Reverse Graphical Approaches for Multiple Test Procedures (DRAFT)

Jiangtao Gou

May 7, 2020

Abstract

The graphical approach has been proposed as a general framework for clinical trial designs involving multiple hypotheses, where decisions are made only based on the observed marginal p-values. The graphical approach starts from a graph that includes all hypotheses as vertices, and gradually remove some vertices when their corresponding hypotheses are rejected. In this paper, we propose a reverse graphical approach, which starts from a set of singleton graphs, and gradually add vertices into graphs until rejection of a set of hypotheses is made. Proofs of familywise error rate control are provided. A simulation study is conducted for statistical power analysis, and a case study is included to illustrate how the proposed approach can be applied to a clinical trial.

Keywords. Graphical approach, Hochberg procedure, Holm procedure, Multiple tests.

1 Introduction

As one of the recent advancements in confirmatory clinical trial design with multiple endpoints, the graphical approach to multiple testing has gained much popularity due to its intuitive and flexible way for constructing, visualizing and communicating a variety of multiple testing procedures Bretz et al. (2009); Burman et al. (2009). The Holm multiple testing procedure, proposed thirty years before the graphical approach, is a simple application of the graphical approach, where the directed graph is complete and the transition weights are all equal Holm (1979). The Holm procedure is often referred to as a step-down procedure, since the observed p-values are firstly ordered from lowest to highest and then compared with their corresponding critical boundaries. There is another widely-used multiple testing procedure called the Hochberg procedure Hochberg (1988). As a step-up procedure, the Hochberg procedure shares exactly the same critical values with the Holm procedure, but the p-values are ordered from highest to lowest in the Hochberg procedure, shifting to an opposite direction comparing with the Holm procedure. A natural question that arises from the comparison between the Holm and Hochberg procedure is, what if we switch the direction of applying the graphical approach? Will the resulting methods still correctly control the type I error rate? To answer these questions, we develop a reverse graphical approach to multiple hypotheses testing. The relation between the graphical approach and reverse graphical approach is similar with the relation between the Holm and Hochberg procedures.

Although the Holm step-down and Hochberg step-up procedures look similar, they are based on different p-value-based tests of the global null hypothesis: the Holm method is a closed testing procedure using the Bonferroni test Dunn (1961) and the Hochberg method is a closed testing procedure using the Simes test Simes (1986). Here the closed testing procedures are constructed using a general method called the closure method Marcus et al. (1976), which guarantees some good properties like coherence Gabriel (1969). Xi and Bretz (2019) provided an explanation of the classical Hochberg procedure using a symmetric complete directed graph with equal vertex weights and equal transition weights. As an extensive generalization of the Hochberg procedure, the reverse graphical approach we developed allows unequal vertex weights and unequal transition weights, and brings in much flexibility over the classical Hochberg procedure. Comparing with the graphical approach, which starts from the biggest graph and gradually reduces the graph, the reverse graphical approach starts from the smallest graphs, which contain only one vertex, and systematically adds vertices back to the graphs if necessary. We include the descriptions of the reverse graphical approaches in Section 2, along with the proof of the type I error control under some assumptions. In Section 3, we investigate the proposed reverse graphical approaches using numerical calculation, and provide examples for illustration. Concluding remarks and further discussion are included in Section 4.

2 Reverse Graphical Approaches

Both the graphical approach and the proposed reverse graphical approach strongly control the familywise error rate (FWER) at level α, which is the probability of rejecting at least one true null hypothesis under arbitrary configurations in which at least one null hypothesis is true Hochberg and Tamhane (1987); Tamhane and Gou (2018); Proschan and Brittain (2020). In this section, we present the reverse graphical approaches with direct proofs of the strong FWER control under independence. When three or more endpoints are involved, it may be necessary to impose additional constraints on vertex weights and transition weights to guarantee the strong FWER control.

2.1 Reverse Graphical Approaches for Two or Three Hypotheses

We first consider the reverse graphical approaches with only two vertices H_1 and H_2, as shown in Figure 1.

Reverse Graphical Approach for Two Hypotheses: The initial weights of H_1 is w_1 and that of H_2 is w_2, which satisfy $0 \leq w_1 + w_2 \leq 1$. The transition probabilities satisfy $0 \leq g_{12} \leq 1$ and $0 \leq g_{21} \leq 1$.

Step 1: Reject both H_1 and H_2 if $p_1 \leq (w_1 + w_2 g_{21})\alpha$ and $p_2 \leq (w_2 + w_1 g_{12})\alpha$ are both satisfied. Otherwise, go to Step 2.

Step 2: Reject H_1 if $p_1 \leq w_1\alpha$, and reject H_2 if $p_2 \leq w_2\alpha$. Stop.

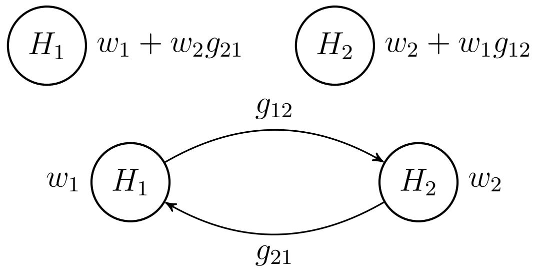

Figure 1: Reverse graphical approach for two hypotheses

Note that in Step 1 we have two choices: reject two hypotheses, or go to Step 2. In Step 2 we have two choices: reject one hypothesis or accept both hypotheses. The control of the type I error rate can be verified directly as stated in Proposition 1.

Proposition 1. *The Reverse Graphical Approach (RGA) strongly controls the FWER for two independent weighted hypotheses.*

Next we describe the reverse graphical approaches with three endpoints, H_1, H_2 and H_3. This procedure starts from graphs with a single vertex, followed by graphs with a pair of hypotheses. The graph including all three endpoints is considered in the last step, as shown in Figure 2.

Reverse Graphical Approach for Three Hypotheses: The initial weights of H_1 is w_1 and that of H_2 and H_3 are w_2 and w_3,

Step 1: We reject all three hypotheses H_1, H_2 and H_3 if

$$p_1 \leq \left(w_1 + \frac{g_{21} + g_{23}g_{31}}{1 - g_{23}g_{32}} \cdot w_2 + \frac{g_{31} + g_{32}g_{21}}{1 - g_{32}g_{23}} \cdot w_3 \right) \alpha,$$
$$p_2 \leq \left(\frac{g_{12} + g_{13}g_{32}}{1 - g_{13}g_{31}} \cdot w_1 + w_2 + \frac{g_{32} + g_{31}g_{12}}{1 - g_{31}g_{13}} \cdot w_3 \right) \alpha,$$
$$p_3 \leq \left(\frac{g_{13} + g_{12}g_{23}}{1 - g_{12}g_{21}} \cdot w_1 + \frac{g_{23} + g_{21}g_{13}}{1 - g_{21}g_{12}} \cdot w_2 + w_3 \right) \alpha$$

are all satisfied. Otherwise, if two inequalities out of three are satisfied, accept the hypothesis whose p-value is greater than its corresponding boundary, and go to Step 2; if only one inequality is satisfied, accept the other two hypotheses which fail to satisfies their corresponding inequalities, and go to Step 3; if all three p-values are greater than the boundaries, accept all hypotheses and stop.

Step 2: Suppose that p_k is greater than its corresponding critical value in Step 1. Consider the index set $\{i, j\} = \{1, 2, 3\} \backslash \{k\}$. We reject the pair of hypotheses H_i and H_j if both $p_i \leq \min\{(w_i + w_k g_{ki})\alpha, (w_i + w_j g_{ji})\alpha\}$ and $p_j \leq \min\{(w_j + w_k g_{kj})\alpha, (w_j + w_i g_{ij})\alpha\}$ are satisfied. Otherwise, if one p-value is greater than its critical value, accept the corresponding hypothesis and go to Step 3; if both p-values fail to satisfy the condition, accept H_i and H_j and stop.

Step 3: Suppose that H_j and H_k have been accepted in the previous steps. Let $i = \{1, 2, 3\} \backslash \{j, k\}$ and reject H_i if $p_i \leq w_i \alpha$. Otherwise, accept H_i and stop.

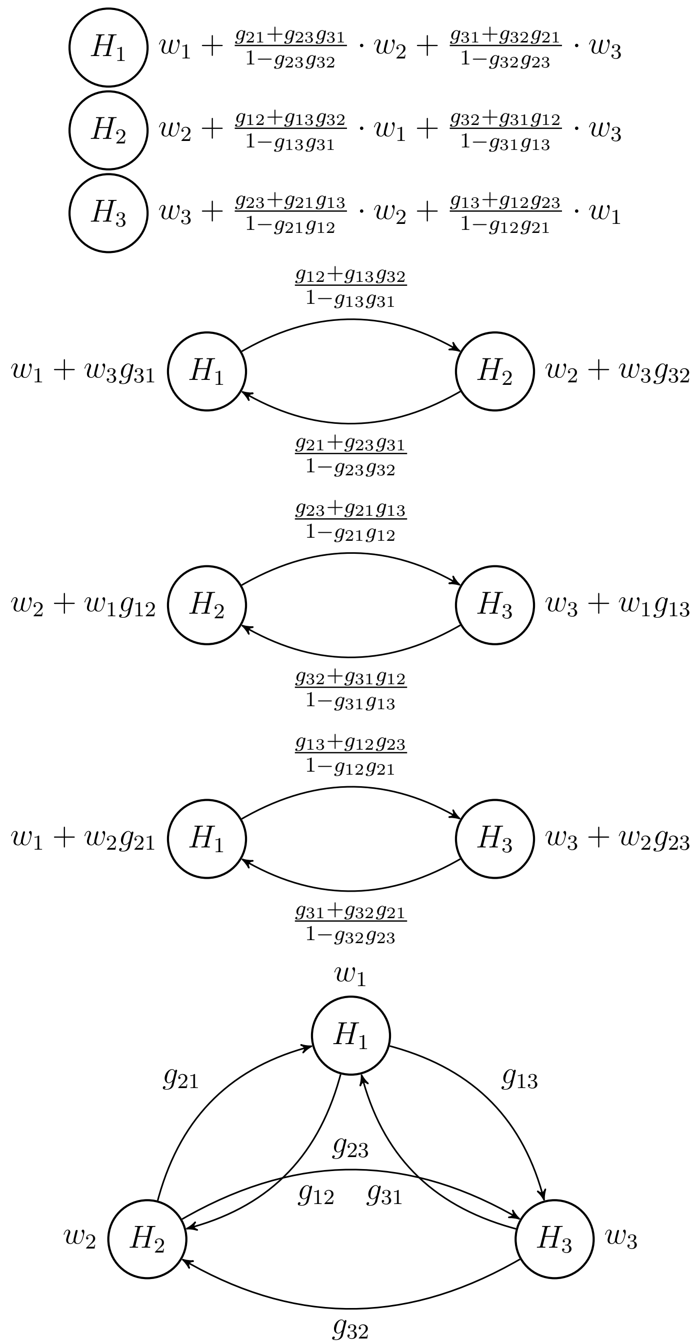

Figure 2: Reverse graphical approach for three hypotheses

In order to demonstrate the strong FWER control of the reverse graphical approach with three hypotheses, we need to bring in additional constraints on vertex weights and transition weights, as presented in Proposition 2. It is worth to note that these constraints are sufficient, not necessary.

Proposition 2. *The Reverse Graphical Approach (RGA) strongly controls the FWER for three independent weighted hypotheses, if any of the following conditions are true:*

1. $w_1w_2 + w_2w_3 + w_3w_1 \geq w_1^2 g_{12}g_{13} + w_2^2 g_{23}g_{21} + w_3^2 g_{31}g_{32}$,

2. $w_1w_2 + w_2w_3 + w_3w_1 \geq (w_1^2 + w_2^2 + w_3^2)/4$,

3. $w_1 + w_2 \geq w_3/2$, $w_2 + w_3 \geq w_1/2$, $w_3 + w_1 \geq w_2/2$,

4. $\max_i\{w_i\} \leq 2/3$ *when* $w_1 + w_2 + w_3 = 1$,

and the following implications are true: (4) ⇒ (3) ⇒ (2) ⇒ (1). Condition (4) is the easiest one to apply, and Condition (1) the most relaxed condition to guarantee the FWER control.

Remark 1. For a symmetric complete graphs with equal weights and equal transition probabilities, the probability under independence is $\Pr\left(\cup_{i=1}^{3}\{\text{reject } H_i\}\right) = \alpha - \alpha^2/4 + \alpha^3/4$, and the reverse graphical approach is reduced to the Hochberg procedure.

2.2 Reverse Graphical Approaches for Four or More Hypotheses

The reverse graphical approaches for two and three hypotheses can be extended to include four or more hypotheses. Similarly, we start from the collection of singleton graphs, gradually add hypotheses back to the graph until either all hypotheses are rejected or the decision has been made based on the largest graph.

Reverse Graphical Approach: We follow the notation in Bretz et al. (2009) and assume that the initial weights assigned to H_k is $w_k(\{1,\ldots,n\})$. Following a valid weighting scheme Maurer and Bretz (2014), the weight for H_k is updated to be $w_k(J)$ when the graph includes hypotheses $\{H_j : j \in J\}$. For notation simplicity, we also denote the initial weights for H_k by w_k, where $w_k = w_k(\{1,\ldots,n\})$.

Step 1: We reject all hypotheses $H_1,\ldots,H_n$ and stop testing if $\cap_{i=1}^{n}\{p_i \leq w_i(\{i\})\}$ is true. Otherwise, accept hypotheses $\{H_i : i \in I\}$ and go to Step $|I|+1$, where $I = \{i : p_i > w_i(\{i\})\}$ is the index set of accepted hypotheses and $|I|$ is the cardinality of the index set I.

Step $i+1$: We have i accepted hypotheses and $n-i$ not-yet-accepted hypotheses when getting to Step $i+1$. Denote the index set of the accepted hypotheses by I and that of the not-yet-accepted hypotheses by $\{1,\ldots,n\}\backslash I$. Consider all H_k where $k \in \{1,\ldots,n\}\backslash I$. We reject all H_k if $\cap_{k\in 1,\ldots,n\backslash J} \cap_{J:|J|=i+1} \{p_k \leq w_k(J)\}$ is true. Equivalently, all $n-i$ not-yet-rejected hypotheses H_k are rejected if $p_k \leq w_k(J)$ holds for any $k \in \{1,\ldots,n\}\backslash I$ and for any index set J with cardinality $i+1$ that contains k. If $\cup_{k\in\{1,\ldots,n\}\backslash I} \cap_{J:|J|=i+1} \{p_k \leq w_k(J)\}$ is false, we accept all hypotheses and stop. Otherwise, update the index set of accepted hypotheses by $I^\star = I \cup \{k : \cup_{J:|J|=i+1}\{p_k > w_k(J)\} \text{ is true}\}$, and go to Step $|I^\star|+1$.

Step n: We have $n-1$ accepted hypotheses and just one not-yet-accepted hypothesis when getting to Step n. Suppose the only not-yet-accepted hypothesis is H_k. Reject H_k if $p_k \leq w_k(\{1,\ldots,n\})$. Otherwise accept it and stop.

We have proved the strong FWER control of the reverse graphical approaches for two and three hypotheses under independence in Proposition 1 and 2. For the reverse graphical approaches for more than three hypotheses, the strong FWER control can be shown at level $\alpha + o(\alpha^2)$ when some sufficient conditions are satisfied, as shown in Propostion 3.

Proposition 3. *The FWER of the Reverse Graphical Approach (RGA) for n hypotheses is bounded from above by $\sum_{i=1}^n w_i\alpha + o(\alpha^2)$ or equivalently $\lim_{\alpha\to 0} \frac{FWER-\sum_{i=1}^n w_i\alpha}{\alpha^2} = 0$, if any of the following conditions are true for any index set $I_m \subset \{1,\dots,n\}$ with cardinality $|I_m| = m$:*

1. $\sum_{i<j} w_i(I_m)w_j(I_m) \geq \sum_{k\in I_m} \sum_{i<j,i\neq k,j\neq k} w_k(I_m)^2 g_{ki}(I_m)g_{kj}(I_m)$,
2. $\sum_{i<j} w_i(I_m)w_j(I_m) \geq \frac{m-2}{2(m-1)} \sum_{k=1}^n w_k(I_m)^2$,
3. $w_i(I_m) + w_j(I_m) \geq \frac{2(m-2)}{(m-1)^2} w_k(I_m)$ *for all triplet (i,j,k) where $i \neq j$, $j \neq k$, $k \neq i$,*

and the following implications are true: (3) ⇒ (2) ⇒ (1). The difference between the FWER and $\sum_{i=1}^n w_i\alpha$ is at most an $O(\alpha^3)$ term.

3 Examples and Numerical Results

We investigate the FWER control and statistical power of the multiple testing procedures using the reverse graphical approach in this section, comparing with the procedures using the graphical approach. W use the average power or A-power, which is the expected proportion of rejected true significances, as the definition of statistical power when multiple hypotheses are involved Gou et al. (2014).

We first consider the procedures using the graphical approaches and the reverse graphical approaches with two hypotheses. The initial weights for H_1 and H_2 are w_1 and $w_2 = 1 - w_1$ respectively. The transition weights are $g_{12} = g_{21} = 1$. Assume the z-statistics (z_1, z_2) follow a standard bivariate normal distribution with mean (δ_1, δ_2), variances equal to 1 and correlation coefficient ρ. The bivariate normal probabilities can be evaluated numerically Genz and Bretz (2009). Figure 3 shows the comparison between the procedures using the graphical approaches and the reverse graphical approaches, where $n = 2$ and nominal significance level $\alpha = 5\%$. We provide the FWER comparison in the top panel and the average power comparison in the bottom panel. The graphical approach is indicated using thin lines and the reverse graphical approach is shown in thick lines. We consider three different correlation coefficients: independence $\rho = 0$ (solid lines), positive dependence $\rho = 0.5$ (dash-dotted lines) and negative dependence $\rho = -0.5$ (dashed lines). The top left panel shows the relation between the FWER and w_1 when both hypotheses are true nulls. For the reverse graphical approach, the FWERs under independence and $\rho = 0.5$ are controlled at or under the nominal significance level 5%. The maximum FWER of the reverse graphical approach under $\rho = -0.5$ is 5.003%, where the difference is negligible for practical purposes. The top right panel presents the FWER when H_1 is a true null and H_2 is a true significance with $\delta_2 = 2$. The FWER increases when w_1 increases, and all FWERs are below the nominal significance level. The reverse graphical approach is less conservative than the graphical approach. The bottom left panel computes the power where H_1 is a true significance with $\delta_1 = 2$ and H_2

is a true null. The reverse graphical approach is slightly more powerful than the graphical approach with average A-power gain 0.2%. The bottom right panel presents the average power when both hypotheses are true significances with $\delta_1 = \delta_2 = 2$. The A-power gain of the reverse graphical approach over the graphical approach is 1.2% on average over different w_1 and ρ, with a maximum value 1.8%.

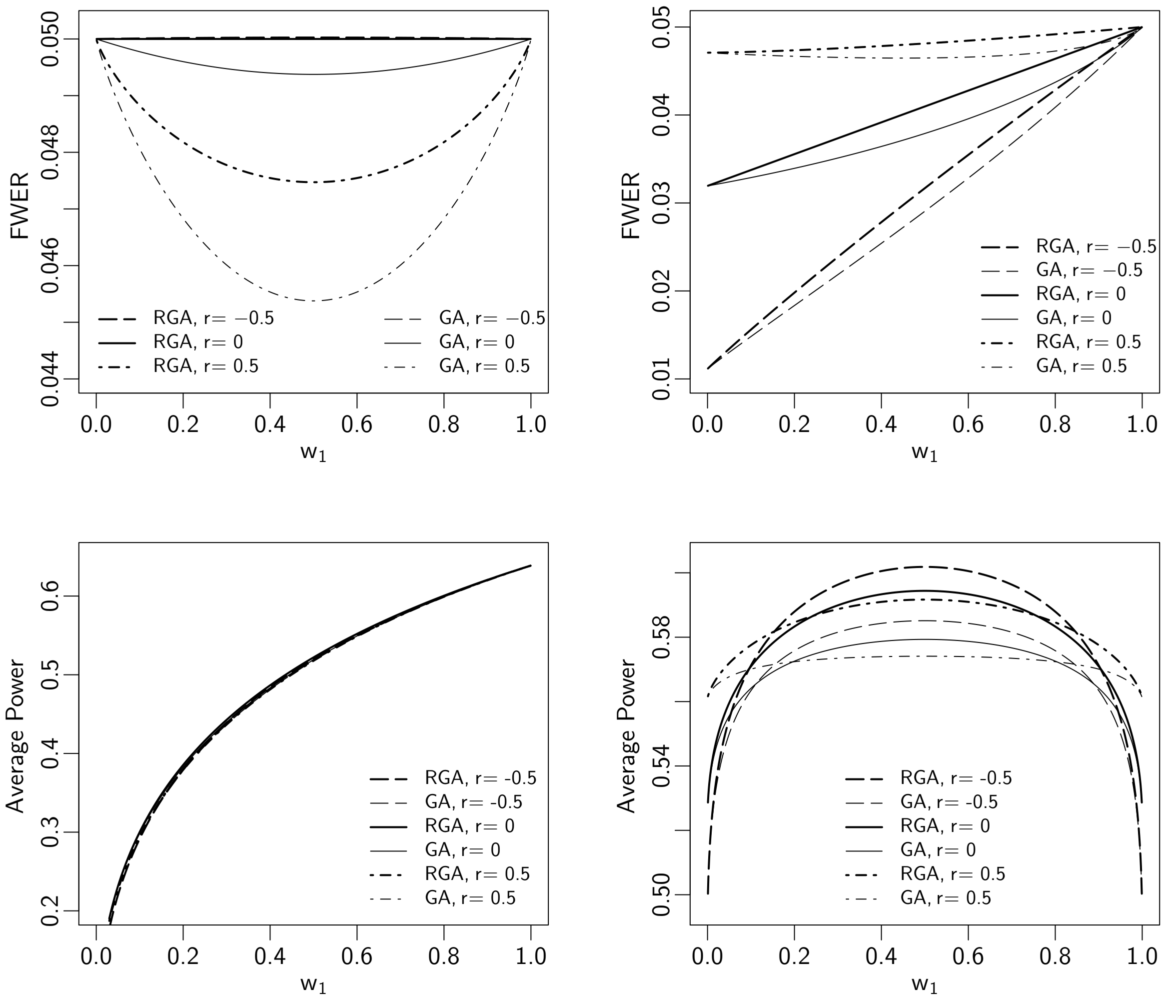

Figure 3: FWER and Power comparisons ($n = 2$, $\alpha = 5\%$). Top left panel: FWER, $(\delta_1, \delta_2) = (0, 0)$. Top right panel: FWER, $(\delta_1, \delta_2) = (0, 2)$. Bottom left panel: Power, $(\delta_1, \delta_2) = (2, 0)$. Bottom right panel: Power, $(\delta_1, \delta_2) = (2, 2)$.

We next investigate the FWER and A-power of graphical and reverse graphical approaches with three hypotheses. We start from a graph with vertex weights $w_1 = 0.40$, $w_2 = 0.25$, $w_3 = 0.35$, and transition weights $g_{12} = 2/3$, $g_{13} = 1/3$, $g_{21} = 1/2$, $g_{23} = 1/2$, $g_{31} = 1/4$, $g_{32} = 3/4$. A simulation study is carried out where the z-statistics are generated from a standard trivariate normal distribution with mean $(\delta_1, \delta_2, \delta_3)$, variances equal to 1 and equi-correlation matrix with coefficient ρ. The p-values under the true null hypotheses follow a standard uniform distribution, and the p-values under the true significances are generated by first generating z-values from a normal distribution with mean 3 and variance 1. Table 1 includes the simulation results under independence ($\rho = 0$), negative dependence ($\rho = -0.2$), positive dependence ($\rho = 0.2$ and 0.5). The number of replica is 2×10^6 for each estimation. The reverse graphical approach (RGA) is less conservative than the graphical approach (GA), and consequently, more powerful than the graphical approach, where the power gain can be as much as 1.2%.

Table 1: FWER and Power comparisons under various configurations ($n = 3$, $\alpha = 5\%$)

ρ		δ_1 / δ_2 / δ_3	0/0/0	0/0/3	0/3/0	0/3/3	3/0/0	3/0/3	3/3/0	3/3/3
-0.2	FWER	GA	4.98	4.44	4.54	3.80	4.38	3.81	3.79	N/A
		RGA	4.98	4.55	4.55	4.19	4.41	4.23	4.20	N/A
	Power	GA	N/A	81.46	78.01	84.42	82.92	84.65	84.97	90.01
		RGA	N/A	81.54	78.13	84.85	82.97	84.87	85.01	90.55
0	FWER	GA	4.91	4.62	4.65	4.26	4.58	4.29	4.28	N/A
		RGA	4.93	4.74	4.66	4.54	4.63	4.58	4.56	N/A
	Power	GA	N/A	81.42	77.90	84.05	82.90	84.44	84.62	89.36
		RGA	N/A	81.49	78.00	84.46	82.94	84.60	84.64	90.08
0.2	FWER	GA	4.78	4.69	4.69	4.61	4.67	4.64	4.62	N/A
		RGA	4.82	4.83	4.70	4.78	4.74	4.81	4.80	N/A
	Power	GA	N/A	81.40	77.77	83.69	82.88	84.23	84.29	88.70
		RGA	N/A	81.44	77.83	84.08	82.91	84.34	84.29	89.59
0.5	FWER	GA	4.31	4.53	4.51	4.93	4.50	4.95	4.93	N/A
		RGA	4.43	4.75	4.59	4.97	4.67	4.99	4.97	N/A
	Power	GA	N/A	81.33	77.64	83.20	82.85	83.95	83.86	87.68
		RGA	N/A	81.34	77.65	83.60	82.86	84.01	83.87	88.87

The reverse graphical approach is uniformly more powerful than the graphical approach with the same graph when dealing with two hypotheses, but not necessary for graphs with three or more vertices. For example, consider a 3-node graph with vertex weights $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$, and transition weights $g_{12} = 3/5$, $g_{13} = 2/5$, $g_{21} = 2/3$, $g_{23} = 1/3$, $g_{31} = 1/2$, $g_{32} = 1/2$. The pre-specified level of significance is 5%. Consider a set of observed p-values $p_1 = 0.020$, $p_2 = 0.025$, $p_3 = 0.060$. When applying the graphical approach, we start from the graph with all three hypotheses. We reject H_1 since $p_1 = 0.020 \leq 0.5 \times 0.05 = 0.025$, and the graph is reduced to a two-vertex graph with vertex weights 0.6 and 0.4 for H_2 and H_3. Based on this two-vertex graph, H_2 is rejected since $p_2 = 0.025 \leq 0.6 \times 0.05 = 0.030$. Finally we accept H_3 since $p_3 = 0.060 > 1 \times 0.05 = 0.05$ based on the single-vertex graph. When applying the reverse graphical approach, we first test hypotheses based on three single-node graphs. Since $p_3 = 0.060 > 1 \times 0.05 = 0.05$, we cannot reject all three hypotheses and get to graphs with two hypotheses. For H_1, we use the graph with H_1 and H_2 and the graph with H_1 and H_3, and have $p_1 = 0.020 \leq \min\{0.6, 0.7\} \times 0.05 = 0.030$. For H_2, we use the graph with H_1 and H_2 and the graph with H_2 and H_3, and have $p_2 = 0.025 > \min\{0.4, 0.6\} \times 0.05 = 0.020$. Therefore we cannot reject both H_1 and H_2, and get to the graph with all three hypotheses. Since $p_1 = 0.020 \leq 0.5 \times 0.05 = 0.025$, we reject H_1. With this set of p-values, the graphical approach rejects both H_1 and H_2 but the reverse graphical approach only reject H_1. For another set of observed p-values $p_1 = 0.030$, $p_2 = 0.035$, $p_3 = 0.040$, however, the graphical approach rejects none of them but the reverse graphical approach rejects all three hypotheses.

Lastly, in order to illustrate the proposed approach, we consider a clinical trial, the-ATMOSPHERE study, where around 7000 patients are balanced, randomized to aliskiren monotherapy (AM), enalapril monotherapy (EM), or the combination (CM) Krum et al. (2011). We only consider the three primary hypotheses in this study, where the null hypotheses are CM is not superior to EM in delaying time to first occurrence of cardiovascular death or heart failure hospitalization (H_1), AM is not non-inferior to EM on this endpoint (H_2), and AM is not superior to EM on this endpoint (H_3). The design on testing H_1, H_2 and H_3 can be visualized using a three-node graph with vertex weights $w_1 = 0.5$, $w_2 = 0.5$, $w_3 = 0$, and transition weights $g_{12} = 1$, $g_{13} = 0$, $g_{21} = 1/4$, $g_{23} = 3/4$, $g_{31} = 1$, $g_{32} = 0$ Maurer and Bretz (2014). With the hypothetical unadjusted p-values in Maurer and Bretz (2014), where $p_1 = 0.100$, $p_2 = 0.007$ and $p_3 = 0.050$, assuming the significance level $\alpha = 0.025$, we reject H_2 and fail to reject H_1 and H_3 using either the graphical approach or the reverse graphical approach. Note here the FWER control of the reverse graphical approach is guaranteed since the conditions in Proposition 2 or 3 are satisfied.

4 Discussion

The graphical approach can be treated as a generalization of the Holm procedure, and the reverse graphical approach can be considered as its counterpart starting from the Hochberg procedure. Assuming independence, we prove the strong FWER control for the reverse graphical approach with two or three hypotheses, and the asymptotic FWER control for the general situation with four or more hypotheses. For statistical power, the classical Hochberg procedure is uniformly more powerful than the Holm procedure. However, when comparing the two graphical approaches, the reverse graphical approach is uniformly more powerful than the graphical approach for multiple testing procedures using the same graph, only when two hypotheses are involved. If we consider a graph with three or more hypotheses, there is no uniformly more powerful approach between these two graphical approaches. This comparison is similar with the comparison between the weighted Holm and weighted Hochberg procedures Tamhane and Liu (2008).

We provide proofs of strong FWER control under independence. It would be interesting to investigate whether the reverse graphical approach controls the FWER strongly under positive dependence Sarkar and Chang (1997); Sarkar (1998) or negative dependence Gou and Tamhane (2018). Some simulation studies support that the weak control of FWER holds under both positive and negative dependence with three or more hypotheses when the test statistics follow a multivariate normal distribution. We may consider calculating the adjusted p-values and simultaneous confidence intervals Wright (1992); Bretz et al. (2011). Another future research direction of interest is to generalize the reverse graphical approach to include repeated testing Maurer et al. (2011), group sequential procedures Maurer and Bretz (2013); Xi and Tamhane (2015); Tamhane et al. (2018), testing families of hypotheses

Maurer and Bretz (2014), and generalized error rate control Robertson et al. (2020).

Acknowledgements

We thank Ajit C. Tamhane and Dong Xi for helpful discussion.

Appendix

Proofs of Proposition 1, 2 and 3 are included in this appendix.

of Proposition 1. We only need to show that the chance of false rejection of true nulls under any configuration is controlled by level α. When both H_1 and H_2 are true nulls, the control of the type I error rate is satisfied when $w_1 + w_2 \leq 1$

$$\begin{aligned}
&\Pr\left(\cup_{i=1}^{2}\{\text{reject } H_i\}\right)\\
&= \Pr\left(\{\{P_1 \leq (w_1 + w_2 g_{21})\alpha\} \cap \{P_2 \leq (w_2 + w_1 g_{12})\alpha\}\} \cup \{P_1 \leq w_1\alpha\} \cup \{P_2 \leq w_2\alpha\}\right)\\
&= (w_1 + w_2 g_{21})(w_2 + w_1 g_{12})\alpha^2 + w_1\alpha(1 - (w_2 + w_1 g_{12})\alpha) + w_2\alpha(1 - (w_1 + w_2 g_{21})\alpha)\\
&= (w_1 + w_2)\alpha - w_1 w_2 (1 - g_{12} g_{21})\alpha^2\\
&\leq (w_1 + w_2)\alpha,
\end{aligned}$$

where the equality holds when $g_{12} = g_{21} = 1$. In addition, when one hypothesis is a true null and the other is a true significance, the type I error rate is also controlled at level α since

$$\begin{aligned}
&\Pr(\text{reject } H_1) = \Pr\left(\{\{P_1 \leq (w_1 + w_2 g_{21})\alpha\} \cap \{P_2 \leq (w_2 + w_1 g_{12})\alpha\}\} \cup \{P_1 \leq w_1\alpha\}\right)\\
&\leq \Pr\left(P_1 \leq (w_1 + w_2 g_{21})\alpha\right) = (w_1 + w_2 g_{21})\alpha \leq \alpha,
\end{aligned}$$

assuming H_1 is the true null and H_2 is the true significance. □

of Proposition 2. First, we assume that the transition probabilities satisfy $0 \leq g_{ij} \leq 1$ and $g_{12} + g_{13} = 1$, $g_{21} + g_{23} = 1$, $g_{31} + g_{32} = 1$. If $\sum_{k \neq i} g_{ik} < 1$ for some i, we can always normalize the transition probabilities using $g_{ij}^{\star} = g_{ij} / \sum_{k \neq i} g_{ik}$, and have $\Pr(E \mid \boldsymbol{w}, \boldsymbol{G}) \leq \Pr(E \mid \boldsymbol{w}, \boldsymbol{G}^{\star})$ where E is an event. Therefore, we only need to calculate the FWER using the graph $(\boldsymbol{w}, \boldsymbol{G}^{\star})$. For notation simplicity, we still denote the normalized transition probability by $\boldsymbol{G}$. Next, we consider the situation when all three hypotheses are true nulls. The probability of rejecting one or more null hypotheses is shown below.

$$\begin{aligned}
&\Pr\left(\cup_{i=1}^{3}\{\text{reject } H_i\}\right)\\
&\leq \Pr\left(\{\{P_1 \leq (w_1 + w_2 + w_3)\alpha\} \cap \{P_2 \leq (w_1 + w_2 + w_3)\alpha\} \cap \{P_3 \leq (w_1 + w_2 + w_3)\alpha\}\}\right.\\
&\quad \cup \{\{P_1 \leq (w_1 + w_3 g_{31})\alpha\} \cap \{P_2 \leq (w_2 + w_3 g_{32})\alpha\}\}\\
&\quad \cup \{\{P_2 \leq (w_2 + w_1 g_{12})\alpha\} \cap \{P_3 \leq (w_3 + w_1 g_{13})\alpha\}\}\\
&\quad \cup \{\{P_3 \leq (w_3 + w_2 g_{23})\alpha\} \cap \{P_1 \leq (w_1 + w_2 g_{21})\alpha\}\}\\
&\quad \left.\cup \{P_1 \leq w_1\alpha\} \cup \{P_2 \leq w_2\alpha\} \cup \{P_3 \leq w_3\alpha\}\right)\\
&= (w_1 + w_2)(w_2 + w_3)(w_3 + w_1)\alpha^3 + w_3^2 g_{31} g_{32} \alpha^2 (1 - (w_1 + w_2 + w_3)\alpha)
\end{aligned}$$

$$
\begin{aligned}
&+ w_1^2 g_{13} g_{12} \alpha^2 (1 - (w_1 + w_2 + w_3)\alpha) + w_2^2 g_{23} g_{21} \alpha^2 (1 - (w_1 + w_2 + w_3)\alpha) \\
&\quad + 1 - (1 - w_1\alpha)(1 - w_2\alpha)(1 - w_3\alpha) \\
&= (w_1 + w_2 + w_3)\alpha + \left[w_1^2 g_{12} g_{13} + w_2^2 g_{23} g_{21} + w_3^2 g_{31} g_{32} - w_1 w_2 - w_2 w_3 - w_3 w_1 \right] \alpha^2 \\
&\quad + \left[(w_1 + w_2)(w_2 + w_3)(w_3 + w_1) + w_1 w_2 w_3 \right. \\
&\qquad \left. - (w_1 + w_2 + w_3) \left(w_1^2 g_{12} g_{13} + w_2^2 g_{23} g_{21} + w_3^2 g_{31} g_{32} \right) \right] \alpha^3 \\
&= (w_1 + w_2 + w_3)\alpha + \left[w_1^2 g_{12} g_{13}) + w_2^2 g_{23} g_{21} + w_3^2 g_{31} g_{32} - w_1 w_2 - w_2 w_3 - w_3 w_1 \right] \alpha^2 \\
&\quad + (w_1 + w_2 + w_3) \left[w_1 w_2 + w_2 w_3 + w_3 w_1 - \left(w_1^2 g_{12} g_{13}) + w_2^2 g_{23} g_{21} + w_3^2 g_{31} g_{32} \right) \right] \alpha^3 \\
&= (w_1 + w_2 + w_3)\alpha \\
&\quad + w g_{12} g_{13} + w_2^2 g_{23} g_{21} + w_3^2 g_{31} g_{32} - w_1 w_2 - w_2 w_3 - w_3 w_1 \alpha^2 \left(1 - (w_1 + w_2 + w_3)\, \alpha \right) \\
&\le (w_1 + w_2 + w_3)\alpha + \left[w_1^2/4 + w_2^2/4 + w_3^2/4 - w_1 w_2 - w_2 w_3 - w_3 w_1 \right] \alpha^2 \left(1 - (w_1 + w_2 + w_3)\, \alpha \right) \\
&\le (w_1 + w_2 + w_3)\alpha.
\end{aligned}
$$

When there are two true nulls and one true significance, we compute the probability of rejecting H_1 or H_2, assuming that H_1 and H_2 are the true nulls.

$$
\begin{aligned}
&\Pr\left(\{\text{reject } H_1\} \cup \{\text{reject } H_2\} \right) \\
&\le \Pr\left(\{\{P_1 \le (w_1 + w_2 + w_3)\alpha\} \cap \{P_2 \le (w_1 + w_2 + w_3)\alpha\} \cap \{P_3 \le (w_1 + w_2 + w_3)\alpha\}\} \right. \\
&\quad \cup \{\{P_1 \le \min\{(w_1 + w_3 g_{31})\alpha, (w_1 + w_2 g_{21})\alpha\}\} \cap \{P_2 \le \min\{(w_2 + w_3 g_{32})\alpha, (w_2 + w_1 g_{12})\alpha\}\}\} \\
&\quad \cup \{\{P_2 \le \min\{(w_2 + w_1 g_{12})\alpha, (w_2 + w_3 g_{32})\alpha\}\} \cap \{P_3 \le (w_3 + w_1 g_{13})\alpha\}\} \\
&\quad \cup \{\{P_3 \le (w_3 + w_2 g_{23})\alpha\} \cap \{P_1 \le \min\{(w_1 + w_2 g_{21})\alpha, (w_1 + w_3 g_{31})\alpha\}\}\} \\
&\quad \left. \cup \{P_1 \le w_1\alpha\} \cup \{P_2 \le w_2\alpha\} \right) \\
&\le \Pr\left(\{\{P_1 \le (w_1 + w_2 + w_3)\alpha\} \cap \{P_2 \le (w_1 + w_2 + w_3)\alpha\}\} \right. \\
&\quad \cup \{\{P_1 \le (w_1 + w_3 g_{31})\alpha\} \cap \{P_2 \le (w_2 + w_3 g_{32})\alpha\}\} \\
&\quad \cup \{P_1 \le (w_1 + w_3 g_{31})\alpha\} \cup \{P_2 \le (w_2 + w_3 g_{32})\alpha\} \\
&\quad \left. \cup \{P_1 \le w_1\alpha\} \cup \{P_2 \le w_2\alpha\} \right) \\
&= \Pr\left(\{\{P_1 \le (w_1 + w_2 + w_3)\alpha\} \cap \{P_2 \le (w_1 + w_2 + w_3)\alpha\}\} \right. \\
&\quad \left. \cup \{P_1 \le (w_1 + w_3 g_{31})\alpha\} \cup \{P_2 \le (w_2 + w_3 g_{32})\alpha\} \right) \\
&\le (w_1 + w_2 + w_3)\, \alpha,
\end{aligned}
$$

where the last inequality follows the proof of Proposition 1 and $(w_1 + w_3 g_{31})\alpha + (w_2 + w_3 g_{32})\alpha = (w_1 + w_2 + w_3)\,\alpha$. When there is only one true null, we assume that H_1 is the true null and calculate the probability of rejecting it.

$$
\begin{aligned}
&\Pr\left(\{\text{reject } H_1\} \right) \\
&= \Pr\left(\{\{P_1 \le (w_1 + w_2 + w_3)\alpha\} \cap \{P_2 \le (w_1 + w_2 + w_3)\alpha\} \cap \{P_3 \le (w_1 + w_2 + w_3)\alpha\}\} \right. \\
&\quad \cup \{\{P_1 \le \min\{(w_1 + w_3 g_{31})\alpha, (w_1 + w_2 g_{21})\alpha\}\} \cap \{P_2 \le \min\{(w_2 + w_3 g_{32})\alpha, (w_2 + w_1 g_{12})\alpha\}\}\} \\
&\quad \cup \{\{P_3 \le \min\{(w_3 + w_2 g_{23})\alpha, (w_3 + w_1 g_{13})\alpha\}\} \cap \{P_1 \le \min\{(w_1 + w_2 g_{21})\alpha, (w_1 + w_3 g_{31})\alpha\}\}\} \\
&\quad \left. \cup \{P_1 \le w_1\alpha\} \right) \\
&\le \Pr\left(P_1 \le (w_1 + w_2 + w_3)\alpha \right) = (w_1 + w_2 + w_3)\alpha.
\end{aligned}
$$

Lastly, we show the implication (3) $\Rightarrow$ (2). Note that the assumptions

$$
\begin{aligned}
w_1 + w_2 \geq w_3/2 &\Rightarrow w_3 w_1 + w_2 w_3 \geq w_3^2/2, \\
w_2 + w_3 \geq w_1/2 &\Rightarrow w_1 w_2 + w_3 w_1 \geq w_1^2/2, \\
w_3 + w_1 \geq w_2/2 &\Rightarrow w_2 w_3 + w_1 w_2 \geq w_2^2/2,
\end{aligned}
$$

yield

$$
2\left(w_1 w_2 + w_2 w_3 + w_3 w_1\right) \geq \left(w_1^2 + w_2^2 + w_3^2\right)/2.
$$

The implications (4) $\Rightarrow$ (3) and (2) $\Rightarrow$ (1) can be verified directly. $\square$

of Proposition 3. We first consider the situation when all n hypotheses are true nulls.

$$
\begin{aligned}
&\Pr\left(\cup_{k=1}^{n}\{\text{reject } H_k\}\right) = \Pr(\cup_{i=1}^{n} \cup_{I:|I|=i} \cap_{k\in I} \cap_{J\ni k:|J|=n+1-i} \{p_k \leq w_k(J)\}) \\
&= \Pr(\cup_{i=1}^{2} \cup_{I:|I|=i} \cap_{k\in I} \cap_{J\ni k:|J|=n+1-i} \{p_k \leq w_k(J)\}) + o(\alpha^2) \\
&\leq \sum_{i=1}^{n} w_i\alpha - \sum_{i<j} w_i w_j \alpha^2 + \sum_{k=1}^{n} \sum_{i<j, i\neq k, j\neq k} w_k^2 g_{ki} g_{kj} \alpha^2 + o(\alpha^2) \\
&= \sum_{i=1}^{n} w_i\alpha - \left[\sum_{i<j} w_i w_j - \sum_{k=1}^{n} \sum_{i<j, i\neq k, j\neq k} w_k^2 g_{ki} g_{kj}\right]\alpha^2 + o(\alpha^2)
\end{aligned}
$$

Therefore, $\sum_{i<j} w_i w_j \geq \sum_{k=1}^{n} \sum_{i<j,i\neq k,j\neq k} w_k^2 g_{ki} g_{kj}$ implies that $\Pr\left(\cup_{k=1}^{n}\{\text{reject } H_k\}\right) \leq \sum_{i=1}^{n} w_i\alpha + o(\alpha^2)$.

Next, suppose there are m true nulls and $n-m$ true significances. Without loss of generality, we assume $H_1, \ldots, H_m$ are true nulls and $H_{m+1}, \ldots, H_n$ are true significances.

$$
\begin{aligned}
&\Pr\left(\cup_{k=1}^{m}\{\text{reject } H_k\}\right) \\
&\leq \Pr(\cup_{i=1}^{m} \cup_{I\subset\{1,\ldots,m\}:|I|=i} \cup_{j=m-1+i}^{n-1+i} \cap_{k\in I} \cap_{J\ni k:|J|=j} \{p_k \leq w_k(J)\}) \\
&= \Pr(\cup_{i=1}^{m} \cup_{I\subset\{1,\ldots,m\}:|I|=i} \cap_{k\in I} \cap_{J\ni k:|J|=m-1+i} \{p_k \leq w_k(J)\}) \\
&\leq \sum_{i=1}^{m} w_i(\{1,\ldots,m\})\alpha - \left[\sum_{i<j} w_i(\{1,\ldots,m\}) w_j(\{1,\ldots,m\})\right. \\
&\quad \left. - \sum_{k=1}^{m} \sum_{i<j,i\neq k,j\neq k} w_k(\{1,\ldots,m\})^2 g_{ki}(\{1,\ldots,m\}) g_{kj}(\{1,\ldots,m\})\right] \alpha^2 + o(\alpha^2),
\end{aligned}
$$

where the last inequality follows that $m < n$, and we consider the index set $\{1,\ldots,m\}$ instead of $\{1,\ldots,n\}$. Denote any index set with cardinality m by I_m. Suppose $\{H_i : i \in I_m\}$ are the collection of true nulls. If follows that

$$
\Pr\left(\cup_{k\in I_m}\{\text{reject } H_k\}\right) \leq \sum_{i\in I_m} w_i(I_m)\alpha + o(\alpha^2) = \sum_{i=1}^{n} w_i\alpha + o(\alpha^2)
$$

when $\sum_{i<j} w_i(I_m) w_j(I_m) \geq \sum_{k\in I_m} \sum_{i<j,i\neq k,j\neq k} w_k(I_m)^2 g_{ki}(I_m) g_{kj}(I_m)$ are satisfied, where the summation indices $i, j \in I_m$. We conclude the strong control of the FWER at level $\alpha + o(\alpha^2)$.

Lastly, we show the implications in Proposition 3. Without loss of generality, we consider the index set $I_n = \{1, \ldots, n\}$. The implications for other index sets I_m can be shown in a similar way. Implication (2) $\Rightarrow$ (1) can be shown by noting that

$$\sum_{i<j, i\neq k, j\neq k} g_{ki} g_{kj} = \frac{1}{2}\left(\left(\sum_{i\neq k} g_{ki}\right)^2 - \sum_{i\neq k} g_{ki}^2\right) \leq \frac{1}{2}\left(\left(\sum_{i\neq k} g_{ki}\right)^2 - \frac{1}{n-1}\left(\sum_{i\neq k} g_{ki}\right)^2\right)$$
$$= \frac{n-2}{2(n-1)}\left(\sum_{i\neq k} g_{ki}\right)^2 \leq \frac{n-2}{2(n-1)},$$

where $\sum_{i\neq k} g_{ki} \leq 1$. For the implication (3) $\Rightarrow$ (2), assuming (3) is satisfied, we combine all inequalities $w_i w_k + w_j w_k \geq \frac{2(n-2)}{(n-1)^2} w_k^2$ for all possible triplets (i, j, k) with distinct indices. It follows that

$$\sum_{i\neq j, j\neq k, k\neq i} w_i w_k + w_j w_k \geq \sum_{i\neq j, j\neq k, k\neq i} \frac{2(n-2)}{(n-1)^2} w_k^2$$

which is equivalent to $2(n-2)\sum_{i\neq j} w_i w_j \geq (n-1)(n-2)\sum_{k=1}^{n} \frac{2(n-2)}{(n-1)^2} w_k^2$. Simplifying this equation, we achieve $\sum_{i\neq j} w_i w_j \geq \frac{n-2}{n-1}\sum_{k=1}^{n} w_k^2$. $\square$

References

Bretz, F., Maurer, W., Brannath, W., and Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* **28,** 586–604.

Bretz, F., Posch, M., Glimm, E., Klinglmueller, F., Maurer, W., and Rohmeyer, K. (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biometrical Journal* **53,** 894–913.

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

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association* **56,** 52–64.

Gabriel, K. R. (1969). Simultaneous test procedures–some theory of multiple comparisons. *The Annals of Mathematical Statistics* **40,** 224–250.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities.* Springer-Verlag, Heidelberg.

Gou, J. and Tamhane, A. C. (2018). Hochberg procedure under negative dependence. *Statistica Sinica* **28,** 339–362.

Gou, J., Tamhane, A. C., Xi, D., and Rom, D. (2014). A class of improved hybrid Hochberg-Hommel type step-up multiple test procedures. *Biometrika* **101,** 899–911.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75,** 800–802.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* John Wiley and Sons, New York, New York.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6,** 65–70.

Krum, H., Massie, B., Abraham, W. T., Dickstein, K., Kober, L., McMurray, J. J., Desai, A., Gimpelewicz, C., Kandra, A., Reimund, B., Rattunde, H., Armbrecht, J., and on behalf of the ATMOSPHERE Investigators (2011). Direct renin inhibition in addition to or as an alternative to angiotensin converting enzyme inhibition in patients with chronic systolic heart failure: rationale and design of the aliskiren trial to minimize outcomes in patients with heart failure (atmosphere) study. *European Journal of Heart Failure* **13,** 107–114.

Marcus, R., Peritz, E., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63,** 655–660.

Maurer, W. and Bretz, F. (2013). Multiple testing in group sequential trials using graphical approaches. *Statistics in Biopharmaceutical Research* **5,** 311–320.

Maurer, W. and Bretz, F. (2014). A note on testing families of hypotheses using graphical procedures. *Statistics in Medicine* **33,** 5340–5346.

Maurer, W., Glimm, E., and Bretz, F. (2011). Multiple and repeated testing of primary, coprimary, and secondary hypotheses. *Statistics in Biopharmaceutical Research* **3,** 336–352.

Proschan, M. A. and Brittain, E. H. (2020). A primer on strong vs weak control of familywise error rate. *Statistics in Medicine* **39,** 1407–1413.

Robertson, D. S., Wason, J. M. S., and Bretz, F. (2020+). Graphical approaches for the control of generalised error rates.

Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *The Annals of Statistics* **26,** 494–504.

Sarkar, S. K. and Chang, C. K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Asociation* **92,** 1601–1608.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73,** 751–754.

Tamhane, A. C. and Gou, J. (2018). Advances in p-value based multiple test procedures. *Journal of Biopharmaceutical Statistics* **28,** 10–27.

Tamhane, A. C., Gou, J., Jennison, C., Mehta, C. R., and Curto, T. (2018). A gatekeeping procedure to test a primary and a secondary endpoint in a group sequential design with multiple interim looks. *Biometrics* **74,** 40–48.

Tamhane, A. C. and Liu, L. (2008). On weighted Hochberg procedures. *Biometrika* **95,** 279–294.

Wright, S. P. (1992). Adjusted p-values for simultaneous inference. *Biometrics* **48,** 1005–1013.

Xi, D. and Bretz, F. (2019). Symmetric graphs for equally weighted tests, with application to the Hochberg procedure. *Statistics in Medicine* **38,** 5268–5282.

Xi, D. and Tamhane, A. C. (2015). Allocating recycled significance levels in group sequential procedures for multiple endpoints. *Biometrical Journal* **57,** 90–107.