concerning a hypothesis has to be made even before the data associated with the remaining hypotheses are observed.

Some progress has been made in testing structured hypotheses. However, it has been primarily focused on controlling the familywise error rate (FWER) [7, 18, 20, 28, 29, 32, 36, 38, 39, 40]. There are few results towards controlling the false discovery rate (FDR), defined as the expected proportion of false discoveries among all discoveries, while accounting for the structure of the tested hypotheses. [3, 17, 30] developed methods for testing hypotheses with a specific hierarchical structure where the structure is limited to only two levels. [41] discussed a method that controls the FDR when the tested hypotheses have a general hierarchical structure. However, that method is shown to control the FDR only under independence.

The primary objective of this paper is to help advance the theory and methods on controlling the FDR for testing structured hypotheses. We do so by focusing on a structure where the hypotheses have a fixed pre-defined testing order since this is the simplest of hierarchical structures, yet it is often seen in real applications such as clinical trials, statistical process control and streaming data analysis. We will refer to such a structure as a fixed sequence structure throughout this paper. Very recently, several methods have been introduced for controlling FDR while testing pre-ordered hypotheses. [9] developed a 'single-step' FDR controlling method for testing hypotheses with the same critical

value α, which tests each hypothesis at level α until a stopping condition is reached. [2, 12, 27] developed several different 'step-up' FDR controlling procedures in the context of high-dimensional regression for testing hypotheses with fixed sequence structure for which hypotheses are tested from highest-ranked to lowest-ranked, and [26] performed asymptotic power analysis for such 'step-up' procedures. In addition, [21] developed procedures for controlling the FDR in an online manner while testing a sequence of possibly infinite pre-ordered hypotheses.

In this paper, we develop 'step-down' FDR controlling methods that fully exploit the fixed sequence structural information, in which hypotheses are tested from lowest-ranked to highest-ranked. We first consider a conventional fixed sequence multiple testing method that keeps rejecting until an acceptance occurs and develop such a method controlling the FDR under arbitrary dependence. It is shown to be optimal in the sense that it cannot be improved by increasing even one of its critical values without losing control over the FDR, or even by imposing a positive dependence condition on the p-values, such as the standard PRDS (positive regression dependence on subset) condition of [6]. This is different from what happens in case of non-fixed sequence multiple testing. For instance, the so-called BY method of [6] that controls the FDR under arbitrary dependence can be improved significantly by the BH method of [4] by imposing this PRDS condition. Since our procedure cannot be improved under positive dependence, we consider the case of negative dependence and develop a more powerful conventional fixed sequence multiple testing method controlling the FDR under negative dependence which includes independence as a special case.

There is a potential for loss of power in a conventional fixed sequence multiple testing method if the ordering of the hypotheses, particularly for the earlier ones, does not match with that of their true effect sizes, potentially leading to some earlier hypothesis being accepted and the follow-up hypotheses having no chance to be tested. To mitigate that, we consider generalizing the conventional fixed sequence multiple testing to one that allows more than one but a prespecified number of acceptances, and develop such generalized fixed sequence multiple testing methods controlling the FDR under both arbitrary dependence and independence.

It is not always the case in real data applications that the hypotheses will have a natural fixed sequence structure or information about how to order them will be available a priori. Nevertheless, the data itself can often provide information on how to order the hypotheses. In this paper, we discuss such a data-driven ordering strategy which can be applied to a broad spectrum of multiple testing problems such as one-sample and two-sample t-tests, and one-sample and two-sample nonparametric tests. Through simulation studies and a real microarray data analysis, this strategy coupled with our proposed fixed sequence methods is seen to perform favorably against the corresponding non-fixed sequence methods under certain settings.

The paper is organized as follows. With some concepts and background information given in Section 2, we present the developments of our conventional and generalized fixed sequence procedures controlling the FDR under various

dependencies in Sections 3 and 4, respectively. Our fixed sequence procedures coupled with a data-driven ordering strategy for the hypotheses are presented and applied to a real microarray data in Section 5. The findings from some simulation studies on the performances of our procedures are given in Section 6. Some concluding remarks are made in Section 7 and proofs of some results are given in the Appendix A.

2. Preliminaries

Suppose that $H_i, i = 1, \ldots, m$, are the m null hypotheses that are ordered a priori and are to be simultaneously tested based on their respective p-values $P_i, i = 1, \ldots, m$. Let m_0 and m_1 of these null hypotheses be true and false, respectively. For notational convenience, we denote the index of the ith true null hypothesis by u_i and the set of indices of the true null hypotheses by I_0, which are both fixed but unknown. Let V and S be the numbers of true and false null hypotheses, respectively, among the R rejected null hypotheses in a multiple testing procedure. Then, the familywise error rate (FWER) and false discovery rate (FDR) of this procedure are defined respectively as $\text{FWER} = \Pr(V > 0)$ and $\text{FDR} = E\left(\frac{V}{\max(R,1)}\right) = E\left(\frac{V}{V+S} 1_{\{V>0\}}\right)$. We say that an error rate (FWER or FDR) is strongly controlled if for any configuration of true and false null hypotheses, the error rate does not exceed some pre-specified significance level, say α.

Typically, the hypotheses are ordered based on their p-values and multiple testing is carried out using a stepwise or single-step procedure. However, when these hypotheses are ordered a prior and not according to their p-values, multiple testing is often performed using a fixed sequence method. Given a non-decreasing sequence of critical constants $0 < \alpha_1 \leq \cdots \leq \alpha_m$, a conventional fixed sequence method is defined as follows:

Definition 1. (Conventional fixed sequence method)

1. *If $P_1 \leq \alpha_1$, then reject H_1 and continue to test H_2; otherwise, stop.*
2. *If $P_i \leq \alpha_i$ then reject H_i and continue to test H_{i+1}; otherwise, stop.*

Thus, a conventional fixed sequence method continues testing in the predetermined order as long as rejections occur. Once an acceptance occurs, it stops testing the remaining hypotheses. In Section 4, we will generalize a conventional fixed sequence method to allow a given number of acceptances. It should be noted that a conventional fixed sequence method with common critical constant α, which is often called the fixed sequence procedure in the literature, strongly controls the FWER at level α [29]. We will refer to it as the FWER fixed sequence procedure in this paper in order to distinguish it from other fixed sequence methods designed to control the FDR.

Regarding assumptions we make about the p-values in this paper, we assume that the true null p-values, which we denote for notational convenience by $\widehat{P}_i$,

for $i = 1, \ldots, m_0$, are marginally distributed as follows:

$$\Pr\left(\widehat{P}_i \leq p\right) \leq p \quad \text{for any } p \in (0,1). \tag{1}$$

One of several types of dependence, like arbitrary dependence, positive dependence, negative dependence, and independence, has been assumed to characterize a dependence structure among the p-values.

By arbitrary dependence, we mean that the p-values do not have any specific form of dependence. The positive dependence is characterized by the positive regression dependence on subset (PRDS) property [6] as defined below:

Definition 2. *(PRDS) The vector of p-values* $\vec{P}$ *is PRDS on the subset of null p-values* $\vec{P}_0 = (\widehat{P}_1, \ldots, \widehat{P}_{m_0})$ *if for every increasing set* D *and for each* $i = 1, \ldots, m_0$*, the conditional probability* $\Pr\left(\vec{P} \in D \;\middle|\; \widehat{P}_i = p\right)$ *is non-decreasing in* p*.*

Several multivariate distributions possess this property (see, for instance, [6, 34]).

The negative dependence condition we consider in this paper is characterized by the following property, to be referred to as the negative association with a subset (NAWS) property:

Definition 3. *(NAWS) The vector of p-values* $\vec{P}$ *is negatively associated with the subset of null p-values* $\vec{P}_0 = (\hat{P}_1, \ldots, \hat{P}_{m_0})$ *if for each* $i = 1, \ldots, m_0$*, the following inequality holds:*

$$\begin{aligned} &\Pr\left(P_j \leq p_j, j = 1, \ldots, m, \text{ with } j \neq i \middle| \widehat{P}_i \leq p_i\right) \\ \leq\ &\Pr\left(P_j \leq p_j, j = 1, \ldots, m, \text{ with } j \neq i\right) \end{aligned} \tag{2}$$

for all fixed p_j*'s and* p_i*.*

This is a weaker notion of negative dependence than the conventional definition of negative association [22] satisfied by several multivariate distributions, including multivariate normal with negative correlation, multinomial, dirichlet, and multivariate hypergeometric. It is easily seen that independence is a special case of the NAWS condition.

3. Conventional fixed sequence procedures

In this section, we present the developments of two simple conventional fixed sequence procedures controlling the FDR under both arbitrary dependence and negative dependence conditions on the p-values.

3.1. Procedure under arbitrary dependence

Since the FDR is more liberal than the FWER, a conventional fixed sequence method controlling the FDR under arbitrary dependence is expected to have

critical values that are at least as large as α, the common critical constant for the FWER fixed sequence method. In the following, we present such a simple conventional fixed sequence FDR controlling procedure.

Theorem 3.1. *Consider a conventional fixed sequence procedure with critical constants*

$$\alpha_i^{(1)} = \min\left(\frac{m\alpha}{m-i+1}, 1\right),\ i = 1, \ldots, m.$$

(i) This procedure strongly controls the FDR at level α under arbitrary dependence of the p-values.

(ii) One cannot increase even one of the critical constants $\alpha_i^{(1)}, i = 1, \ldots, m$, while keeping the remaining fixed without losing control of the FDR under some parameter configuration. This is true even when $\vec{P}$ is assumed to be PRDS on $\vec{P}_0$.

Proof. of (i). Our proof is similar to the arguments used by [5] for the FDR control of their step-down procedure. Let u_1 be the fixed but unknown index of the first true null hypothesis, then the first $u_1 - 1$ null hypotheses are all false. Note that the event $\{V > 0\}$ implies that $S \geq u_1 - 1$ and $\hat{P}_1 \leq \alpha_{u_1}^{(1)}$, and therefore we have

$$\begin{aligned}\text{FDR} &= E\left(\frac{V}{V+S}1_{\{V>0\}}\right) \leq E\left(\frac{m_0}{m_0+u_1-1}1_{\{V>0\}}\right)\\ &= \frac{m_0}{m_0+u_1-1}Pr(V>0) \leq \frac{m_0}{m_0+u_1-1}Pr(\hat{P}_1 \leq \alpha_{u_1}^{(1)})\\ &\leq \frac{m-u_1+1}{m}\alpha_{u_1}^{(1)} \leq \alpha.\end{aligned}$$

The first inequality follows from the fact that $V/(V+S)$ is an increasing function of V and a decreasing function of S. The third inequality follows from the fact that $m_0/(m_0+u_1-1)$ is an increasing function of m_0 and $m_0 \leq m-u_1+1$ since there are at least $u_1 - 1$ false nulls. This proves part (i).

For a proof of part (ii), see Appendix A. □

Remark 3.1. Theorem 3.1 shows that when controlling the FDR under arbitrary dependence, the operating characteristic of the proposed fixed sequence method is much different from that of the usual stepwise procedure of Benjamini and Yekutieli [6] that relys on p-value based ordering of the hypotheses. It cannot be further improved, even by imposing the PRDS assumption on the p-values, unlike the BY procedure that is known to be greatly improved by the Benjamini-Hochberg (BH) procedure under such positive dependence. Also, as shown in our proof of Theorem 3.1(ii) under arbitrary dependence (see Appendix A), the least favorable configuration (the configuration which leads to the largest error rate, see [10]) of the newly suggested fixed sequence FDR controlling procedure is when the ordering of the hypotheses is perfect (i.e, when all the false null hypotheses are tested before the true ones), 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 least favorable configuration is much different from that given in [16] for the BY procedure under arbitrary dependence.

Since the procedure in Theorem 3.1 cannot be improved under the PRDS condition, we consider in the next subsection the condition of negative dependence which includes independence as a special case, and under such condition, develop a more powerful conventional fixed sequence method that controls the FDR.

3.2. Procedure under negative dependence

When the p-values are negatively associated as defined in Section 2, the critical constants of the conventional fixed sequence procedure in Theorem 3.1 can be further improved as in the following:

Theorem 3.2. *The conventional fixed sequence method with critical constants*

$$\alpha_i^{(2)} = \frac{i\alpha}{1+(i-1)\alpha}, i = 1, \ldots, m$$

strongly controls the FDR at level α when the p-values satisfy the NAWS condition.

To prove Theorem 3.2, we use the following lemma, with a proof given in Appendix:

Lemma 3.1. *Let $m_{0,i}$ and $m_{1,i}$ respectively denote the numbers of true and false null hypotheses among the first i null hypotheses, and*

$$d_i = \begin{cases} 1_{\{i\in I_0\}} & \text{if } i = 1 \\ (m_{1,i-1}1_{\{i\in I_0\}} - m_{0,i-1}1_{\{i\notin I_0\}})/(i(i-1)) & \text{if } i > 1. \end{cases}$$

Then, the FDR of any fixed sequence procedure can be expressed as

$$\mathrm{FDR} = \sum_{i=1}^{m} d_i \Pr(R \geq i).$$

Proof of Theorem 3.2. If $u_1 = 1$, then

$$\mathrm{FDR} \leq \mathrm{FWER} = \Pr(V > 0) = \Pr\left(\widehat{P}_1 \leq \alpha_1^{(2)}\right) \leq \alpha.$$

If $u_1 > 1$, then by Lemma 3.1,

$$\begin{aligned} \mathrm{FDR} &= \sum_{i=1}^{m} d_i \Pr(R \geq i) = \sum_{i=u_1}^{m} d_i \Pr(R \geq i) \\ &= \sum_{i=u_1}^{m} \left(d_i + \frac{m_{1,i}\alpha}{i}\right) \Pr(R \geq i) - \sum_{i=u_1}^{m} \frac{m_{1,i}\alpha}{i} \Pr(R \geq i). \end{aligned} \tag{3}$$

The second equality follows from the fact that $d_i = 0$ for $i = 1, \ldots, u_1 - 1$.

For each $i = u_1, \ldots, m$, the following inequality holds, i.e.,

$$\left(d_i + \frac{m_{1,i}\alpha}{i}\right) \Pr\left(R \geq i\right) \leq \frac{m_{1,i-1}\alpha}{i-1} Pr(R \geq i-1). \tag{4}$$

To see this, we consider, separately, the case when $i \in I_0$ and when $i \notin I_0$. Suppose $i \in I_0$, then $m_{1,i-1} = m_{1,i}$ and

$$\begin{aligned}
& \left(d_i + \frac{m_{1,i}\alpha}{i}\right) \Pr\left(R \geq i\right) \\
= \; & \left(\frac{m_{1,i-1}}{i(i-1)} + \frac{m_{1,i-1}\alpha}{i}\right) \Pr\left(P_1 \leq \alpha_1^{(2)}, \ldots, P_{i-1} \leq \alpha_{i-1}^{(2)}, P_i \leq \alpha_i^{(2)}\right) \\
\leq \; & \frac{m_{1,i-1}\left(1 + (i-1)\alpha\right)}{i(i-1)} \Pr\left(P_1 \leq \alpha_1^{(2)}, \ldots, P_{i-1} \leq \alpha_{i-1}^{(2)}\right) \Pr\left(P_i \leq \alpha_i^{(2)}\right) \\
\leq \; & \frac{m_{1,i-1}\alpha}{i-1} Pr(R \geq i-1).
\end{aligned}$$

The first and second inequalities follow from (2) and (1), respectively.

Now suppose $i \notin I_0$, then $m_{1,i} = m_{1,i-1} + 1$ and

$$\begin{aligned}
& \left(d_i + \frac{m_{1,i}\alpha}{i}\right) \Pr\left(R \geq i\right) = \left(-\frac{m_{0,i-1}}{i(i-1)} + \frac{(m_{1,i-1}+1)\alpha}{i}\right) \Pr\left(R \geq i\right) \\
\leq \; & \left(-\frac{m_{0,i-1}\alpha}{i(i-1)} + \frac{(m_{1,i-1}+1)\alpha}{i}\right) \Pr\left(R \geq i\right) \\
= \; & \frac{m_{1,i-1}\alpha}{i-1} \Pr\left(R \geq i\right) \leq \frac{m_{1,i-1}\alpha}{i-1} \Pr\left(R \geq i-1\right).
\end{aligned}$$

In the second equality, we use the fact that $m_{0,i-1} + m_{1,i-1} = i - 1$.

Applying (4) to (3), we have

$$\begin{aligned}
\text{FDR} \quad = \quad & \sum_{i=u_1}^{m} \left(d_i + \frac{m_{1,i}\alpha}{i}\right) \Pr\left(R \geq i\right) - \sum_{i=u_1}^{m} \frac{m_{1,i}\alpha}{i} \Pr\left(R \geq i\right) \\
\leq \quad & \sum_{i=u_1}^{m} \frac{m_{1,i-1}\alpha}{i-1} \Pr\left(R \geq i-1\right) - \sum_{i=u_1}^{m} \frac{m_{1,i}\alpha}{i} \Pr\left(R \geq i\right) \\
= \quad & \alpha \Pr\left(R \geq u_1 - 1\right) - \frac{m_1\alpha}{m} \Pr\left(R = m\right) \\
\leq \; & \alpha.
\end{aligned}$$

The equality follows from that fact that $m_{1,u_1-1} = u_1 - 1$, since the first $u_1 - 1$ hypotheses are false. □

Remark 3.2. The conventional fixed sequence procedure in Theorem 3.2 is nearly optimal in the sense that the upper bound of the FDR of this procedure is very close to α under certain configurations. Consider the following configuration: All the false null hypotheses are tested before the true null hypotheses, the false null p-values are all 0 with probability 1, and the true null p-values are independently distributed as $U(0,1)$. Under this configuration, it is easy to check that (4) becomes an equality. Following the proof of Theorem 3.2, the FDR of this procedure is exactly $\alpha - \frac{m_1\alpha}{m} \Pr\left(R = m\right)$. When m_1/m converges

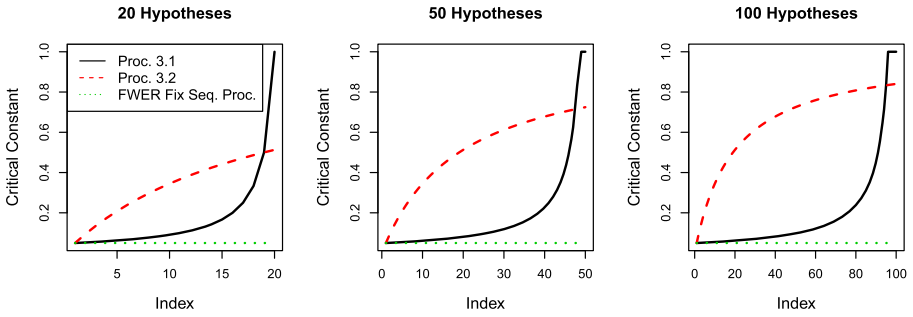

FIG 1. *A plot of the critical constants of the procedures in Theorems 3.1 (solid line), 3.2 (dashed line), and the FWER fixed sequence procedure (dotted line) for m =20, 50, and 100.*

to π_1 as $m \to \infty$ with $0 \le \pi_1 < 1$, an approximate lower bound of the FDR is $\alpha - \pi_1 \alpha e^{-(1-\pi_1)\frac{1-\alpha}{\alpha}}$.

To see why, we first note that

$$\frac{m_1\alpha}{m}\Pr(R = m) = \frac{m_1\alpha}{m}\prod_{i=m_1+1}^{m}\alpha_i^{(2)} < \frac{m_1\alpha}{m}(\alpha_m^{(2)})^{m-m_1}.$$

Then, by simple calculation, we have

$$\begin{aligned}\lim_{m\to\infty}\mathrm{FDR} &\ge \alpha - \lim_{m\to\infty}\frac{m_1\alpha}{m}(\alpha_m^{(2)})^{m-m_1}\\ &= \alpha - \lim_{m\to\infty}\frac{m_1\alpha}{m}\left(1+\frac{1-\alpha}{\alpha}\frac{1}{m}\right)^{-m(1-\frac{m_1}{m})}\\ &= \alpha - \pi_1\alpha e^{-(1-\pi_1)\frac{1-\alpha}{\alpha}}.\end{aligned}$$

This lower bound of the FDR is very close to the pre-specified level α. For example, for large m, if $m_1/m = 0.2$, then with $\alpha = 0.05$, the lower bound under the above configuration is about 0.04999975.

Remark 3.3. Figure 1 compares the critical constants for the proposed procedures along with the FWER fixed sequence procedure at level α. It should be noted that the first few critical constants are the most important ones. If the first few critical constants are too small, then the procedure might stop too early and the remaining hypotheses will not have a chance to be tested. With this in mind, it can be seen that the critical constants of the procedure introduced in Theorem 3.2 are by far the best, and the critical constants of the procedure in Theorem 3.1 are slightly better than those of the conventional fixed sequence procedure.

4. Fixed sequence procedures that allow more acceptances

A conventional fixed sequence method might potentially lose power if an early null hypothesis fails to be rejected, with the remaining hypotheses having no

chance of being tested. To remedy this, we generalize a conventional fixed sequence method to one that allows a certain number of acceptances. The procedure will keep testing hypotheses until a pre-specified number of acceptance has been reached. The same idea has also been used by [19] to develop FWER controlling procedures in fixed-sequence multiple testing.

Suppose k is a pre-specified positive integer and $\alpha_1 \leq \cdots \leq \alpha_m$ is a non-decreasing sequence of critical constants. A fixed sequence method that allows more acceptances is defined below.

Definition 4. (Fixed sequence method stopping on the k^{th} acceptance)

1. *If $P_1 \leq \alpha_1$, then reject H_1; otherwise, accept H_1. If $k > 1$ or H_1 is rejected, then continue to test H_2; otherwise stop.*
2. *If $P_i \leq \alpha_i$, then reject H_i; otherwise, accept H_i. If the number of accepted hypotheses so far is less than k, then continue to test H_{i+1}; otherwise stop.*

It is easy to see that when $k = 1$, the fixed sequence method stopping on the k^{th} acceptance reduces to the conventional one.

Theorem 4.1. *The fixed sequence method stopping on the k^{th} acceptance with critical constants*

$$\alpha_i^{(3)} = \begin{cases} \frac{\alpha}{k} & \text{if} \quad i = 1, \ldots, k \\ \frac{(m-k+1)\alpha}{(m-i+1)k} & \text{if} \quad i = k+1, \ldots, m \end{cases}$$

strongly controls the FDR at level α under arbitrary dependence of the p-values.

Proof. Let U be the index of the first rejected true null hypothesis. If no true null hypotheses are rejected, then set $U = 0$. We will show that for $i = 1, \ldots, m_0$,

$$E\left(\frac{V}{V+S}1_{\{U=u_i\}}\right) \leq \frac{\alpha}{k}. \tag{5}$$

If $i \leq k$, then

$$E\left(\frac{V}{V+S}1_{\{U=u_i\}}\right) \leq \Pr\left(U = u_i\right) \leq \Pr\left(\hat{P}_i \leq \frac{\alpha}{k}\right) \leq \frac{\alpha}{k}.$$

Now, assume $i > k$. Note that the event $\{U = u_i\}$ implies $V \leq m - u_i + 1$ and $S \geq u_i - k$, because the first $u_i - 1$ hypotheses were either false nulls or not rejected true nulls and among the first $u_i - 1$ hypotheses tested, there can be at most $k - 1$ acceptances. Thus,

$$\begin{aligned} & E\left(\frac{V}{V+S}1_{\{U=u_i\}}\right) \leq E\left(\frac{m-u_i+1}{(m-u_i+1)+(u_i-k)}1_{\{U=u_i\}}\right) \\ \leq \quad & \frac{m-u_i+1}{m-k+1}\Pr\left(\hat{P}_i \leq \alpha_{u_i}^{(3)}\right) \leq \frac{\alpha}{k}. \end{aligned}$$

The first inequality follows from the fact that $V/(V+S)$ is an increasing function of V and a decreasing function of S.

From (5), we have

$$\text{FDR} = \sum_{i=1}^{\min(m_0,k)} E\left(\frac{V}{S+V}1_{\{U=u_i\}}\right) \leq \sum_{i=1}^{k}\frac{\alpha}{k} = \alpha,$$

where the first equality follows from the fact that if none of the first k true null hypotheses are rejected, then $V = 0$. □

We should point out that the result in Theorem 4.1 is weaker than that in Theorem 3.1, although the method in Theorem 4.1 reduces to that in Theorem 3.1 when $k = 1$. More specifically, we cannot prove that the procedure in Theorem 4.1 is optimal in the sense that its critical constants cannot be further improved without losing control of the FDR under arbitrary dependence of the p-values. However, under certain distributional assumptions on the p-values, the critical constants of this procedure can potentially be improved. In particular, we have the following result.

Theorem 4.2. *Consider a fixed sequence method stopping on the k^{th} acceptance with critical constants*

$$\alpha_i^{(4)} = \frac{(r_{i-1}+1)\alpha}{k+(i-k)\alpha}, i = 1,\ldots,m,$$

where r_i is the number of rejections among the first i tested hypotheses, with $r_0 = 0$, for $i = 1,\ldots,m$. This procedure strongly controls the FDR at level α if the true null p-values are mutually independent and are independent of the false null p-values.

Before presenting a proof of the above theorem, let us introduce a few more notations. Supposing that we have r rejections, let V_i and S_i be the numbers of false rejections and true rejections among the first $i \leq r$ rejections and J_i be the index of the ith rejected hypothesis. For $i > r$, let $V_i = V_r$, $S_i = V_r$ and $J_i = m+1$. In addition, for notational convenience, define $V_0 = S_0 = J_0 = 0, V_0/0 = 0$, and $S_0/0 = 1$.

We use the following two lemmas, with proofs given in Appendix, to prove the theorem.

Lemma 4.1. *The FDR of any fixed sequence method stopping on the k^{th} acceptance can be expressed as*

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

Lemma 4.2. *Consider the procedure defined in Theorem 4.2, the following inequality holds for $i = 1,\ldots,m$,*

$$\begin{aligned} & E\left(\left(\frac{V_i}{i}-\frac{V_{i-1}}{i-1}\right)1_{\{J_i<m+1\}}\right) \\ \leq\ & E\left(\frac{(k-J_{i-1}+i-1)\alpha}{k}\frac{S_{i-1}}{i-1}1_{\{J_{i-1}<m+1\}} - \frac{(k-J_i+i)\alpha}{k}\frac{S_i}{i}1_{\{J_i<m+1\}}\right). \end{aligned}$$

Proof of Theorem 4.2. By Lemmas 4.1 and 4.2, we have

$$\begin{aligned}
\text{FDR} &= E\left(\sum_{i=1}^{m}\left(\frac{V_i}{i}-\frac{V_{i-1}}{i-1}\right)1_{\{J_i<m+1\}}\right)\\
&\le E\left(\alpha-\frac{(k-J_1+1)\alpha}{k}S_1 1_{\{J_i<m+1\}}\right)\\
&\quad + E\left(\sum_{i=2}^{m}\left(\frac{(k-J_{i-1}+i-1)\alpha}{k}\frac{S_{i-1}}{i-1}1_{\{J_{i-1}<m+1\}}\right.\right.\\
&\quad \left.\left.-\frac{(k-J_i+i)\alpha}{k}\frac{S_i}{i}1_{\{J_i<m+1\}}\right)\right)\\
&= E\left(\alpha-\frac{(k-J_m+m)\alpha}{k}\frac{S_m}{m}1_{\{J_m<m+1\}}\right)\le\alpha. \qquad \square
\end{aligned}$$

Remark 4.1. When $k=1$, the generalized fixed sequence method in the above theorem reduces to the conventional fixed sequence method in Theorem 3.2, since in this case $r_{i-1}=i-1$, and thus continues to control the FDR when the p-values are negatively associated. However, when $k>1$, it can only control the FDR under the independence assumption made in the theorem. It can be shown that this method, when $k>1$, may no longer control the FDR when the p-values are negatively associated. Consider, for example, the problem of simultaneously testing two hypotheses H_1 and H_2 for which both of them are true, the associated p-values $\widehat{P}_1$ and $\widehat{P}_2$ are both $U(0,1)$, and $\widehat{P}_2=1-\widehat{P}_1$. It is easy to see that under such configuration, when $k=2$, the FDR of this procedure is equal to $\frac{\alpha}{2-\alpha}+\frac{\alpha}{2}>\alpha$.

Remark 4.2. Unlike our other fixed sequence procedures, the critical values $\alpha_i^{(4)}$ of the procedure introduced in Theorem 4.2 are random as the i^{th} critical value depends on the number of rejected hypotheses among the first $i-1$ tested hypotheses. As more hypotheses are rejected, the critical values become larger allowing for even more rejections. Also, it should be noted that technically, $\alpha_i^{(4)}$ is measurable with respect to

$$\mathcal{F}_{i-1}=\sigma\{P_j,\mathbf{1}(P_j\le\alpha_{i-1}^{(4)}),\ \text{for all } j=1,\ldots,m\}.$$

Remark 4.3. To apply the fixed sequence methods introduced in Theorems 4.1 and 4.2 in practice, it is important to determine an appropriate value for the tuning parameter k. Some results on how the value of k affects the number of rejections are given in our real data analysis in Section 5 and our simulation study in Section 6. For the procedures in this section, our simulation study suggests a value of k which is between 5% to 10% of m tends to give the best power.

5. Data driven ordering

The applicability of the aforementioned fixed sequence methods depends on the availability of ordering among the hypotheses. When the hypotheses cannot

be pre-ordered, one can use pilot data available to establish a good ordering among the hypotheses in some cases. For example, in replicated studies, the hypotheses for the follow-up study can be ordered using the data from the primary study. However, when prior information is unavailable, ordering can usually be assessed from the data itself. Such data-driven ordering has been used by several authors in fixed sequence methods controlling the FWER and generalized FWER [11, 19, 23, 24, 37].

Assume that X_i is the variable of interest corresponding to the hypothesis H_i, for $i = 1, \ldots, m$, and that we have n independent observations $X_{i1}, \ldots, X_{in}$ on each X_i. An ordering statistic, Y_i, and a test statistic, T_i, are determined for each $i = 1, \ldots, m$. The Y_i's are used to order all of the hypotheses $H_1, \ldots, H_m$, T_i is used to test the hypothesis $H_i, i = 1, \ldots, m$, and P_i is the corresponding p-value. In addition, Y_i is chosen such that it is independent of the T_i's under H_i and tends to be larger as the effect size increases. The approach is outlined below.

Definition 5. *Data-Driven Ordering Procedure*

1. *The hypotheses are ordered based on $Y_1, \ldots, Y_m$ where the hypothesis corresponding to the largest Y_i is placed first, the hypothesis corresponding to the second largest is placed second, and so on.*
2. *The hypotheses are tested using a fixed sequence procedure based on the p-values $P_1, \ldots, P_m$ and the testing order established in Step 1.*

We give a few examples to further illustrate the approach.

Example 1. *One sample T-test.* Given X_i following a $N(\mu_i, \sigma^2)$, consider testing $H_i : \mu_i = 0$ against $H_i' : \mu_i \neq 0$ simultaneously for $i = 1, \ldots, m$, based on the observations $X_{i1}, \ldots, X_{in}$. Let $\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n$ and $s_i^2 = \sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2/(n-1)$ be the sample mean and variance. Let $Y_i = \sum_{j=1}^{n} X_{ij}^2$ be the ordering statistics, that is, the hypotheses are ordered according to the values of the corresponding sums of squares, and $T_i = \sqrt{n}\bar{X}_i/s_i$ be the usual t-test statistic for testing H_i. Then, $P_i = 2\,(1 - F(|T_i|))$, $i = 1, \ldots, m$, where $F(\cdot)$ is the cdf of the t-distribution with $n-1$ degrees of freedom, are the p-values. When $\mu_i = 0$, T_i and Y_i are independent (see, for instance, [25]; p. 156). Furthermore,

$$E(Y_i) = n(\mu_i^2 + \sigma^2),$$

which suggests that $|\mu_i|$ tends to larger values as Y_i increases.

Example 2. *Two sample T-test.* Suppose $X_{ij}^{(l)}$ follows a $N(\mu_i^{(l)}, \sigma^2)$ distribution, for $j = 1, \ldots, n_l$, and $l = 1, 2$. Consider testing $H_i : \mu_i^{(1)} = \mu_i^{(2)}$ against $H_i' : \mu_i^{(1)} \neq \mu_i^{(2)}$ simultaneously for $i = 1, \ldots, m$ using the $n = n_1 + n_2$ data vectors. The hypotheses can be tested using the two-sample t-test statistics T_i and are ordered through the values of the 'total sum of squares,' which is $Y_i = \sum_{l=1}^{2}\sum_{j=1}^{n_l}(X_{ij}^{(l)} - \bar{X}_i)^2$, where $\bar{X}_i = \sum_{l=1}^{2}\sum_{j=1}^{n_l} X_{ij}^{(l)}/n$, for $i = 1, \ldots, m$. The rationale behind this is independence between the Y_i's and T_i's under H_i (see, for instance, [37], and the following result: $E[Y_i] = (n-1)\sigma^2 + n_1 n_2(\mu_i^{(1)} - \mu_i^{(2)})^2/n$.

Example 3. *Nonparametric test.* [24] describe a data-driven ordering strategy for nonparametric tests. In the one sample case, we are interested in testing $H_i : \mu_i = 0$ against $H_i' : \mu_i \neq 0$ where $X_{ij}, j = 1, \ldots, n$ are assumed to be symmetric about μ_i. The hypotheses are tested using the one-sample Wilcoxon test and ordered based on $Y_i = \text{med}(|X_{i1}|, \ldots, |X_{in}|)$. In the two sample case, we are interested in testing $H_i : \mu_i^{(1)} = \mu_i^{(2)}$ against $H_i' : \mu_i^{(1)} \neq \mu_i^{(2)}$ using $n = n_1 + n_2$ data vectors, where $X_{ij}^{(l)}, j = 1, \ldots, n_l$ are assumed to be symmetric about $\mu_i^{(l)}$ for $l = 1, 2$. The hypotheses are tested using the two-sample Wilcoxon test and ordered based on the interquartile range $Y_i = q_{3i} - q_{1i}$, where q_{1i} and q_{3i} are respectively the 1^{st} and 3^{rd} quartile of the mixture of $X_{ij}^{(1)}$'s and $X_{ij}^{(2)}$'s.

When our proposed fixed sequence procedures are used in applications coupled with the aforementioned data-driven ordering strategy, FDR control is still maintained under the independence assumption if the ordering statistics are chosen to be independent of the test statistics even though the same data used for ordering as well as testing the hypotheses. More specifically, we have the following result.

Theorem 5.1. *Suppose $X_1, \ldots, X_m$ are mutually independent. If the hypotheses $H_1, \ldots, H_m$ are ordered based on the ordering statistics $Y_i, i = 1, \ldots, m$, and simultaneously tested using the test statistics $T_i, i = 1, \ldots, m$, then, if T_i is independent of Y_i under H_i, the fixed sequence multiple testing procedures introduced in Theorems 3.1-4.2 strongly control the FDR at level α.*

Proof. Assume without any loss of generality that $Y_1 \geq \cdots \geq Y_m$, so that conditional on the Y_i's, H_i is the ith hypotheses to be tested in our fixed sequence multiple testing methods. When H_i is true, P_i is independent of both Y_i and $X_j, j = 1, \ldots, m$ with $j \neq i$. This follows from independence of the X_i's and that of Y_i and T_i under H_i. Thus, conditional on the Y_i's, each true null p-value P_i still satisfies (1) and is independent of all other p-values P_j with $j \neq i$. Therefore, we have for each of the procedures in Theorems 3.1, 3.2, 4.1, and 4.2,

$$E\left(\frac{V}{\max(R,1)} \,\middle|\, Y_1, \ldots, Y_m\right) \leq \alpha. \tag{6}$$

This proves the desired result. □

Although the fixed sequence methods introduced in Theorems 3.1-4.2 coupled with data-driven ordering strategy are developed under independence, however, our simulation study in the following Section 6 shows that these methods also control the FDR well under various dependence structures. So, we applied all of our proposed four methods to the HIV microarray data [35] used by [8]. These data consist of $m = 7680$ gene expression levels across eight subjects, four HIV infected and four uninfected. The data were log-transformed and normalized. Our goal is to determine which genes are differentially expressed by testing $H_i : \mu_i^{(1)} = \mu_i^{(2)}$ versus $H_i' : \mu_i^{(1)} \neq \mu_i^{(2)}$ simultaneously for $i = 1, \ldots, 7680$, where $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are the gene specific mean expressions for HIV infected and uninfected subjects, respectively.

TABLE 1
The Number of Discoveries out of $m = 7680$ Genes in the HIV Data from [35] by the procedure from Theorem 4.2 and the BH Procedure

	Procedure from Theorem 4.2				BH Procedure
	k = 1	k/m = 0.05	k/m = 0.1	k/m = 0.15	
$\alpha = 0.001$	11	13	9	8	8
$\alpha = 0.01$	11	18	17	16	13
$\alpha = 0.025$	11	18	18	18	13
$\alpha = 0.05$	11	20	19	19	18
$\alpha = 0.1$	20	21	24	20	22

TABLE 2
The Number of Discoveries out of $m = 7680$ Genes in the HIV Data from [35] by the procedure from Theorem 4.1 and the BY Procedure

	Procedure from Theorem 4.1				BY Procedure
	k = 1	k/m = 0.05	k/m = 0.1	k/m = 0.15	
$\alpha = 0.001$	11	10	8	8	0
$\alpha = 0.01$	11	13	13	11	8
$\alpha = 0.025$	11	15	13	13	10
$\alpha = 0.05$	11	16	15	13	10
$\alpha = 0.1$	11	18	16	16	13

We applied our proposed procedures with the p-values generated from two sample t-tests for the genes. Since there is no natural ordering among the genes, we used the ordering statistics for two sample t-tests in Example 2 to order these tested hypotheses. We compared the procedure in Theorem 4.2 with the BH procedure. The results are summarized in Table 1 for different values of k where $k/m = 0.05, 0.1$, and 0.15. As seen from Table 1, for all values of k except $k = 1$, the procedure in Theorem 4.2 generally has more rejections than the BH procedure. When α is small, $k/m = 0.05$ tends to have the most rejections, but for large α, $k/m = 0.1$ has the most rejections. Also, we compared the procedure in Theorem 4.1 with the BY procedure. The results are displayed in Table 2. As seen from Table 2, for most values of k, our procedure outperforms the BY procedure in terms of the number of rejections. When $\alpha = 0.001$, the BY procedure cannot reject any hypotheses, but the procedure in Theorem 4.1 has at least 8 rejections for all the values of k considered.

6. Simulation study

A simulation study was conducted to address the performances of the proposed procedures. We will refer to the procedures in Theorems 3.1, 3.2, 4.1, and 4.2 as Procedures 1–4, respectively. Specifically, we addressed the following two questions:

1. How do Procedures 1, 2, 3, and 4 compare against the BH and BY procedures in terms of FDR and power?
2. For Procedures 3 and 4, how should k be chosen so that the power is large?

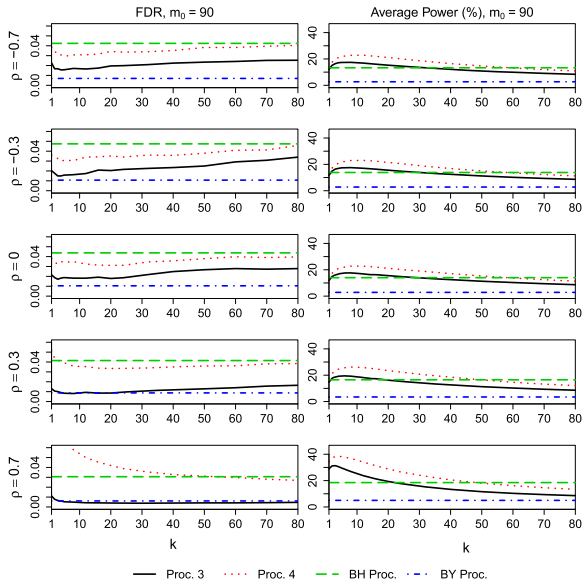

FIG 2. *The FDR (left) and average power (right) of Procedure 3 (solid) in Theorem 4.1, Procedure 4 (dotted) in Theorem 4.2, the BH procedure (dashed), and the BY procedure (dotted dash) for 100 hypotheses with* $m_0 = 90$ *and* $n = 10$ *under common correlation.*

In each simulation, n independent m dimensional random normal vectors with covariance matrix Σ and components $Z_i \sim N(\mu_i, 1), i = 1, \ldots, m$, 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. The μ_i corresponding to each false null hypothesis is set to the value at which the power of one-sample t-test at level 0.05 is 0.75. As for the joint dependence, we considered two types of correlation structure: (i) equal correlation structure where Σ had off-diagonal components equal to ρ and diagonal components equal to 1; (ii) block dependence structure where Σ had a block diagonal matrix structure with 2×2 blocks along the diagonal with blocks constructed as $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

We set $\alpha = 0.05$ and $m = 100$. The hypotheses were ordered using the 'sum of squares ordering' used in Example 1 from Section 5. We had 5,000 runs of simulation for each of the procedures considered. We noted the false discovery proportion and the proportion of correctly rejected false null hypotheses for each procedure in each of these runs. The simulated FDR and average power (the expected proportion of correctly rejected false null hypotheses) were obtained by averaging out the corresponding 5,000 values.

We first compared Procedures 1–4 with the BH and BY procedures when $m_0 = 90$ and $n = 10$. Figure 2 displays these comparisons in terms of the simulated FDR and average power, respectively, under block correlation ($\rho = -0.3$ and $\rho = -0.7$) and equal correlation ($\rho = 0, \rho = 0.3$, and $\rho = 0.7$) while k varies from 1 to 80. We did not explicitly label Procedures 1 and 2 as these are

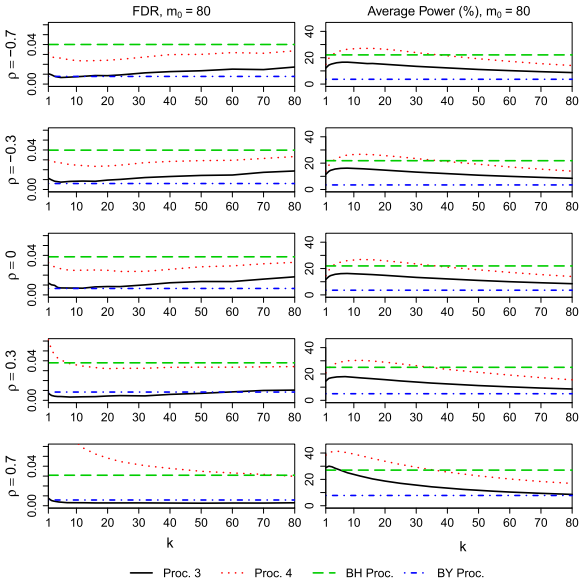

FIG 3. *The FDR (left) and average power (right) of Procedure 3 (solid) in Theorem 4.1, Procedure 4 (dotted) in Theorem 4.2, the BH procedure (dashed), and the BY procedure (dotted dash) for 100 hypotheses with* $m_0 = 80$ *and* $n = 10$ *under common correlation.*

both special cases of Procedures 3 and 4 when $k = 1$, respectively. As seen from Figure 2, Procedure 3 controls the FDR at level 0.05 under every dependence configuration. Procedure 4 controls the FDR under block correlation ($\rho = -0.3$ and $\rho = -0.7$) or independence ($\rho = 0$), generally maintains control of FDR under mild correlation ($\rho = 0.3$), but loses control of FDR for small values of k under strong correlation ($\rho = 0.7$). From Figure 2, one can also see that Procedure 3 tends to have its largest power when k is between about 3 and 10, and Procedure 4 tends to have its largest power when k is between about 5 and 15. In addition, Figure 2 shows that for a well chosen k, Procedures 3 and 4 outperform the BH procedure, and for all values of k, Procedures 3 and 4 outperform the BY procedure.

Next, we compared the FDR and average power of these procedures when $m_0 = 80$ and $n = 10$. Figure 3 displays these comparisons for $\rho = -0.7$, $\rho = -0.3$, $\rho = 0$, $\rho = 0.3$, and $\rho = 0.7$ while k varies from 1 to 80. In terms of FDR, the results shown in Figure 3 are similar to the results shown in Figure 2. In terms of power, Figure 3 shows that the powers of all the procedures tend to show an increase over the powers shown in Figure 2. Figure 3 also shows that for a well chosen k, Procedure 4 can outperform the BH procedure, but, in general, Procedure 3 cannot. Again, for all values of k, Procedures 3 and 4 outperform the BY procedure.

It should be noted that in many cases the FDR of Procedure 3 and 4 is well below 0.05, which suggests there is room for improvement. But, Theorem 3.1

and Remark 3.2 show that there is no or little room for improvement in the critical values at $k = 1$. The reason for these seemingly contradictory results is that by the proof of Theorem 3.1(ii) and Remarks 3.1 and 3.2, the critical values of the procedures introduced in Theorem 3.1 and 3.2 are determined on the basis of that the ordering of the p-values is correct (all false null p-values is tested before true null p-values) and that all false null p-values are rejected with probability 1. These settings are far from true in our simulation study.

7. Concluding remarks

In this paper, we have developed 'step-down' procedures which control the FDR and exploit the structure of pre-ordered hypotheses. We have been able to produce the desired methods in the most simple as well as a general setting covering different dependence scenarios. Our simulation study and real data analysis show that in some cases, the proposed procedures can be powerful alternatives to existing FDR controlling procedures.

Using some of the techniques developed in this paper, it is possible to develop other types of fixed sequence procedures controlling the FDR, such as a fallback-type procedure. Unlike the conventional and generalized fixed sequence procedures developed in this paper, the fallback-type procedure tests the remaining hypotheses no matter how many earlier hypotheses are accepted, which is needed for analyzing stream data in sequential change detection problems.

Although we have only considered the simplest hierarchical structure – fixed sequence structure – by using the similar techniques presented in this paper, we were able to develop simple and powerful procedures that control the FDR under various dependencies when testing multiple hypotheses with a more complex hierarchical structure. We plan to present these procedures in a future communication.

Appendix A: Technical proofs

A.1. Proof of Theorem 3.1 (ii)

For any u_1, $1 \le u_1 \le m$, consider a joint distribution of the p-values such that the first $u_1 - 1$ hypotheses are false null hypotheses whose corresponding p-values are 0 with probability one. The remaining $m - u_1 + 1$ hypotheses are true null hypotheses such that $\hat{P}_1 \sim U(0,1)$ and $\hat{P}_i = \hat{P}_1$ for $i = 2, \ldots, m_0$. Under such joint distribution of the p-values, the FDR of the conventional fixed sequence procedure is exactly $\alpha_{u_1}^{(1)}(m - u_1 + 1)/m$. If $\alpha_{u_1}^{(1)} = 1$ then the critical constant is already at its largest and cannot be improved. Otherwise, if $\alpha_{u_1}^{(1)} < 1$, then FDR $= \alpha$ and thus $\alpha_{u_1}^{(1)}$ cannot be made any larger.

In the above construction $\vec{P}$ is PRDS on $\vec{P}_0$. We only need to note that for $i = 1, \ldots, m_0$, $\widehat{P}_i = p$ implies $\vec{P} = (0, \ldots, 0, p, \ldots, p)$. Thus, for any increasing set D,

$$\Pr\left(\vec{P} \in D | \widehat{P}_i = p\right) = \begin{cases} 1 & \text{if } (0,\ldots,0,p,\ldots,p) \in D \\ 0 & \text{otherwise,} \end{cases}$$

which is an increasing function in p. □

A.2. Proof of Lemma 3.1

Lemma 3.1 can be regarded as a special case of Lemma 4.1 with $k = 1$. Note that for $i = 1,\ldots,m$, $m_{1,i} = m_{1,i-1}$ and $m_{0,i} = m_{0,i-1} + 1$ when $i \in I_0$, $m_{0,i} = m_{0,i-1}$ when $i \notin I_0$, and $m_{0,i-1} + m_{1,i-1} = i - 1$. Thus, when $k = 1$, the event $\{R \geq i\}$ implies $V_i = m_{0,i}$ and hence

$$\frac{V_i}{i} - \frac{V_{i-1}}{i-1} = \begin{cases} 1_{\{1 \in I_0\}} & \text{for } i = 1 \\ \frac{m_{1,i-1} 1_{\{i \in I_0\}} - m_{0,i-1} 1_{\{i \notin I_0\}}}{i(i-1)} & \text{for } i = 2,\ldots,m. \end{cases} \tag{7}$$

By (7) and Lemma 4.1, the desired result follows. □

A.3. Proof of Lemma 4.1

It is easy to see that

$$\begin{aligned} \text{FDR} &= E\left(\sum_{i=1}^{m} \frac{V_i}{i} 1_{\{R=i\}}\right) = E\left(\sum_{i=1}^{m} \left(\frac{V_i}{i} 1_{\{R \geq i\}} - \frac{V_i}{i} 1_{\{R \geq i+1\}}\right)\right) \\ &= E\left(\sum_{i=1}^{m} \left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1}\right) 1_{\{R \geq i\}}\right) \\ &= E\left(\sum_{i=1}^{m} \left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1}\right) 1_{\{J_i < m+1\}}\right), \end{aligned}$$

the desired result. □

A.4. Proof of Lemma 4.2

It is easy to see that for $i = 1,\ldots,m$, if there are at least i rejections, then $i \leq J_i \leq \min(i + k - 1, m)$. For ease of notation, let $j_i^* = \min(i + k - 1, m)$. For $i, j = 1,\ldots,m$, define $f_i(j) = \frac{(k-j+i)\alpha}{k} \frac{S_i}{i}$ and $W_i(j) = 1_{\{J_{i-1} \leq j, J_i > j\}}$. Regarding the relationship between J_i and $W_i(j)$, there are the following two equalities available:

$$1_{\{J_i = j\}} = W_i(j-1) 1_{\{P_j \leq \alpha_j^{(4)}\}} \tag{8}$$

and

$$W_i(j) - W_i(j-1) = 1_{\{J_{i-1} = j\}} - 1_{\{J_i = j\}}. \tag{9}$$

The first equality follows from the fact that for $i = 1, \ldots, m$ and $j = i, \ldots, j_i^*$, when $J_i = j$, there are $i-1$ rejections among the first $j-1$ tested hypotheses and the i^{th} rejection is exactly the j^{th} tested hypothesis, thus

$$1_{\{J_i=j\}} = 1_{\{J_{i-1}\leq j-1, J_i > j-1, P_j \leq \alpha_j^{(4)}\}} = W_i(j-1)1_{\{P_j \leq \alpha_j^{(4)}\}}.$$

The second equality follows from the fact that the event $\{W_i(j) = 1\}$ implies that there are exactly $i-1$ rejections among the first j tested hypotheses, thus for $j = i-1, \ldots, j_{i-1}^*$,

$$\begin{aligned} & W_i(j) - W_i(j-1) = 1_{\{J_{i-1}\leq j, J_i > j\}} - W_i(j-1) \\ = \; & 1_{\{J_{i-1}=j\}} + 1_{\{J_{i-1}\leq j-1, J_i > j-1, P_j > \alpha_j^{(4)}\}} - W_i(j-1) \\ = \; & 1_{\{J_{i-1}=j\}} - W_i(j-1)1_{\{P_j \leq \alpha_j^{(4)}\}} \\ = \; & 1_{\{J_{i-1}=j\}} - 1_{\{J_i=j\}}, \end{aligned}$$

where, the third equality follows from the fact that $1_{\{P_j > \alpha_j^{(4)}\}} = 1 - 1_{\{P_j \leq \alpha_j^{(4)}\}}$ and the fourth follows from (8).

By using the above two equalities, we can prove two inequalities below, which are needed in the proof of this lemma. Firstly, we show by using (8) that the following inequality holds:

$$E\left(\left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} + f_i(j) - f_{i-1}(j)\right)1_{\{J_i=j\}}\right) \leq E\left(\frac{\alpha}{k}\frac{S_{i-1}}{i-1}W_i(j-1)\right). \quad (10)$$

Proof of (10). To see this, we consider, separately, the case when $j \in I_0$ and when $j \notin I_0$.

Suppose $j \in I_0$, then $S_i = S_{i-1}$ and $V_i = V_{i-1} + 1$ when $J_i = j$. Using the fact that $V_{i-1} + S_{i-1} = i - 1$, the left hand side of (10), after some algebra, becomes

$$\begin{aligned} & E\left(\frac{k+(j-k)\alpha}{ki}\frac{S_{i-1}}{i-1}W_i(j-1)1_{\{P_j \leq \alpha_j^{(4)}\}}\right) \\ = \; & E\left(\frac{k+(j-k)\alpha}{ki}\frac{S_{i-1}}{i-1}W_i(j-1)\right)\Pr\left(P_j \leq \alpha_j^{(4)}\right) \\ \leq \; & E\left(\frac{k+(j-k)\alpha}{ki}\frac{S_{i-1}}{i-1}W_i(j-1)\right)\frac{i\alpha}{k+(j-k)\alpha} \\ = \; & E\left(\frac{\alpha}{k}\frac{S_{i-1}}{i-1}W_i(j-1)\right). \end{aligned}$$

The first equality follows from these two facts: (i) When $J_i = j$, i.e. the i^{th} rejection is H_j, S_{i-1} is only dependent on the first $j-1$ p-values, since S_{i-1} is the number of false null hypotheses among the first $i-1$ rejections; (ii) $W_i(j-1)$ is also only dependent on the first $j-1$ p-values, since $W_i(j-1)$ is 1 if and only if there are exactly $i-1$ rejections among the first $j-1$ hypotheses tested. The inequality follows from (1).

Now suppose $j \notin I_0$, then $S_i = S_{i-1} + 1$ and $V_i = V_{i-1}$. Similarly, using the fact that $V_{i-1} + S_{i-1} = i - 1$, the left hand side of (10), after some algebra, becomes

$$
\begin{aligned}
& E\left(\left(\frac{-V_{i-1}}{i(i-1)} + \frac{((j-k)S_{i-1} + (i-1)(k-j+i))\alpha}{ki(i-1)}\right) W_i(j-1) 1_{\{P_j \le \alpha_j^{(4)}\}}\right) \\
& \le E\left(\left(\frac{-V_{i-1}}{i(i-1)}\frac{(k-j+i)\alpha}{k} + \frac{((j-k)S_{i-1} + (i-1)(k-j+i))\alpha}{ki(i-1)}\right) W_i(j-1)\right) \\
& = E\left(\frac{\alpha}{k}\frac{S_{i-1}}{i-1} W_i(j-1)\right).
\end{aligned}
$$

The inequality follows by the fact that $j \ge i$ so that $k - j + i \le k$. In addition, in the last line we use the fact that $V_{i-1} + S_{i-1} = i - 1$. □

Next, we show by using (9) that the following inequality holds:

$$
E\left(f_{i-1}(J_{i-1}) 1_{\{J_{i-1} < m+1\}} - f_{i-1}(J_i) 1_{\{J_i < m+1\}}\right) \ge E\left(\sum_{j=i}^{j_i^*} \frac{\alpha}{k}\frac{S_{i-1}}{i-1} W_i(j-1)\right). \tag{11}
$$

Proof of (11). By using (9), we have

$$
\begin{aligned}
& E\left(f_{i-1}(J_{i-1}) 1_{\{J_{i-1} < m+1\}} - f_{i-1}(J_i) 1_{\{J_i < m+1\}}\right) \\
= \; & E\left(\sum_{j=i-1}^{j_{i-1}^*} f_{i-1}(j) 1_{\{J_{i-1}=j\}} - \sum_{j=i}^{j_i^*} f_{i-1}(j) 1_{\{J_i=j\}}\right) \\
= \; & E\left(\sum_{j=i-1}^{j_{i-1}^*} f_{i-1}(j) 1_{\{J_{i-1}=j\}} - \sum_{j=i}^{j_{i-1}^*} f_{i-1}(j) 1_{\{J_i=j\}}\right) \\
= \; & E\left(\sum_{j=i}^{j_{i-1}^*} f_{i-1}(j)\left(1_{\{J_{i-1}=j\}} - 1_{\{J_i=j\}}\right) + f_{i-1}(i-1) 1_{\{J_{i-1}=i-1\}}\right) \\
= \; & E\left(\sum_{j=i}^{j_{i-1}^*} f_{i-1}(j)(W_i(j) - W_i(j-1)) + f_{i-1}(i-1) W_i(i-1)\right) \\
= \; & E\left(\sum_{j=i}^{j_{i-1}^*} \left(f_{i-1}(j-1) - f_{i-1}(j)\right) W_i(j-1) + f_{i-1}(j_{i-1}^*) W_i(j_{i-1}^*)\right) \\
\ge \; & E\left(\sum_{j=i}^{j_i^*} \frac{\alpha}{k}\frac{S_{i-1}}{i-1} W_i(j-1)\right),
\end{aligned}
$$

the desired result. Here, the second equality follows from the fact that if $j_{i-1}^* = m$, then $j_i^* = m$; otherwise, $j_{i-1}^* = i+k-2$ and $j_i^* = i+k-1$ so that $f_{i-1}(j_i^*) = 0$.

The fourth equality follows from (9) and the fact that $W_i(i-1) = 1_{\{J_{i-1}=i-1\}}$. The inequality follows from the definition of $f_{i-1}(j)$. □

Proof of Lemma 4.2. Finally, by combining these two inequalities, we can get the desired result as follows.

$$\begin{aligned}
& E\left(\left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1}\right) 1_{\{J_i<m+1\}}\right) \\
= \; & E\left(\left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} + f_i(J_i) - f_{i-1}(J_i)\right) 1_{\{J_i<m+1\}}\right) \\
& -E\left(f_{i-1}(J_{i-1})1_{\{J_{i-1}<m+1\}} - f_{i-1}(J_i)1_{\{J_i<m+1\}}\right) \\
& +E\left(f_{i-1}(J_{i-1})1_{\{J_{i-1}<m+1\}} - f_i(J_i)1_{\{J_i<m+1\}}\right) \\
= \; & E\left(\sum_{j=i}^{j_i^*}\left(\frac{V_i}{i} - \frac{V_{i-1}}{i-1} + f_i(j) - f_{i-1}(j)\right) 1_{\{J_i=j\}}\right) \\
& -E\left(f_{i-1}(J_{i-1})1_{\{J_{i-1}<m+1\}} - f_{i-1}(J_i)1_{\{J_i<m+1\}}\right) \\
& +E\left(f_{i-1}(J_{i-1})1_{\{J_{i-1}<m+1\}} - f_i(J_i)1_{\{J_i<m+1\}}\right) \\
\leq \; & E\left(\sum_{j=i}^{j_i^*}\frac{\alpha}{k}\frac{S_{i-1}}{i-1}W_i(j-1) - \sum_{j=i}^{j_i^*}\frac{\alpha}{k}\frac{S_{i-1}}{i-1}W_i(j-1)\right) \\
& +E\left(f_{i-1}(J_{i-1})1_{\{J_{i-1}<m+1\}} - f_i(J_i)1_{\{J_i<m+1\}}\right) \\
= \; & E\left(f_{i-1}(J_{i-1})1_{\{J_{i-1}<m+1\}} - f_i(J_i)1_{\{J_i<m_1\}}\right),
\end{aligned}$$

The inequality follows from (10) and (11). □

Acknowledgements

Our sincere thanks go to the referee and the AE whose helpful and insightful comments and suggestions have significantly improved the presentation of the paper.

References

[1] Aharoni, E. and Rosset, S. (2014). Generalized α-investing: definitions, optimality results and application to public databases. *Journal of the Royal Statistical Society: Series B* **76** 771–794. MR3248676

[2] Barber, R. and Candes, E. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics* **43** 2055–2085. MR3375876

[3] Benjamini, Y. and Heller, R. (2007). False discovery rates for spatial signals. *J. Amer. Satist. Assoc.* **102** 1272–1281. MR2412549

[4] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* **57** 289–300. MR1325392

[5] Benjamini, Y. and Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J. Statist. Plann. Inference* **82** 163–170. MR1736441

[6] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

[7] Dmitrienko, A., D'Agostino, R., and Huque, M. (2013). Key multiplicity issues in clinical drug development. *Statistics in Medicine* **32** 1079–1111. MR3045885

[8] Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statistical Science* **23** 1–22. MR2431866

[9] Farcomeni, A. and Finos, L. (2013). FDR control with pseudo-gatekeeping based on a possibly data driven order of the hypotheses. *Biometrics* **69** 606–613. MR3106588

[10] Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. *Biometrical Journal* **43** 985–1005 MR1878272

[11] Finos, L. and Farcomeni, A. (2011). k-FWER control without multiplicity correction, with application to detection of genetic determinants of multiple sclerosis in Italian twins. *Biometrics* **67** 174–181. MR2898829

[12] G'Sell, M. G., Wager, S., Chouldechova, A., and Tibshirani, R. (2016). Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society: Series B* **78** 423–444. MR3454203

[13] Goeman, J. and Finos, L. (2012). The inheritance procedure: Multiple testing of tree-structured hypotheses. *Statistical Applications in Genetics and Molecular Biology* **11** 1–18. MR2924200

[14] Goeman, J. and Mansmann, U. (2008). Multiple testing on the directed acyclic graph of gene ontology. *Bioinformatics* **24** 537–544.

[15] Goeman, J. and Solari, A. (2010). The sequential rejection principle of familywise error control. *Ann. Statist.* **38** 3782–3810. MR2766868

[16] Guo, W. and Rao, M. (2008). On control of the false discovery rate under no assumption of dependency. *Journal of Statistical Planning and Inference* **28** 3176–3188. MR2526229

[17] Heller, R., Manduchi, E., Grant, G., and Ewens, W. (2009). A flexible two-stage procedure for identifying gene sets that are differentially expressed. *Bioinformatics* **25** 929–942.

[18] Hommel, G., Bretz, F., and Maurer, W. (2007). Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine* **26** 4063–4074. MR2405792

[19] Hommel, G. and Kropf, S. (2005). Testing for differentiation in gene expression using a data driven order or weights for hypotheses. *Biometrical Journal* **47** 554–562. MR2190478

[20] Huque, M. and Alosh, M. (2008). A flexible fixed-sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials. *Journal of Statistical Planning and Inference* **138** 321–335. MR2412590

[21] Javanmard, A. and Montanari, A. (2015). On online control of false discovery rate. *arXiv preprint arXiv:1502.06197*.

[22] Joag Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Ann. Statist.* **11** 286–295.

[23] Kropf, S. and Läuter, J. (2002). Multiple tests for different sets of variables using a data-driven ordering of hypotheses, with an application to gene expression data. *Biometrical Journal* **44** 789–800.

[24] Kropf, S., Läuter, J., Eszlinger, M., Krohn, K., and Paschkeb, R. (2004). Nonparametric multiple test procedures with data-driven order of hypotheses and with weighted hypotheses. *Journal of Statistical Planning and Inference* **125** 31–47.

[25] Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer, New York.

[26] Lei, L. and Fithian, W. (2016). Power of ordered hypothesis testing. *arXiv preprint arXiv:1606.01969*.

[27] Li, A. and Barber, R. (2017). Accumulation tests for FDR control in ordered hypothesis testing. *J. Amer. Statist. Assoc.* **112** 837–849.

[28] Li, J. and Mehrotra, D. (2008). An efficient method for accommodating potentially underpowered primary endpoints. *Statistics in Medicine* **27** 5377–5391.

[29] Maurer, W., Hothorn, L., and Lehmacher, W. (1995). *Multiple comparisons in drug clinical trials and preclinical assays: A-priori ordered hypotheses*. Vol. 6, Fischer-Verlag, Stuttgart, Germany.

[30] Mehrotra, D. and Heyse, J. (2004). Use of the false discovery rate for evaluating clinical safety data. *Statistical Methods in Medical Research* **13** 227–238.

[31] Millen, B. and Dmitrienko, A. (2011). Chain procedures: A class of flexible closed testing procedures with clinical trial applications. *Statistics in Biopharmaceutical Reseach* **3** 14–30.

[32] Rosenbaum, P. (2008). Testing hypotheses in order. *Biometrika* **95** 248–252.

[33] Ross, G. J., Tasoulis, D., and Adams, N. (2011). Nonparametric monitoring of data streams for changes in location and scale. *Technometrics* **53** 379–389.

[34] Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30** 239–257.

[35] van't Wout, A., Lehrma, G., Mikheeva, S., OKeeffe, G., Katze, M., Bumgarner, R., Geiss, G., and Mullins, J. (2003). Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4(+)-T-cell lines. *Journal of Virology* **77** 1392–1402.

[36] Westfall, P. and Krishen, A. (2001). Optimally weighted, fixed sequence and gate-keeper multiple testing procedures. *Journal of Statistical Planning and Inference* **99** 25–41.

[37] Westfall, P., Kropf, S., and Finos, L. (2004). Weighted FWE-controlling methods in highdimensional situations. In *Recent Developments in Multiple Comparison Procedures*, eds. Y. Benjamini, F. Bretz, and S.

Sarkar, Vol. 47, Beachwood, OH: Institute of Mathematical Statistics, pp. 143–154.

[38] Wiens, B. (2003). A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics* **2** 211–215.

[39] Wiens, B. and Dmitrienko, A. (2005). The fallback procedure for evaluating a single family of hypotheses. *J. Biopharm. Stat.* **15** 929–942.

[40] Wiens, B. and Dmitrienko, A. (2010). On selecting a multiple comparison procedure for analysis of a clinical trial: Fallback, fixed sequence, and related procedures. *Statistics in Biopharmaceutical Research* **2** 22–32.

[41] Yekutieli, D. (2008). Hierarchical false discovery rate-controlling methodology. *J. Amer. Statist. Assoc.* **103** 309–316. MR2420235