Received: 18 January 2019 | Revised: 2 May 2019 | Accepted: 2 September 2019

DOI: 10.1002/sim.8375

RESEARCH ARTICLE

WILEY Statistics in Medicine

Symmetric graphs for equally weighted tests, with application to the Hochberg procedure

Dong Xi[1] | **Frank Bretz**[2,3]

[1]Statistical Methodology, Novartis Pharmaceuticals, East Hanover, New Jersey

[2]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

[3]Section for Medical Statistics, Medical University of Vienna, Vienna, Austria

Correspondence
Dong Xi, Statistical Methodology, Novartis Pharmaceuticals Corporation, East Hanover, NJ 07936.
Email: dong.xi@novartis.com

The graphical approach to multiple testing provides a convenient tool for designing, visualizing, and performing multiplicity adjustments in confirmatory clinical trials while controlling the familywise error rate. It assigns a set of weights to each intersection null hypothesis within the closed test framework. These weights form the basis for intersection tests using weighted individual p-values, such as the weighted Bonferroni test. In this paper, we extend the graphical approach to intersection tests that assume equal weights for the elementary null hypotheses associated with any intersection hypothesis, including the Hochberg procedure as well as omnibus tests such as Fisher's combination, O'Brien's, and F tests. More specifically, we introduce symmetric graphs that generate sets of equal weights so that the aforementioned tests can be applied with the graphical approach. In addition, we visualize the Hochberg and the truncated Hochberg procedures in serial and parallel gatekeeping settings using symmetric component graphs. We illustrate the method with two clinical trial examples.

KEYWORDS

closed test procedure, component graph, familywise error rate, graphical approach, multiple testing, weighting scheme

1 | INTRODUCTION

Regulatory guidelines[1,2] mandate the strong control of the familywise error rate (FWER) in confirmatory clinical trials with primary and secondary objectives. That is, the probability to erroneously reject at least one true null hypothesis is controlled at a prespecified significance level $\alpha \in (0, 1)$ for any configuration of true and false null hypotheses. Graphical approaches provide a general framework for designing, visualizing, and performing FWER-controlling multiple test procedures.[3,4] Under this framework, null hypotheses are represented by nodes associated with weights denoting their local significance levels. Weighted directed edges between nodes specify the propagation of the local significance levels: Once a null hypothesis is rejected, its local weight is split and propagated to other hypotheses. The graphical approach covers many common multiple test procedures based on weighted Bonferroni tests such as the fixed sequence (or hierarchical) test,[5] the Holm procedure,[6] and several gatekeeping procedures.[7-9]

The graphical approach can also be used to generate the weighting scheme (or strategy)[10] for the underlying closed test procedure[11] by generating a local weight for each elementary hypothesis for a given intersection hypothesis. This allows the extension of the graphical approach to weighted Simes tests[10,12] and weighted parametric tests.[10,13] These tests naturally fit into the extended scope of the graphical approach with weighting schemes, as they compare weighted individual p-values with local significance levels. The latter is calculated as the product of the local hypothesis weights and the overall significance level for a given intersection hypothesis.

The choice of hypothesis weights and the order in which multiple hypotheses are tested is typically specific to a particular clinical trial. While there are practical guidelines for some situations,[14,15] elementary hypotheses are often weighted equally and tested with appropriate methods such as the Hochberg procedure.[16,17] To the best of our knowledge, however, graphical visualizations of these test procedures have not yet been described in the literature. One reason is that the weighted version of the Hochberg procedure[18] uses ratios of weights that are incompatible with the underlying weighting scheme. In addition, many omnibus tests do not rely on weighted individual p-values, including Fisher's combination,[19] O'Brien's,[20] and F tests. They combine in different ways the information from the elementary hypotheses but the common underlying assumption is that all elementary hypotheses are equally weighted. Thus, in order to use the Hochberg procedure and the aforementioned omnibus tests with the graphical approach, a weighting scheme needs to be developed that assigns equal weights to the elementary hypotheses. For more complex settings where these aforementioned tests are only applicable to some but not all hypotheses, it is critical to understand the property of the graphs, which generates these weighting schemes so that one does not have to investigate every intersection hypothesis in the closed test framework.

In Section 2, we introduce symmetric graphs and show that symmetry is preserved after hypotheses are rejected and removed from the graph. This property leads to weighting schemes that assign a set of equal weights to each intersection hypothesis within the closed test framework. This allows us to test each intersection hypothesis at its significance level, namely, the sum of its local hypothesis weights multiplied by the overall significance level α. Therefore, we can apply equally weighted tests to any intersection hypothesis and construct a closed test procedure. In Section 3, we formulate the Hochberg procedure as a closed test procedure and visualize it using a suitable symmetric graph. In Section 4, we investigate symmetric component graphs in serial and parallel gatekeeping settings. In Section 5, we provide two clinical trial examples to illustrate the truncated Hochberg procedure in a parallel gatekeeping setting and the Hochberg procedure in a successive graph. In Section 6, we provide concluding remarks and further discussions.

2 | SYMMETRIC GRAPHS FOR EQUALLY WEIGHTED TESTS

Consider testing m elementary null hypotheses $H_i, i \in I = \{1, \dots, m\}$, while strongly controlling the FWER at an overall significance level α. The graphical approach[3] defines a closed test procedure for all intersection hypotheses $H_J = \cap_{j\in J} H_j$, where $\emptyset \neq J \subseteq I$. An intersection test is applied to each intersection hypothesis H_J. The closed test procedure rejects an elementary hypothesis H_i if all H_J with $i \in J \subseteq I$ are rejected.

Using the graphical approach, nodes represent the elementary hypotheses and weighted directed edges determine how the local hypothesis weights are propagated between the nodes. To be more precise, we assign the local weight w_i for each $H_i, i \in I$, in the initial graph. In addition, we denote by g_{ij} the transition weight for the edge from H_i to H_j: If H_i is rejected, we propagate the fraction g_{ij} of its local weight w_i to H_j. We assume the regularity conditions $0 \leq w_i \leq 1$, $\sum_{i\in I} w_i \leq 1$, $0 \leq g_{ij} \leq 1$, $g_{ii} = 0$, and $\sum_{j\in I} g_{ij} \leq 1$.

Let $\boldsymbol{w}_J = (w_j(J), j \in J)$ denote the vector of local weights for all elementary hypotheses H_j with $j \in J$. We call the collection of these vectors across all intersection hypotheses $H_J, J \subseteq I$, a weighting scheme.[7,10] The set of local weights $w_j(J)$ can be determined using Algorithm 1 in the work of Bretz et al.[10] We reproduce this algorithm in Appendix A for convenience and refer to it in the following as Algorithm 1.

We can perform a weighted Bonferroni test at level α for any weight vector $\boldsymbol{w}_J$ generated by Algorithm 1. Let p_i denote the individual, unadjusted p-values for $H_i, i \in I$. The weighted Bonferroni test then rejects the intersection hypothesis H_J, $J \subseteq I$, if $p_j \leq w_j(J)\alpha$ for at least one $j \in J$. The resulting closed test procedure admits a shortcut that is exploited by the original graphical approach from the work of Bretz et al.[3] Alternatively, closed test procedures based on weighted Simes and weighted parametric tests can be applied.[10,12,13]

In order to use the Hochberg procedure or any of the aforementioned omnibus tests with the graphical approach, a weighting scheme needs to be developed that assigns equal weights to the elementary hypotheses. That is, if $w_i(J) = w_j(J)$ for all $i \neq j \in J$ and all $J \subseteq I$, then we can apply any of these tests at the significance level $w_\cdot(J)\alpha$ without relying on the individual local weights, where $w_\cdot(J) = \sum_{j\in J} w_j(J)$.

One way to generate such weighting schemes is through symmetric graphs introduced in the following. An initial graph is symmetric if $w_i = w_j$ and $g_{ij} = g_{ik}$ for all $i \neq j \neq k \in I$ such that every node has the same local weight and all outgoing edges from a same node have the same transition weight. In Appendix B.1, we show that symmetry is preserved after hypotheses are rejected and removed from the graph. That is, updating a symmetric initial graph according to Algorithm 1 leads to symmetric reduced graphs that assign for each intersection hypothesis a set of equal local weights to its elementary hypotheses. Note that we did not require $g_{ij} = g_{ji}$ for the aforementioned considerations. However, because $g_{ij} = g_{ik}$ for

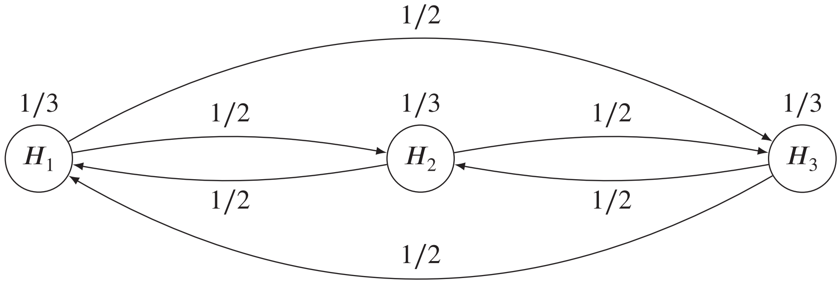

FIGURE 1 A symmetric graph for three hypotheses

TABLE 1 Weighting scheme for the symmetric graph in Figure 1 and the rejection rules for Fisher's combination test as well as for the procedures by Holm and Hochberg. Note that $w.(J) = 1$ for all $J \subseteq I = \{1, 2, 3\}$

	Local weight			Rejection rule		
J	$w_1(J)$	$w_2(J)$	$w_3(J)$	Holm	Fisher	Hochberg (for at least one $j \in J$)
$\{1, 2, 3\}$	1/3	1/3	1/3	$\min_{j\in\{1,2,3\}}\{p_j\} \le \alpha/3$	$t_{\{1,2,3\}} \ge q_{\chi^2_6}(\alpha)$	$p_{(j)}(\{1, 2, 3\}) \le \alpha/(4-j)$
$\{1, 2\}$	1/2	1/2	–	$\min_{j\in\{1,2\}}\{p_j\} \le \alpha/2$	$t_{\{1,2\}} \ge q_{\chi^2_4}(\alpha)$	$p_{(j)}(\{1, 2\}) \le \alpha/(3-j)$
$\{1, 3\}$	1/2	–	1/2	$\min_{j\in\{1,3\}}\{p_j\} \le \alpha/2$	$t_{\{1,3\}} \ge q_{\chi^2_4}(\alpha)$	$p_{(j)}(\{1, 3\}) \le \alpha/(3-j)$
$\{2, 3\}$	–	1/2	1/2	$\min_{j\in\{2,3\}}\{p_j\} \le \alpha/2$	$t_{\{2,3\}} \ge q_{\chi^2_4}(\alpha)$	$p_{(j)}(\{2, 3\}) \le \alpha/(3-j)$
$\{1\}$	1	–	–	$p_1 \le \alpha$	$t_1 \ge q_{\chi^2_2}(\alpha)$	$p_1 \le \alpha$
$\{2\}$	–	1	–	$p_2 \le \alpha$	$t_2 \ge q_{\chi^2_2}(\alpha)$	$p_2 \le \alpha$
$\{3\}$	–	–	1	$p_3 \le \alpha$	$t_3 \ge q_{\chi^2_2}(\alpha)$	$p_3 \le \alpha$

all $i \neq j \neq k \in I$ and $\sum_{j\in I} g_{ij} \le 1$, it is usually less conservative to take $\sum_{j\in I} g_{ij} = 1$ and, thus, $g_{ij} = 1/(|I| - 1) = 1/(m - 1)$, where $|A|$ denotes the cardinality of a set A. In Section 3, we will see that the condition $\sum_{j\in I} g_{ij} = 1$ is a sufficient condition for the Hochberg procedure to be consonant.

As an example of a symmetric graph, consider three equally weighted null hypotheses H_i with $w_i = 1/3$, $i \in I = \{1, 2, 3\}$. Let the transition weights be 1/2 between any pair of hypotheses. Figure 1 visualizes the resulting graph, which is symmetric. Furthermore, we show in Table 1 the weighting scheme as a result of applying Algorithm 1 and conclude that every intersection hypothesis has indeed equal weights for its elementary hypotheses.

In fact, Figure 1 visualizes the Holm procedure[6] if the equally weighted Bonferroni test is used. That is, we reject an intersection hypothesis H_J if $p_j \le \alpha/|J|$ for at least one $j \in J$ (see also Figure 4 in the work of Bretz et al[3]). We can also use other intersection tests instead. Consider, for example, Fisher's combination test[19] that uses $t_J = -2\sum_{j\in J} \log p_j$ to test H_J. Under the assumption of independent uniformly distributed p-values, the distribution of t_J is $\chi^2_{2|J|}$ and we reject H_J if $t_J \ge q_{\chi^2_{2|J|}}(w.(J)\alpha)$, where $q_{\chi^2_{2|J|}}(\cdot)$ denotes the upper quantile of the χ^2 distribution with $2|J|$ degrees of freedom. For convenience, we provide the rejection rules for both the Holm procedure and Fisher's combination test side by side in Table 1. While the Holm procedure uses the individual local weights $w_i(J)$, Fisher's combination test only uses the significance level $w.(J)\alpha$ in the rejection rule for an intersection hypothesis H_J. Following the same idea of using symmetric graphs and weighting schemes, other omnibus tests such as O'Brien's and F tests can be used instead of Fisher's combination test, as their rejection decisions similarly refer only to the significance levels of the intersection hypotheses. Note that, throughout this paper, we do not consider the weighted version of Fisher's combination test[19,21] because the test statistic under the null hypothesis is a weighted sum of independent χ^2 statistics, which is less convenient to work with.

3 | GRAPHICAL REPRESENTATION OF THE HOCHBERG PROCEDURE

The original approach by Hochberg[16] is a popular multiple test procedure that assigns equal weights to the elementary hypotheses. This equally weighted Hochberg procedure strongly controls the FWER at level α under certain dependence conditions of the test statistics.[22-24] Weighted versions of the Hochberg procedure are not considered here as they have pitfalls such as lack of FWER control and lack of monotonicity in rejection decisions in terms of p-values.[18]

As before, assume individual unadjusted p-values $p_1, \dots, p_m$ for $H_1, \dots, H_m$, respectively. Let $p_{(1)} \le p_{(2)} \le \dots \le p_{(m)}$ denote the ordered p-values corresponding to $H_{(1)}, H_{(2)}, \dots, H_{(m)}$, respectively. The Hochberg procedure operates as follows.

1. If $p_{(m)} \leq \alpha$, reject all hypotheses and stop. Otherwise, retain $H_{(m)}$ and test $H_{(m-1)}$.
2. In general, if $p_{(j)} \leq \alpha/(m-j+1)$, reject $H_{(j)}, \dots, H_{(1)}$ and stop. Otherwise, retain $H_{(j)}$ and test $H_{(j-1)}$ until $j = 1$.

Liu[25] showed that the Hochberg procedure performed at level α is equivalent to a closed test procedure, which rejects an intersection hypothesis H_J at level α if

$$p_{(j)}(J) \leq \alpha/(|J| - j + 1) \text{for at least one } j = 1, \dots, |J|, \tag{1}$$

where $p_{(1)}(J) \leq \dots \leq p_{(|J|)}(J)$ are the ordered p-values based on $p_j, j \in J$. Because the rejection rule (1) relies on the significance level of the intersection instead of the local weights of its elementary hypotheses, Figure 1 also visualizes the Hochberg procedure with three hypotheses. The corresponding rejection rules are provided in Table 1 based on (1). In the general case of testing m hypotheses, the Hochberg procedure can be visualized as a symmetric graph with m nodes with weights $w_i = 1/m$ and $g_{ij} = 1/(m-1)$ for all $i \neq j \in I$ (see Figure 1 for an illustration with $m = 3$). Note that Figure 1 also visualizes the weighting scheme for the Hommel procedure[26] with three hypotheses if the Simes test[27] is used for each intersection hypothesis instead of the test based on (1). Although the graphical visualization is the same for the procedures by Holm, Hochberg, and Hommel, the latter two require certain dependence conditions of the test statistics to control the FWER.[22-24]

A desirable property for closed test procedures is consonance, which means that the rejection of an intersection hypothesis implies the rejection of an elementary hypothesis associated with it.[28,29] As a result, consonant closed test procedures admit shortcuts that simplify substantially their implementation and interpretation.[30] For omnibus tests such as Fisher's combination, O'Brien's, and F tests, consonance is difficult to achieve and one usually has to perform the entire closed test procedure.

For the Hochberg procedure, however, Liu[25] provided a sufficient condition for consonance, where each intersection hypothesis has the same critical values for its largest p-value, its second largest p-value, etc. Using the intersection test in (1), this is equivalent to the requirement that each intersection hypothesis should have the same sum of local weights, ie, $w.(J) = w.(J')$ for all $J \neq J' \subseteq I$. In the symmetric graph, we have $w_i = w_j$ and $g_{ij} = g_{ik}$ for all $i \neq j \neq k \in I$. To guarantee $w.(J) = w.(J')$, we need $\sum_{j \in I} g_{ij} = 1$ for all $i \neq j \in I$. In other words, a sufficient condition for consonance using the Hochberg test (1) for a given symmetric weighting scheme is that, for any node, the transition weights of outgoing edges sum to 1. Under this condition, the closed test procedure admits a shortcut that corresponds exactly to the original Hochberg step-up procedure[16] at level $\alpha \sum_{i=1}^{m} w_i = \alpha$. In this symmetric graph, we have $w_i = w_j = 1/m$ and $g_{ij} = 1/(m-1)$ for all $i \neq j \in I$.

To visualize the Hochberg procedure using the graphical approach, assume that $H_{(j)}, H_{(j-1)}, \dots, H_{(1)}$ are rejected by the Hochberg procedure for some $j \in I$ but $H_{(j+1)}, H_{(j+2)}, \dots, H_{(m)}$ are not. Let the index set $J = \{(j+1), \dots, (m)\}$ contain the hypotheses not rejected and $J^c = I \setminus J = \{(j), \dots, (1)\}$ denote the complementary index set of rejected hypotheses.

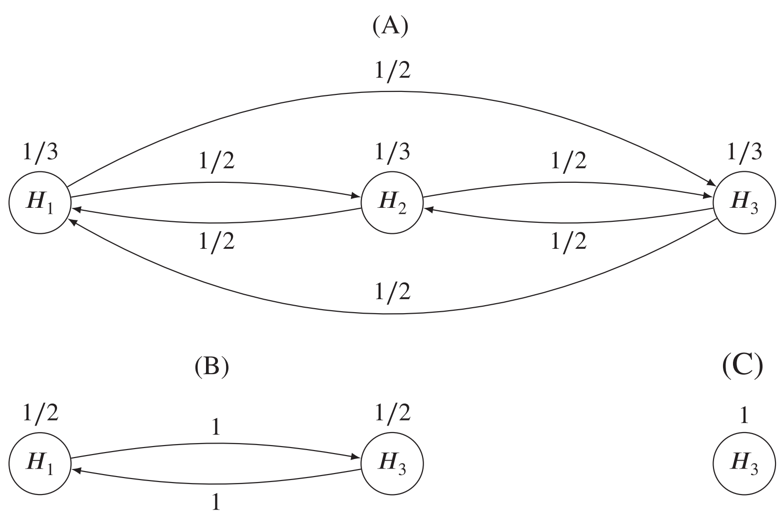

FIGURE 2 Symmetric graphs for three hypotheses using the Hochberg procedure. A, Initial graph; B, Reduced graph after removing H_2; C, Reduced graph after removing H_1, H_2

Then, one can apply Algorithm 1 for each index in J^c and obtain the reduced graphs. As shown in the work of Bretz et al,[3] the weighting scheme is unique, regardless of the sequence in which hypotheses H_j are removed from J^c. The reduced graphs, on the other hand, provide a graphical illustration of the sequentially rejective Hochberg procedure and improve the communication with clinical teams, especially when the Hochberg procedure is part of a more complex multiple test procedure, as illustrated in the next section.

We conclude this section with a numerical example. Let $\alpha = 0.025$ (one-sided) and assume the unadjusted p-values 0.01, 0.011, 0.03 for H_1, H_2, H_3, respectively. The Hochberg procedure fails to reject H_3 but rejects H_2 and H_1 because $p_{(3)} = p_3 = 0.03 > 0.025 = \alpha$ but $p_{(2)} = p_2 = 0.011 < 0.0125 = \alpha/2$. Thus, $J = \{(3)\} = \{3\}$ and $J^c = \{(2), (1)\} = \{2, 1\}$. Figure 2 illustrates the graphical Hochberg procedure with the reduced graphs after rejecting, in sequence, H_2 and H_1. Because the initial graph is symmetric, the reduced graphs are also symmetric. Although the Hochberg procedure rejects $H_{(i)}$ based on the ordered p-values $p_{(i)}$, its graphical representation uses the original order of hypotheses H_i because the propagation of weights is between the original hypotheses to reflect clinical considerations and is not data driven.

4 | EXTENSIONS TO GATEKEEPING PROCEDURES

Symmetric graphs extend the original graphical approach to closed test procedures using intersection tests such as Fisher's combination, O'Brien's, and F tests as well as the Hochberg procedure in the form of (1). In practice, many graphical multiple test procedures are not necessarily symmetric but may have one or more symmetric components, so that the proposed methods can be applied to these component graphs. One application is gatekeeping procedures that test families of hypotheses under certain gatekeeping conditions. In this section, we illustrate the application of symmetric graphs in serial and parallel gatekeeping settings. While the testing procedures are the same as the original proposals in the literature, the graphical illustration facilitates the generalization of serial and parallel gatekeeping procedures to more general settings.

4.1 | Symmetric components in serial gatekeeping procedures

Consider the situation where a family of primary hypotheses is tested, followed by a family of secondary hypotheses. A serial gatekeeping procedure requires that the secondary family is tested only if all hypotheses of the primary family have previously been rejected.[5,31] Figure 3 visualizes such a serial gatekeeping procedure for two primary hypotheses H_1, H_2 and two secondary hypotheses H_3, H_4. Here, ε denotes an infinitesimally small transition weight that effectively ensures the serial gatekeeping condition.[3]

Although Figure 3 is not symmetric, the primary hypothesis family with H_1 and H_2 has a symmetric component graph and so does the secondary family with H_3 and H_4. In addition, the propagation from the primary family to the secondary family is equally split between H_3 and H_4 with a transition weight of $\varepsilon/2$. This ensures symmetry for H_3 and H_4 after the propagation. We show in Table 2 the weighting scheme as a result of applying Algorithm 1. We conclude that each intersection hypothesis with H_1 and H_2 assigns equal weights of $1/2$ to these two hypotheses. A similar conclusion holds for H_3 and H_4. As a result, one can perform a closed test procedure by applying intersection tests such as Fisher's combination, O'Brien's, or F tests to the weighting scheme in Table 2. For the Hochberg procedure, the component graph for H_1, H_2 is a symmetric graph with equal initial weights $w_1(I) = w_2(I) = 1/2$ and transition weights 1. Consonance is satis-

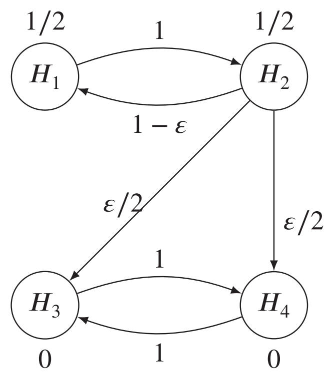

FIGURE 3 Graphical representation of a serial gatekeeping procedure with two symmetric component graphs

	Local weight					Local weight			
J	$w_1(J)$	$w_2(J)$	$w_3(J)$	$w_4(J)$	J	$w_1(J)$	$w_2(J)$	$w_3(J)$	$w_4(J)$
$\{1,2,3,4\}$	1/2	1/2	0	0	$\{2,3\}$	–	1	0	–
$\{1,2,3\}$	1/2	1/2	0	–	$\{2,4\}$	–	1	–	0
$\{1,2,4\}$	1/2	1/2	–	0	$\{3,4\}$	–	–	1/2	1/2
$\{1,3,4\}$	1	–	0	0	$\{1\}$	1	–	–	–
$\{2,3,4\}$	–	1	0	0	$\{2\}$	–	1	–	–
$\{1,2\}$	1/2	1/2	–	–	$\{3\}$	–	–	1	–
$\{1,3\}$	1	–	0	–	$\{4\}$	–	–	–	1
$\{1,4\}$	1	–	–	0					

TABLE 2 Weighting scheme for the serial gatekeeping procedure visualized in Figure 3. Note that $w_{\cdot}(J) = 1$ for all $J \subseteq I = \{1, 2, 3, 4\}$

fied and we can therefore apply the shortcut version of the Hochberg procedure to H_1, H_2. Because one has to complete testing in the primary family of hypotheses before testing the secondary family, the component graph for H_3, H_4 is also a symmetric graph with equal initial weights $w_1(I) = w_2(I) = 1/2$ and transition weights 1, after having rejected H_1, H_2. Thus, the overall test procedure is consonant. In conclusion, Figure 3 visualizes the serial gatekeeping procedure with two symmetric component graphs.

We conclude this section with a remark on consonance for serial gatekeeping procedures. In general, serial gatekeeping procedures test the families of hypotheses in a fixed sequence, ie, in a hierarchical order. Because the fixed sequence test is consonant,[5] a serial gatekeeping procedure is consonant and thus admits a shortcut if we use a consonant multiple test procedure for each family (see Appendix B.2). Consequently, the graphical serial gatekeeping procedure using the Hochberg test (1) is consonant if (i) each family has a symmetric component graph for the consonant Hochberg procedure (see Section 3), and (ii) the propagation from a given family h splits equally among all hypotheses in the subsequent family $h + 1$.

4.2 | Symmetric components in parallel gatekeeping procedures

A parallel gatekeeping procedure requires that the secondary family is tested only if at least one primary hypothesis has previously been rejected.[7] We use entangled graphs[32] to visualize parallel gatekeeping procedures using truncated tests.[8,33] For the sake of illustration, consider testing m primary hypotheses and only one secondary hypothesis. The entangled graph is a combination of two subgraphs. The first subgraph is a graph for serial gatekeeping with a symmetric component graph for the primary hypotheses. Propagation from the primary hypotheses to the secondary hypothesis occurs only when all primary hypotheses are rejected. The second subgraph is a symmetric graph for the Bonferroni test to allow the parallel gatekeeping condition. Propagation of a transition weight 1 exists between each of the primary hypotheses and the secondary hypothesis. To combine the two subgraphs, we assign the weights γ and $1-\gamma$ to the first and second subgraph, respectively. The local weight for a hypothesis in the entangled graph is then a weighted sum of its local weights in each subgraph.

In panels (A) to (C) of Figure 4, we depict the entangled graph for a parallel gatekeeping procedure with two primary hypotheses (H_1 and H_2) and one secondary hypothesis (H_3). The dashed edges with weight 0 in the second subgraph (B) are shown only to emphasize in the entangled graph (C) that there is no propagation between these two hypotheses in the second subgraph. Although the entangled graph is not symmetric, the component graphs for the primary hypotheses H_1 and H_2 are symmetric in both subgraphs 1 and 2. We show in Table 3 the weighting schemes for subgraphs 1 and 2 as a result of applying Algorithm 1. Here, $w_{hi}(J)$ denotes the local weight of H_i in the intersection hypothesis H_J from subgraph h. We conclude that both subgraphs have symmetric weighting schemes for H_1 and H_2. The weighting scheme of the entangled graph is a weighted sum of the weighting schemes of subgraph 1 and 2 using the weights γ and $1 - \gamma$, respectively. As a result, the weighting scheme of the entangled graph is also symmetric and one can perform a closed test procedure by applying intersection tests such as Fisher's combination, O'Brien's, and F tests. For an intersection hypothesis H_J, its significance level for the aforementioned tests is $[\gamma w_{1\cdot}(J) + (1-\gamma)w_{2\cdot}(J)]\alpha$, where $w_{1\cdot}(J)$ and $w_{2\cdot}(J)$ are the sum of local weights for H_J in subgraph 1 and 2, respectively.

In order to apply the Hochberg procedure to the primary hypotheses, we consider the truncated Hochberg procedure[8] that uses a convex combination of the Hochberg and the Bonferroni critical values with the weights γ and $1 - \gamma$, respectively. The truncated Hochberg procedure operates as follows.

1. If $p_{(m)} \leq (\gamma + \frac{1-\gamma}{m})\alpha$, reject all hypotheses and stop. Otherwise, retain $H_{(m)}$ and test $H_{(m-1)}$.
2. In general, if $p_{(j)} \leq (\frac{\gamma}{m-j+1} + \frac{1-\gamma}{m})\alpha$, reject $H_{(j)}, \ldots, H_{(1)}$ and stop. Otherwise, retain $H_{(j)}$ and test $H_{(j-1)}$ until $j = 1$.

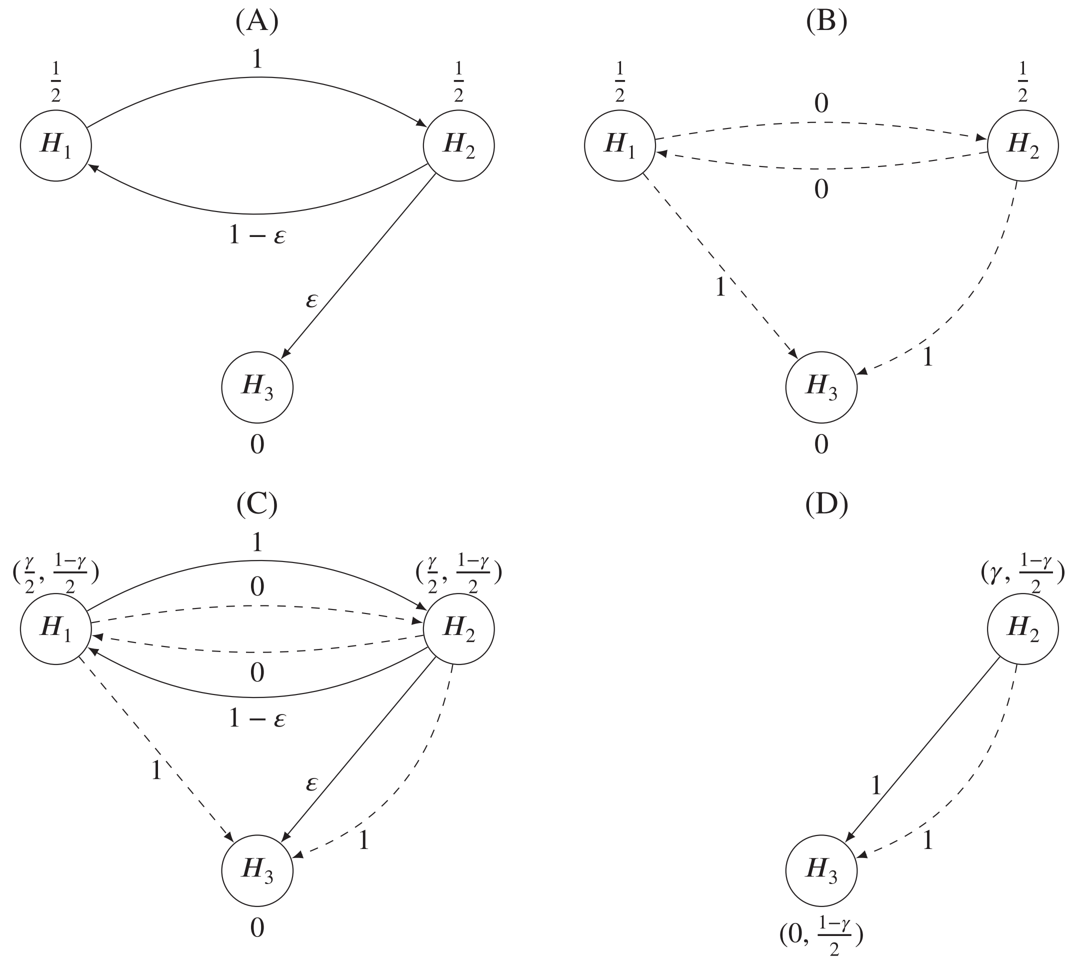

FIGURE 4 Entangled graph of a parallel gatekeeping procedure with symmetric component graphs and weights γ and $1-\gamma$ for subgraphs 1 and 2, respectively. A, Subgraph 1 with weight γ; B, Subgraph 2 with weight $1-\gamma$; C, Entangled graph with weights $\gamma, 1-\gamma$; D, Reduced graph after removing H_1 using truncated Hochberg

TABLE 3 Weighting schemes for the parallel gatekeeping procedure from Figure 4 with two primary hypotheses and one secondary hypothesis

	Subgraph 1 Local weight				Subgraph 2 Local weight			
J	$\gamma w_{11}(J)$	$\gamma w_{12}(J)$	$\gamma w_{13}(J)$	$\gamma w_{1\cdot}(J)$	$(1-\gamma)w_{21}(J)$	$(1-\gamma)w_{22}(J)$	$(1-\gamma)w_{23}(J)$	$(1-\gamma)w_{2\cdot}(J)$
$\{1,2,3\}$	$\gamma/2$	$\gamma/2$	0	γ	$(1-\gamma)/2$	$(1-\gamma)/2$	0	$1-\gamma$
$\{1,2\}$	$\gamma/2$	$\gamma/2$	–	γ	$(1-\gamma)/2$	$(1-\gamma)/2$	–	$1-\gamma$
$\{1,3\}$	γ	–	0	γ	$(1-\gamma)/2$	–	$(1-\gamma)/2$	$1-\gamma$
$\{2,3\}$	–	γ	0	γ	–	$(1-\gamma)/2$	$(1-\gamma)/2$	$1-\gamma$
$\{1\}$	γ	–	–	γ	$(1-\gamma)/2$	–	–	$(1-\gamma)/2$
$\{2\}$	–	γ	–	γ	–	$(1-\gamma)/2$	–	$(1-\gamma)/2$
$\{3\}$	–	–	γ	γ	–	–	$1-\gamma$	$1-\gamma$

As a result, the truncated Hochberg procedure corresponds to an entangled graph based on the Hochberg procedure and the Bonferroni test, in analogy to the truncated Holm procedure that uses an entangled graph based on the Holm procedure and the Bonferroni test.[32] Furthermore, subgraph 1 in panel (A) of Figure 4 represents a serial gatekeeping procedure using the graphical Hochberg procedure for the primary family. Thus, it is consonant since there is only one secondary hypothesis (see Section 4.1). Subgraph 2 in panel (B) of Figure 4 is also consonant since it represents a graphical sequentially rejective procedure using the Bonferroni test for the primary family.[3] As a result, their entangled graph is consonant[32] and admits a shortcut procedure using the truncated Hochberg procedure for the primary family. In the aforementioned example with two primary hypotheses and one secondary hypothesis, panel (D) of Figure 4 illustrates a reduced graph after rejecting H_1 using the truncated Hochberg procedure, assuming $p_2 \geq p_1, p_2 > [\gamma + (1-\gamma)/2]\alpha$ but $p_1 \leq [\gamma/2 + (1-\gamma)/2]\alpha$.

We conclude this section with a remark on consonance for parallel gatekeeping procedures. In general, a two-family parallel gatekeeping procedure using a truncated multiple test procedure (eg, the truncated Holm or Hochberg procedure) for the primary family can be represented as an entangled graph of two subgraphs. The first subgraph is a graphical serial gatekeeping procedure with equal split of the propagation from the primary family to the secondary hypotheses. The second subgraph uses a Bonferroni test for the primary family and propagates equally from each primary hypothesis to all secondary hypotheses. Since the weighting schemes for both subgraphs are consonant, the weighting scheme of the entangled graph is also consonant.[32] Furthermore, the multiple test procedure defined by these two subgraphs is consonant if the local multiple test procedure for each family is consonant in the respective subgraph. The proof is provided in Appendix B.2. Therefore, if we use the truncated Hochberg procedure for the primary family and the Hochberg procedure for the secondary family, subgraph 1 is then the graphical serial gatekeeping procedure using the Hochberg procedure for both families and subgraph 2 uses the Bonferroni test for the primary family and the Hochberg procedure for the secondary family. If we use the respective symmetric component graph for each family in each subgraph and use equal split of the propagation from the primary family to all secondary hypotheses, the resulting graphical parallel gatekeeping procedure is consonant and the shortcut procedure uses the truncated Hochberg procedure for the primary family. We illustrate this with a clinical trial example in Section 5.1.

5 | CLINICAL TRIAL EXAMPLES

5.1 | Parallel gatekeeping procedure

We discuss a clinical trial application to illustrate the graphical representation of the truncated Hochberg procedure in a parallel gatekeeping procedure. Dmitrienko et al[7] described a clinical trial in patients with acute lung injury and the acute respiratory distress syndrome to show superiority of an investigational treatment versus control. There are two primary hypotheses on the number of days alive and on mechanical ventilation during a 28-day study period (H_1) and 28-day all-cause mortality rate (H_2). There are two secondary hypotheses on the number of days the patients were out of the intensive care unit (H_3) and general quality of life (H_4). Since regulatory claims can be made with the rejection of either H_1 or H_2, the parallel gatekeeping structure allows testing of H_3 and H_4 if at least one primary hypothesis is rejected.

Dmitrienko et al[7] used weighted Bonferroni and Simes tests for the primary family in a closed test framework. We modify their multiple test strategy and use the truncated Hochberg procedure as an illustration of a powerful test procedure implemented in a shortcut fashion. This will also provide a graphical visualization of the general parallel gatekeeping procedure by Dmitrienko et al.[8] As a result, we assign equal local weights and allow directed edges between the primary as well as the secondary hypotheses. We use an entangled graph to visualize this parallel gatekeeping procedure. The first subgraph visualizes the serial gatekeeping procedure in Figure 3 with symmetric component graphs for primary and secondary hypotheses, respectively. The second subgraph uses a symmetric component graph with no propagation for primary hypotheses and a symmetric component graph with propagation of transition weight 1 for secondary hypotheses. The transition weight is $1/2$ from a primary hypothesis to each of the two secondary hypotheses to enable the parallel gatekeeping strategy. To combine the two subgraphs, we assign the weights γ and $1-\gamma$ to the first and second subgraph, respectively. We illustrate the resulting subgraphs in panels (A) and (B) and the entangled graph in panel (C) of Figure 5. By applying Algorithm 1, we can confirm that both subgraphs have symmetric weighting schemes for H_1, H_2 and H_3, H_4, respectively. As a result, the weighting scheme of the entangled graph is also symmetric.

We apply the truncated Hochberg procedure for H_1, H_2 and the Hochberg procedure for H_3, H_4. Note that subgraph 1 illustrates the serial gatekeeping procedure using the shortcut procedure of the Hochberg procedure for both primary and secondary hypotheses (see Section 4.1). At the same time, subgraph 2 also admits a shortcut if the Bonferroni test is applied to H_1, H_2 and the Hochberg procedure to H_3, H_4 (see Section 4.2). As a result, the entangled graph admits a shortcut procedure,[32] which uses the truncated Hochberg procedure for H_1, H_2 and the Hochberg procedure for H_3, H_4.

As an illustration, consider the unadjusted p-values 0.01, 0.015, 0.011, 0.03 for H_1, H_2, H_3, H_4, respectively. The overall significance level is $\alpha = 0.025$ (one-sided) and we set $\gamma = 0.5$. First, the truncated Hochberg procedure rejects both H_1 and H_2 because $p_2 > p_1$ and $p_2 = 0.015 < 0.01875 = (0.5 + 0.5/2) \cdot 0.025$. Thus, $J_{\{1,2\}} = \emptyset$ and $J^c_{\{1,2\}} = \{1, 2\}$. Panels (D) and (E) of Figure 5 illustrate the reduced graphs after rejecting, in sequence, H_1 and H_2 by applying Algorithm 1 to each subgraph, or equivalently, by applying Algorithm 2 in the work of Maurer and Bretz.[32] After rejecting H_1 and H_2, the secondary hypotheses H_3 and H_4 can be tested using the Hochberg procedure at level $\alpha = 0.025$, as illustrated by the symmetric graph

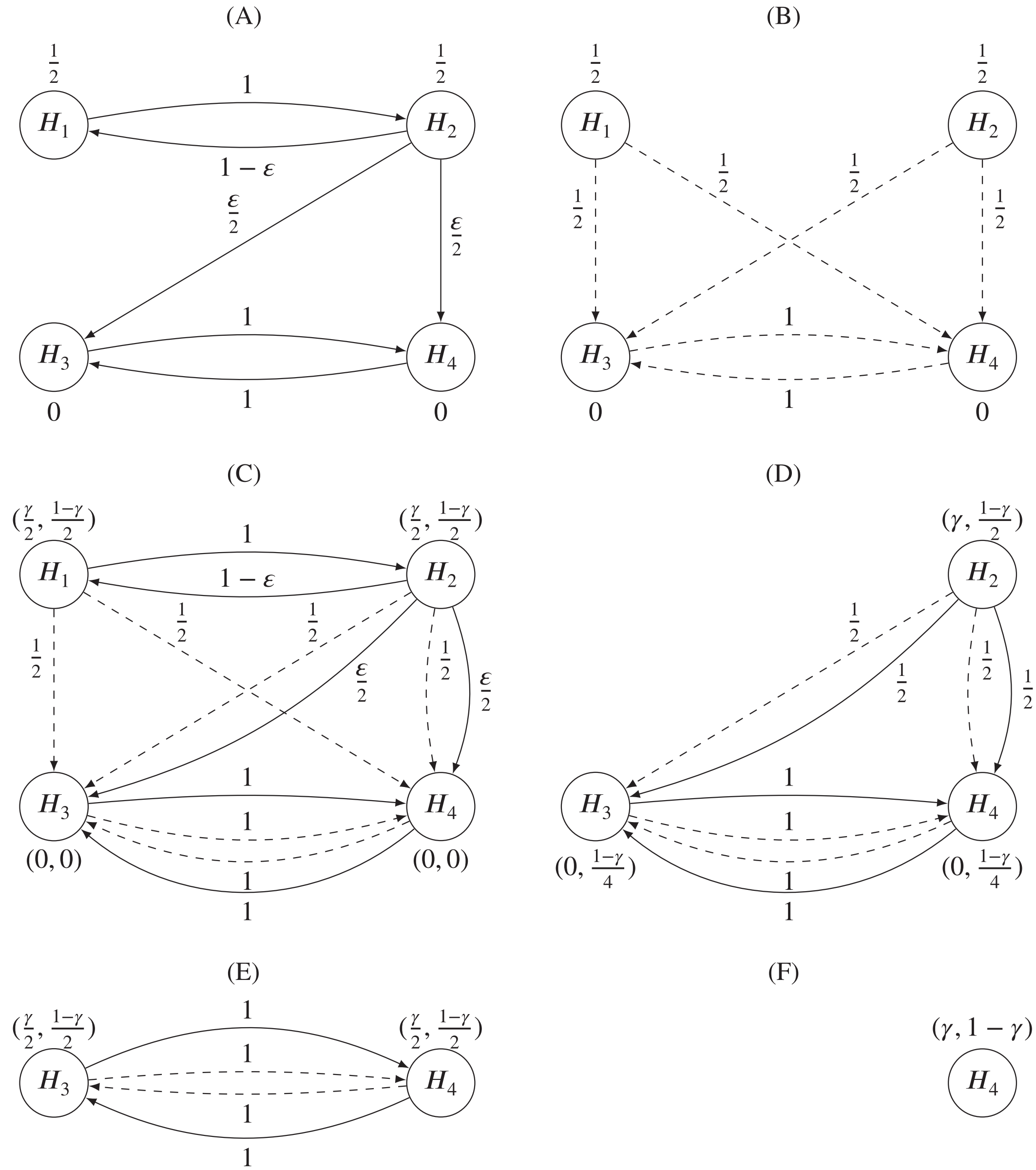

FIGURE 5 Entangled graph of a parallel gatekeeping procedure using the truncated Hochberg procedure for H_1, H_2 and the Hochberg procedure for H_3, H_4 with weights γ and $1-\gamma$ for subgraphs 1 and 2, respectively. A, Subgraph 1 with weight γ; B, Subgraph 2 with weight $1-\gamma$; C, Entangled graph with weights $\gamma, 1-\gamma$; D, Reduced graph after removing H_1; E, Reduced graph after removing H_1, H_2; F, Reduced graph after removing H_1, H_2, H_3

in panel (E) of Figure 5. Thus, we reject H_3 but not H_4 because $p_4 = 0.03 > 0.025$ and $p_3 = 0.011 \leq 0.025/2 = 0.0125$ (see panel (F) of Figure 5). In conclusion, we can reject H_1, H_2, H_3 but not H_4 while controlling the FWER in the strong sense at level α.

5.2 | Successive graph with symmetric components

In the second example, we consider a successive graph in which a secondary hypothesis is tested only if its parent primary hypothesis has been rejected before. Following Maurer et al,[34] we consider testing four hypotheses: two primary hypotheses H_1, H_2 and two secondary hypotheses H_3, H_4, where H_3 is the descendant hypothesis of H_1 and H_4 is the descendant hypothesis of H_2. That is, $H_3(H_4)$ will not be rejected if $H_1(H_2)$ has not been rejected before (successiveness property). As a result, we assign initially the weights 0 to each of H_3 and H_4 and 1/2 to each of H_1 and H_2. When $H_1(H_2)$ is rejected, one half of its weight is propagated to $H_2(H_1)$ and the other half is propagated to $H_3(H_4)$. When the secondary hypothesis $H_3(H_4)$ is rejected, its weight is propagated to the other primary hypothesis $H_2(H_1)$. Panel (A) of Figure 6 visualizes the resulting graph. Note that this graph is similar to a mixture-based gatekeeping procedure of multiple sequences of hypotheses,[35] although the graphical approach is more flexible in fine-tuning hypothesis weights and transition weights. If the Bonferroni test is applied to the graph in panel (A) of Figure 6, one can implement a sequentially rejective test procedure following Algorithm 1 in Bretz et al.[3]

In one of our consulting cases, the clinical trial team was interested in including an additional secondary endpoint. As a result, H_3 is replaced by two secondary hypotheses H_{31}, H_{32}, and H_4 by H_{41}, H_{42}. When a primary hypothesis is rejected, one half of its weight is propagated to the other primary hypothesis and the other half is split between its two associated descendant secondary hypotheses. When both secondary hypotheses (H_{31}, H_{32} or H_{41}, H_{42}) are rejected, their weights are propagated to the other primary hypotheses (H_2 or H_1, respectively). Panel (B) of Figure 6 visualizes the corresponding graph. Note that panels (A) and (B) of Figure 6 illustrate two graphs with different levels of details. While the two graphs have the same structure, panel (A) collapses the two secondary hypotheses into one node. Since both secondary hypotheses must be rejected to propagate the weight to the other primary hypothesis, the two graphs in Figure 6 can be used interchangeably to facilitate communication and implementation.

In this particular consulting case, the positive dependence condition[22] is satisfied between the two secondary endpoints. Thus, the team preferred using the Hochberg procedure for each of the two secondary families of hypotheses. The graph in panel (B) of Figure 6 allows using the Hochberg procedure because the component graphs for H_{31}, H_{32} and H_{41}, H_{42} are both symmetric. More specifically, (i) the incoming transition weights are the same for H_{31} and H_{32}; (ii) the transition weights between H_{31} and H_{32} are the same; and (iii) when both hypotheses are rejected, their weights are propagated to the other primary hypothesis H_2. Thus, the component graph for H_{31} and H_{32} is symmetric and it can be verified that the weighting scheme for H_{31} and H_{32} is also symmetric. Similar arguments hold also for H_{41} and H_{42} and the Hochberg procedure can be applied again.

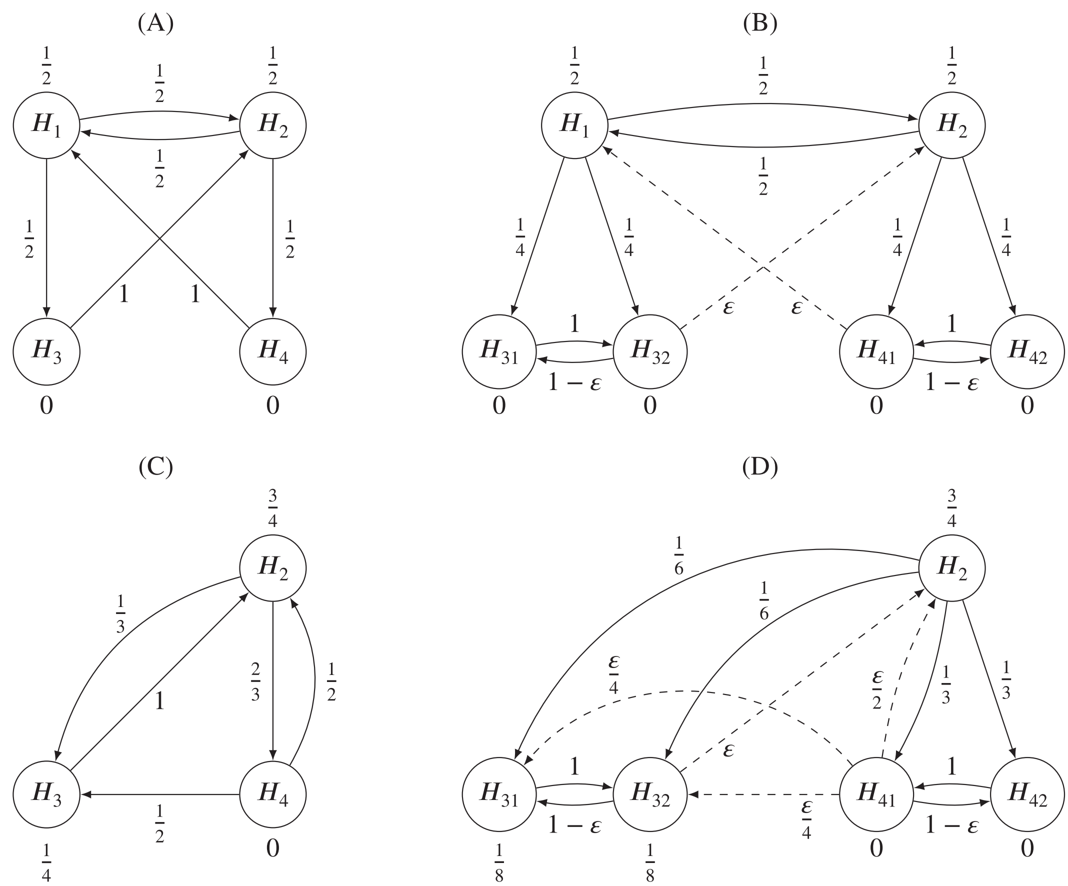

FIGURE 6 Successive graphs with symmetric components. A, Successive graph with four hypotheses; B, Successive graph with two symmetric components; C, Reduced graph of panel (A) after removing H_1; D, Reduced graph of panel (B) after removing H_1

There are two approaches to implement the multiple test procedure based on panel (B) of Figure 6. With the first approach, we derive the underlying closed test procedure. Using Algorithm 1, one can derive the weighting scheme and then apply the intersection test (1) to H_{31}, H_{32} and H_{41}, H_{42}, respectively, at the significance level of each of the two families of secondary hypotheses. Using the idea of Bretz et al,[10] one can then combine the Hochberg intersection test with the Bonferroni test applied to the remaining hypotheses. For example, the intersection hypothesis $H_{\{2,31,32,41,42\}}$ has the weight vector $(3/4, 1/8, 1/8, 0, 0)$, as illustrated in panel (D) of Figure 6. Note that symmetry is preserved for H_{31}, H_{32} and H_{41}, H_{42}, respectively. One can reject $H_{\{2,31,32,41,42\}}$ if the Bonferroni test rejects H_2 at level $3\alpha/4$ or the Hochberg test in (1) rejects $H_{31} \cap H_{32}$ at level $\alpha/8 + \alpha/8 = \alpha/4$. With the second approach, we perform the multiple test procedure in two steps using a shortcut. In the first step, we implement the sequentially rejective procedure to panel (A) of Figure 6. The p-values are $p_3 = \max\{p_{31}, p_{32}\}$ for H_3 and $p_4 = \max\{p_{41}, p_{42}\}$ for H_4. The rationale is that the rejection of $H_3(H_4)$ is equivalent to the rejection of both $H_{31}, H_{32}(H_{41}, H_{42})$, which happens only if the maximum of $p_{31}, p_{32}(p_{41}, p_{42})$ is below the significance level for the family of $H_{31}, H_{32}(H_{41}, H_{42})$. In the second step, we can then apply the Hochberg procedure at the respective level to H_{31} and H_{32} and similarly to H_{41} and H_{42}, depending on the results from the first step. Note that both approaches earlier lead to the same test decisions because the propagation from one family of secondary hypotheses to a primary hypothesis happens only when both hypotheses in this family are rejected.

To illustrate the methodology, assume the unadjusted p-values 0.01, 0.02, 0.002, 0.01, 0.02, 0.03 for $H_1, H_2, H_{31}, H_{32}, H_{41}, H_{42}$, respectively. Thus, $p_3 = \max\{0.002, 0.01\} = 0.01$ and $p_4 = \max\{0.02, 0.03\} = 0.02$. Then, we can apply Algorithm 1 in the work of Bretz et al[3] to panel (A) of Figure 6 at the one-sided significance level 0.025. As a result, only H_1 is rejected because $p_1 < 0.025/2 = 0.0125$, $p_2 > 0.025/2 + 0.025/4 = 0.01875$, and $p_3 > 0.025/4 = 0.00625$, as illustrated in panel (C) of Figure 6. Then, in the second step, we can apply the Hochberg procedure to H_{31} and H_{32} at level $0.025/4 = 0.00625$. Since $p_{32} > 0.00625$ but $p_{31} < 0.00625/2 = 0.003125$, we can reject H_{31}. In summary, we reject H_1 and H_{31} using the successive graph in Figure 6 with the Hochberg procedure applied to each of the two families of the secondary hypotheses.

6 | DISCUSSION

In this paper, we extended the graphical approach from the work of Bretz et al[3] to visualize the Hochberg procedure as well as closed versions of various omnibus tests such as Fisher's combination, O'Brien's, and F tests. The key is to develop a weighting scheme that assigns equal weights to the elementary hypotheses for every intersection hypothesis within the closed testing framework. We proposed the use of symmetric graphs and showed that Algorithm 1 provides such weighting schemes under conditions that are easy to verify. One main application of the proposed approach is the use of symmetric graphs as components of complex multiple test strategies. In particular, we illustrated how to visualize the Hochberg and the truncated Hochberg procedures in serial and parallel gatekeeping settings. In both cases, we derived shortcut procedures that simplify substantially their implementation and interpretation.

For more complex graphs, the serial gatekeeping procedure can be generalized to a graph where hypotheses are grouped into nonhierarchical families and the propagation is implemented among families if all hypotheses are rejected within a family.[36] The second clinical trial example in Section 5 incorporates the Hochberg procedure in such a setting and it allows the use of shortcuts. Other gatekeeping procedures, such as k-out-of-n gatekeeping procedures,[37] can be represented by entangled graphs so that the ideas developed in this paper remain applicable.

In some applications, shortcut procedures may not exist. For example, Bauer et al[38] proposed a multiple test procedure to demonstrate the superiority of three doses of an investigational treatment against a control for an efficacy and a safety endpoint. The authors used the Bonferroni test for all intersection hypotheses, leading to the Bonferroni-based graphical multiple test procedures shown in figure 8 in the work of Bretz et al.[3] Using the methods proposed in this paper, one can obtain a less conservative test procedure by applying, for example, the Hochberg intersection test (1) to test, separately, the efficacy and safety hypotheses. However, the resulting graphical test procedure is an example of a successive graph in which the hierarchical requirement between a primary and a secondary hypothesis is mixed with other considerations.[34] Such applications do not admit shortcuts in general and the closed test procedure has to be implemented. Likewise, omnibus tests such as Fisher's combination, O'Brien's, and F tests, in general, do not admit shortcut procedures and the closed test procedure has to be used along with the graph generating the weighting scheme.

Another future direction of research is to extend the graphical Hochberg procedure proposed in this paper to group sequential and adaptive designs. Maurer and Bretz[39] proposed a group sequential design for graphical approaches and derived a shortcut procedure, which includes the group sequential Holm procedure[40] and can be extended to allow different timing for propagation.[41] Extensions of the original graphical approach to adaptive designs also exist that allow for midtrial changes,[42,43] including enrichment of a certain subgroup.[44,45] All these approaches use Bonferroni or parametric tests. It will be interesting and relevant to clinical trials to explore the graphical Hochberg procedure in these settings.

ACKNOWLEDGEMENTS

Parts of this manuscript were written while the second author (F.B.) was on a sabbatical leave at University of Canterbury in Christchurch, New Zealand. He would like to thank Dr. Daniel Gerhard for his support. The authors are grateful to two referees and an associate editor for their valuable comments, which improved the presentation of this paper.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Dong Xi https://orcid.org/0000-0003-1482-234X
Frank Bretz https://orcid.org/0000-0002-2008-8340

REFERENCES

1. Food and Drug Administration. Guidance for industry multiple endpoints in clinical trials (draft). https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm536750.pdf. Published January 2017.
2. European Medicines Agency. Guideline on multiplicity issues in clinical trials (draft). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/03/WC500224998.pdf. Published April 2017.
3. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Statist Med*. 2009;28(4):586-604.
4. Burman C-F, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. *Statist Med*. 2009;28(5):739-761.
5. Maurer W, Hothorn LA, Lehmacher W. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. *Biomed Chem Pharm Ind*. 1995;6:3-18.
6. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6(2):65-70.
7. Dmitrienko A, Offen WW, Westfall PH. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statist Med*. 2003;22(15):2387-2400.
8. Dmitrienko A, Tamhane AC, Wiens BL. General multistage gatekeeping procedures. *Biometrical Journal*. 2008;50(5):667-677.
9. Li Y, Wang X, Wang D, Offen W. Applications of general multistage gatekeeping and graphical multiple testing strategies in a clinical trial setting. *Contemp Clin Trials*. 2019;76:9-15.
10. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biometrical Journal*. 2011;53(6):894-913.
11. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655-660.
12. Lu K. Graphical approaches using a Bonferroni mixture of weighted Simes tests. *Statist Med*. 2016;35(22):4041-4055.
13. Xi D, Glimm E, Maurer W, Bretz F. A unified framework for weighted parametric multiple test procedures. *Biomertical Journal*. 2017;59(5):918-931.
14. Bretz F, Maurer W, Hommel G. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. *Statist Med*. 2011;30(13):1489-1501.
15. Wiens BL, Dmitrienko A, Marchenko O. Selection of hypothesis weights and ordering when testing multiple hypotheses in clinical trials. *J Biopharm Stat*. 2013;23(6):1403-1419.
16. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988;75(4):800-802.

17. Gou J, Tamhane AC, Xi D, Rom D. A class of improved hybrid Hochberg–Hommel type step-up multiple test procedures. *Biometrika*. 2014;101(4):899-911.
18. Tamhane AC, Liu L. On weighted Hochberg procedures. *Biometrika*. 2008;95(2):279-294.
19. Fisher RA. *Statistical Methods for Research Workers*. London, UK: Oliver and Boyd; 1922.
20. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984;40:1079-1087.
21. Brannath W, Posch M, Bauer P. Recursive combination tests. *J Am Stat Assoc*. 2002;97(457):236-244.
22. Sarkar SK. Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Ann Stat*. 1998;26:494-504.
23. Huque MF. Validity of the Hochberg procedure revisited for clinical trial applications. *Statist Med*. 2016;35(1):5-20.
24. Gou J, Tamhane AC. Hochberg procedure under negative dependence. *Statistica Sinica*. 2018;28:339-362.
25. Liu W. Multiple tests of a non-hierarchical finite family of hypotheses. *J R Stat Soc Series B Stat Methodol*. 1996;58:455-461.
26. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*. 1988;75(2):383-386.
27. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 1986;73(3):751-754.
28. Gabriel KR. Simultaneous test procedures–some theory of multiple comparisons. *Ann Math Stat*. 1969:224-250.
29. Brannath W, Bretz F. Shortcuts for locally consonant closed test procedures. *J Am Stat Assoc*. 2010;105(490):660-669.
30. Hommel G, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statist Med*. 2007;26(22):4063-4073.
31. Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *J Stat Plan Inference*. 2001;99(1):25-40.
32. Maurer W, Bretz F. Memory and other properties of multiple test procedures generated by entangled graphs. *Statist Med*. 2013;32(10):1739-1753.
33. Strassburger K, Bretz F. Compatible simultaneous lower confidence bounds for the Holm procedure and other Bonferroni-based closed tests. *Statist Med*. 2008;27(24):4914-4927.
34. Maurer W, Glimm E, Bretz F. Multiple and repeated testing of primary, coprimary, and secondary hypotheses. *Stat Biopharm Res*. 2011;3(2):336-352.
35. Dmitrienko A, Kordzakhia G, Brechenmacher T. Mixture-based gatekeeping procedures for multiplicity problems with multiple sequences of hypotheses. *J Biopharm Stat*. 2016;26(4):758-780.
36. Maurer W, Bretz F. A note on testing families of hypotheses using graphical procedures. *Statist Med*. 2014;33(30):5340-5346.
37. Xi D, Tamhane AC. A general multistage procedure for *k*-out-of-*n* gatekeeping. *Statist Med*. 2014;33(8):1321-1335.
38. Bauer P, Brannath W, Posch M. Multiple testing for identifying effective and safe treatments. *Biometrical Journal*. 2001;43(5):605-616.
39. Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. *Stat Biopharm Res*. 2013;5(4):311-320.
40. Ye Y, Li A, Liu L, Yao B. A group sequential Holm procedure with multiple primary endpoints. *Statist Med*. 2013;32(7):1112-1124.
41. Xi D, Tamhane AC. Allocating recycled significance levels in group sequential procedures for multiple endpoints. *Biometrical Journal*. 2015;57(1):90-107.
42. Klinglmueller F, Posch M, König F. Adaptive graph-based multiple testing procedures. *Pharm Stat*. 2014;13(6):345-356.
43. Sugitani T, Bretz F, Maurer W. A simple and flexible graphical approach for adaptive group-sequential clinical trials. *J Biopharm Stat*. 2016;26(2):202-216.
44. Rosenblum M, Qian T, Du Y, Qiu H, Fisher A. Multiple testing procedures for adaptive enrichment designs: combining group sequential and reallocation approaches. *Biostatistics*. 2016;17(4):650-662.
45. Sugitani T, Posch M, Bretz F, König F. Flexible alpha allocation strategies for confirmatory adaptive enrichment clinical trials with a prespecified subgroup. *Statist Med*. 2018;37(24):3387-3402.

How to cite this article: Xi D, Bretz F. Symmetric graphs for equally weighted tests, with application to the Hochberg procedure. *Statistics in Medicine*. 2019;38:5268–5282. https://doi.org/10.1002/sim.8375

APPENDIX A

ALGORITHM 1

We reproduce algorithm 1 in the work of Bretz et al[10] for convenience and refer to it as Algorithm 1. Let $I = \{1, \dots, m\}$ and $w_i = w_i(I), i \in I$. For a given proper subset $J \subset I$, let $J^c = I \setminus J$ denote the set of indices that are not contained in J. The following algorithm then determines the weights $w_j(J), j \in J$. It has to be repeated for each $J \subseteq I$ to generate the $m2^{m-1}$ weights of the complete weighting scheme. As shown in the work of Bretz et al,[3] the weighting scheme is unique, regardless of the sequence in which hypotheses H_j are removed from J^c.

(1) Select $j \in J^c$.
(2) Update the graph

$$I \to I \setminus \{j\}, J^c \to J^c \setminus \{j\}$$
$$w_\ell(I) \to \begin{cases} w_\ell(I) + w_j(I) g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \to \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{jk}}{1 - g_{\ell j} g_{j\ell}}, & \ell \neq k \in I, g_{\ell j} g_{j\ell} < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(3) If $|J^c| \geq 1$, go to step (1); otherwise, set $w_j(J) = w_j(I)$ and stop.

APPENDIX B

PROOFS

B.1 | Proof for Section 2

We prove that updating a symmetric graph according to Algorithm 1 leads to reduced graphs that are again symmetric. To be more precise, assume an initial symmetric graph with $w_i = w_j$ and $g_{ij} = g_{ik}$ for all $i \neq j \neq k \in I$ for the null hypotheses $H_i, i \in I = \{1, \dots, m\}$. We need to show that a reduced graph remains symmetric after removing any index $k \in J^c$ from I, where $J^c = I \setminus J$ denotes the set of indices that are not contained in J for a given proper subset $J \subset I$.

Assume that $J^c = \{k\}$ and k is removed from I. According to Algorithm 1, the local weights $w_i(J) = w_i + w_k g_{ki}$ and $w_j(J) = w_j + w_k g_{kj}$ for $J = I \setminus \{k\}$ are the same because $w_i = w_j$ and $g_{ki} = g_{kj}$. For the transition weights, we have for $i, j, h \in J$

$$g_{ij}(J) = \frac{g_{ij} + g_{ik} g_{kj}}{1 - g_{ik} g_{ki}} = \frac{g_{ih} + g_{ik} g_{kh}}{1 - g_{ik} g_{ki}} = g_{ih}(J).$$

Thus, the reduced graph for $J = I \setminus \{k\}$ is also symmetric with the local weights being equal for all elementary hypotheses $H_j, j \in J$.

By induction, we can show that any reduced graph of an initial symmetric graph is symmetric and, thus, its local weights are equal for all elementary hypotheses involved.

B.2 | Proofs for Section 4

First, we show that a serial gatekeeping procedure in Section 4.1 is consonant if we use a consonant multiple test procedure for each family. To this end, assume that we reject the intersection hypothesis $H_J = \cap_{j \in J} H_j$ in the hypotheses family h.

Assume further that H_j has a positive local weight so that the hypotheses in any other family have local weight 0. Since only hypotheses in family h have positive local weights, the intersection test for H_J reduces to the intersection test used by the multiple test procedure for the family h, which is consonant. Thus, at least one of the hypotheses in the family h must be rejected by this multiple test procedure and therefore by the overall serial gatekeeping procedure.

Second, we show that the two subgraphs of the entangled graph for the two-family parallel gatekeeping procedure in Section 4.2 are consonant if we use a consonant multiple test procedure for each family in each subgraph. From Section 4.1, subgraph 1 is consonant if we use a consonant multiple test procedure for each family. In addition, subgraph 2 is consonant if the multiple test procedure for the secondary family is consonant. To see this, note that primary hypotheses always have the same local weights in any intersection hypothesis. Thus, we focus on an intersection hypothesis involving all primary hypotheses not rejected by the Bonferroni test and all secondary hypotheses. Then, the rejection of this intersection hypothesis leads to the rejection of a secondary hypothesis by its multiple test procedure, which is consonant. Thus, a secondary hypothesis must be rejected by the closed test procedure represented by subgraph 2.