Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

On Generalized Fixed Sequence Procedures for Controlling the FWER

Zhiying Qiu,[a] Wenge Guo[b]* and Gavin Lynch[c]

Testing a sequence of pre-ordered hypotheses to decide which of these can be rejected or accepted while controlling the familywise error rate (FWER) is of importance in many scientific studies such as clinical trials. In this paper, we first introduce a generalized fixed sequence procedure whose critical values are defined by using a function of the numbers of rejections and acceptances, and which allows follow-up hypotheses to be tested even if some earlier hypotheses are not rejected. We then construct the least favorable configuration for this generalized fixed sequence procedure and present a sufficient condition for the FWER control under arbitrary dependence. Based on the condition, we develop three new generalized fixed sequence procedures controlling the FWER under arbitrary dependence. We also prove that each generalized fixed sequence procedure can be described as a specific closed testing procedure. Through simulation studies and a clinical trial example, we compare the power performance of these proposed procedures with those of the existing FWER controlling procedures. Finally, when the pairwise joint distributions of the true null p-values are known, we further improve these procedures by incorporating pairwise correlation information while maintaining the control of the FWER.

Keywords: Critical values; Fallback procedure; Familywise error rate; Fixed sequence procedure; Multiple testing; Power.

1. Introduction

In applications of clinical trials, the hypotheses to be tested are often hierarchically ordered based on their importance, clinical relevance, or dose concentration, etc., and thus are tested in a pre-defined sequential order. Although the problem of fixed sequence multiple testing has received much attention and several popular familywise error rate (FWER) controlling procedures, such as the conventional fixed sequence procedure and fallback procedure, have been introduced, further progress is still needed for advancing its theory and methods (Dmitrienko *et al.* [2, 3]).

In this paper, we focus on developing new multiple testing procedures to deal with the situation in which the hypotheses are pre-ordered based on prior knowledge and tested based on the p-values. Throughout the whole paper, methods which operate on such fixed sequence multiple testing problems are called fixed sequence procedures. Fixed sequence procedures are unlike stepwise methods, such as the Holm procedure [4] and Hochberg procedure [5], where the ordering and testing of the hypotheses are both based on the corresponding p-values. In the literature, Maurer, Hothorn and Lehmacher [6] introduced the first fixed sequence multiple testing procedure, which we will refer to as the conventional fixed sequence procedure. In this procedure, each hypothesis is tested at pre-specified level α as long as all of the previous hypotheses have been rejected. It is proved that the procedure strongly controls the FWER at level α under no dependence assumptions on the p-values. However, the main issue with this procedure is that it does not allow any acceptances. Once a hypothesis is not rejected, the remaining hypotheses will have no chance to be tested and thus accepted. Therefore, the procedure will perform poorly if one of the early hypotheses is insignificant. To deal with this issue, Wiens [7] and Wiens and Dmitrienko [8] introduced another popular fixed sequence procedure – the fallback procedure, in which the remaining hypotheses have a chance to be tested, even if an acceptance occurs. And later, several authors have proposed various extensions of the

[a] *Biostatistics and Programming, Sanofi, Bridgewater, NJ 08807, U.S.A.*
[bc] *New Jersey Institute of Technology, Newark, NJ 07102, U.S.A.*
* *Correspondence to: Wenge Guo, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102, U.S.A. E-mail:wenge.guo@njit.edu.*

fallback procedure in order to improve its power including Li and Mehrotra [9] and Huque and Alosh [10]. Recently, Bretz et al. [1] introduced a general graphical approach, by which the conventional fixed sequence procedure and fallback procedure can be visualized.

Compared to the conventional fixed sequence procedure, the fallback procedure is more flexible in the sense that every hypothesis has a chance to be tested, even some previous hypotheses have been accepted. However, Hommel and Bretz [11] showed that in certain situations, the fallback procedure might violate the inherent hierarchical relationships among the hypotheses. For any two hypotheses, the earlier important hypothesis may have less chance to be rejected than the later one, even if their p-values are the same. This is not desired for a good multiple testing procedure. Subsequent works on developing more desirable and more powerful procedures for addressing the problem of fixed sequence multiple testing have been done by many authors, including Bretz *et al.* [1], Rosenbaum [12], Millen and Dmitrienko [13], etc. In addition, Hommel and Kropf [14] introduced a specific fixed sequence procedure, which allows a pre-specified number k of acceptances and has the same critical value α/k. For a detailed review of recent developments in this area of research, see Dmitrienko *et al.* [3] and Wiens and Dmitrienko [15], and for applications of fixed sequence multiple testing procedures in different fields, see Alosh and Huque [16, 17] and Tu *et al.* [18].

In this paper, a main goal is to develop new theory and methods for addressing the problem of fixed sequence multiple testing. We firstly introduce a more general procedure, termed as the generalized fixed sequence procedure, whose critical values are defined by using a function of the numbers of rejections and acceptances, and which allows each hypothesis to be tested even if earlier hypotheses are not rejected. We then discuss a configuration, which we call the Dirac-Ordered configuration, under which the FWER of the procedure attains the maximum among all the configurations having the same joint distribution for the true null p-values. Based on this configuration, we present a sufficient condition for the FWER control of a generalized fixed sequence procedure for any dependence of the p-values. Based on the condition, we develop three new fixed sequence procedures controlling the FWER. To better evaluate the proposed procedure, we illustrate generalized fixed sequence procedures as closed testing procedures.

The rest of the paper is organized as follows. We present some basic notations and generalize the conventional fixed sequence procedure in Section 2. We construct the least favorable configuration for the aforementioned procedure and present a sufficient condition for the FWER control of such a procedure in Section 3. In Section 4, we introduce three new fixed sequence procedures based on this condition. In Section 5, we illustrate our proposed procedures as closed testing procedures. Simulation studies and a real data analysis are respectively performed in Section 6 and 7 to evaluate the performance of the proposed procedures. In Section 8, we further improve the aforementioned procedures by incorporating pairwise correlation information of the true null p-values. Some concluding remarks are made in Section 9 and proofs of some results are given in the Appendix.

2. Preliminary

In this section we present some basic notations and generalize the concept of the conventional fixed sequence procedure. Suppose $H_i, i = 1, \cdots, n$, are n null hypotheses which are pre-ordered based on prior knowledge and are to be tested based on their respective p-values $P_i, i = 1, \ldots, n$. Among these n hypotheses, let n_0 of them be true null hypotheses and n_1 be false. For notational convenience, let $\widehat{H}_i$ denote the i^{th} true null hypothesis and $\widehat{P}_i$ denote the corresponding p-value. Likewise, let $\tilde{H}_i$ denote the i^{th} false null hypothesis and $\tilde{P}_i$ denote the corresponding p-value. Define the FWER as the probability of incorrectly rejecting at least one true null hypothesis. In this paper, we assume that marginally, the true null p-values are stochastically greater than or equal to uniform distribution on [0,1], i.e., for $u \in [0, 1]$,

$$\Pr\{\widehat{P}_i \leq u\} \leq u,\ i = 1, \ldots, n_0, \tag{1}$$

and jointly, the p-values are arbitrarily dependent, i.e., the p-values do not have any type of known dependence structure.

Note that for the conventional fixed sequence procedure, a main drawback is that it does not allow any acceptance. In the following, we generalize the concept of the conventional fixed sequence multiple testing procedure so that even though some acceptances occur, the remaining hypotheses still have a chance to be tested.

Definition 1 *[Generalized Fixed Sequence Procedure] Given a function $\alpha(s,t)$ defined on $s = 0, \cdots, n-1$ and $t = 0, \cdots, n-1$, consider testing H_i, $i = 1, \ldots, n$. H_i is rejected iff $P_i \leq \alpha(s_{i-1}, t_{i-1})$, where s_{i-1} and t_{i-1} are, respectively, the numbers of rejected and accepted hypotheses when testing $H_1, \ldots, H_{i-1}$, with $s_0 = t_0 = 0$.*

The function $\alpha(s,t)$ is termed the critical value function throughout the manuscript.

Remark 1 It is easy to see that when $\alpha(s,t) = \alpha$ if $t = 0$ and $\alpha(s,t) = 0$ if $t > 0$, the generalized fixed sequence procedure reduces to the conventional fixed sequence procedure in Maurer *et al.* [6]. Besides, when the critical value

function is given in the form of

$$\alpha(s,t) = \begin{cases} \frac{\alpha}{k}, & \text{if } t = 0, \cdots, k-1, \\ 0, & \text{if } t = k, \cdots, n-1, \end{cases}$$

where k is a pre-specified integer with $0 < k < n$, the corresponding procedure reduces to the specific fixed sequence procedure introduced by Hommel and Kropf [14], which allows a pre-specified number k of acceptances. For the fallback procedure in Wiens and Dmitrienko [8], since the critical value for each hypothesis depends on the specific profile of previously tested hypotheses rather than the number of rejections or acceptances among the previous hypotheses, we cannot use a critical value function to define the critical values of the fallback procedure. That is, the fallback procedure is not a special case of the generalized fixed sequence procedure.

3. Main theoretical results

We will introduce in this section a sufficient condition on the critical value function for which the generalized fixed sequence procedure strongly controls the FWER at level α under arbitrary dependence. Before presenting the condition, for any configuration P of the tested hypotheses $(H_1, \ldots, H_n)$ and the corresponding p-values $(P_1, \ldots, P_n)$, we introduce a corresponding configuration described as follows: (i) the true null p-values $\widehat{P}_i, i = 1, \ldots, n_0$, have the same joint distribution as in the configuration P, (ii) the false null p-values $\tilde{P}_i = 0, i = 1, \ldots, n_1$, with probability 1, (iii) the order of the hypotheses to be tested, $H_1, \ldots, H_n$, is rearranged such that the false null hypotheses are tested before the true null hypotheses so that the order is $\tilde{H}_1, \cdots, \tilde{H}_{n_1}, \widehat{H}_1, \cdots, \widehat{H}_{n_0}$. This configuration is termed as a Dirac-Ordered (DO_P) configuration of P throughout the paper and the FWER under this configuration is denoted by $\text{FWER}_{\text{DO}_P}$. The following proposition shows that the FWER of the generalized fixed sequence procedure is larger under the Dirac-Ordered configuration DO_P than the original configuration P. Thus, in order to prove the FWER control of the generalized fixed sequence procedure, it is enough to show its FWER control under the Dirac-Ordered configuration.

Proposition 1 *Consider a generalized fixed sequence procedure with a critical value function $\alpha(s,t), s = 0, \cdots, n-1, t = 0, \cdots, n-1$. If $\alpha(s,t)$ is increasing in s and decreasing in t, then the FWER of this procedure under any configuration P, $FWER_P$, satisfies the following inequality:*

$$FWER_P \leq FWER_{DO_P}. \tag{2}$$

For the proof of Proposition 1, see Appendix.

Remark 2 The aforementioned Dirac-Ordered configuration is similar to the Dirac-Uniform configuration introduced in Finner and Roters [19]. The Dirac-Uniform configuration assumes independent p-values where the true null p-values are U(0, 1) and the false null p-values are zero with probability 1. However, in the Dirac-Ordered configuration, no independence assumption is made on the p-values but instead the order of hypotheses are taken into account.

Based on the Dirac-Ordered configuration, we now present a sufficient condition of a given generalized fixed sequence procedure to strongly control the FWER under arbitrary dependence.

Theorem 1 *Consider a generalized fixed sequence procedure with the critical value function $\alpha(s,t)$, where $\alpha(s,t)$ is increasing in s and decreasing in t.*
(i) The generalized fixed sequence procedure strongly controls the FWER at level α under arbitrary dependence if

$$\sum_{t=0}^{n-s-1} \alpha(s,t) \leq \alpha \quad for\ s = 0, \cdots, n-1. \tag{3}$$

(ii) If (3) becomes an equality for a given value s^* of s, then the FWER control is sharp in the sense that there exists a joint distribution for $(P_1, \cdots, P_n)$ for which the FWER of this procedure is exactly α.

Proof. By Proposition 1, it is enough to show that for any configuration P, $\text{FWER}_{\text{DO}_P} \leq \alpha$. With the probabilities evaluated under the Dirac-Ordered configuration DO_P of P, we have

$$\begin{aligned}
\text{FWER}_{\text{DO}_P} &= \Pr\{\widehat{P}_1 \leq \alpha(n_1, 0)\} \\
&\quad + \sum_{t=1}^{n_0-1} \Pr\{\widehat{P}_1 > \alpha(n_1, 0), \cdots, \widehat{P}_t > \alpha(n_1, t-1), \widehat{P}_{t+1} \leq \alpha(n_1, t)\} \\
&\leq \sum_{t=0}^{n_0-1} \Pr\{\widehat{P}_{t+1} \leq \alpha(n_1, t)\} \\
&\leq \sum_{t=0}^{n_0-1} \alpha(n_1, t) = \sum_{t=0}^{n-n_1-1} \alpha(n_1, t) \leq \alpha. \qquad (4)
\end{aligned}$$

The second inequality follows from (1) and the last one follows from (3).

For the proof of (ii), see Appendix. □

4. Procedures under arbitrary dependence

Theorem 1 provides a general approach for constructing FWER controlling fixed sequence procedures under arbitrary dependence. We can develop different kinds of fixed sequence procedures by choosing various kinds of critical value functions satisfying (3). In the following, we propose three special fixed sequence procedures based on three different types of critical value functions.

First, we consider the case where the critical value function $\alpha(s, t)$ increases with s but stays constant with respect to t. Thus, the procedure rewards the successful rejection of a hypothesis by increasing the critical values for the remaining hypotheses to be tested. But once the hypothesis fails to be rejected, no penalty towards those critical values is made.

PROCEDURE A1. Test the hypotheses according to the generalized fixed sequence procedure with the critical value function

$$\alpha(s, t) = \frac{\alpha}{n - s} \text{ for } 0 \leq s, t \leq n - 1. \qquad (5)$$

Remark 3 It is easy to see that Procedure A1 is similar to the Holm procedure in the sense that they have similar critical value functions. However, the Holm procedure does not require a pre-specified order of the null hypotheses and stops on the first accepted hypothesis; whereas, Procedure A1 requires the tested hypotheses to be pre-ordered but continues to test all the remaining hypotheses even if a hypothesis fails to be rejected.

Second, we consider the case where the critical value function $\alpha(s, t)$ is constant in s but decreasing in t. Specifically, we let $\alpha(s, t)$ decrease in t at a constant rate β. Thus, in contrast to Procedure A1, this procedure punishes the failure to reject a hypothesis by decreasing the critical values for the remaining hypotheses to be tested, but no reward is made for successful rejections.

PROCEDURE A2. Test the hypotheses according to the generalized fixed sequence procedure with critical value function

$$\alpha(s, t) = \frac{1 - \beta}{1 - \beta^n} \beta^t \alpha \text{ for } 0 \leq s, t \leq n - 1, \qquad (6)$$

where β is a pre-specified constant satisfying $0 \leq \beta < 1$.

Remark 4 In Procedure A2, when $\beta = 0$, its critical values are always equal to α for $t = 0$, the critical value of the conventional fixed sequence procedure. On the other hand, as β approaches to 1, its critical values approach to α/n, the critical value of the Bonferoni procedure.

Finally, we develop a fixed sequence procedure which combines the ideas of Procedures A1 and A2 so that this procedure rewards rejections and punishes acceptances. To construct its critical value function $\alpha(s, t)$, we start by assuming $\alpha(s, t)$ decreases by a constant c for each extra acceptance such that $\alpha(s, t-1) - \alpha(s, t) = c$ for $1 \leq t \leq n - 1$. Thus, $\alpha(s, t) = \alpha(s, 0) - tc$. In order to satisfy (3), it must be the case that

$$\alpha(s, 0) \leq \frac{\alpha}{n - s} + \frac{n - s - 1}{2} c. \qquad (7)$$

Furthermore, by taking the derivative of (7) with respect to s, one can see that $\alpha(s,0)$, and hence $\alpha(s,t)$, is increasing in s if and only if $c \leq 2\alpha/(n-s)^2$. By taking $c = 2\alpha/n^2$, we obtain the following procedure.

PROCEDURE A3. Test the hypotheses according to the generalized fixed sequence procedure with critical value function

$$\alpha(s,t) = \left(\frac{1}{n-s} + \frac{n-s-1}{n^2} - \frac{2t}{n^2}\right)\alpha \text{ for } 0 \leq s,t \leq n-1. \tag{8}$$

Remark 5 In Procedure A3, when a hypothesis is accepted, the critical values for the remaining hypotheses are reduced by constant $2\alpha/n^2$. On the other hand, when a hypothesis is rejected, the critical values increase by $\frac{\alpha}{(n-s+1)(n-s)} - \frac{\alpha}{n^2}$, which depends on the corresponding number of rejections s.

It is easy to see that $\alpha(s,t)$ defined in (5), (6) and (8) are all increasing in s, decreasing in t, and satisfy (3) with equality for all values of s. Thus, we have the following result.

Proposition 2 *Procedure A1, A2, and A3 all strongly control the FWER at level α under arbitrary dependence and their FWER controls are sharp in the sense that for each of these procedures, there exists a joint distribution for $(P_1, \cdots, P_n)$ for which its FWER is exactly equal to α.*

Remark 6 Although Bretz et al. [1] proposed a general graphical approach for developing new multiple testing procedures, in which the pre-specified transition coefficients are used to determine the allocation of the critical values to other hypotheses once a hypothesis is rejected, all of our proposed procedures cannot be described by using the graphical approach. For example, consider using Procedure A2 for testing two pre-ordered hypotheses, H_1 and H_2. If H_1 is rejected, then there is no critical value transferred from H_1 to H_2, which implies the transition coefficient from H_1 to H_2 is equal to zero. However, if H_1 is accepted, Procedure A2 implies that the critical value for H_2 will decrease. Obviously, the transition coefficient of zero from H_1 to H_2 cannot describe such outcome. Therefore, the graphical approach, although it is pretty general, cannot define Procedure A2. Similarly, for Procedure A1 and A3, neither of them can be described as a special case of the general graphical approach, since their critical values are also the functions of the numbers of rejections and acceptances.

5. Generalized fixed sequence procedure as a closed test

Suppose the critical value function $\alpha(s,t)$ is given. The generalized fixed sequence procedure with the critical value function $\alpha(s,t)$ can be illustrated as a closed testing procedure defined as follows. For any non-empty index set $I \subseteq \{1, \cdots, n\}$, consider an intersection hypothesis defined as $H_I = \bigcap_{i \in I} H_i$ and a local test based on the p-values for testing H_I: H_I is rejected if $P_j \leq \alpha(s^*_{j-1}, t^*_{j-1})$ for at least one $j \in I$, where

$$s^*_{j-1} = \sum_{k=1}^{j-1} I(k \notin I),$$
$$t^*_{j-1} = \sum_{k=1}^{j-1} I(k \in I),$$

and $I(\cdot)$ is indicator function. Here, $\alpha(s^*_{j-1}, t^*_{j-1})$ is termed as the local critical values of the above local test. Based on such local tests, we can define a closed testing procedure by using the closure principle (Marcus *et al.* [23]). For these two procedures, we have the following theorem.

Theorem 2 *The generalized fixed sequence procedure is equivalent to the aforementioned closed testing procedure for an arbitrary number of hypotheses.*

For the proof of Theorem 2, see Appendix.

To better evaluate the performance of the aforementioned three proposed procedures A1-A3, we illustrate them as closed testing procedures and compare their local critical values of testing intersection hypotheses with those of three commonly used multiple testing procedures, Holm's procedure, the conventional fixed sequence procedure, and the fallback procedure, which can also be illustrated as closed testing procedures. Table 1 lists the local critical values of

Table 1. Local critical values for any intersection hypotheses of three hypotheses using Procedure A1-A3 (PA1-PA3), conventional fixed sequence procedure (FS), fallback procedure (FB) and Holm's procedure (HM). For PA2, $\beta = 0.5$. For FB, initial weights are equal.

Hypothesis	PA1	PA2	PA3	FS	FB	HM
$H_1 \cap H_2 \cap H_3$	$\frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3}$	$\frac{4\alpha}{7}, \frac{2\alpha}{7}, \frac{\alpha}{7}$	$\frac{5\alpha}{9}, \frac{\alpha}{3}, \frac{\alpha}{9}$	$\alpha, 0, 0$	$\frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3}$	$\frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3}$
$H_1 \cap H_2$	$\frac{\alpha}{3}, \frac{\alpha}{3}$	$\frac{4\alpha}{7}, \frac{2\alpha}{7}$	$\frac{5\alpha}{9}, \frac{\alpha}{3}$	$\alpha, 0$	$\frac{\alpha}{3}, \frac{\alpha}{3}$	$\frac{\alpha}{2}, \frac{\alpha}{2}$
$H_1 \cap H_3$	$\frac{\alpha}{3}, \frac{\alpha}{2}$	$\frac{4\alpha}{7}, \frac{2\alpha}{7}$	$\frac{5\alpha}{9}, \frac{7\alpha}{18}$	$\alpha, 0$	$\frac{\alpha}{3}, \frac{2\alpha}{3}$	$\frac{\alpha}{2}, \frac{\alpha}{2}$
$H_2 \cap H_3$	$\frac{\alpha}{2}, \frac{\alpha}{2}$	$\frac{4\alpha}{7}, \frac{2\alpha}{7}$	$\frac{11\alpha}{18}, \frac{7\alpha}{18}$	$\alpha, 0$	$\frac{2\alpha}{3}, \frac{\alpha}{3}$	$\frac{\alpha}{2}, \frac{\alpha}{2}$
H_1	$\frac{\alpha}{3}$	$\frac{4\alpha}{7}$	$\frac{5\alpha}{9}$	α	$\frac{\alpha}{3}$	α
H_2	$\frac{\alpha}{2}$	$\frac{4\alpha}{7}$	$\frac{11\alpha}{18}$	α	$\frac{2\alpha}{3}$	α
H_3	α	$\frac{4\alpha}{7}$	α	α	α	α

the aforementioned six procedures in the case of three hypotheses. For the fallback procedure, the weights are set to be equal and for Procedure A2, β is set 0.5. It is easy to see from Table 1 that there is no procedure which is uniformly more powerful than others. For Procedure A1, its local critical values are smaller than those of Holm's procedure but are comparable with those of fallback procedure. For Procedure A2 and A3, their local critical values are generally larger in the most cases for higher-rank hypotheses and smaller for lower-rank hypotheses compared with the fallback procedure and Procedure A1. Thus, contrary to those two procedures, the local critical values for higher-rank hypotheses are always larger than the lower-rank hypotheses. Of course, it can also be seen from Table 1 that the proposed procedures are not α-exhaustive, i.e., not all intersection hypotheses are tested at the full α level. It implies a potential improvement upon the proposed procedures is possible.

6. Numerical findings

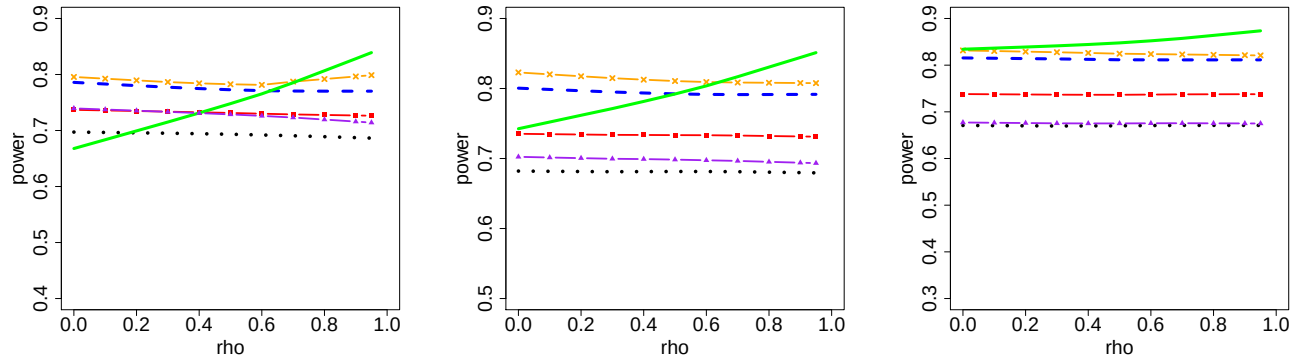
In this section, simulation studies were performed to investigate the power performance of the proposed Procedures A1-A3 compared to the existing Holm, conventional fixed-sequence and fallback procedures with respect to the correlation ρ among test statistics, the proportion π_0 of true null hypotheses among all tested hypotheses. For fixed sequence procedures, we consider a pre-specified testing order for which early hypotheses are a mixture of n_1 false null hypotheses and a fixed m true null hypotheses. When $m = 0$, it implies an ideal order for fixed sequence procedures in which all the false null hypotheses are ordered in front of true null hypotheses. When $m > 0$, it implies m true null hypotheses are mistakenly ordered compared to the aforementioned ideal order and we say there are m ordering mistakes in the testing order.

To simulate the values of average power (Westfall and Krishen [22]), which is the expected proportion of rejected false nulls among all false null hypotheses, for each of the aforementioned procedures, we first generated n dependent normal random variables $T_i \sim N(\mu_i, 1), i = 1, \ldots, n$, with n_0 $(= \pi_0 n)$ of the μ_i's being equal to 0 and the rest being equal to $d = \sqrt{10}$, and an equicorrelation matrix with correlation ρ. We then applied each aforementioned procedure to the generated data to test $H_i : \mu_i = 0$ against $K_i : \mu_i \neq 0$ simultaneously for $i = 1, \ldots, n$, at level $\alpha = 0.05$. The above steps were repeated for 100,000 times.

In the simulation, the p-value P_i corresponding to the hypothesis H_i was calculated by $P_i = 2(1 - \Phi(T_i)), i = 1, \cdots, n$, where $\Phi(\cdot)$ is the cdf of $N(0, 1)$. For those fixed sequence procedures, the order of the tested hypotheses was specified as follows: let the first n_1 hypotheses be false nulls, randomly insert m true null hypotheses among the n_1 false nulls indicating m ordering mistakes, and let the last $n_0 - m$ hypotheses be true nulls. Specifically, for the fallback procedure, the pre-specified weights $w_i, i = 1, \ldots, n$, for the n hypotheses are chosen to be an equally decreasing geometric sequence with a decreasing rate γ and a sum equal to one, that is, $w_i = \frac{\gamma^{i-1}(1-\gamma)}{1-\gamma^n}$. Note that when γ approaches to 1, the hypotheses are equally weighted and when $\gamma = 0$, the fallback procedure reduces to the conventional fixed sequence procedure. Finally, for notational convenience, the proposed Procedures A1-A3 are labeled PA1, PA2, and PA3, and the existing Holm, conventional fixed-sequence and fallback procedures are labeled HM, FS and FB, respectively.

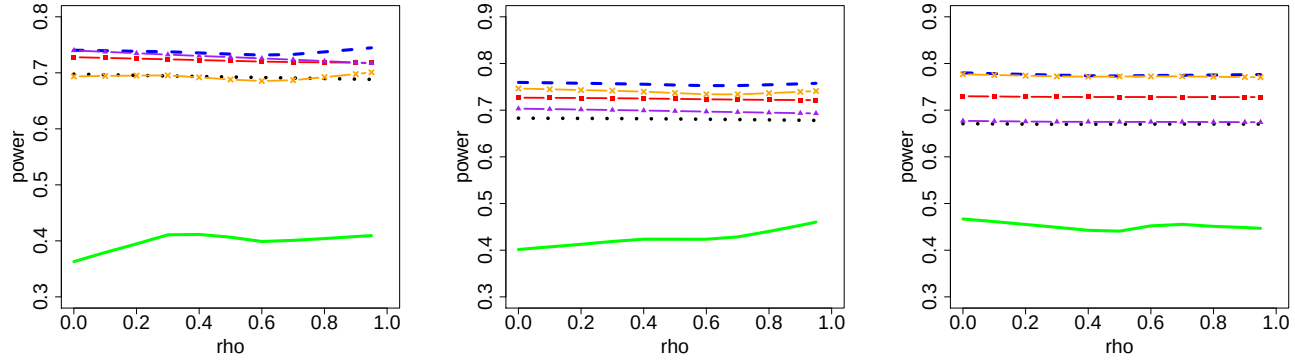
In the simulation, we set $n = 8$, $\pi_0 = 0.25, 0.5$ or 0.75, and $m = 0, 1$ or 2 for all aforementioned procedures. Specifically, we set $\beta = 0.5$ for Procedure A2 and $\gamma = 0.5$ for the fallback procedure.

Figures 1-3 present a comparison of the simulated average powers of the aforementioned 6 procedures. When $m = 0$, as seen from Figure 1, the power of conventional fixed sequence procedures is increasing in ρ and other 5 procedures perform steadily for different ρ. Among these 6 procedures, both Procedure A2 and the fallback procedure are comparably powerful; however, when ρ is large, they are slightly less powerful than the conventional fixed-sequence procedure. When



(a) $\pi_0 = 0.25$ **(b)** $\pi_0 = 0.5$ **(c)** $\pi_0 = 0.75$

Figure 1. Simulated average powers of 6 procedures (PA1 - · · ·; PA2 - – – –; PA3 - –■– ; FB - – × –; FS - ——; HM - –▲–) under equal correlation ρ for $n = 8, d = \sqrt{10}, m = 0, \alpha = 0.05, \beta = 0.5, \gamma = 0.5$.



(a) $\pi_0 = 0.25$ **(b)** $\pi_0 = 0.5$ **(c)** $\pi_0 = 0.75$

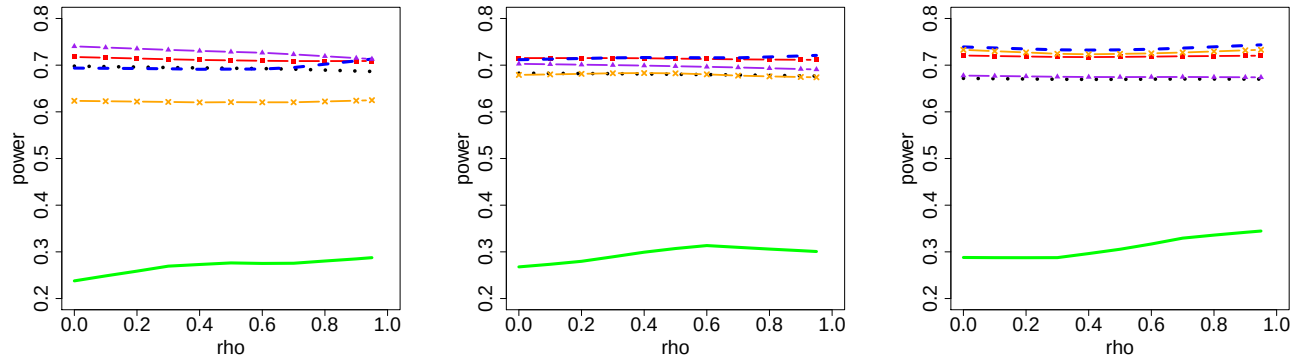
Figure 2. Simulated average powers of 6 procedures (PA1 - · · ·; PA2 - – – –; PA3 - –■– ; FB - – × –; FS - ——; HM - –▲–) under equal correlation ρ for $n = 8, d = \sqrt{10}, m = 1, \alpha = 0.05, \beta = 0.5, \gamma = 0.5$.

$m > 0$, as seen from Figure 2 and 3, the proposed Procedures A1-A3 perform very well for different values of π_0 and ρ. Among Procedures A1-A3, Procedure A2 or A3 are always slightly more powerful than Procedure A1 under different scenarios and Procedure A2 and A3 are comparable.

Summarizing the above observations, there is not an uniformly powerful procedure among the aforementioned six procedures, and except for the fixed sequence procedure, all other procedures are almost unaffected by the level of correlation of the test statistics. Compared to the conventional Holm's procedure with equal weights, Procedure A1 is slightly less powerful, but with unequal weights, their power relation may change. Also, in almost all scenarios, Procedure A2 is more powerful than Procedure 3 (except in the case of a relatively large number of ordering mistakes where the two procedures are comparable).

7. A clinical trial example

We revisited a hypertension trial example analyzed in Dmitrienko *et al.* [20]. The purpose of this clinical trial was to test the efficacy and safety of four doses of an investigational drug versus placebo. The four doses, from the lowest to highest doses, were respectively labeled D1, D2, D3, and D4, and the placebo was labeled P. The primary endpoint was the reduction in diastolic blood pressure (measured in mm Hg). There are 8 two-sided hypotheses including four dose-placebo contrasts and four pairwise contrasts. Since high doses were expected to be more efficacious than low doses, high dose-placebo contrasts (D4 vs. P, D3 vs. P) were tested before testing low dose-placebo contrasts (D2 vs. P, D1 vs. P).



(a) $\pi_0 = 0.25$ **(b)** $\pi_0 = 0.5$ **(c)** $\pi_0 = 0.75$

Figure 3. Simulated average powers of 6 procedures (PA1 - ···; PA2 - – – –; PA3 - –■–; FB - – × –; FS - ——; HM -–▲–) under equal correlation ρ for $n = 8, d = \sqrt{10}, m = 2, \alpha = 0.05, \beta = 0.5, \gamma = 0.5$.

Table 2. Comparison results of 6 procedures in the hypertension trial example (P = Placebo and D1-D4 denote four doses of the investigational drug). PA1 = proposed Procedure A1, $PA2_1$ - $PA2_3$ = proposed Procedure A2 with $\beta = 0.1, 0.5$ and 0.9, PA3 = proposed Procedure A3, HM = Holm procedure, FS = conventional fixed sequence procedure, FB_1 - FB_3 = fallback procedure with $\gamma = 0.1, 0.5$ and 0.9. For FB_1 - FB_3, γ denotes the equally decreasing rate of weights assigned to 8 hypotheses. The overall Type I error rate is $\alpha = 0.05$.

	Raw p-value	PA1	$PA2_1$	$PA2_2$	$PA2_3$	PA3	HM	FS	FB_1	FB_2	FB_3
D4-P	0.0008	R	R	R	R	R	R	R	R	R	R
D3-P	0.0135	NR	R	R	NR	R	NR	R	R	R	NR
D2-P	0.0197	NR	R	R	NR	NR	NR	R	R	R	NR
D1-P	0.7237	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
D4-D1	0.0003	R	R	R	R	R	R	NR	NR	R	R
D4-D2	0.2779	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
D3-D1	0.0054	R	NR	R	R	NR	R	NR	NR	NR	R
D3-D2	0.8473	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Rejection number		3	4	5	3	3	3	3	3	4	3

After testing these four dose-placebo comparisons, four pairwise comparisons were tested in an order of D4 vs D1, D4 vs. D2, D3 vs. D1, and D3 vs. D2. We pre-specified $\alpha = 0.05$ and applied the three newly proposed Procedures A1-A3 and three existing procedures Holm, conventional fixed sequence and fallback procedures to this example. Same as in Section 6, the pre-specified weights for fallback procedure are chosen to be a decreasing sequence with equally decreasing rate γ and γ is set to 0.1, 0.5 or 0.9, respectively. For Procedure A2, β is also set to 0.1, 0.5 or 0.9, respectively. Table 2 lists the raw p-values of the 8 hypotheses and the test results using the aforementioned six procedures.

As seen from Table 2, Procedure A2 with $\beta = 0.5$ performs the best rejecting 5 null hypotheses at level 0.05. But when $\beta = 0.1$ and 0.9, it rejects 4 and 3 hypotheses, respectively. In contrast, the conventional fixed sequence procedure, Holm procedure, Procedure A1, and Procedure A3 only reject 3 null hypotheses. For the fallback procedure, its testing results depend on the pre-specified weights. When the equally decreasing rate γ of weights is set to 0.1, 0.5 or 0.9, it rejects 3, 4 or 3 hypotheses, respectively.

8. Further improvement

In the preceding sections, only the marginal distributions of the null p-values are used when developing the newly introduced Procedures A1-A3. However, in practice, the null p-values often have a known common pairwise joint distribution, and it would be worthwhile to consider further improving the aforementioned procedures by explicitly

utilizing such additional dependence information, which could potentially produce more powerful FWER controlling procedures than Procedures A1-A3. So, with that in mind, we present some improved results here under the following assumption.

Assumption 1 *The null p-values* $\widehat{P}_1, \ldots, \widehat{P}_{n_0}$ *have a known common pairwise joint distribution function* $F(u,v) = \Pr\left(\widehat{P}_i \le u, \widehat{P}_j \le v\right)$.

Under Assumption 1, Theorem 1 can be further strengthened as follows.

Theorem 3 *Under Assumption 1, the generalized fixed sequence procedure with critical value function* $\alpha(s,t)$ *strongly controls the FWER at level* α *if for any* $0 \le s, t \le n-1$,

$$\sum_{t=0}^{n-s-1} \alpha(s,t) - \sum_{t=1}^{n-s-1} F\left(\alpha(s,t-1), \alpha(s,t)\right) \le \alpha, \tag{9}$$

where $\alpha(s,t)$ *is increasing in* s *and decreasing in* t.

For proof of Theorem 3, see Appendix.

Remark 7 The amount of improvement of the critical values of the aforementioned procedure depends on the pairwise joint cdf $F(u,v)$. Assume $\widehat{P}_i \sim U(0,1), i = 1, \ldots, n_0$, then under perfect positive correlation where $F(\alpha(s,t-1), \alpha(s,t)) = \alpha(s,t)$, (9) reduces to $\alpha(s,t) \le \alpha$, which is a remarkable improvement on (3). On the other hand, under independence where $F(\alpha(s,t-1), \alpha(s,t)) = \alpha(s,t-1)\alpha(s,t)$, there is only a limited improvement.

Based on Theorem 3, we can respectively develop improved versions of Procedures A1-A3 as follows.

PROCEDURE B1. Test the hypotheses according to the generalized fixed sequence procedure with critical value function $\alpha(s,t) = \alpha(s,0)$ for $s,t = 0, \ldots, n-1$ and $\alpha(s,0)$ satisfies the following equation for $s = 0, \ldots, n-1$,

$$(n-s)\alpha(s,0) - (n-s-1)F(\alpha(s,0), \alpha(s,0)) = \alpha. \tag{10}$$

PROCEDURE B2. Test the hypotheses according to the generalized fixed sequence procedure with critical value function $\alpha(s,t) = \alpha(0,0)\beta^t$ for $s,t = 0, \ldots, n-1$ and $\alpha(0,0)$ satisfies the following equation,

$$\frac{1-\beta^n}{1-\beta}\alpha(0,0) - \sum_{t=0}^{n-1} F\left(\alpha(0,0)\beta^{t-1}, \alpha(0,0)\beta^t\right) = \alpha, \tag{11}$$

where β is a pre-specified constant satisfying $0 \le \beta < 1$.

PROCEDURE B3. Test the hypotheses according to the generalized fixed sequence procedure with critical value function $\alpha(s,t) = \alpha(s,0) - 2t\alpha/n^2$ for $s,t = 0, \ldots, n-1$ and $\alpha(s,0)$ satisfies the following equation for $s = 0, \ldots, n-1$,

$$(n-s)\left(\alpha(s,0) - \frac{(n-s-1)\alpha}{n^2}\right) - \sum_{t=1}^{n-s-1} F\left(\alpha(s,0) - \frac{2(t-1)\alpha}{n^2}, \alpha(s,0) - \frac{2t\alpha}{n^2}\right) = \alpha. \tag{12}$$

It is easy to see that the critical value functions for Procedures B1-B3 are all decreasing in t. And, it can be shown that (10)-(12) all have solutions for any cdf $F(u,v)$, and even have unique solutions if $F(u,v)$ is assumed to satisfy certain conditions, for example, $F(u,v)$ is the cdf of a bivariate normal distribution. There is no guarantee that a solution for $\alpha(s,0)$ in (9) and (11) is increasing in s. If it is not, a minor adjustment of $\alpha(s,0)$ can always be made to force $\alpha(s,0)$ to be increasing, although the resulting $\alpha(s,t)$ becomes a little smaller. For simplicity, in the following discussions, we assume that (10)-(12) all have unique solutions and the resulting critical value functions $\alpha(s,t)$ of Procedures B1-B3 are increasing in s and decreasing in t. Finally, we need to point out that it is typically not possible to obtain closed form solutions for (10)-(12). Instead, these solutions can be approximated numerically by using the bisection method [21].

Remark 8 The critical value functions in Procedures B1-B3 maintain the same monotonicity properties as their corresponding Procedures A1-A3, respectively. For example, the critical value function in Procedure B1 is increasing in s and constant in t, and the critical value function in Procedure B2, like Procedure A2, decreases by the constant rate β for every unit increase in t. Also, for Procedures B1-B3, all of their critical value functions $\alpha(s,t)$ satisfy (9). Thus, by Theorem 3, we have the following result holds.

Table 3. Critical values (percentage change) of Procedure A1 and Procedure B1 with $n = 8$.

s	Procedure A1	Procedure B1		
		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
0	0.006250	0.006336 (1.4%)	0.006756 (8.1%)	0.008794 (40.7%)
1	0.007143	0.007250 (1.5%)	0.007813 (9.4%)	0.010052 (40.7%)
2	0.008333	0.008469 (1.6%)	0.009055 (8.7%)	0.011719 (40.6%)
3	0.01	0.010178 (1.8%)	0.010894 (9.0%)	0.013978 (39.8%)
4	0.0125	0.012746 (2.0%)	0.013643 (9.1%)	0.017266 (38.1%)
5	0.016666	0.017027 (2.2%)	0.018178 (9.1%)	0.0224 (34.4%)
6	0.025	0.025546 (2.2%)	0.026958 (7.8%)	0.031362 (25.5%)
7	0.05	0.05 (0.0%)	0.05 (0.0%)	0.05 (0.0%)

Table 4. Critical values (percentage change) of Procedure A2 and Procedure B2 with $n = 8$ and $\beta = 0.5$.

t	Procedure A2	Procedure B2		
		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
0	0.025098	0.025631 (2.1%)	0.027171 (8.3%)	0.033173 (32.2%)
1	0.012549	0.012815 (2.1%)	0.013586 (8.3%)	0.016586 (32.2%)
2	0.006275	0.006408 (2.1%)	0.006793 (8.3%)	0.008293 (32.2%)
3	0.003137	0.003204 (2.1%)	0.003396 (8.3%)	0.004147 (32.2%)
4	0.001569	0.001602 (2.1%)	0.001698 (8.3%)	0.002073 (32.2%)
5	0.000784	0.000801 (2.1%)	0.000849 (8.3%)	0.001037 (32.2%)
6	0.000392	0.000400 (2.1%)	0.0004252 (8.3%)	0.000518 (32.2%)
7	0.000196	0.000200 (2.1%)	0.000212 (8.3%)	0.000259 (32.2%)

Proposition 3 *Under Assumption 1, Procedures B1-B3 strongly control the FWER at* α.

In order to show the improvements of critical values of Procedures B1-B3 over Procedures A1-A3, we performed some numerical calculations to illustrate the gains of critical values of Procedures B1-B3 over Procedures A1-A3, respectively. We consider n two-sided hypothesis tests and assume that any pair of test statistics associated with true null hypotheses follows bivariate normal distribution with common pairwise correlation ρ. We set the parameter β in Procedure A2 and B2 to be $\beta = 0.5$. Table 3 summarizes the numerical results of calculating the critical values of Procedure A1 and B1 for $n = 8, \rho = 0.2, 0.5, 0.8$ and the improvement percentage of the critical values of Procedure B1 over A1. Table 4 and 5 show the similar comparison results for Procedure A2 vs B2 with the same values of n and ρ as in Table 3, and Procedure A3 vs B3 with $n = 5, \rho = 0.5$ and 0.8. As seen from these three tables, when ρ is small, the percentages of improvement of critical values are pretty small and are generally no more than 2%. However, when ρ is large, the improvements are remarkable and some are even over 30%.

9. Conclusion

The main focus of this paper has been to advance the theory and methods of fixed sequence multiple testing for controlling the FWER. We have introduced a generalized fixed sequence procedure and given sufficient conditions for its FWER control under arbitrary dependence. We have proposed several new fixed sequence procedures by considering different critical value functions. Through simulation studies, it has been shown that some advantages of our proposed generalized fixed sequence procedures over the existing FWER controlling procedures in some situations can be achieved. When the pairwise joint distributions of the true null p-values are known, we have improved the aforementioned procedures by incorporating the distributional information into the construction of these procedures while maintaining the control of the FWER. Specifically, in the case of bivariate normal distribution with common correlation, we have numerically shown improvements of the critical values of the improved procedures over the aforementioned procedures.

To use the fixed sequence methods, prior knowledge of the ordering of the tested hypotheses is required. When the ordering is completely correct, i.e., the false null hypotheses are ordered ahead of the true null hypotheses, even the

Table 5. Critical values (percentage change) of Procedure A3 and Procedure B3 with $n = 5$.

Procedure A3					
s \ t	0	1	2	3	4
0	0.018	0.014	0.01	0.006	0.002
1	0.0185	0.0145	0.0105	0.0065	-
2	0.0207	0.0167	0.0127	-	-
3	0.027	0.023	-	-	-
4	0.05	-	-	-	-
Procedure B3 with $\rho = 0.5$					
s \ t	0	1	2	3	4
0	0.02 (11.1%)	0.016 (14.3%)	0.012 (20%)	0.008 (33.3%)	0.004 (100%)
1	0.02 (8.1%)	0.016 (10.3%)	0.012 (14.3%)	0.008 (23.1%)	-
2	0.0222 (7.2%)	0.0182 (9.00%)	0.0142 (11.8%)	-	-
3	0.0289 (7.0%)	0.0249 (8.3%)	-	-	-
4	0.05 (0.0%)	-	-	-	-
Procedure B3 with $\rho = 0.8$					
s \ t	0	1	2	3	4
0	0.0219 (21.7%)	0.0179 (27.9%)	0.0139 (39.0%)	0.0099 (65.0%)	0.0059 (195.0%)
1	0.0232 (25.4%)	0.0192 (32.4%)	0.0152 (44.8%)	0.0112 (72.3%)	-
2	0.0264 (27.5%)	0.0223 (33.5%)	0.0184 (44.9%)	-	-
3	0.0333 (23.3%)	0.0293 (27.4%)	-	-	-
4	0.05 (0.0%)	-	-	-	-

conventional fixed sequence procedure, which does not allow any acceptance, has a natural advantage over the existing p-value based stepwise FWER controlling procedures such as the Holm and Hochberg procedures. However, when the ordering is not completely correct, the conventional fixed sequence procedure usually loses its edge over those stepwise procedures, whereas our proposed fixed sequence procedures can still perform well. Of course, when the ordering information is completely incorrect, our proposed fixed sequence procedures no longer have the advantage over those p-value based stepwise procedures. Therefore, a natural extension might be to use a combination of the a-priori ordering information and the p-values to order the hypotheses to be tested and then develop FWER controlling procedures based on such ordering. Moreover, it was pointed out in Section 5 that the proposed procedures are not α-exhaustive, and thus it is possible to develop newer procedures that are more powerful than the proposed procedures.

Acknowledgments

The research of the second author was supported in part by NSF grants DMS-1006021 and DMS-1309162.

Appendix

A.1 Proof of Proposition 1

For any configuration P, let E_1 and E_2 respectively denote the events of no false rejection under configurations P and DO_P for a generalized fixed sequence procedure. E_1 and E_2 both can be expressed as set functions in terms of the p-values,

$$E_1(P_1, \cdots, P_n) = \bigcap_{j=1}^{n_0} \left\{ \widehat{P}_j > \alpha(\widehat{s}_{j-1}, \widehat{t}_{j-1}) \right\} \tag{13}$$

and

$$E_2(P_1, \cdots, P_n) = \bigcap_{j=1}^{n_0} \left\{ \widehat{P}_j > \alpha(\tilde{s}_{j-1}, \tilde{t}_{j-1}) \right\}, \tag{14}$$

where $\widehat{s}_j$ (or $\tilde{s}_j$) and $\widehat{t}_j$ (or $\tilde{t}_j$) are the total numbers of rejections and acceptances after testing $\widehat{H}_j$ under configuration P (or DO_P), respectively.

It is easy to see that when event E_2 occurs, $\tilde{s}_j = n_1$ and $\tilde{t}_j = j$ for any $j = 1, \ldots, n_0$. Thus, if both events E_1 and E_2 occur, we have $\widehat{s}_j \leq n_1$ and $\widehat{t}_j \geq j$ and hence $\alpha(\widehat{s}_{j-1}, \widehat{t}_{j-1}) \leq \alpha(\tilde{s}_{j-1}, \tilde{t}_{j-1})$ by using the fact that $\alpha(s,t)$ is increasing in s and decreasing in t.

Thus, we have

$$E_2(P_1, \cdots, P_n) \subseteq E_1(P_1, \cdots, P_n) \tag{15}$$

and hence

$$\text{FWER}_P \leq 1 - \Pr\{E_1(P_1, \cdots, P_n)\} \leq 1 - \Pr\{E_2(P_1, \cdots, P_n)\} = \text{FWER}_{\text{DO}_P}, \tag{16}$$

the desired result is proved. □

A.2 Proof of Theorem 1.(ii)

Considering a Dirac-Ordered configuration where the number of false null hypotheses is $n_1 = s^*$, the false null p-values $\tilde{P}_i = 0$ with probability 1 for $i = 1, \ldots, n_1$, and the order of the hypotheses is $\tilde{H}_1, \cdots, \tilde{H}_{n_1}, \widehat{H}_1, \cdots, \widehat{H}_{n_0}$. Thus,

$$\sum_{j=1}^{n_0} \alpha(n_1, j-1) = \sum_{j=0}^{n-s^*-1} \alpha(s^*, j) = \alpha. \tag{17}$$

In the following, we will specify the joint distribution of the true null p-values, $\widehat{P}_1, \ldots, \widehat{P}_{n_0}$, such that $\widehat{P}_i \sim U(0,1)$ and the FWER of the generalized fixed sequence procedure under the aforementioned Dirac-Ordered configuration is exactly α. Let $\widehat{P}_1 \sim U(0,1)$ and for $j = 2, \ldots, n_0$, let

$$\widehat{P}_j \sim \begin{cases} U\left(0, 1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1)\right), & \text{if } \widehat{P}_i > \alpha(n_1, i-1) \text{ for } i = 1, \ldots, j, \\ U\left(1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1), 1\right), & \text{otherwise.} \end{cases} \tag{18}$$

Note that by (3), we have $\alpha(n_1, j-1) \leq 1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1)$ for $j = 1, \ldots, n_0$. Thus, when event $\{\widehat{P}_j \leq \alpha(n_1, j-1)\}$ occurs, event $\bigcap_{i=1}^{j-1} \{\widehat{P}_i > \alpha(n_1, i-1)\}$ occurs with probability 1. Hence, we have

$$\begin{aligned} & \Pr\left\{ \widehat{P}_1 > \alpha(n_1, 0), \ldots, \widehat{P}_{j-1} > \alpha(n_1, j-2), \widehat{P}_j \leq \alpha(n_1, j-1) \right\} \\ = \quad & \Pr\left\{ \widehat{P}_j \leq \alpha(n_1, j-1) \right\}. \end{aligned} \tag{19}$$

In the following, we use mathematical induction to show that $\widehat{P}_j \sim U(0,1)$ for $j = 1, \ldots, n_0$. Trivially, $\widehat{P}_1 \sim U(0,1)$. Assume $\widehat{P}_i \sim U(0,1)$ for $i = 1, \ldots, j-1$. Thus,

$$\begin{aligned}
&\Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} = 1 - \Pr\left\{\bigcup_{i=1}^{j-1}\left\{\widehat{P}_i \le \alpha(n_1, i-1)\right\}\right\} \\
&= 1 - \sum_{i=1}^{j-1} \Pr\left\{\widehat{P}_1 > \alpha(n_1, 0), \ldots, \widehat{P}_{i-1} > \alpha(n_1, i-2), \widehat{P}_i \le \alpha(n_1, i-1)\right\} \\
&= 1 - \sum_{i=1}^{j-1} \Pr\left\{\widehat{P}_i \le \alpha(n_1, i-1)\right\} = 1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1). \qquad (20)
\end{aligned}$$

The third equality follows from (19) and the last one follows from the assumption of $\widehat{P}_i \sim U(0,1)$ for $i = 1, \ldots, j-1$.

For $0 \le u \le 1 - \sum_{i=1}^{j-1} \alpha(n_1, j-1)$, by (18) and (20), we have

$$\begin{aligned}
\Pr\left\{\widehat{P}_j \le u\right\} &= \Pr\left\{\widehat{P}_j \le u \,\middle|\, \bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} \Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} \\
&= \frac{u}{1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1)} \Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} \\
&= u. \qquad (21)
\end{aligned}$$

For $1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1) < u \le 1$, similarly, by (18) and (20), we have

$$\begin{aligned}
\Pr\left\{\widehat{P}_j \le u\right\} &= \Pr\left\{\widehat{P}_j \le u \,\middle|\, \bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} \Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} \\
&\quad + \Pr\left\{\widehat{P}_j \le u \,\middle|\, \bigcup_{i=1}^{j-1}\left\{\widehat{P}_i \le \alpha(n_1, i-1)\right\}\right\} \left(1 - \Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\}\right) \\
&= \Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\} \\
&\quad + \frac{u - (1 - \sum_{i=1}^{j-1} \alpha(n_1, i-1))}{\sum_{i=1}^{j-1} \alpha(n_1, i-1)} \left(1 - \Pr\left\{\bigcap_{i=1}^{j-1}\left\{\widehat{P}_i > \alpha(n_1, i-1)\right\}\right\}\right) \\
&= u. \qquad (22)
\end{aligned}$$

Combining (21) and (22), we have that $\widehat{P}_j \sim U(0,1)$ for $j = 1, \ldots, n_0$.

Finally, under the aforementioned Dirac-Ordered configuration specified as above, we have

$$\begin{aligned}
\text{FWER} &= \sum_{j=1}^{n_0} \Pr\left\{\widehat{P}_1 > \alpha(n_1, 0), \ldots, \widehat{P}_{j-1} > \alpha(n_1, j-2), \widehat{P}_j \le \alpha(n_1, j-1)\right\} \\
&= \sum_{j=1}^{n_0} \Pr\left\{\widehat{P}_j \le \alpha(n_1, j-1)\right\} = \sum_{j=1}^{n_0} \alpha(n_1, j-1) = \alpha.
\end{aligned}$$

The second equality follows from (19), the third follows from $\widehat{P}_j \sim U(0,1)$, and the last one follows from (17). □

A.3 Proof of Theorem 2

To prove these two procedures are equivalent, we only need to prove the following two results for any individual hypothesis $H_i, i = 1, \ldots, n$:

Result 1. If H_i is accepted by the generalized fixed sequence procedure, then it is also accepted by the closed testing procedure.

Result 2. If H_i is accepted by the closed testing procedure, then it is also accepted by the generalized fixed sequence procedure.

Proof of Result 1. Let $I = \{j : P_j > \alpha(s_{j-1}, t_{j-1})\}$ and consider $\widetilde{H}_i = \bigcap_{j\in I} H_j$. Since H_i is accepted by the generalized fixed sequence procedure, $P_i > \alpha(s_{i-1}, t_{i-1})$ and thus H_i is contained in $\widetilde{H}_i$. Note that

$$s_{i-1} = \sum_{j=1}^{i-1} I\{P_j \le \alpha(s_{j-1}, t_{j-1})\} = \sum_{j=1}^{i-1} I\{j \notin I\} = s^*_{i-1}.$$

Similarly, we have $t^*_{i-1} = t_{i-1}$. Thus, $\alpha(s^*_{i-1}, t^*_{i-1}) = \alpha(s_{i-1}, t_{i-1})$ and $P_i > \alpha(s^*_{i-1}, t^*_{i-1})$. Therefore, $\widetilde{H}_i$ is not rejected by the local test and thus H_i is accepted by the closed testing procedure.

Proof of Result 2. Since H_i is accepted by the closed testing procedure, thus there exists an index set I with $i \in I$ such that the intersection hypothesis, $\bigcap_{j\in I} H_j$ is not rejected by the corresponding local test. Thus, for any $j \in I$, we have $P_j > \alpha(s^*_{j-1}, t^*_{j-1})$. In the following, we use mathematical induction to prove that $s^*_{j-1} \ge s_{j-1}$ and $t^*_{j-1} \le t_{j-1}$ for each $j = 1, \ldots, n$. It is easy to see that the result holds for $j = 1$. Assume that the result holds for $j = k$. We prove that the result also holds for $j = k + 1$. Note that

$$s^*_k = \sum_{j=1}^{k} I\{j \notin I\} \ge \sum_{j=1}^{k} I\{P_j \le \alpha(s^*_{j-1}, t^*_{j-1})\} \ge \sum_{j=1}^{k} I\{P_j \le \alpha(s_{j-1}, t_{j-1})\} = s_k.$$

The first inequality follows from the fact that $P_j \le \alpha(s^*_{j-1}, t^*_{j-1})$ implies $j \notin I$ for each $j = 1, \ldots, n$ and the second follows from the assumption of induction. Similarly, we can prove $t^*_k \le t_k$. Hence, this result holds by induction. Thus, $P_i > \alpha(s^*_{i-1}, t^*_{i-1}) \ge \alpha(s_{i-1}, t_{i-1})$ follows due to monotonicity of the critical value function. Therefore, H_i is accepted by the generalized fixed sequence procedure. □

A.4 Proof of Theorem 3

Based on Proposition 1, it is enough to show that for any configuration P, the FWER of the generalized fixed sequence procedure under the corresponding Dirac-Ordered configuration is less than or equal to α. With the probabilities evaluated under the Dirac-Ordered configuration, we have

$$\begin{aligned}
\text{FWER}_{\text{DO}_P} &= \Pr\{\widehat{P}_1 \le \alpha(n_1, 0)\} \\
&\quad + \sum_{t=1}^{n_0-1} \Pr\{\widehat{P}_1 > \alpha(n_1, 0), \cdots, \widehat{P}_t > \alpha(n_1, t-1), \widehat{P}_{t+1} \le \alpha(n_1, t)\} \\
&\le \Pr\{\widehat{P}_1 \le \alpha(n_1, 0)\} + \sum_{t=1}^{n_0-1} \Pr\{\widehat{P}_t > \alpha(n_1, t-1), \widehat{P}_{t+1} \le \alpha(n_1, t)\} \\
&= \sum_{t=0}^{n_0-1} \Pr\{\widehat{P}_{t+1} \le \alpha(n_1, t)\} - \sum_{t=1}^{n_0-1} \Pr\{\widehat{P}_t \le \alpha(n_1, t-1), \widehat{P}_{t+1} \le \alpha(n_1, t)\} \\
&\le \sum_{t=0}^{n_0-1} \alpha(n_1, t) - \sum_{t=1}^{n_0-1} F(\alpha(n_1, t-1), \alpha(n_1, t)) \\
&\le \sum_{t=0}^{n-n_1-1} \alpha(n_1, t) - \sum_{t=1}^{n-n_1-1} F(\alpha(n_1, t-1), \alpha(n_1, t)) \le \alpha. \qquad (23)
\end{aligned}$$

The second inequality follows from Assumption 1 and the last one follows from (9). □

References

1. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* 2009; **28**:586-604.
2. Dmitrienko A, Bretz F, Westfall P, Troendle J, Wiens B, Tamhane, A, Hsu, J. Multiple testing methodology. In *Multiple Testing Problems in Pharmaceutical Statistics*, Dmitrienko A, Tamhane A, and Bretz F(eds). Chapman & Hall/CRC Press: New York, 2009; 35-98.
3. Dmitrienko A, D'Agostino R, Huque M. Key multiplicity issues in clinical drug development. *Statistics in Medicine* 2013; **32**:1079–1111.
4. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979; **6**: 65–70.
5. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; **75**: 800–802.
6. Maurer W, Hothorn L, Lehmacher W. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. *Biometrie in der Chemisch-harmazeutischen Industrie* 1995; **6**:3–18.

7. Wiens B. A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics* 2003; **2**: 211–215.
8. Wiens B, Dmitrienko A. The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics* 2005; **15**: 929–942.
9. Li J, Mehrotra D. An efficient method for accommodating potentially underpowered primary endpoints. *Statistics in Medicine* 2008; **27**: 5377–5391.
10. Huque M, Alosh M. A flexible fixed sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials. *Journal of Statistical Planning and Inference* 2008; **138**: 321–335.
11. Hommel G, Bretz F. Aesthetics and power considerations in multiple testing - a contradiction? *Biometrical Journal* 2008; **50**: 657–666.
12. Rosenbaum P. Testing hypotheses in order. *Biometrika* 2008; **95**: 248–252.
13. Millen B, Dmitrienko A. Chain procedures: a class of flexible closed testing procedures with clinical trial applications. *Statistics in Biopharmaceutical Reseach* 2011; **3**: 14–30.
14. Hommel G, Kropf S. Testing for differentiation in gene expression using a data-driven order or weights for hypotheses. *Biometrical Journal* 2005; **47**: 554–562.
15. Wiens B, Dmitrienko A. On selecting a multiple comparison procedure for analysis of a clinical trial: fallback, fixed sequence, and related procedures. *Statistics in Biopharmaceutical Research* 2010; **2**: 22–32.
16. Alosh M, Huque M. A flexible strategy for testing subgroups and overall population. *Statistics in Medicine* 2009; **28**: 3–23.
17. Alosh M, Huque M. A consistency-adjusted alpha-adaptive strategy for sequential testing. *Statistics in Medicine* 2010; **29**: 1559–1571.
18. Tu Y, Cheng B, Cheung Y. A note on confidence bounds after fixed-sequence multiple tests. *Journal of Statistical Planning and Inference* 2012; **142**: 2993–2998.
19. Finner H, Roters M. On the false discovery rate and expected type I errors. *Biometrical Journal* 2001; **43**: 985–1005.
20. Dmitrienko A, Wiens B, Westfall P. Fallback tests in dose-response clinical trials. *Journal of Biopharmaceutical Statistics* 2006; **16**: 745–755.
21. Arfken G. *Mathematical Methods for Physicists.* Academic Press: Orlando, 1985.
22. Westfall P, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *Journal of Statistical Planning and Inference* 2001; **99**: 25–40.
23. Marcus R., Peritz E., Gabriel K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 1976; **63**: 655–660.