Testing partial conjunction hypotheses under dependency, with applications to meta-analysis[*]

Marina Bogomolov

*Faculty of Data and Decision Sciences, Technion - Israel Institute of Technology,
Haifa 3200003, Israel*
e-mail: marinabo@technion.ac.il

Abstract: In many statistical problems the hypotheses are naturally divided into groups, and the investigators are interested to perform group-level inference, possibly along with inference on individual hypotheses. We consider the goal of discovering groups containing u or more signals with group-level false discovery rate (FDR) control. This goal can be addressed by multiple testing of partial conjunction hypotheses with a parameter u, which reduce to global null hypotheses for $u = 1$. We consider the case where the partial conjunction p-values are combinations of within-group p-values, and obtain sufficient conditions on (1) the dependencies among the p-values within and across the groups, (2) the combining method for obtaining partial conjunction p-values, and (3) the multiple testing procedure, for obtaining FDR control on partial conjunction discoveries. We consider separately the dependencies encountered in the meta-analysis setting, where multiple features are tested in several independent studies, and the p-values within each study may be dependent. Based on the results for this setting, we generalize the procedure of Benjamini, Heller, and Yekutieli (2009) for assessing replicability of signals across studies, and extend their theoretical results regarding FDR control with respect to replicability claims.

Keywords and phrases: False discovery rate, global null, meta-analysis, partial conjunction hypothesis, replicability analysis.

1. Introduction

In modern research, the investigators often face complex statistical problems, which involve testing a large number of hypotheses. In many applications the hypotheses have a natural group structure, and identifying groups containing signals, (i.e. false null hypotheses), is often of interest. For example, in functional magnetic resonance imaging, several types of division into groups have been considered. If the goal is to find locations (voxels) of activation while a subject performs a certain cognitive task, each hypothesis corresponds to a single location. The division of the brain into anatomical or functional regions may define groups of hypotheses, and the researchers may be interested to find brain

---

regions containing activated locations (see, for example, [45] and [2]). If several cognitive tasks are studied, one may be interested to find brain locations that are activated in at least two (or, more generally, at least a pre-specified number u of) cognitive tasks. In this case the hypotheses regarding the activation in the same location in different tasks define a group, and the goal is to find groups (i.e. locations) with at least u signals. Motivated by this application, Benjamini and Heller [3] introduced the *partial conjunction hypothesis* testing problem, which is defined for group g as

$$H_0^{u/n}(g) : k_g < u \text{ versus } H_1^{u/n}(g) : k_g \geq u,$$

where k_g is the number of false null hypotheses in group g. In the case where $u = 1$, the partial conjunction null reduces to the global null, or the intersection hypothesis, which holds if all the null hypotheses in the group are true. Benjamini and Heller [3] developed a general method for combining the p-values for the elementary hypotheses in group g, so that the combined p-value is valid for testing $H_0^{u/n}(g)$. For the fMRI application above, [3] suggested identifying the brain locations which are activated in at least u out of n cognitive tasks by multiple testing of the family $\{H_0^{u/n}(g), \ g = 1, \ldots G\}$, where G is the number of brain locations considered. In this setting, the false discovery corresponds to rejecting $H_0^{u/n}(g)$ for some group g which contains less than u signals, which, in this application, translates to erroneously stating that the brain location g, with $k_g < u$, is activated in at least u cognitive tasks. The false discovery rate (FDR, [5]) and the familywise error rate (FWER) for the family of partial conjunction hypotheses are defined with respect to these erroneous group discoveries. Other examples of inference at the group level in fMRI research were given in [23] and [1], who developed different algorithms incorporating group-level inference. Besides fMRI research, there are many other areas where multiple testing of partial conjunction hypotheses has been considered, for example in genomic research, where several hypotheses are tested for each genetic variant (see, e.g., [46], [31]).

An important statistical challenge where multiple testing of partial conjunction hypotheses may be very useful is that of meta-analysis. Consider the setting where the same set of features is tested in several independent studies. The hypotheses tested in different studies regarding the same feature may be considered as a group. The global null hypothesis corresponding to a single feature addresses the question: "Is this feature non-null (i.e. is it corresponding to a false null hypothesis) in at least one of the studies?" Identifying the features with a positive answer to this question corresponds to the typical goal of meta-analysis. However, sometimes the evidence that the feature is non-null in at least one study is not enough, and the researchers wish to identify features with consistent findings across studies, which corresponds to the goal of replicability analysis (see [15] and references therein). In this case the interest may be to find features which are non-null in at least $u \geq 2$ studies. Benjamini et al. [4] suggested addressing this goal by testing the family $\{H_0^{u/n}(g), \ g = 1, \ldots, G\}$, where n is the number of studies and G is the number of features considered. The larger is the value of u, the stronger is our requirement for claiming replicability. Since it may be not

clear what value of u should be chosen, [4] suggested to select the features with potential replicability findings by applying the Benjamini-Hochberg procedure (BH, [5]) on the family of global null p-values for all the features, and assess the replicability strength for each selected feature g by a lower bound on the true number of studies where the feature is non-null, $\hat{k}(g)$, obtained from testing the partial conjunction hypotheses in order with $u = 1, \ldots, n$. The authors proved that their procedure controls the expected proportion of features with erroneously reported lower bounds (i.e. features with $\hat{k}(g) > k(g)$), out of all features selected, under independence within each study. For real data examples illustrating replicability analysis, see, e.g., [4], [30], [32], [58], [18].

In this paper, we consider multiple testing of partial conjunction hypotheses with FDR or weighted FDR ([6]) control. The main goal of this paper is to develop sufficient conditions on (1) the multiple testing procedure, (2) the dependency structure among the p-values for the elementary hypotheses (referred to as elementary p-values hereafter), and (3) the method for obtaining partial conjunction p-values, for (weighted) FDR control on the family of partial conjunction hypotheses. This goal is addressed by building upon the results of [12], [13], [48], [3], and [4]. We consider separately testing the partial conjunction hypotheses with (weighted) FDR control under the dependencies of the meta-analysis setting. We generalize the theoretical results of [4] for their procedure for assessing the strength of replicability for the selected features, showing that this procedure provides the above error rate control regarding false replicability claims not only under independence within each study, but also under a certain form of positive dependence among the p-values within each study, which is typically considered in multiple testing literature and is defined below. Recall that a set $D \subset \mathbb{R}^m$ is called non-decreasing if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ such that $\boldsymbol{x} \leq \boldsymbol{y}$ (i.e. $x_i \leq y_i$ for $i = 1, \ldots, m$), $\boldsymbol{x} \in D$ implies that $\boldsymbol{y} \in D$.

Definition 1.1 (PRDS, [8]). *The vector $\boldsymbol{X} = (X_1, \ldots, X_m)$ is positive regression dependent on a subset $I_0 \subseteq \{1, \ldots, m\}$ if for any non-decreasing set D, and for each $i \in I_0$, $\mathbb{P}[\boldsymbol{X} \in D \,|\, X_i = x]$ is non-decreasing in x.*

In addition, we extend the procedure of [4] for addressing (1) a wide class of selection rules for selecting the promising features, (2) weights reflecting prior knowledge on the replicability extent of the features, as well as weights reflecting different penalties for erroneous replicability claims (similar to [24] and [6], respectively), and obtain the same error control guarantee for the extended procedure. For arbitrary dependence within each study, a more conservative variant of this procedure is suggested. The remaining contributions of this paper are described in more detail below.

For common multiple testing procedures, the question "What are the dependencies among the p-values, for which the procedure guarantees the control of its target error rate?" was typically addressed. For example, according to the results of [8], the BH procedure guarantees FDR control when applied on p-values which satisfy the PRDS property on the subset of indices of true null hypotheses (its FDR control under independence was initially proved by [5]). However, when this type of dependence holds among the elementary p-values, it

is not clear whether it holds among the partial conjunction p-values, which may be complicated functions of the within-group elementary p-values. As a special case, consider global null p-values based on the method of Simes [50] (see Section 2.1 for the definition). Based on the results of [8], the BH procedure applied on Simes' p-values controls the FDR on the family of global null hypotheses if Simes' p-values satisfy the PRDS property, however it is not clear for what dependency structures among the elementary p-values this condition holds. In fact, in the invited lecture devoted to Prof. Yoav Benjamini's 70th birthday, which was held in Israel in 2018, Prof. Abba Krieger showed that when the elementary p-values satisfy the PRDS property, the global null Simes' p-values may not satisfy this property. He presented a counterexample, which we give in Appendix B. However, this does not yield that the BH procedure applied on Simes' global null p-values does not control the FDR in this case, because the PRDS property is a sufficient, not a necessary condition, for FDR control of the BH procedure. Interestingly, indeed, the results of [48] yield that the BH procedure applied on Simes' p-values guarantees FDR control on the family of global null hypotheses if the set of elementary p-values satisfies the PRDS property. Benjamini and Heller [3] showed a similar result for the BH procedure applied on partial conjunction p-values based on Fisher's or Stouffer's methods, assuming a certain dependency structure among the elementary p-values. In this work these results are generalized with respect to several aspects, which are described below.

Sufficient conditions on the dependency among the elementary p-values for FDR control on partial conjunction hypotheses are developed for a general class of *self-consistent* procedures, defined by [12], which includes the BH procedure, as well as its weighted variants (see [6], [24], and [12]). This class also includes several more conservative procedures, which may be of interest due to structural constraints on the rejection set, possibly induced by spatial structure or graph structure of the hypotheses (see [17] and [39]).

Besides the PRDS dependence among the p-values for the initial set of elementary hypotheses, more complex forms of dependencies are considered. Specifically, we consider a certain form of positive dependence across the groups and independence within each group, as well as positive dependence across the groups and arbitrary dependence within the groups. Independence and arbitrary dependence across the groups are also considered. Finally, we consider the dependencies which may hold in meta-analysis, where each feature is tested in several independent studies, addressing independence, positive dependence, and arbitrary dependence within each study. The results of this paper show that for most of the dependency structures above, the BH procedure, as well as its weighted variants, guarantees FDR control when applied on certain valid partial conjunction p-values. Therefore, despite the complex dependencies among the partial conjunction p-values in these situations, which may not satisfy the PRDS property, there is no need for conservative adjustments that were suggested for arbitrary dependence, such as the adjustment of [8], or its generalization given by [12].

The sufficient conditions that are obtained for FDR control depend on the method for obtaining partial conjunction p-values. We consider partial conjunc-

tion p-values which are connected to self-consistent multiple testing procedures, such as those based on Simes' and Bonferroni methods. In particular, we address the global null p-values connected to the adaptive self-consistent procedures, such as the adaptive BH procedure incorporating Storey's estimator for the proportion of nulls [52], and their extensions to general partial conjunction p-values, based on the method of [3]. For the meta-analysis setting, we also consider the popular Fisher's ([22]) and Stouffer's ([53]) methods for combining p-values for testing the partial conjunction null.

For obtaining the results above, several lemmas in the spirit of the group-level super-uniformity lemma of [48] are obtained. These lemmas are also used for generalizing the results of [4] regarding their procedure for replicability analysis, as described above. The lemmas may be useful for extending the theoretical results for other procedures, such as p-filter [48], DAGGER [47], and TreeBH [16]. Although the paper addresses FDR control on partial conjunction hypotheses, its results may also be used for obtaining FWER control on these hypotheses, under certain dependency structures among the elementary p-values.

This paper is organized as follows. Section 2 includes several global null tests which are addressed in this work, as well as their extensions for testing partial conjunction hypotheses. Sections 3 and 4 consider the setting where the hypotheses are divided into natural groups, and give several sufficient conditions for (weighted) FDR control on a family of partial conjunction hypotheses defined for these groups, for several dependency structures. In Section 5 we consider the meta-analysis setting, and give some results for partial conjunction testing in this setting, as well as a generalization of the procedure and the results of [4] addressing replicability analysis. Sections 4 and 5 contain the main results of this paper, addressing different dependency structures among the elementary p-values. Several helpful results and key lemmas are given in Section 6. The paper ends by a discussion in Section 7. Most of the proofs are given in Appendix A.

Remark 1.1. *Throughout this paper, vectors and matrices are denoted by bold letters, and p-values for composite hypotheses comprising combinations of elementary p-values are denoted by capital letters. In addition, we define* $0/0 = 0$, *and for two random variables X and Y such that if $Y = 0$, then $X = 0$ almost surely, we define $X/Y = 0$.*

2. Aggregate-level testing of a single hypothesis

2.1. Testing a global null hypothesis

Consider a family of m elementary null hypotheses, $H_1, \ldots, H_m$, with p-values $p_1, \ldots, p_m$. The global null hypothesis is defined as the intersection of these null hypotheses, stating that all these null hypotheses are true, $H_0 = \cap_{i=1}^m H_i$. We consider the cases where the p-value for the global null is obtained by combining the p-values for the individual hypotheses, $H_1, \ldots, H_m$, via a certain combining function $f : [0, 1]^m \to [0, 1]$. The combination of p-values $f(p_1, \ldots, p_m)$ is a valid p-value for the global null if for any $x \in (0, 1)$, $\mathbb{P}[f(p_1, \ldots, p_m) \leq x] \leq x$ under

the global null, i.e. when all the null hypotheses are true. Several methods for combining the p-values $p_1, \ldots, p_m$ for global null testing have been developed and studied, see, e.g., [40], [43], [57], [19], [59], and references therein. In this work the focus is on the following methods.

- Fisher's method [22]:

$$f_{\text{Fisher}}(p_1, \ldots, p_m) = \mathbb{P}\left[\chi^2_{(2m)} \geq -2\sum_{i=1}^{m} \log(p_i)\right], \qquad (2.1)$$

  where $\chi^2_{(2m)}$ is a chi-square variable with $2m$ degrees of freedom.
- Stouffer's method [53]:

$$f_{\text{Stouffer}}(p_1, \ldots, p_m) = 1 - \Phi\left(\frac{\sum_{i=1}^{m} \Phi^{-1}(1 - p_i)}{\sqrt{m}}\right), \qquad (2.2)$$

  with $\Phi(\cdot)$ denoting the cumulative distribution function of a standard normal variable, and $\Phi^{-1}(\cdot)$ denoting its inverse function.
- Simes' method [50]:

$$f_{\text{Simes}}(p_1, \ldots, p_m) = \min_{k=1,\ldots,m} \frac{m p_{(k)}}{k}, \qquad (2.3)$$

  where $p_{(1)} \leq \ldots \leq p_{(m)}$ is the ordered sequence of p-values.

All the methods above give valid global null p-values under independence of individual p-values. Simes' method is known to be also valid under certain forms of positive dependence, see [49], [8], [21]. For general dependence, the following more conservative methods are considered:

- Bonferroni's method:

$$f_{\text{Bonferroni}}(p_1, \ldots, p_m) = \min\{m p_{(1)}, \ 1\} \qquad (2.4)$$

- Hommel's method [36]:

$$f_{\text{Hommel}}(p_1, \ldots, p_m) = \min\left\{\left(1 + \frac{1}{2} + \ldots + \frac{1}{m}\right) \min_{k=1,\ldots,m} \frac{m p_{(k)}}{k}, \ 1\right\} \qquad (2.5)$$

Some of the methods for global null testing are connected to certain multiple testing procedures. For example, if one uses Bonferroni's method (2.4) for testing the global null, then the global null is rejected at level α if and only if Bonferroni's procedure, applied at level α on the p-values $p_1, \ldots, p_m$, rejects at least one hypothesis. There is a similar connection between Simes' p-value (2.3) and the BH procedure, as well as Hommel's p-value (2.5) and the procedure of [8] (BY), which is a conservative variant of the BH procedure with FDR control guarantees under arbitrary dependence. In all these cases, the global null p-value is the minimal adjusted p-value of a certain multiple testing procedure. Recall that for a given multiple testing procedure, an adjusted p-value

for H_i, denoted by P_i^{adj}, is the minimum level α of the procedure such that H_i is rejected at level α ([60]). The set of rejected hypotheses by a given level-α procedure is $\{i : P_i^{adj} \leq \alpha\}$, therefore at least one rejection is made if and only if the minimum adjusted p-value, $P_{(1)}^{adj}$, is bounded above by α. Under the global null, each rejected hypothesis corresponds to a false discovery, therefore in this case $\mathbb{P}\left[P_{(1)}^{adj} \leq \alpha\right] = \mathbb{P}\left[(V > 0)\right]$, where V is the number of false discoveries of the given level-α multiple testing procedure. In addition, under the global null, $\mathbb{P}\left[(V > 0)\right] = \text{FWER} = \text{FDR}$, which gives the following corollary.

Corollary 2.1. *Assume we are given a multiple testing procedure which guarantees control of FWER (or FDR) when all the null hypotheses are true (i.e. in the weak sense) under a certain dependency structure among the p-values. Then, under the same dependency structure, the minimum adjusted p-value of this procedure is a valid p-value for testing the global null.*

Remark 2.1. *A natural requirement for a function $f(p_1, \ldots, p_m)$ which gives a global null p-value is that f is non-decreasing in each coordinate. In fact, Birnbaum [10] showed that when the p-values are independent, for each combination function $f(p_1, \ldots, p_m)$ satisfying this requirement one can find an alternative hypothesis against which the combination function $f(p_1, \ldots, p_m)$ gives a best test of the global null. This requirement is satisfied when f is based on a certain multiple testing procedure in the sense of Corollary 2.1, assuming that the multiple testing procedure satisfies the monotonicity property defined below.*

Definition 2.1. *A multiple testing procedure is monotone if it satisfies the following condition: given a set of rejected hypotheses $\mathcal{R}_1$, decreasing a certain p-value will result in rejecting the set $\mathcal{R}_2$, where $\mathcal{R}_1 \subseteq \mathcal{R}_2$.*

Based on Corollary 2.1, one can obtain validity of global null tests based on the results for the multiple testing procedures which induce these tests. The fact that Bonferroni's and Hommel's global null p-values are valid under any dependence of elementary p-values may be obtained from the facts that Bonferroni and BY control the FWER in the weak sense under arbitrary dependence. Similarly, one can obtain the validity of Simes' method and its weighted versions (see [6] and [35]) under PRDS dependence based on the result of [12], showing that the doubly-weighted BH procedure guarantees FDR control under this type of positive dependence. Similarly, the minimum adjusted p-value of an adaptive variant of the BH procedure, incorporating an estimator for the proportion of true null hypotheses, denoted by π_0, is a valid global null p-value if the dependency among the p-values is such that the procedure guarantees FDR control. See [7] and [13] for several adaptive procedures and their performance comparisons. Let us consider the minimum adjusted p-value of the adaptive BH procedure incorporating a modification of Storey's estimator ([51]) for the proportion of nulls, suggested by [52]. The estimator is based on a tuning parameter $\lambda \in (0, 1)$, and is given by

$$\hat{\pi}_0(\lambda) = \frac{W(\lambda) + 1}{(1 - \lambda)m}, \tag{2.6}$$

with $W(\lambda) = \sum_{i=1}^{m} \mathbb{1}(p_i > \lambda)$, where $\mathbb{1}(\cdot)$ is the indicator function. The minimum adjusted p-value of this procedure is

$$f_{Simes-Storey}(p_1, \ldots, p_m) = \begin{cases} \min\left\{\min_{k:p_{(k)}\leq\lambda} \left\{m\hat{\pi}_0(\lambda)p_{(k)}/k\right\}, \ 1\right\} & p_{(1)} \leq \lambda \\ 1, & p_{(1)} > \lambda \end{cases}$$

$$(2.7)$$

It was proved by [52] that the adaptive BH procedure with Storey's estimator for π_0 controls the FDR under independence, therefore the test based on its minimum adjusted p-value, $f_{Simes-Storey}(p_1, \ldots, p_m)$, is valid under independence. In fact, we address a more general class of global null tests, induced by a certain class of multiple testing procedures, including certain adaptive FDR-controlling procedures. According to Remark 2.1, the tests induced by monotone multiple testing procedures satisfy the desirable monotonicity property.

2.2. Testing a partial conjunction hypothesis

Consider a family of m elementary null hypotheses $H_1, \ldots, H_m$, and let m_1 be the number of false null hypotheses in this family. Benjamini and Heller [3] considered testing the partial conjunction null hypothesis $H_0^{u/m} : m_1 < u$ versus the alternative $H_1^{u/m} : m_1 \geq u$, for a fixed $u \in \{1, \ldots, m\}$, addressing the question whether the family contains at least u false null hypotheses. For $u = 1$, the partial conjunction null reduces to the global null, and the tests developed in [3] reduce to familiar tests, while for $u = n$, their tests lead to rejecting the null if the maximum p-value is below the given significance level α. In general, the p-value for testing $H_0^{u/m}$, denoted by $P^{u/m}$, is valid if under this partial conjunction null, i.e. in the case where $m_1 < u$, $\mathbb{P}\left[P^{u/m} \leq x\right] \leq x$ for all $x \in [0, 1]$. Benjamini and Heller [3] developed the following method for constructing valid partial conjunction p-values $P^{u/m}$: given a certain combining function f for the global null hypothesis, which is valid for the dependency among the p-values in any subgroup $A \subseteq \{1, \ldots, m\}$ of size $m - u + 1$ and which is non-decreasing in each of its components, obtain a pooled p-value $P^{u/m}$ by applying the function f on the $m - u + 1$ largest p-values. Since the combining function f is non-decreasing in each of its components, this method is equivalent to computing the global null combined p-value for each subgroup $A \subseteq \{1, \ldots, m\}$ of size $m - u + 1$ and taking the maximum of all the combined p-values. Denoting by $\boldsymbol{p}_A$ the set of p-values for the hypotheses in group A, $P^{u/m}$ can be written as follows:

$$P^{u/m} = \max\{f(\boldsymbol{p}_A) : A \subseteq \{1, \ldots, m\}, |A| = m - u + 1\}.$$

See [3] for partial conjunction p-values based on Fisher's, Stouffer's, Simes', and Bonferroni's methods. Let us obtain the partial conjunction p-value for testing $H_0^{u/m}$, based on the combining function (2.7), which is non-decreasing

in each coordinate. For obtaining $P^{u/m}$, we apply the combining function (2.7) on the largest $m - u + 1$ p-values. Let

$$\hat{\pi}_0^{u/m}(\lambda) = \frac{1 + \sum_{i=u}^m \mathbb{1}(p_{(i)} > \lambda)}{(m - u + 1)(1 - \lambda)},$$

and define $S(\lambda) = \{k \in \{1, \ldots, m - u + 1\} : p_{(u-1+k)} \leq \lambda\}$. Then

$$P_{Simes-Storey}^{u/m} = \begin{cases} \min\left\{ \min_{k \in S(\lambda)} \left\{ (m - u + 1)\hat{\pi}_0^{u/m}(\lambda)p_{(u-1+k)}/k \right\}, \ 1 \right\} & p_{(u)} \leq \lambda \\ 1, & p_{(u)} > \lambda \end{cases}$$

$$(2.8)$$

Since the combining function (2.7) gives a valid global null p-value under independence among $p_1, \ldots, p_m$, the partial conjunction p-value given above is valid for testing $H_0^{u/m}$ under independence.

3. Testing a family of partial conjunction hypotheses: preliminaries and general results

3.1. The setting

Assume one has M elementary null hypotheses, $H_1, \ldots, H_M$, with p-values $p_1, \ldots, p_M$, respectively. The set of hypotheses is divided into G disjoint groups. Let $A_g \subseteq \{1, \ldots, M\}$ be the set of indices of hypotheses belonging to group g, and let $n_g = |A_g|$ be the number of hypotheses in group g. Denote by k_g the (unknown) number of false null hypotheses in group g.

The researcher is interested in testing the family of G partial conjunction hypotheses for the given groups, $\{H_0^{u_g/n_g}, \ g = 1, \ldots, G\}$, based on partial conjunction p-values $\{P_g^{u_g/n_g}, \ g = 1, \ldots, G\}$, where for each $g \in \{1, \ldots, G\}$, $P_g^{u_g/n_g}$ is a combination of p-values for the hypotheses in group g. When $u_g = 1$ for $g = 1, \ldots, G$, this is the family of global null hypotheses. When the group sizes are equal, i.e. $n_g = n$ for $g = 1, \ldots, G$, it seems natural to test partial conjunction hypotheses with a common value $u_g = u$ for each group. However, when the groups are of different sizes and the researcher wishes to identify groups in which at least a proportion $p \in (0, 1]$ of the null hypotheses are false, the values of u_g may differ across groups. Let $\mathcal{G}_0 \subseteq \{1, \ldots, G\}$ be the set of indices for which the partial conjunction null hypotheses are true, i.e. $\mathcal{G}_0 = \{g \in \{1, \ldots, G\} : k_g < u_g\}$.

We consider the case where a certain multiple testing procedure receives as input the partial conjunction p-values $\{P_g^{u_g/n_g}, \ g = 1, \ldots, G\}$, and outputs the set indices of rejected partial conjunction hypotheses, $\mathcal{R} = \mathcal{R}(p_1, \ldots, p_M) \subseteq \{1, \ldots, G\}$. The FWER for the family of partial conjunction hypotheses is

$$\mathbb{P}\left[|\mathcal{R} \cap \mathcal{G}_0| > 0\right],$$

while the weighted FDR, introduced in [6], for the family of partial conjunction hypotheses is

$$\text{FDR}_v^{\text{PC}} := \mathbb{E} \left[\frac{\sum_{g \in \mathcal{G}_0} v_g I(g \in \mathcal{R})}{\sum_{g=1}^{G} v_g I(g \in \mathcal{R})} \right], \tag{3.1}$$

where the penalty weights $(v_1, \ldots, v_G)$ reflect the penalties for false rejections of partial conjunction hypotheses, and the rewards for true rejections. The weighted FDR reduces to the regular FDR when all the penalty weights are equal to unity. Benjamini and Hochberg [6] developed a variant of the BH procedure controlling the weighted FDR. The partial conjunction hypotheses may also be associated with prior weights $(w_1, \ldots, w_G)$, reflecting prior beliefs regarding their truth status, which may be used for increasing the power of a multiple testing procedure, as suggested by ([24]). Unlike the penalty weights, prior weights do not play any role in the definition of the weighted FDR. Following ([12]), we consider the general case where the partial conjunction hypotheses are associated with both types of weights.

3.2. *Preliminary definitions*

Blanchard and Roquain [12] developed two sufficient conditions for FDR control of a multiple testing procedure. In this section we review the definitions of [12] and [13] which shall be used for obtaining sufficient conditions for FDR control on partial conjunction hypotheses. Consider a family of m hypotheses $H_1, \ldots, H_m$, with p-values $\boldsymbol{p} = (p_1, \ldots, p_m)$, which are valid, in the sense that for any true null hypothesis H_i, $\mathbb{P}\left[p_i \leq x\right] \leq x$ for all $x \in [0, 1]$. Assume that these hypotheses are associated with a vector of prior weights $\boldsymbol{w} = (w_1, \ldots, w_m)$, and a vector of penalty weights $\boldsymbol{v} = (v_1, \ldots, v_m)$. Recall that given a rejection set $\mathcal{R}$ and the set of indices of true null hypotheses $M_0 \subseteq \{1, \ldots, m\}$, the weighted FDR with penalty weights given by $\boldsymbol{v}$ is defined as

$$\text{FDR}_v = \mathbb{E} \left[\frac{\sum_{i \in M_0} v_i I(i \in \mathcal{R})}{\sum_{i=1}^{m} v_i I(i \in \mathcal{R})} \right].$$

Define $|A|_{\boldsymbol{v}} \equiv \sum_{i=1}^{m} v_i I(i \in \mathcal{A})$ as the volume of the set A, for any $A \subseteq \{1, \ldots, m\}$. Then the weighted FDR can be written as

$$\mathbb{E} \left[\frac{|M_0 \cap \mathcal{R}|_{\boldsymbol{v}}}{|\mathcal{R}|_{\boldsymbol{v}}} \right].$$

Note that for $\boldsymbol{v} = (1, \ldots, 1)$, $|\cdot|_{\boldsymbol{v}}$ is simply the cardinality measure, and the weighted FDR reduces to the regular FDR. A threshold collection of a multiple testing procedure used for testing the family $\{H_1, \ldots, H_m\}$ was defined by Blanchard and Roquain [12] as follows.

Definition 3.1 (Blanchard and Roquain, [12])**.** *A threshold collection Δ is a function*

$$\Delta : (i, r) \in \{1, \ldots, m\} \times \mathbb{R}^+ \mapsto \Delta(i, r) \in \mathbb{R}^+,$$

which is non-decreasing in its second variable.

Following [12] we consider the factorized threshold collection of the following form: $\forall (i, r) \in \{1, \ldots, m\} \times \mathbb{R}^+$,

$$\Delta(i, r) = \alpha w_i \beta(r)/m, \tag{3.2}$$

where w_i is a prior weight for H_i, $i = 1, \ldots, m$, and $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ is a non-decreasing function called a shape function. Following [13], we also consider an adaptive threshold collection of the following form: $\forall (i, r) \in \{1, \ldots, m\} \times \{1, \ldots, m\}$,

$$\Delta(i, r) = \Delta(r) = \alpha r \hat{\pi}_0^{-1}/m, \tag{3.3}$$

where $\hat{\pi}_0$ is the estimator for π_0, the proportion of nulls. This threshold collection does not incorporate prior or penalty weights. In all the definitions below, we set $\boldsymbol{v} = (1, \ldots, 1)$ for multiple testing procedures with thresholds of form (3.3). Given a threshold collection Δ, define $L_\Delta(r)$ as in [12], i.e.

$$L_\Delta(r) = \{i \in \{1, \ldots, m\} : p_i \leq \Delta(i, r)\}.$$

We are now ready to review several properties of multiple testing procedures, which shall be considered in the following sections.

Definition 3.2. *Let $\mathcal{R} = \mathcal{R}(\boldsymbol{p}) \subseteq \{1, \ldots, m\}$ be the set of indices of rejected hypotheses by a given multiple testing procedure. For each $i \in \{1, \ldots, m\}$, let $\boldsymbol{p}^{-i}$ be the vector of p-values excluding p_i.*

1. *A multiple testing procedure is **self-consistent** with respect to threshold collection Δ of form (3.2) or (3.3) if the following inclusion holds almost surely:*

$$\mathcal{R} \subseteq L_\Delta(|\mathcal{R}|_v). \tag{3.4}$$

2. *The multiple testing procedure is **non-increasing** if $|\mathcal{R}(\boldsymbol{p})|_v$ is non-increasing in each p-value.*
3. *The multiple testing procedure is **stable** if for each $i \in \mathcal{R}$, fixing $\boldsymbol{p}^{-i}$ and changing p_i as long as $i \in \mathcal{R}$ will not change the set $\mathcal{R}$. For such procedure, we denote by $\mathcal{R}^{-i} = \mathcal{R}^{-i}(\boldsymbol{p}^{-i})$ the rejection set which is obtained when the p-values for all the hypotheses except H_i are given by $\boldsymbol{p}^{-i}$, and p_i has any value such that H_i is rejected.*
4. *A stable multiple testing procedure is **concordant** if for each $i \in \{1, \ldots, m\}$, $|\mathcal{R}^{-i}|_v$ is non-increasing in each coordinate of $\boldsymbol{p}^{-i}$.*

The definitions in items 1 and 2 were given in [12] and [13]. The property in item 3 was defined in [14]. Properties similar to those in items 3 and 4 were defined in [1] and [9]. Blanchard and Roquain [12] showed that step-up and step-down procedures, or more generally, step-up-down procedures ([54]) are self-consistent. The BH procedure and its weighted variants, suggested by

[6] and [24], are special cases of the doubly-weighted BH procedure considered by [12], which is self-consistent with respect to thresholds of form (3.2). More conservative procedures satisfying self-consistency property with respect to the same collection of thresholds were developed for addressing certain constraints, such as structural constraints on the rejection set, which may be natural in some applications. Examples are the procedure in [17] which was suggested for controlling the FDR in genome-wide association studies, and its generalization, Focused BH, suggested in [39] for FDR control on the filtered set of discoveries. The monotonicity defined in item 2 is a very natural property, satisfied by many commonly used multiple testing procedures. Stability and concordance are satisfied for any step-up procedure with thresholds of form (3.2), including the BH procedure and its weighted variants (see Appendix A.3 for a proof).

The adaptive BH procedure incorporating Storey's estimator (2.6) for the proportion of nulls is non-increasing and self-consistent with respect to thresholds of form (3.3). For a general adaptive procedure with thresholds of form (3.3), let us consider the following conditions on the estimator for the proportion of nulls $\hat{\pi}_0$, following [13]:

Condition 3.1. *(a) The estimator* $\hat{\pi}_0 : [0,1]^m \to (0,\infty)$ *is a measurable, coordinate-wise non-decreasing function. (b) If the p-values* $(p_1,\ldots,p_m)$ *are independent, then for any true null hypothesis* H_i, $\mathbb{E}\left[1/\hat{\pi}_0(\boldsymbol{p}_{0,i})\right] \leq 1/\pi_0$, *where* $\boldsymbol{p}_{0,i}$ *is the vector of p-values with* p_i *replaced by 0.*

The condition that the estimator $\hat{\pi}_0$ is non-decreasing in each *p*-value is very natural, and it is satisfied by Storey's estimator (2.6). Benjamini, Krieger, and Yekutieli [7] proved that this estimator satisfies part (b) of Condition 3.1 as well. See Corollary 13 of [13] for other estimators satisfying Condition 3.1. In the next sections we rely on the definitions above for formulating sufficient conditions for FDR control on partial conjunction hypotheses.

3.3. Results for general valid partial conjunction p-values

We assume hereafter that for each $g = 1,\ldots,G$, the partial conjunction *p*-value $P_g^{u_g/n_g}$ is a combination of the *p*-values for the hypotheses belonging to A_g. Then the dependence among the partial conjunction *p*-values is induced by the combining method and the dependence among the elementary *p*-values. The following proposition shows several sufficient conditions for FDR control on partial conjunction hypotheses for general valid partial conjunction *p*-values. It addresses the following assumptions on the dependency structure:

A1 The *p*-values in each group are independent of the *p*-values in any other group.

A2 The set of partial conjunction *p*-values $\{P_g^{u_g/n_g}, g = 1,\ldots,G\}$ satisfies the PRDS property on the subset $\mathcal{G}_0$ (i.e. the subset of indices of true partial conjunction null hypotheses).

Proposition 3.1. *Let $\{H_0^{u_g/n_g}, g = 1, \ldots, G\}$ be a family of partial conjunction hypotheses associated with penalty weights $\{v_g, g = 1, \ldots, G\}$ and prior weights $\{w_g, g = 1, \ldots, G\}$ satisfying $\sum_{g \in \mathcal{G}} v_g w_g = G$. Let $\{P_g^{u_g/n_g}, g = 1, \ldots, G\}$ be the set of corresponding partial conjunction p-values, and assume that the dependency among the elementary p-values within each group is such that these partial conjunction p-values are valid for testing the family $\{H_0^{u_g/n_g}, g = 1, \ldots, G\}$. Consider the following assumptions:*

1. *Assumption A1 is satisfied, and the multiple testing procedure satisfies either of the following conditions:*

    1.1 The procedure is non-increasing and is self-consistent with respect to thresholds of form (3.2) with identity shape function $\beta(x) = x$.

    1.2 The procedure is non-increasing and is self-consistent with respect to thresholds of form (3.3), where $\hat{\pi}_0$ satisfies Condition 3.1.

2. *Assumption A2 is satisfied, and the multiple testing procedure is non-increasing and is self-consistent with respect to thresholds of form (3.2) with identity shape function $\beta(x) = x$.*

3. *The p-values across groups are arbitrarily dependent, and the multiple testing procedure is self-consistent with respect to thresholds of form (3.2) with the shape function of the form*

$$\beta_\nu(r) = \int_0^r x d\nu(x), \tag{3.5}$$

   where ν is an arbitrary probability distribution on $(0, \infty)$.

If the assumptions of either of the three items above are satisfied, then the multiple testing procedure guarantees $FDR_v^{PC} \leq \alpha$.

Remark 3.1. *For item 1 along with 1.1, as well as for items 2 and 3, one has a tighter bound for FDR_v^{PC}. Specifically, one has*

$$FDR_v^{PC} \leq \frac{\alpha}{G} \sum_{g \in \mathcal{G}_0} v_g w_g \leq \alpha.$$

This proposition follows immediately from the results of [12] and [13]. Since the partial conjunction *p*-values are combinations of within-group elementary *p*-values, under the assumptions of item 1, the partial conjunction *p*-values are independent. Therefore, the result of item 1 along with item 1.1 follows from Propositions 2.7 and 3.3 of [12]. The result of item 1 along with item 1.2 follows from Theorem 11 of [13]. In the case of item 2, satisfy the PRDS property, while in the case of item 3, the partial conjunction *p*-values may be arbitrarily dependent. Therefore, items 2 and 3 follow from Propositions 3.6 and 3.7 in [12] respectively, in conjunction with Proposition 2.7 in that paper.

As mentioned in Section 3.2, many multiple testing procedures with thresholds of form (3.2) are non-increasing and self-consistent, including the BH procedure and its generalizations incorporating weights, unified by the doubly-weighted BH procedure in [12]. In addition, several adaptive variants of the BH

procedure satisfy the conditions of item 1.2, including the adaptive BH procedure with Storey's estimator (2.6). Therefore, under independence across groups, all these procedures guarantee FDR control on partial conjunction hypotheses, if valid partial conjunction p-values are given as input to the procedure. However, as discussed in the introduction, when the p-values belonging to different groups are dependent, it may be not clear whether the partial conjunction p-values satisfy the PRDS property on the subset of true partial conjunction nulls, as required in item 2. In these cases, one may revert to replacing the identity shape function $\beta(r) = r$ by the shape function of form (3.5), which gives FDR control guarantees under arbitrary dependence among the partial conjunction p-values, according to item 3. However, this replacement results in lower thresholds leading to loss of power, because $\beta_\nu(r) \leq r$. For example, the adjustment of the BH procedure for addressing arbitrary dependence, suggested by [8], replaces the identity shape function by $\beta_\nu(r) = r/(\sum_{j=1}^m 1/j)$. In Section 4 it is shown that for several dependency structures among the elementary p-values, this conservative adjustment is not needed if certain methods for constructing partial conjunction p-values are used.

4. Results for partial conjunction p-values connected to multiple testing procedures

This section addresses the setting described in Section 3.1. According to Corollary 2.1 and Remark 2.1, the minimum adjusted p-value of a multiple testing procedure is a valid global null p-value, which is non-decreasing in each elementary p-value, provided that the multiple testing procedure guarantees FDR control in the weak sense and is monotone according to Definition 2.1. In Section 2.2, we showed the method of [3] for obtaining a partial conjunction p-value based on a certain coordinate-wise non-decreasing combining function for testing the global null. Taking these two results together, we may obtain partial conjunction p-values connected to multiple testing procedures. Examples are partial conjunction p-values based Simes', Hommel's, Bonferroni methods, which are connected (in the above sense) to BH, BY, and Bonferroni procedures, respectively. Similarly, the partial conjunction p-value in (2.8) is connected to the adaptive BH method with Storey's estimator (2.6) for the proportion of nulls.

We consider the case where for each $g \in \{1, \ldots, G\}$, the partial conjunction p-value $P_g^{u_g/n_g}$ is connected to a certain multiple testing procedure $\mathcal{M}_g$ which is monotone according to Definition 2.1. Formally,

$$P_g^{u_g/n_g} = \max\{P^{1/(n_g-u_g+1)}[A] : A \subseteq A_g, |A| = n_g - u_g + 1\}, \qquad (4.1)$$

where $P^{1/(n_g-u_g+1)}[A]$ is the minimum adjusted p-value according to $\mathcal{M}_g$ applied on the p-values with indices in A. The multiple testing procedures $\mathcal{M}_g$ may be different for different groups g, for example for certain groups one may use the partial conjunction p-value based on Simes' method, while for the other

groups one may revert to Bonferroni's method. The next theorem addresses such partial conjunction p-values.

Let $\boldsymbol{p}^{-g}$ be the vector of p-values for the hypotheses not belonging to group g, and for each $i \in A_g$, let $\boldsymbol{p}_g^{-i}$ be the vector of p-values for the hypotheses in group g, excluding H_i. We consider the following dependency structures among the elementary p-values:

D1 Independence within each group: for each $g \in \{1, \ldots, G\}$, the p-values in the set $\{p_i, \; i \in A_g\}$ are independent.

D2 Overall positive dependence: the set of all the p-values, $p_1, \ldots, p_M$, is PRDS on the subset of true null hypotheses.

D3 Conditional positive dependence across groups: for each $g \in \{1, \ldots, G\}$, and each true null hypothesis H_i in group g, the vector $\boldsymbol{p}$ is PRDS with respect to p_i, conditionally on $\boldsymbol{p}_g^{-i}$, i.e. $\mathbb{P}\left[\boldsymbol{p} \in D \,|\, p_i = x, \boldsymbol{p}_g^{-i} = \boldsymbol{y}\right]$ is non-decreasing in x, for any non-decreasing set $D \in [0, 1]^M$, and any vector $\boldsymbol{y} \in [0, 1]^{n_g - 1}$.

D4 Positive dependence across groups: for each $g \in \{1, \ldots, G\}$, and each true null hypothesis H_i in group g, the vector $(\boldsymbol{p}^{-g}, p_i)$, consisting of p_i along with p-values for the hypotheses not belonging to group g, is PRDS with respect to p_i, i.e. $\mathbb{P}\left[(\boldsymbol{p}^{-g}, p_i) \in D \,|\, p_i = x\right]$ is non-decreasing in x, for any non-decreasing set $D \in [0, 1]^{M - n_g + 1}$.

In addition, we consider the following assumptions on multiple testing procedures:

T1 A multiple testing procedure is non-increasing.

T2 A multiple testing procedure is self-consistent with respect to thresholds of form (3.2) with prior and penalty weights equal to unity and identity shape function $\beta(x) = x$.

T3 A multiple testing procedure is stable and concordant.

T4 A multiple testing procedure is self-consistent with respect to thresholds of form (3.3), where $\hat{\pi}_0$ satisfies Condition 3.1.

Theorem 4.1. *Let $\mathcal{M}$ be a self-consistent multiple testing procedure with respect to thresholds of form (3.2) with identity shape function $\beta(x) = x$. Consider a family of partial conjunction hypotheses $\{H_0^{u_g/n_g}, \; g = 1, \ldots, G\}$, associated with prior weights $\{w_g, \; g = 1, \ldots, G\}$ and penalty weights $\{v_g, \; g = 1, \ldots, G\}$, satisfying $\sum_{g=1}^{G} w_g v_g = G$. Let $\mathcal{M}_g, g = 1, \ldots, G$ be multiple testing procedures satisfying the monotonicity property given in Definition 2.1. Assume that for each $g \in \{1, \ldots, G\}$, the partial conjunction p-value $P_g^{u_g/n_g}$ is connected to $\mathcal{M}_g$ in the sense of (4.1). In each of the following three cases, requiring additional assumptions on the procedure $\mathcal{M}$, the procedures $\mathcal{M}_g, g = 1, \ldots, G$, and the dependency among the elementary p-values, the procedure $\mathcal{M}$ applied on $\{P_g^{u_g/n_g}, \; g = 1, \ldots, G\}$ at level α guarantees*

$$FDR_v^{PC} \leq \frac{\alpha}{G} \sum_{g \in \mathcal{G}_0} v_g w_g \leq \alpha.$$

1. (a) *The p-values $(p_1, \ldots, p_M)$ satisfy the overall positive dependence property, as defined in item D2.*
   (b) *The multiple testing procedure $\mathcal{M}$ satisfies assumption T1.*
   (c) *For each $g \in \{1, \ldots, G\}$, the multiple testing procedure $\mathcal{M}_g$ satisfies assumption T2.*
2. (a) *The p-values $(p_1, \ldots, p_M)$ satisfy conditional positive dependence across groups, and independence within each group, as defined in items D3 and D1.*
   (b) *The multiple testing procedure $\mathcal{M}$ satisfies assumption T3.*
   (c) *For each $g \in \{1, \ldots, G\}$, the multiple testing procedure $\mathcal{M}_g$ satisfies assumption T4.*
3. (a) *The p-values $(p_1, \ldots, p_M)$ satisfy positive dependence across groups, as defined in item D4, and the dependence among the p-values within each group is unspecified.*
   (b) *The multiple testing procedure $\mathcal{M}$ satisfies assumption T3.*
   (c) *The partial conjunction p-values are connected to the Bonferroni procedure, i.e. for each $g \in \{1, \ldots, G\}$, $P_g^{u_g/n_g} = (n_g - u_g + 1)p_{g(u_g)}$, where $p_{g(1)} \leq \ldots \leq p_{g(n_g)}$ is the sequence of ordered p-values for the hypotheses in group g.*

The results of Theorem 4.1 rely on the sufficient conditions for FDR control obtained by [12], as well as an additional pair of such sufficient conditions, which we obtain using the techniques of [12] (see Section 6.1). Each pair of conditions addresses the multiple testing procedure and the dependency structure. The condition addressing the dependency structure follows from Lemma 6.1, which generalizes the group-level super-uniformity lemma of [48]. The cases of independence across groups and arbitrary dependence across groups are not considered in this theorem, since they are covered by Proposition 3.1.

The doubly-weighted BH procedure considered by ([12]) is self-consistent with respect to thresholds of form (3.2), is non-increasing, stable, and concordant. Therefore, for this procedure we obtain the following corollary:

Corollary 4.1. *The doubly-weighted BH procedure at level α applied on the partial conjunction p-values $\{P_g^{u_g/n_g}, g = 1, \ldots, G\}$ guarantees $FDR_v^{PC} \leq \alpha$ in each of the following three cases:*

1 *The p-values are positively dependent in the sense of item D2, and the partial conjunction p-values are based on Simes' combining method (2.3).*
2 *The p-values are independent within each group and are conditionally positively dependent across the groups, in the sense of items D1 and D3, and the partial conjunction p-values are computed using Simes-Storey function given in (2.8), i.e. they are connected in the sense of (4.1) to the adaptive BH procedure with Storey's estimator for the proportion of nulls.*
3 *The p-values are positively dependent across the groups, in the sense of item D4, are arbitrarily dependent within the groups, and the partial conjunction p-values are based on Bonferroni's method.*

The result of item 1 with respect to the global null hypotheses (i.e. for the case where $u_g = 1$ for $g = 1, \ldots, G$) follows from the results of [48]. Item 1 generalizes this result, addressing in addition partial conjunction hypotheses with $u_g > 1$. It is easy to see that the dependency structure in item 2 satisfies the overall positive dependence requirement of item 1, so under the condition of item 2, the doubly-weighted BH procedure is valid when applied on Simes' partial conjunction p-values. However, the more strict requirement of item 2 allows to replace these p-values by Simes-Storey partial conjunction p-values, which are more liberal for the groups which are enriched with signal. The result of item 2 is useful for the meta-analysis setting, see Section 5. Finally, item 3 shows that when relaxing the independence within groups requirement in item 2, and requiring unconditional positive dependence across groups, the error control is still guaranteed when the doubly-weighted BH procedure is applied on the Bonferroni partial conjunction p-values. Theorem 4.1 shows that when the doubly-weighted BH procedure is replaced by a more conservative self-consistent procedure with respect to thresholds of form (3.2), possibly due to the structural constraints on the set of rejected partial conjunction hypotheses (see Section 3.2 for examples), we have the same results as in Corollary 4.1 under lenient assumptions on the procedure. Under the assumptions of item 2, one can also replace Storey's estimator by another estimator satisfying Condition 3.1, see Section 3.2.

It may be of interest to consider the results of Theorem 4.1 for two extreme cases: $G = M$ and $G = 1$. In the first case, each group consists of one hypothesis, so $n_g = u_g = 1$, and FDR_v^{PC} reduces to the weighted FDR for the elementary hypotheses. The partial conjunction p-value based on Simes' method satisfies the assumption of part (c) in item 1, and in this extreme case, for each group g this partial conjunction p-value reduces to the elementary p-value of the single hypothesis in group g. Both types of positive dependence across groups reduce to overall positive dependence in this case, therefore the result of Theorem 4.1 under the assumptions of item 1 follows from combining Proposition 3.6 and Proposition 2.7 of [12]. In the second case, where $G = 1$, the entire set of p-values constitutes one group. Consider Theorem 4.1 with the BH procedure taking the role of $\mathcal{M}$. In this extreme case, the BH procedure reduces to rejecting the single partial conjunction null when the given partial conjunction p-value is upper bounded by α. Therefore, FDR_v^{PC} is simply the significance level of the partial conjunction test. Thus, Theorem 4.1 shows that in each of the cases 1, 2, and 3, a partial conjunction p-value satisfying assumption (c) is valid under the dependence assumption (a) within the single group. Particularly, the partial conjunction p-values based on Simes', Simes-Storey, and Bonferroni methods, are valid under PRDS dependence, independence, and arbitrary dependence, respectively. These are also known results, see Section 2.

Finally, we would like to note that Corollary 4.1 may be used for obtaining several results on Simes' p-value, which is the minimum adjusted p-value of the BH procedure. Obviously, each of the dependency structures considered in the corollary holds when we restrict the problem to any subset of groups, i.e. when

we consider testing $\{H_0^{u_g/n_g}, \ g \in \mathcal{G}_s\}$, where $\mathcal{G}_s \subseteq \{1, \ldots, G\}$. Therefore, the BH procedure controls the FDR in each of the cases considered not only for the entire family of partial conjunction hypotheses, but also for any subset of those hypotheses. Using this fact along with Corollary 2.1, we obtain the following result.

Corollary 4.2. *Let $\mathcal{G}_s \subseteq \{1, \ldots, G\}$ be a subset of group indices, and consider applying Simes' combining function (2.3) on the partial conjunction p-values in the set $\{P_g^{u_g/n_g}, \ g \in \mathcal{G}_s\}$. The combined p-value is valid for testing the intersection of partial conjunction hypotheses $\cap_{g \in \mathcal{G}_s} H_0^{u_g/n_g}$ in each of the three cases considered in Corollary 4.1.*

Obviously, the same result holds for the weighted variant of Simes' *p*-value, which is the minimum adjusted *p*-value of the doubly-weighted BH procedure. The fact that Simes' *p*-value is valid when it receives as input Simes' global null *p*-values and all the elementary *p*-values are PRDS has been proven by [48] (see item (d) of their group-level superuniformity lemma). Corollary 4.2 generalizes this result, addressing general Simes' partial conjunction *p*-values, as well as other dependency structures and combining methods. Consider a procedure based on the closed testing principle of [41], which considers the family $\{H_0^{u_g/n_g}, g = 1, \ldots, G\}$ as the family of elementary hypotheses, and tests the intersections of these hypotheses using Simes' test applied on the corresponding partial conjunction *p*-values. Based on Corollary 4.2, we obtain that in each of the three cases of Corollary 4.1, all these Simes' tests are valid, therefore this procedure guarantees FWER control on the partial conjunction hypotheses, as well as on the entire set of all their intersections. In fact, it reduces to Hommel's procedure ([37]) applied on partial conjunction *p*-values. We conclude that Hommel's procedure, as well as its shortcuts (see, e.g., [34], [42], [28], [27]) are valid for FWER control on the partial conjunction hypotheses in cases 1–3 of Corollary 4.1. Moreover, Goeman and Solari ([28]) showed that closed testing with valid local tests can be used for controlling the false discovery proportions (FDP) over all subsets, and suggested using this approach for more flexible multiple testing. In our setting, this approach gives the flexibility to select a subset of groups post-hoc, and to estimate the number of false partial conjunction hypotheses for those selected groups. We discuss this goal and the implications of Corollary 4.2 in more detail in Section 7.

5. The meta-analysis setting

5.1. Preliminaries

We consider the setting where the set of m features are tested in $n \geq 2$ independent studies. The corresponding hypotheses can be arranged in a matrix with m rows and n columns, where the (i, j)th entry of the matrix is H_{ij}, the null hypothesis for feature i in study j. Consider the corresponding $m \times n$ matrix of *p*-values for these hypotheses, $\mathbf{p}$, where p_{ij} is the *p*-value for H_{ij}, $i = 1, \ldots, m$;

$j = 1, \ldots, n$. The hypotheses belonging to the same row are those corresponding to the same feature in different studies.

For feature $i \in \{1, \ldots, m\}$, let $k(i)$ be the number of studies where feature i has an effect, i.e. the number of false null hypotheses in the set $\{H_{ij}, j = 1, \ldots, n\}$. For $u \in \{1, \ldots, n\}$, consider the family of partial conjunction nulls $\{H_i^{u/n}, i = 1, \ldots, m\}$, where $H_i^{u/n}$ states that feature i has an effect in less than u studies, i.e. $H_i^{u/n} : k(i) < u$. For $u = 1$, identifying false null hypotheses in this family corresponds to identifying features with an effect in at least one study, which corresponds to the typical goal of meta-analysis. Testing this family for $u \geq 2$ corresponds to identifying features with an effect in at least u studies, which is considered as the goal of replicability analysis (see, e.g., [4], [58], [33], [18]).

Denote by $\boldsymbol{p}_{\cdot j}$ the vector of m p-values in column j, and by $\mathbf{p}_{\cdot(-j)}$ the $m \times (n-1)$ matrix which is obtained from $\mathbf{p}$ by excluding column j. Similarly, $\boldsymbol{p}_{i\cdot}$ and $\mathbf{p}_{(-i)\cdot}$ denote the vector of n p-values in row i, and the $(m-1) \times n$ matrix which is obtained from $\mathbf{p}$ by excluding row i, respectively. We denote by $\boldsymbol{p}$, $\boldsymbol{p}_{\cdot(-j)}$, and $\boldsymbol{p}_{(-i)\cdot}$ the vectors obtained by stacking the rows of $\mathbf{p}$, $\mathbf{p}_{\cdot(-j)}$, and $\mathbf{p}_{(-i)\cdot}$, respectively.

5.2. *Multiple testing of partial conjunction hypotheses in the meta-analysis setting*

We address testing the family $\{H_i^{u/n}, i = 1, \ldots, m\}$ with (weighted) FDR control. A typical meta-analysis addresses independent studies, so we assume independence across columns: for each $j \in \{1, \ldots, n\}$, the vectors $\boldsymbol{p}_{\cdot j}$ and $\boldsymbol{p}_{\cdot(-j)}$ are independent. Thus, for obtaining valid partial conjunction p-values, one may use any combining method which is valid under independence, such as Stouffer's, Fisher's, or Simes-Storey (2.8) methods. The dependency within the columns, in combination with the chosen partial conjunction p-values, guides the choice of the multiple testing procedure, which should guarantee the desired error rate control for the induced dependencies among those combined p-values. For the results below, the following dependencies are considered within each study $j \in \{1, \ldots, n\}$, in conjunction with the assumption of independence across studies:

M1 Positive dependence: the vector of p-values for the hypotheses in study j, $\boldsymbol{p}_{\cdot j}$, satisfies the PRDS property on the subset of true null hypotheses.
M2 Independence: the p-values within $\boldsymbol{p}_{\cdot j}$ are independent.
M3 Arbitrary dependence: the dependence among $\{p_{ij}, i = 1, \ldots, m\}$ is unspecified.

To obtain an analogy to the dependency structures in Section 4, it is useful to note that in the current matrix setting, the hypotheses may be viewed as divided into m groups, where the hypotheses within each row of the matrix form a group. As in Sections 3 and 4, a partial conjunction hypothesis is considered for each group (corresponding to a specific feature), and the goal is (weighted)

FDR control for the tested family of m partial conjunction hypotheses. The independence across the studies translates to independence of p-values within each group. Assuming in addition independence within each study (as defined in item M2), we obtain independence across the groups, as defined in item A1 in Section 3.3. Similarly, arbitrary dependence within each study translates to arbitrary dependence across groups. Therefore, the theoretical results for the dependencies in items M2 and M3 are covered by items 1 and 3 of Proposition 3.1, respectively. Let us consider the dependency structure in item M1, where the p-values are positively dependent within each study, and are independent across the studies. We show in Appendix A.4 that this dependency structure satisfies conditional positive dependence across groups and independence within groups, defined in items D3 and D1 in Section 4, respectively, as well as overall positive dependence, defined in item D2 in Section 4. Therefore, the results of Theorem 4.1 addressing the dependencies above hold for the dependency structure in M1. The following theorem states this result, and an additional result addressing Fisher's and Stouffer's methods for obtaining partial conjunction p-values. The theorem addresses assumptions T1–T4 given in Section 4.

Theorem 5.1. *Consider the meta-analysis setting, where the interest lies in testing the family $\{H_i^{u/n}, i = 1, \ldots, m\}$ for a specific $u \in \{1, \ldots, n\}$. Assume that this family is associated with prior weights $\{w_i, i = 1, \ldots, m\}$, and penalty weights $\{v_i, i = 1, \ldots, m\}$, satisfying $\sum_{i=1}^{m} w_i v_i = m$. Let $\mathcal{M}$ be a self-consistent multiple testing procedure with respect to thresholds of form (3.2) with identity shape function $\beta(x) = x$. Assume independence of p-values across the studies, meaning that for each $j \in \{1, \ldots, n\}$, the vectors $\boldsymbol{p}_{\cdot j}$ and $\boldsymbol{p}_{\cdot(-j)}$ are independent. In addition, assume positive dependence within each study, in the sense of item M1. In each of the following cases, procedure $\mathcal{M}$ applied on $\{P_i^{u/n}, i = 1, \ldots, m\}$ at level α guarantees*

$$FDR_v^{PC} \leq \frac{\alpha}{m} \sum_{i \in M_0} v_i w_i \leq \alpha,$$

where $M_0 = \{i \in \{1, \ldots, m\} : H_i^{u/n}$ is true$\}$, the index set of features for which the partial conjunction null is true.

1. *The partial conjunction p-values are connected to a monotone multiple testing procedure $\mathcal{M}_g$, in the sense of (4.1), and the multiple testing procedures $\mathcal{M}, \mathcal{M}_g$ satisfy either (a) or (b):*

   (a) *$\mathcal{M}$ satisfies assumption T1 and $\mathcal{M}_g$ satisfies assumption T2*

   (b) *$\mathcal{M}$ satisfies assumption T3 and $\mathcal{M}_g$ satisfies assumption T4*

2. *The multiple testing procedure $\mathcal{M}$ satisfies assumption T1, the partial conjunction p-values are based on Fisher's or Stouffer's methods, and the elementary p-values for true null hypotheses are $U(0, 1)$ random variables.*

Remark 5.1. *All our theoretical results are obtained under the assumption that the elementary p-values are valid, i.e. for each true null elementary hypothesis,*

the p-value is either $U(0, 1)$ random variable, or is stochastically larger than this variable. For the result of item 2 of Theorem 5.1, we make a stronger assumption, which holds when all the test statistics are continuous.

As a corollary, we obtain that under independence across studies and positive dependence within each study, the doubly-weighted BH procedure guarantees weighted FDR control on the family $\{H_i^{u/n}, i = 1, \ldots, m\}$, when the p-values are based on Simes', Simes-Storey (2.8), Fisher's, or Stouffer's methods, and the test statistics are continuous. The result of item 2 is a generalization of the result of [3], showing that in the matrix setting, where the p-values within each column are positively dependent, and the p-values across the columns are independent, the BH procedure guarantees FDR control when applied on partial conjunction p-values based on Stouffer's or Fisher's methods. Item 2 shows that this result remains true for the weighted variants of the BH procedure, as well as other self-consistent procedures with respect to thresholds of form (3.2), possibly incorporating structural constraints on the rejection set (see Section 3.2 for examples). This result is based on Lemma 6.2, which may be of independent interest.

5.3. Assessing the replicability extent for each feature

As discussed above, identifying features with replicated signals can be made by testing $\{H_i^{u/n}, i = 1, \ldots, m\}$ for $u \geq 2$. However, as noted by [4], it may be not clear what value of u should be chosen to establish replicability. For example, if $n > 2$, the replicability strength for features with an effect in all studies, i.e. those with $k(i) = n$, is stronger than for features with an effect in only two studies. Thus, it may be of interest to assess the replicability extent for certain features. For this purpose, Benjamini et al. [4] developed a procedure, which selects the features with evidence for having an effect in at least one study by applying the BH procedure on the global null p-values $\{P_i^{1/n}, i = 1, \ldots, m\}$, and estimates, for each selected feature i, a lower bound $\hat{k}(i)$ for $k(i)$, the number of studies where feature i has an effect. This procedure is based on sequential testing of partial conjunction hypotheses for the selected features. Benjamini et al. [4] proved that it controls the expected proportion of features with false replicability claims, i.e. $\hat{k}(i) > k(i)$, out of all the selected features, under independence within each study.

In this section an extension of the procedure of [4] is proposed, allowing general selection rules based on the entire matrix $\mathbf{p}_{m \times n}$ for selecting the features for which assessment of replicability strength will be made, as well as incorporating prior and/or penalty weights for the features. Allowing general selection rules seems to be valuable in practice: for example, a researcher may wish to assess the replicability of findings of a specific study, addressing the BH procedure on one column of the p-value matrix as the selection rule. Alternatively, one may be interested in establishing replicability for features with global null p-values smaller than a pre-specified threshold, such as Bonferroni's threshold. Prior

weights may be used for incorporating prior knowledge regarding the features, both in the selection step and in assessing replicability. Penalty weights may be incorporated in the error rate of [4] for assigning different prices for erroneous replicability claims. Based on the theoretical results we developed in previous sections, we also extend the theoretical results of [4], addressing positive and arbitrary dependencies within the studies, in addition to independence.

Assume we are given a vector of penalty weights $(v_1, \ldots, v_m)$ and a vector of prior weights $(w_1, \ldots, w_m)$ satisfying $\sum_{i=1}^{m} v_i w_i = m$, as well as a selection rule $\mathcal{S}$, which receives as input the matrix $\mathbf{p}$ and outputs the indices of selected features $\mathcal{S} = \mathcal{S}(\mathbf{p}) \subseteq \{1, \ldots, m\}$. The target error rate of the generalized procedure is

$$\mathbb{E}\left[\frac{\sum_{i \in \mathcal{S}} v_i \mathbb{1}(\hat{k}(i) > k(i))}{|\mathcal{S}|_v,}\right]. \tag{5.1}$$

This error rate reduces to the target error rate of [4] when all the penalty weights are equal to unity. Given a shape function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$, the generalized procedure targeting control of (5.1) at level q is given by the following two-step algorithm:

1. Apply a selection rule $\mathcal{S}$, which receives as input the matrix of p-values $\mathbf{p}$ and outputs the indices of selected features, $\mathcal{S} \subseteq \{1, \ldots, m\}$.
2. For each selected feature $i \in \mathcal{S}$, test sequentially the partial conjunction hypotheses with $u = 1, \ldots, n$ at level $w_i \beta(|\mathcal{S}|_v) q/m$, in order to find

$$\hat{k}(i) = \max\left\{u : \max\{P_i^{1/n}, P_i^{2/n}, \ldots, P_i^{u/n}\} \leq w_i \beta(|\mathcal{S}|_v) q/m\right\},$$

and claim that $k(i) \geq \hat{k}(i)$, i.e. feature i has an effect in at least $\hat{k}(i)$ studies. The maximum of an empty set is defined as 0.

When the selection rule is the BH procedure applied on the global null p-values, the prior and penalty weights are equal to unity, and the shape function β is the identity function $\beta(x) = x$, the above procedure reduces to the procedure of [4]. This procedure guarantees that each selected feature i has a non-trivial lower bound, i.e. $\hat{k}(i) > 0$, provided that the global null p-values which are used in Step 1 are computed using the same method as the partial conjunction p-values used in Step 2. The same property holds for the generalized procedure, when the BH selection rule is replaced by any multiple testing procedure applied on global null p-values, provided that the procedure is self-consistent with respect to thresholds of form (3.2), with the same shape function and weights as those used in Step 2.

For theoretical results, we consider the following assumptions on the selection rule in Step 1:

C1 The selection rule $\mathcal{S}$ is *non-increasing*, in the sense that $|\mathcal{S}(\mathbf{p}^{(1)})|_{\boldsymbol{v}} \geq |\mathcal{S}(\mathbf{p}^{(2)})|_{\boldsymbol{v}}$, if $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ are $m \times n$ matrices satisfying $p_{ij}^{(1)} \leq p_{ij}^{(2)}$, for each $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.

C2 The selection rule $\mathcal{S}$ is *stable* and *concordant* with respect to rows of $\mathbf{p}_{m \times n}$, according to the following definitions:

(a) A selection rule $\mathcal{S}$ is stable if for each $i \in \mathcal{S}(\mathbf{p})$, fixing all p-values in $\mathbf{p}_{(-i)\cdot}$, and changing the p-values in row i so that feature i is still selected, will not change the set $\mathcal{S}$. Let $\mathcal{S}^{(-i)\cdot} = \mathcal{S}^{(-i)\cdot}(\mathbf{p}_{(-i)\cdot})$ be the set of selected features when the p-values not corresponding to feature i are defined by $\mathbf{p}_{(-i)\cdot}$, and the n p-values for feature i change as long as feature i is selected.

(b) A stable selection rule $\mathcal{S}$ is concordant if for any two $(m-1) \times n$ matrices $\mathbf{p}^{(1)}_{(-i)\cdot}$ and $\mathbf{p}^{(2)}_{(-i)\cdot}$ satisfying $\mathbf{p}^{(1)}_{(-i)\cdot} \leq \mathbf{p}^{(2)}_{(-i)\cdot}$, where the inequality is understood coordinate-wise, it holds that $|\mathcal{S}^{(-i)\cdot}(\mathbf{p}^{(1)}_{(-i)\cdot})|_{\boldsymbol{v}} \geq |\mathcal{S}^{(-i)\cdot}(\mathbf{p}^{(2)}_{(-i)\cdot})|_{\boldsymbol{v}}$.

When $n = 1$ and the features are selected if the corresponding hypotheses are rejected by a given multiple testing procedure, each of the above properties reduces to the corresponding property of the given multiple testing procedure, as defined in Section 3.2. Moreover, for the general case where $n > 1$, assume that the selection of features is made by computing a certain combination of p-values within each row (e.g. the global null p-values for each feature), and applying a multiple testing procedure on these m combined p-values. Assume that the combining function is non-decreasing in each coordinate, as it happens for common global null combining functions. Then the selection rule is non-increasing if the corresponding multiple testing procedure is non-increasing, and the same is true with respect to stability and concordance. Therefore, the selection rule of [4], which selects the features by applying BH on global null p-values, satisfies assumptions C1 and C2 provided that the global null p-values are non-decreasing in each elementary p-value. Moreover, if BH is replaced by its weighted variant, these conditions are still satisfied. Finally, let us consider selecting the features by applying a multiple testing procedure on the p-values of a specific study j. Since for each row i, the combining function $f(p_{i1}, \ldots, p_{in}) = p_{ij}$ is non-decreasing in each p-value, any non-increasing, stable and concordant multiple testing procedure, such as those given above, will yield a selection rule with the same properties. The following theorem addresses assumptions C1 and C2, as well as assumptions T1–T4 given in Section 4.

Theorem 5.2. *Assume independence of p-values across the studies, meaning that for each $j \in \{1, \ldots, n\}$, the vectors $\boldsymbol{p}_{\cdot j}$ and $\boldsymbol{p}_{\cdot(-j)}$ are independent. The generalized procedure consisting of two steps above guarantees*

$$\mathbb{E}\left[\frac{\sum_{i \in \mathcal{S}} v_i \mathbb{1}(\hat{k}(i) > k(i))}{|\mathcal{S}|_v}\right] \leq q,$$

i.e. it controls the expected weighted proportion of selected features for which false replicability claims were made, in each of the following cases.

1. *The shape function β is the identity function $\beta(x) = x$, the p-values within each study are positively dependent in the sense of item M1 in Section 5.2, and either of the following conditions hold:*

   (a) *The selection rule $\mathcal{S}$ is non-increasing in the sense of item C1, the partial conjunction p-values are based on Fisher's or Stouffer's methods, and the elementary p-values for true null hypotheses are $U(0,1)$ random variables.*

   (b) *The selection rule $\mathcal{S}$ is non-increasing in the sense of item C1 and the partial conjunction p-values are connected in the sense of (4.1) to a monotone multiple testing procedure which satisfies assumption T2, such as those based on Simes' method.*

   (c) *The selection rule satisfies assumption C2, and the partial conjunction p-values are connected in the sense of (4.1) to a monotone multiple testing procedure which satisfies assumption T4, such as those based on Simes-Storey method (2.8).*

2. *The shape function β is the identity function $\beta(x) = x$, the p-values within each study are independent, the selection rule $\mathcal{S}$ is stable, and the partial conjunction p-values may be arbitrary, as long as they are valid under independence.*

3. *The shape function β is of form (3.5), e.g. $\beta(x) = x/(\sum_{i=1}^{m} 1/i)$, and the method for obtaining partial conjunction p-values is valid under independence. The dependence within each study and the selection rule $\mathcal{S}$ may be arbitrary.*

As mentioned above, the procedure of [4] is a special case of the procedure considered in Theorem 5.2, and since the selection rule considered in [4] is stable, the theoretical result of [4] is covered by item 2. Theorem 5.2 generalizes the result of [4], showing that it remains to hold under positive dependence within each study for several choices of partial conjunction p-values, such as those based on Fisher's, Stouffer's, Simes' and Simes-Storey (2.8) methods. Moreover, the BH procedure may be replaced by other selection rules, and one may incorporate prior and penalty weights, provided that the above conditions hold. Finally, according to item 3, one may address arbitrary selection rules and arbitrary dependencies within the studies by replacing the threshold $w_i|\mathcal{S}|_v q/m$ in Step 2 by a more conservative threshold, e.g. $w_i|\mathcal{S}|_v q/(m \sum_{i=1}^{m} 1/i)$.

6. Sufficient conditions for FDR control and key lemmas

6.1. Sufficient conditions for FDR control

Several results in this paper are proved by showing that the two sufficient conditions for FDR control, given by [12], are satisfied in the context of multiple testing of partial conjunction hypotheses. For completeness, these sufficient conditions are given in item 1 of the proposition below. An additional pair of similar conditions is given in item 2 of this proposition, and is also useful for obtaining

some results in this paper. We address here the setting and the definitions of Section 3.2.

For each pair of sufficient conditions, the first condition is on the multiple testing procedure, requiring self-consistency with respect to thresholds of form (3.2). The second condition addresses primarily the dependency structure among the p-values, requiring that for each true null hypothesis, its p-value and $|\mathcal{R}(\boldsymbol{p})|_v$ (replaced by another quantity for stable procedures), satisfy the dependency control condition, defined below.

Definition 6.1 (Dependency control condition, [12])**.** *Let $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing function. A couple (U, V) of non-negative real random variables satisfy the dependency control condition with shape function β if the following inequalities hold:*

$$\forall c > 0, \ \ \mathbb{E}\left[\frac{\mathbb{1}(U \leq c\beta(V))}{V}\right] \leq c.$$

Proposition 6.1. *Let $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing shape function, and α a positive number. Consider a multiple testing procedure $\mathcal{M}$ with thresholds of form (3.2), with shape function β and target FDR level α, which receives as input the p-values $\boldsymbol{p}$ and outputs the set of rejected hypotheses $\mathcal{R}(\boldsymbol{p})$. In each of the following two cases, the multiple testing procedure $\mathcal{M}$ guarantees*

$$FDR_v = \mathbb{E}\left[\frac{\sum_{i \in M_0} v_i \mathbb{1}(i \in \mathcal{R})}{\sum_{i=1}^m v_i \mathbb{1}(i \in \mathcal{R})}\right] \leq \frac{\alpha}{m} \sum_{i \in M_0} w_i v_i \leq \alpha.$$

1. *The multiple testing procedure $\mathcal{M}$ is self-consistent with respect to its thresholds, and for each $i \in M_0$, the couple $(p_i, |\mathcal{R}|_{\boldsymbol{v}})$ satisfies the dependency control condition with respect to the shape function β.*
2. *The multiple testing procedure $\mathcal{M}$ is stable and self-consistent with respect to its thresholds, and for each $i \in M_0$, the couple $(p_i, |\mathcal{R}^{-i}|_{\boldsymbol{v}})$ satisfies the dependency control condition with respect to the shape function β.*

Item 1 was proved by [12], see their Proposition 2.7. The proof of item 2 is similar.

Proof of item 2 of Proposition 6.1. Assume the multiple testing procedure is stable and self-consistent with respect to thresholds of form (3.2) with a certain shape function β, and the couple $(p_i, |\mathcal{R}^{-i}|_{\boldsymbol{v}})$ satisfies the dependency control condition with respect to the same shape function β. Then, for this multiple testing procedure,

$$\text{FDR}_v = \mathbb{E}\left[\frac{\sum_{i \in M_0} v_i \mathbb{1}(i \in \mathcal{R})}{\sum_{i=1}^m v_i \mathbb{1}(i \in \mathcal{R})}\right] = \sum_{i \in M_0} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{R})}{|\mathcal{R}|_{\boldsymbol{v}}}\right]$$

$$= \sum_{i \in M_0} v_i \mathbb{E}\left[\frac{\mathbb{1}\{i \in \mathcal{R}, p_i \leq w_i \beta(|\mathcal{R}|_{\boldsymbol{v}})\alpha/m\}}{|\mathcal{R}|_{\boldsymbol{v}}}\right] \tag{6.1}$$

$$= \sum_{i \in M_0} v_i \mathbb{E}\left[\frac{\mathbb{1}\{i \in \mathcal{R}, p_i \leq w_i \beta(|\mathcal{R}^{-i}|_{\boldsymbol{v}})\alpha/m\}}{|\mathcal{R}^{-i}|_{\boldsymbol{v}}}\right] \tag{6.2}$$

$$\leq \sum_{i \in M_0} v_i \mathbb{E}\left[\frac{\mathbb{1}\{p_i \leq w_i\beta(|\mathcal{R}^{-i}|_{\boldsymbol{v}})\alpha/m\}}{|\mathcal{R}^{-i}|_{\boldsymbol{v}}}\right], \tag{6.3}$$

where the equality in (6.1) follows from the fact that the multiple testing procedure is self-consistent with respect to thresholds of form (3.2), i.e. $\mathcal{R} \subseteq \{i : p_i \leq w_i\beta(|\mathcal{R}|_v)\alpha/m\}$. The equality in (6.2) follows from the fact that the procedure is stable, yielding that if $i \in \mathcal{R}$, then $\mathcal{R} = \mathcal{R}^{-i}$. Now, using the fact that for each $i \in M_0$, the pair $(p_i, |\mathcal{R}^{-i}|_{\boldsymbol{v}})$ satisfies the dependency control condition with respect to shape function β, we obtain for each $i \in M_0$,

$$\mathbb{E}\left[\frac{\mathbb{1}\{p_i \leq w_i\beta(|\mathcal{R}^{-i}|_{\boldsymbol{v}})\alpha/m\}}{|\mathcal{R}^{-i}|_{\boldsymbol{v}}}\right] \leq \frac{w_i\alpha}{m}.$$

Combining this inequality with (6.3), we obtain

$$\mathrm{FDR}_v \leq \frac{\alpha}{m} \sum_{i \in M_0} v_iw_i. \qquad \square$$

6.2. Key lemmas

The main results in this paper are based on the following lemmas, in conjunction with Proposition 2.7 in [12], which gives two sufficient conditions for (weighted) FDR control, and item 2 of Proposition 6.1 addressing stable procedures. The lemmas below yield that the second sufficient condition for (weighted) FDR control, addressing primarily the dependency structure among the *p*-values, is satisfied under the conditions of Theorems 4.1 and 5.1. Lemma 6.2 is used for proving Theorem 5.2. Given a function $g : [0,1]^k \to [0,\infty)$, we say that g is non-increasing if for any two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [0,1]^k$ satisfying $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$ where the inequality is understood coordinate-wise, it holds that $g(\boldsymbol{x}_1) \geq g(\boldsymbol{x}_2)$. The lemmas address dependency structures D1–D4 and assumptions T1–T4 given in Section 4.

Lemma 6.1. *Consider the setting of Section 3.1, where $\boldsymbol{p}$ is the vector of all the elementary p-values, and $\boldsymbol{p}^{-g}$ is the vector of all the p-values excluding those belonging to group g, for $g = 1, \ldots, G$. Let $g \in \mathcal{G}_0$, so the partial conjunction null for group g, $H_0^{u_g/n_g}$, is true. Assume that $P_g^{u_g/n_g}$ is connected to a monotone multiple testing procedure $\mathcal{M}_g$ in the sense of (4.1). Let $f : [0,1]^M \to [0,\infty)$ and $h : [0,1]^{M-n_g} \to [0,\infty)$ be two non-increasing functions.*

1. *Assume that the elementary p-values satisfy the overall positive dependence condition, as defined in item D2. Assume that $\mathcal{M}_g$ satisfies assumption T2, which holds, for example, if $P_g^{u_g/n_g}$ is based on Simes' method. Then the pair $(P_g^{u_g/n_g}, f(\boldsymbol{p}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$.*

2. *Assume that the elementary p-values satisfy conditional positive dependence across groups and independence within each group, in the sense of*

items D3 and D1. Assume that $\mathcal{M}_g$ satisfies assumption T4, which holds, for example, if $P_g^{u_g/n_g}$ is based on Simes-Storey method (2.8). Then the pair $(P_g^{u_g/n_g}, h(\boldsymbol{p}^{-g}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$.

3. *Assume that the elementary p-values satisfy positive dependence across groups in the sense of item D4. Assume that $\mathcal{M}_g$ is the Bonferroni procedure, i.e. $P_g^{u_g/n_g} = (n_g - u_g + 1)p_{g(u_g)}$, where $p_{g(1)} \leq \ldots \leq p_{g(n_g)}$ is the sequence of ordered p-values for the hypotheses in group g. Then, under any dependency within each group, the pair $(P_g^{u_g/n_g}, h(\boldsymbol{p}^{-g}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$.*

Let us fix a certain partial conjunction p-value $P_g^{u_g/n_g}$. The result that the dependency control condition holds for $(P_g^{u_g/n_g}, f(\boldsymbol{p}))$ with the identity shape function $\beta(x) = x$ for any non-increasing function $f : [0, 1]^M \to [0, \infty)$ is equivalent to the result that the inequality

$$\mathbb{E}\left[\frac{\mathbb{1}(P_g^{u_g/n_g} \leq f(\boldsymbol{p}))}{f(\boldsymbol{p})}\right] \leq 1 \tag{6.4}$$

is satisfied for any such function f. The equivalence is obtained as follows. Assume that (6.4) is satisfied for any non-increasing function $f : [0, 1]^M \to [0, \infty)$. Given such function f and a constant $c > 0$, using the inequality in (6.4) for $\tilde{f} = cf$, which is also a non-increasing function, we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(P_g^{u_g/n_g} \leq cf(\boldsymbol{p}))}{f(\boldsymbol{p})}\right] \leq c, \tag{6.5}$$

which implies that the pair $(P_g^{u_g/n_g}, f(\boldsymbol{p}))$ satisfies the dependency control condition. On the other hand, if the inequality in (6.5) holds for any non-increasing function $f : [0, 1]^M \to [0, \infty)$ and for any $c > 0$, then it holds in particular for $c = 1$, implying (6.4) for any such function f. The inequality (6.4) is similar in form to the inequalities in the group-level superuniformity lemma of [48]. Moreover, based on the equivalence we showed above, we obtain that item 1 of Lemma 6.1 for the case where $u_g = 1$ and P_g^{1/n_g} is based on Simes' method reduces to item (b) of the group-level superuniformity lemma of [48]. Therefore, Lemma 6.1 may be viewed as complementing the group-level superuniformity lemma of [48], extending the results for Simes' global null p-values to general partial conjunction p-values, and addressing additional combining methods and dependency structures. Since the latter lemma was useful for proving FDR control results for some methods addressing groups of hypotheses, such as p-filter [48], DAGGER [47], TreeBH [16], and Focused BH [39], Lemma 6.1 may be useful for extending those results.

The following lemma addresses the meta-analysis setting given in Section 5.1, and is useful for obtaining theoretical results addressing positive dependence within each study. The results addressing independence and arbitrary dependence follow directly from the results of [12].

Lemma 6.2. *Consider the meta-analysis setting given in Section 5.1. Assume independence of p-values across the studies, meaning that for each $j \in \{1, \ldots, n\}$, the vectors $\boldsymbol{p}_{\cdot j}$ and $\boldsymbol{p}_{\cdot(-j)}$ are independent. In addition, assume positive dependence within each study, in the sense of item M1 in Section 5.2. Let i be a feature for which the partial conjunction null $H_i^{u/n}$ is true. Let $f : [0,1]^{m \times n} \to [0, \infty)$ and $h : [0,1]^{(m-1) \times n} \to [0, \infty)$ be two non-increasing functions.*

1. *Assume that either of the following conditions holds: (i) $P_i^{u/n}$ is based on Fisher's or Stouffer's methods, and the elementary p-values for true null hypotheses are $U(0,1)$ random variables, or (ii) $P_i^{u/n}$ is connected in the sense of (4.1) to a monotone multiple testing procedure, which satisfies assumption T2, which holds, for example, if $P_i^{u/n}$ is based on Simes' method. Then the pair $(P_i^{u/n}, f(\boldsymbol{p}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$.*

2. *Assume that $P_i^{u/n}$ is connected in the sense of (4.1) to a monotone multiple testing procedure which satisfies assumption T4, for example $P_i^{u/n}$ is based on Simes-Storey method (2.8). Then the pair $(P_i^{u/n}, h(\boldsymbol{p}_{(-i)\cdot}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$.*

The result regarding Fisher's and Stouffer's method is shown based on the proof techniques of [3]. Similarly to [3], we rely on the theorem of Efron [20]. The remaining results follow from the fact that the hypotheses corresponding to the same feature in different studies may be viewed as a group in the setting of Section 3.1, and the dependency structure of Lemma 6.2 with respect to this group structure satisfies the conditions of items 1 and 2 of Lemma 6.1. Similarly to the results of Lemma 6.1, the results of Lemma 6.2 may be given in the form of (6.4), and may also be viewed as generalizing the group-level superuniformity lemma of [48].

7. Discussion

We obtained sufficient conditions for FDR control on partial conjunction hypotheses, addressing different dependency structures across and within the groups. For a given division into groups, we addressed two goals: (1) testing a pre-specified family of partial conjunction hypotheses $\{H_0^{u_g/n_g}, \ g = 1, \ldots, G\}$ with FDR control, and (2) selecting a subset of interesting groups via a pre-specified selection rule, and estimating the number of false null hypotheses in each selected group, so that the expected fraction of groups with over-estimated number of false null hypotheses, among all the selected groups, is bounded by a pre-specified level α. The latter goal was addressed in the meta-analysis setting, however, as we show in Section 5.2, this setting can be considered as a group setting of Section 3. Therefore, these two goals can be unified as follows. For a single set of hypotheses $\mathcal{R} \subseteq \{1, \ldots, G\}$, defined by an a-priori fixed selection

rule, it is required that

$$\mathbb{E}\left[\frac{\sum_{g \in \mathcal{R}} v_g \mathbb{1}(|A_g \cap M_1| < r_g)}{|\mathcal{R}|_v}\right] \leq \alpha, \tag{7.1}$$

where $M_1 \subseteq \{1, \ldots, M\}$ is the index set of all the false null hypotheses, $r_g = u_g$ for the procedures in Section 3, and r_g is the estimator for the number of false null hypotheses in A_g, which is an additional output of the procedure in Section 5.3. In the setting of Section 3, the selection rule is defined by the multiple testing procedure which is applied on the partial conjunction p-values, while in the setting of Section 5.3, the selection rule may be more general, however it should be fixed in advance.

Goeman and Solari ([28]) suggested a flexible approach to multiple testing, where a researcher may select any subset of hypotheses $S \subseteq \{1, \ldots, M\}$, based on the data and possibly prior knowledge, and obtain a lower confidence bound $d^M(S)$ for the number of false null hypotheses in this set. For valid post-hoc inference, the lower bounds are required to satisfy the following simultaneity property:

$$\mathbb{P}\left[d^M(S) \leq |S \cap M_1| \text{ for all } S \subseteq \{1, \ldots, M\}\right] \geq 1 - \alpha. \tag{7.2}$$

These lower bounds can be used for obtaining simultaneous upper bounds for the false discovery proportions (FDP) for all the groups, therefore procedures for obtaining these bounds are often referred to as FDP-controlling procedures. In our group setting, the approach of Goeman and Solari ([28]) may be translated to having the flexibility to select any subset of groups $K \subseteq \{1, \ldots, G\}$, and obtain a lower bound on the number of false null hypotheses within each selected group $g \in K$, rather than obtaining lower bounds only for one set $\mathcal{R}$, outputted by a fixed in advance selection rule. This flexibility can be obtained by replacing the requirement in (7.1) by the following:

$$\mathbb{P}\left[r_g \leq |A_g \cap M_1| \text{ for all } g \in \{1, \ldots, G\}\right] \geq 1 - \alpha, \tag{7.3}$$

where r_g are random bounds, and $r_g \in \{0, u_g\}$ in the setting of Section 3, while $r_g \in \{0, 1, \ldots, n_g\}$ in the setting of Section 5.3. It is easy to see that (7.3) is a more strict requirement than our requirement (7.1), i.e. (7.3) implies (7.1). Indeed, (7.3) is equivalent to

$$\mathbb{E}\left[\mathbb{1}\left\{\sum_{g=1}^{G} \mathbb{1}(|A_g \cap M_1| < r_g) > 0\right\}\right] \leq \alpha, \tag{7.4}$$

which implies (7.1). Obviously, given simultaneous lower bounds for all the subsets, i.e. $d^M(S)$ satisfying (7.2), we obtain lower bounds r_g satisfying (7.3) for the given set of groups with index set $\{1, \ldots, G\}$ in the setting of Section 5.3. In the setting of Section 3, the lower bounds $r_g = u_g \mathbb{1}(d^M(A_g) \geq u_g)$ satisfy (7.3), and a procedure which rejects $H_0^{u_g/n_g}$ if $r_g = u_g$ controls the FWER on any subset of partial conjunction hypotheses. In addition, for any subset of

groups with index set $K \subseteq \{1, \ldots, G\}$, one may obtain a lower bound for the number of false partial conjunction hypotheses in the set $\{H_0^{u_g/n_g}, \ g \in K\}$, by computing $d^G(K) = \sum_{g \in K} \mathbb{1}(d^M(A_g) \geq u_g)$, and these are simultaneous $(1 - \alpha)$-confidence lower bounds for all subsets $K \in \{1, \ldots, G\}$ provided that $d^M(S)$ satisfy (7.2). Therefore, flexible inference on a-priori given groups of hypotheses $\{A_g, \ g = 1, \ldots, G\}$ can be obtained by an FDP-controlling procedure, outputting $d^M(S)$ satisfying (7.2). However, if only FWER or FDP control on the family of $\{H_0^{u_g/n_g}, \ g = 1, \ldots, G\}$ is of interest, which can be obtained via control of (7.3) with $r_g \in \{0, u_g\}$, then one may apply an FDP-controlling procedure on the set of partial conjunction p-values $\{P_g^{u_g/n_g}, \ g = 1, \ldots, G\}$ rather than on all the elementary p-values $\{p_1, \ldots, p_M\}$, which may offer a power gain. The results of this paper may be useful for obtaining validity of such methods under different dependency structures within and across the groups. We discuss this next.

Goeman and Solari ([28]) showed that closed testing can be used for obtaining FDP control (7.2). Several other procedures for controlling the FDP have been developed (see, e.g., [25], [11], and [38]). However, Goeman et al. [26] showed that these procedures, as well as other procedures targeting control of tail probabilities of the FDP, such as FWER-controlling procedures, are either equivalent to a closed testing procedure, or can be uniformly improved by one. The properties of the closed testing procedure rely on the properties of the local tests used for testing the intersection hypotheses: the procedure is valid for FDP control if all its local tests are valid, and the more powerful the local tests are, the more powerful is the FDP-controlling procedure. As noted in Section 4, if the family of hypotheses of interest is $\{H_0^{u_g/n_g}, \ g = 1, \ldots, G\}$, then closed testing may receive as input the partial conjunction p-values for these hypotheses, and test the intersections of these hypotheses. This approach can be used for control of FWER or FDP on the partial conjunction hypotheses, or for obtaining lower bounds $r_g \in \{0, u_g\}$ satisfying (7.3), as long as the local tests for testing intersections of partial conjunction hypotheses are valid. While full closed testing applied on all the elementary hypotheses $H_1, \ldots, H_M$ can also be used for these purposes (because $r_g \in \{0, u_g\}$ satisfying (7.3) may be obtained from $d^M(S)$ satisfying (7.2), as discussed above), the approach of applying closed testing directly on the partial conjunction hypotheses may be more powerful, because of restricting attention only to a certain subset of intersections of the elementary hypotheses. Corollary 4.2 shows three cases in which Simes' test is valid for testing an intersection of partial conjunction hypotheses, yielding validity of closed testing procedure on partial conjunction p-values with Simes' local tests. For example, under arbitrary dependence within each group and positive dependence across the groups in the sense of item D4 in Section 4, closed testing procedure with Simes' local tests applied on Bonferroni partial conjunction p-values is valid for FDP control. Considering alternative valid approaches which account for arbitrary dependence, such as applying Bonferroni local tests on partial conjunction p-values, or applying full closed testing with Bonferroni local tests, would result in less powerful procedures. The results of

Corollary 4.2 could be useful even if the goal is FDP control on all subsets of elementary hypotheses, given by (7.2): one could gain power by replacing certain local tests by more powerful valid tests, possibly applied on global null group *p*-values rather than on elementary *p*-values.

Similarly to the power of the FDP-controlling procedures, the power of the FDR-controlling procedures applied on partial conjunction *p*-values strongly depends on the choice of partial conjunction tests. In this work we developed sufficient conditions for FDR control on partial conjunction hypotheses for the cases where their *p*-values are connected to self-consistent multiple testing procedures, or are constructed using Fisher's or Stouffer's methods, showing that in several cases arbitrary dependence adjustments are not needed for FDR control. It may be of interest to address other partial conjunction tests, such as those based on combining methods of Pearson ([44], [43]), Tippett ([55]), Vovk and Wang ([57]), Wilson ([59]), the Higher Criticism method ([19]), or methods based on *e*-values (see [29], [56]). Obviously, when the partial conjunction *p*-values are combinations of within-group *p*-values, and the *p*-values belonging to different groups are independent or are arbitrarily dependent, the results of [12] can be used for obtaining sufficient conditions for FDR control on partial conjunction hypotheses, as long as the dependency structure within the groups is such that the combined *p*-values are valid for partial conjunction testing. The more interesting cases are when there are dependencies across the groups, or dependencies within the studies in the meta-analysis setting, possibly implying complex dependencies among the partial conjunction *p*-values. The arbitrary dependence adjustments may lead to a substantial power loss, therefore searching for realistic dependency structures which do not require these adjustments for certain powerful partial conjunction tests may be an interesting direction for future research. In particular, it may be of interest to extend the results of this paper for certain dependencies across the studies in the meta-analysis setting, e.g. dependencies between case-control studies with shared controls.

Some of our results address Simes-Storey *p*-value (2.8) for the partial conjunction hypothesis, which is based on the minimum adjusted *p*-value of the adaptive BH procedure with Storey's plug-in estimator. In particular, we showed that in the meta-analysis setting, when there is positive dependence within each study in the sense of item M1 in Section 5.2, the doubly-weighted BH procedure controls the weighted FDR on the family of partial conjunction hypotheses when the partial conjunction *p*-values are based on Simes-Storey combining function. In addition, we showed that in this case the procedure for assessing replicability of findings in Section 5.3 controls the expected fraction of false replicability claims for the selected features. Similar results were shown for Stouffer's and Fisher's partial conjunction *p*-values. It may be of interest to study the power of Simes-Storey partial conjunction test in comparison to Stouffer's and Fisher's tests, as well as possibly other tests which are valid under independence. While there are several papers which compared the power of different methods for global null testing under independence (see, e.g., [40], [43]), and in a recent work by [33] several partial conjunction tests were compared under different scenarios, it seems that Simes-Storey partial conjunction test has not been studied and com-

pared to other tests. In addition, following the empirical results of [13], showing that the adaptive BH procedure incorporating Storey's plug-in estimator with $\lambda = \alpha$ controls the FDR under positive dependence, one may expect that the partial conjunction test, based on Simes-Storey p-value with this choice of λ, is valid under positive dependence. Obtaining this result may give the possibility to replace Simes' method by Simes-Storey method for combining positively dependent p-values. The latter method is expected to be more powerful than the former for the groups which contain a large proportion of false null hypotheses, so investigation of the validity of Simes-Storey partial conjunction test under positive dependence is also an interesting direction for future research. It may be also worth to investigate the performance of partial conjunction tests based on adaptive FDR-controlling procedures incorporating other plug-in estimators satisfying Condition 3.1, such as those considered in Corollary 13 of [13].

When the hypotheses have a group structure, testing the global null hypotheses for the groups may be performed for selecting promising groups of hypotheses, which may be only a first step, after which the elementary hypotheses within the selected groups are tested. The global null hypotheses may be replaced by partial conjunction hypotheses with $u \geq 2$, if one is in search for groups with at least u signals. This two-step approach, with global null testing in the first step, was addressed by [1], and can be viewed as hierarchical testing of a tree, which was introduced in the context of FDR control by [62], and was further developed by [61]. The latter papers considered a general tree of hypotheses, where each hypothesis is associated with a single parent hypothesis, except for the hypotheses at the first level of the tree. In our setting, each group of elementary hypotheses may be associated with a parent, its partial conjunction hypothesis, which gives us a two-level tree structure, with partial conjunction hypotheses at the first level and elementary hypotheses at the second level. Yekutieli et al. ([62]) considered several types of FDR which may be relevant when addressing a tree of hypotheses, including level-l FDR, which addresses only discoveries at a specific level l, chosen in advance. For the above two-level structure, the results of this work are useful for obtaining sufficient conditions for level-1 FDR control. While [61] considered independence of the p-values across the entire tree, we consider trees with dependencies within and across the levels. The dependence between the level-1 and level-2 p-values follows from the fact that the partial conjunction p-values are combinations of within-group elementary p-values, and the dependence between level-1 p-values is induced by the dependencies within the level-2 p-values. Several recently developed methods ([47], [48], [16], [39]) can be used for controlling different variants of FDR on general trees, where each parent hypothesis is the intersection of its child hypotheses, and their p-values are computed either by combining the p-values of their child hypotheses, or by combining the p-values of all their descendants residing at the finest level of the tree (which are referred to as leaf hypotheses). The lemmas developed in this paper generalize the results of the group-level super-uniformity lemma of [48], and may be used for extending the theoretical guarantees of the methods developed in the above papers. Finally, an interesting research direction is generalizing the results of this work for obtaining sufficient conditions for overall

and level-restricted FDR control on a tree of hypotheses with an arbitrary number of levels, addressing various combining methods for obtaining *p*-values for parent hypotheses, in conjunction with different dependency structures among the elementary *p*-values.

Appendix A: Proofs

A.1. Proof of Proposition 3.1

Proof of item 1.1. It is assumed that for each group $g = 1, \ldots, G$, the partial conjunction *p*-value $P_g^{u_g/n_g}$ is a combination of *p*-values within group g. In addition, it is assumed that the p-values in each group are independent of the p-values in any other group. Combining these two facts, we obtain that the partial conjunction *p*-values $P_g^{u_g/n_g}, g = 1, \ldots, G$ are independent. Now, since the multiple testing procedure is assumed to be non-increasing, Proposition 3.3 of [12] yields that for each group g with a true null hypothesis $H_0^{u_g/n_g}$, the pair $(P_g^{u_g/n_g}, |\mathcal{R}|_{\boldsymbol{v}})$ satisfies the dependency control with identity shape function $\beta(x) = x$. Now, since the multiple testing procedure is self-consistent with respect to thresholds of form (3.2), we obtain that the two sufficient conditions in Proposition 2.7 of [12] are satisfied, and the result follows.

Proof of item 1.2. As noted in the proof of item 1.1, since there is independence across the groups, the partial conjunction *p*-values are independent. Therefore, all the assumptions of Theorem 11 of [13] hold, and the result follows.

Proofs of items 2 and 3. The result of item 2 follows from Propositions 3.6 and 2.7 of [12]. The result of item 3 follows from Propositions 3.7 and 2.7 of [12].

A.2. Proof of Theorem 4.1

Proof of item 1. Let $|\mathcal{R}(\boldsymbol{p})|_{\boldsymbol{v}}$ be the weighted number of rejected hypotheses when the given multiple testing procedure $\mathcal{M}$ is applied on $P_g^{u_g/n_g}, g = 1, \ldots, G$. Since it is assumed that for each group g, the procedure $\mathcal{M}_g$ satisfies the monotonicity property given in Definition 2.1, $P_g^{u_g/n_g}$ is non-decreasing in each coordinate of $\boldsymbol{p}$. In addition, it is assumed that the procedure $\mathcal{M}$ is non-increasing. Using these facts we obtain that $|\mathcal{R}(\boldsymbol{p})|_{\boldsymbol{v}}$ is non-increasing in each coordinate of $\boldsymbol{p}$. Therefore, using item 1 of Lemma 6.1 we obtain that for each group g such that $H_0^{u_g/n_g}$ is a true null hypothesis, the pair $(P_g^{u_g/n_g}, |\mathcal{R}(\boldsymbol{p})|_{\boldsymbol{v}})$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$. Since it is assumed that $\mathcal{M}$ is self-consistent with respect to thresholds of form (3.2) with $\beta(x) = x$, we obtain that the two sufficient conditions of Proposition 2.7 of [12], given in item 1 of Proposition 6.1, are satisfied, and the result follows.

Proof of item 2. Consider the rejection set $\mathcal{R} \subseteq \{1, \ldots, G\}$ of the multiple testing procedure $\mathcal{M}$ when it receives as input the partial conjunction p-values $P_g^{u_g/n_g}$, $g = 1, \ldots, G$. Since $\mathcal{M}$ is stable, one may define for each $g \in \{1, \ldots, G\}$ the set $\mathcal{R}^{-g}$ as the rejection set which is obtained when all the partial conjunction p-values except $P_g^{u_g/n_g}$ are fixed, and $P_g^{u_g/n_g}$ has any value such that $H_0^{u_g/n_g}$ is rejected. Since each partial conjunction p-value $P_g^{u_g/n_g}$ is a combination of p-values in group g, and $\mathcal{R}^{-g}$ is a function of all the partial conjunction p-values except $P_g^{u_g/n_g}$, we obtain that $\mathcal{R}^{-g}$ is a function of all the elementary p-values which do not belong to group g, i.e. $\mathcal{R}^{-g} = \mathcal{R}^{-g}(\boldsymbol{p}^{-g})$. Since $\mathcal{M}$ is concordant, for each $g \in \{1, \ldots, G\}$, $|\mathcal{R}^{-g}|_{\boldsymbol{v}}$ is non-increasing in each partial conjunction p-value in the set $\{P_k^{u_k/n_k}, k \neq g\}$. Each partial conjunction p-value is non-decreasing in each coordinate of $\boldsymbol{p}$, as shown in the proof of item 1 of Theorem 4.1. Therefore, for each $g \in \{1, \ldots, G\}$, $|\mathcal{R}^{-g}|_{\boldsymbol{v}}$ is non-increasing in each coordinate of $\boldsymbol{p}^{-g}$. Hence, according to item 2 of Lemma 6.1, for each group $g \in \mathcal{G}_0$, the pair $(P_g^{u_g/n_g}, |\mathcal{R}^{-g}|_{\boldsymbol{v}})$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$. In addition, the multiple testing procedure $\mathcal{M}$ is self-consistent with respect to thresholds of form (3.2) with $\beta(x) = x$. Therefore, the conditions of item 2 of Proposition 6.1 are satisfied, and the result follows.

Proof of item 3. The proof is similar to the proof of item 2 above. As shown in that proof, for each $g \in \{1, \ldots, G\}$, $|\mathcal{R}^{-g}|_{\boldsymbol{v}}$ is non-increasing in each coordinate of $\boldsymbol{p}^{-g}$. Hence, according to item 3 of Lemma 6.1, for each group $g \in \mathcal{G}_0$, the pair $(P_g^{u_g/n_g}, |\mathcal{R}^{-g}|_{\boldsymbol{v}})$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$. Now the result follows from item 2 of Proposition 6.1.

A.3. Proof of Corollary 4.1

The corollary follows from Theorem 4.1. It is enough to show that any step-up procedure with thresholds of form (3.2), in particular the doubly-weighted BH procedure, is self-consistent, non-increasing, stable, and concordant. Blanchard and Roquain ([12]) showed that a step-up procedure with thresholds $\Delta(i, r)$ of form (3.2) rejects the hypotheses in the set $L_\Delta(\hat{r})$, where $\hat{r} = \max\{r : |L_\Delta(i, r)|_{\boldsymbol{v}} \geq r\}$, and $|L_\Delta(\hat{r})|_{\boldsymbol{v}} = \hat{r}$, therefore, it satisfies the self-consistency condition (3.4) with equality. Obviously, for each $r \geq 0$, $|L_\Delta(i, r)|_{\boldsymbol{v}}$ is non-increasing in each p-value, therefore $\hat{r} = |L_\Delta(\hat{r})|_{\boldsymbol{v}}$ is non-increasing in each p-value. It remains to show stability and concordance. Let us fix $i \in \{1, \ldots, m\}$. For each $r \geq 0$, let

$$L_\Delta^{(i)}(r) = \{j \neq i : p_j \leq \Delta(i, r)\}.$$

Define

$$\hat{r}^{(i)} = \max\{r : |L_\Delta^{(i)}(r)|_{\boldsymbol{v}} \geq r - v_i\}.$$

Note that $\hat{r}^{(i)}$ is a function of $\boldsymbol{p}^{-i}$. Let us show that if $i \in \mathcal{R}$, then $\hat{r} = \hat{r}^{(i)}$. Assume that $i \in \mathcal{R}$. Then $p_i \leq \Delta(i, \hat{r})$, where $\hat{r} = \max\{r : |L_\Delta(r)|_{\boldsymbol{v}} \geq r\}$.

For any $r \geq \hat{r}$, $p_i \leq \Delta(i, r)$, which yields that $|L_\Delta(r)|_{\boldsymbol{v}} = v_i + |L_\Delta^{(i)}(r)|_{\boldsymbol{v}}$. Therefore, according to the definition of $\hat{r}$, $|L_\Delta^{(i)}(\hat{r})|_{\boldsymbol{v}} \geq \hat{r} - v_i$, and for any $r > \hat{r}$, $|L_\Delta^{(i)}(r)|_{\boldsymbol{v}} < r - v_i$. Thus we obtain that $\hat{r} = \hat{r}^{(i)} = \max\{r : |L_\Delta^{(i)}(r)|_{\boldsymbol{v}} \geq r - v_i\}$. Therefore, fixing $\boldsymbol{p}^{-i}$ and changing p_i as long as $i \in \mathcal{R}$ will not change the set $\mathcal{R}$, which will be equal to $\mathcal{R}^{-i}(\boldsymbol{p}^{-i}) = \{i\} \cup L_\Delta^{(i)}(\hat{r}^{(i)})$. Note that for each $r \geq 0$, $|L_\Delta^{(i)}(r)|_{\boldsymbol{v}}$ is non-increasing in each coordinate of $\boldsymbol{p}^{-i}$, therefore $\hat{r}^{(i)}$ is non-increasing in each coordinate of $\boldsymbol{p}^{-i}$. Since $|L_\Delta^{(i)}(r)|_{\boldsymbol{v}}$ is non-decreasing in r, we obtain that $|L_\Delta^{(i)}(\hat{r}^{(i)})|_{\boldsymbol{v}}$ is non-increasing in each coordinate of $\boldsymbol{p}^{-i}$. Therefore, this is also true for $|\mathcal{R}^{-i}(\boldsymbol{p}^{-i})|_{\boldsymbol{v}} = v_i + |L_\Delta^{(i)}(\hat{r}^{(i)})|_{\boldsymbol{v}}$. Thus we have shown that any step-up procedure with thresholds of form (3.2) is stable and concordant. This completes the proof.

A.4. Proof of Theorem 5.1

The proof of item 1 is based on the following lemma.

Lemma A.1. *Assume the meta-analysis setting of Section 5.1, where the hypotheses are arranged in a matrix with m rows and n columns. Assume that for each $j \in \{1, \ldots, n\}$, $\boldsymbol{p}_{\cdot j}$ is independent of $\boldsymbol{p}_{\cdot(-j)}$, and $\boldsymbol{p}_{\cdot j}$ satisfies the PRDS property on the subset of true null hypotheses. Then the following results hold:*

1. *The vector of all the $m \times n$ p-values $\boldsymbol{p}$ satisfies the PRDS property on the subset of indices of true null hypotheses.*
2. *When the rows of the matrix define groups of hypotheses, the elementary p-values satisfy positive dependence across groups, as defined in item D4 in Section 4.*
3. *When the rows of the matrix define groups of hypotheses, the elementary p-values satisfy conditional positive dependence across groups and independence within each group, as defined in items D3 and D1 in Section 4, respectively.*

The proof of this lemma is given in the end of this section. The results of this lemma show that under independence across the studies and positive dependence within each study, as defined in item M1 in Section 5.2, the p-values satisfy the dependency structure assumed in items 1, 2, and 3 of Theorem 4.1. The other assumptions of items 1 and 2 of Theorem 4.1 are assumed to be satisfied, therefore the result of item 1 of Theorem 5.1 follows from items 1 and 2 of Theorem 4.1, when we address the hypotheses within each row of the matrix as a group.

Let us now prove item 2 of Theorem 5.1. According to Lemma 6.2, for each $i \in \{1, \ldots, m\}$ such that $H_i^{u/n}$ is true, the pair $(P_i^{u/n}, f(\boldsymbol{p}))$ satisfies the dependency control condition with respect to the identity shape function $\beta(x) = x$, for any non-increasing function $f : [0,1]^M \to [0, \infty)$. Since the multiple testing procedure $\mathcal{M}$ is assumed to be non-increasing, and both Stouffer's and Fisher's partial conjunction p-values are non-decreasing in each elementary p-value, we obtain that for each $i \in \{1, \ldots, m\}$ such that $H_i^{u/n}$ is true, the pair

$(P_i^{u/n}, |\mathcal{R}(\boldsymbol{p})|_{\boldsymbol{v}})$ satisfies the dependency control condition with respect to the identity shape function $\beta(x) = x$. Since $\mathcal{M}$ is assumed to be self-consistent with respect to thresholds of form (3.2) with identity shape function $\beta(x) = x$, we obtain that the two sufficient conditions for FDR control, given in Proposition 2.7 of [12] (and item 1 of Proposition 6.1), are satisfied, which completes the proof.

Proof of Lemma A.1. Items 1 and 2 of Lemma A.1 are based on the following auxiliary lemma.

Lemma A.2. *Let $\boldsymbol{p}_1 \in [0,1]^{n_1}$ and $\boldsymbol{p}_2 \in [0,1]^{n_2}$ be two independent vectors, and assume that $\boldsymbol{p}_1$ satisfies the PRDS property on the subset $I_0 \subseteq \{1, \ldots, n_1\}$. Then the vector obtained by stacking $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, denoted by $(\boldsymbol{p}_1, \boldsymbol{p}_2)$, satisfies the PRDS property on the subset I_0.*

Let us first show how item 1 of Lemma A.1 follows from Lemma A.2. Let H_{ij} be a true null hypothesis, and let $\boldsymbol{p}$ be the vector of all the $m \times n$ p-values. Note that $\boldsymbol{p}$ can be partitioned into two vectors: $\boldsymbol{p}_{\cdot j}$, the vector of p-values in study j, and $\boldsymbol{p}_{\cdot(-j)}$, the vector of p-values for the hypotheses in all the studies except study j. Since $\boldsymbol{p}_{\cdot j}$ is PRDS on the subset of true null hypotheses in study j, and $\boldsymbol{p}_{\cdot j}$ is independent of $\boldsymbol{p}_{\cdot(-j)}$, (because of independence across studies), we obtain based on Lemma A.2 that $\boldsymbol{p}$ satisfies the PRDS property on the subset of indices corresponding to true null hypotheses in study j. Therefore, for any non-decreasing set D in $[0,1]^{mn}$, $\mathbb{P}[\boldsymbol{p} \in D \,|\, p_{ij} = x]$ is non-decreasing in x. Thus we have proved item 1 of Lemma A.1. Item 2 of Lemma A.1 follows similarly. Since the groups are defined by rows, independence within each group follows from the independence across studies: for each $i \in \{1, \ldots, m\}$, the vector $\boldsymbol{p}_{i\cdot}$ contains independent p-values. Let us now show positive dependence across groups. Let $i \in \{1, \ldots, m\}$ be a row (group) which contains at least one true null hypothesis, and let H_{ij} be a true null hypothesis within this row. Let D be a non-decreasing set in $[0,1]^{mn-m+1}$. We have to show that $\mathbb{P}\left\{(\boldsymbol{p}_{(-i)\cdot}, p_{ij}) \in D \,|\, p_{ij} = x\right\}$ is non-decreasing in x. Note that $(\boldsymbol{p}_{(-i)\cdot}, p_{ij}) = (\boldsymbol{p}_{(-i)(-j)}, \boldsymbol{p}_{\cdot j})$, where $\boldsymbol{p}_{(-i)(-j)}$ is the vector obtained by stacking the rows of matrix $\mathbf{p}$ after removing its row i and column j. According to our assumptions, $\boldsymbol{p}_{(-i)(-j)}$ is independent of $\boldsymbol{p}_{\cdot j}$, while $\boldsymbol{p}_{\cdot j}$ is PRDS on the subset of true null hypotheses in study j. Therefore, according to Lemma A.2, $(\boldsymbol{p}_{(-i)(-j)}, \boldsymbol{p}_{\cdot j}) = (\boldsymbol{p}_{(-i)\cdot}, p_{ij})$ is PRDS on the subset of true null hypotheses in study j, which yields that $\mathbb{P}\left\{(\boldsymbol{p}_{(-i)\cdot}, p_{ij}) \in D \,|\, p_{ij} = x\right\}$ is non-decreasing in x. This completes the proof of item 2 of Lemma A.1.

Let us now prove item 3 of Lemma A.1. Let $\boldsymbol{y} \in [0,1]^{n-1}$, and let $D \in [0,1]^{mn}$ be a non-decreasing set. Let H_{ij} be a true null hypothesis, and let $\boldsymbol{p}_{i(-j)}$ be the vector of p-values for the hypotheses in $\{H_{ik}, k \neq j\}$. Finally, let $(\boldsymbol{p}_{(-i)\cdot}, p_{ij}, \boldsymbol{y})$ be the vector $\boldsymbol{p}$ such that the sub-vector corresponding to the hypotheses not belonging to row i is given by $\boldsymbol{p}_{(-i)\cdot}$, the p-value for H_{ij} is p_{ij}, and the sub-vector of p-values for the remaining hypotheses, belonging to the set $\{H_{ik}, k \neq j\}$, is

$\boldsymbol{y}$. Note that

$$\mathbb{P}\left[\boldsymbol{p} \in D \,|\, p_{ij} = x, \boldsymbol{p}_{i(-j)} = \boldsymbol{y}\right] = \frac{\mathbb{P}\left[(\boldsymbol{p}_{(-i)\cdot}, p_{ij}) \in \tilde{D} \,|\, p_{ij} = x\right]}{f_{\boldsymbol{p}_{i(-j)}|p_{ij}=x}(\boldsymbol{y})}, \qquad (\text{A.1})$$

where

$$\tilde{D} = \left\{(\boldsymbol{p}_{i(-j)}, p_{ij}) : (\boldsymbol{p}_{i(-j)}, p_{ij}, \boldsymbol{y}) \in D\right\},$$

and $f_{\boldsymbol{p}_{i(-j)}|p_{ij}=x}$ is the conditional density of $\boldsymbol{p}_{i(-j)}$ given $p_{ij} = x$. Since the p-values for the hypotheses in row i are independent, this conditional density is the density of $\boldsymbol{p}_{i(-j)}$. Therefore, the denominator in (A.1) does not depend on x. Since D is a non-decreasing set, $\tilde{D}$ is also a non-decreasing set. According to item 2, the p-values satisfy positive dependence across groups, therefore the nominator of (A.1) is non-decreasing in x. We obtain that the fraction in (A.1) is non-decreasing in x. Thus we have proved conditional positive dependence across groups. The independence within each group, i.e. row, follows from the independence across studies.

Let us now prove Lemma A.2. Let D be a non-decreasing set in $[0,1]^{n_1+n_2}$. Let $i \in I_0$. Since $\boldsymbol{p}_2$ is independent of $\boldsymbol{p}_{1i}$, the conditional distribution of $\boldsymbol{p}_2$ given p_{1i} is identical to the distribution of $\boldsymbol{p}_2$. Therefore, for any $x \in [0,1]$,

$$\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}_2) \in D \,|\, p_{1i} = x\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}_2) \in D \,|\, p_{1i} = x, \boldsymbol{p}_2\right]\right] \qquad (\text{A.2})$$

where the external expectation on the right hand-side of (A.2) is with respect to the distribution of $\boldsymbol{p}_2$. Let $\boldsymbol{y} \in [0,1]^{n_2}$. Since $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are independent, for any $x \in [0,1]$,

$$\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}_2) \in D \,|\, p_{1i} = x, \boldsymbol{p}_2 = \boldsymbol{y}\right] = \mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{y}) \in D \,|\, p_{1i} = x\right]. \qquad (\text{A.3})$$

The set $\{\boldsymbol{p}_1 \,:\, (\boldsymbol{p}_1, \boldsymbol{y}) \in D\}$ is non-decreasing, because the set D is non-decreasing. Therefore, since $\boldsymbol{p}_1$ is PRDS with respect to p_{1i},

$$\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{y}) \in D \,|\, p_{1i} = x\right]$$

is a non-decreasing function of x. Based on (A.2) and (A.3), we obtain

$$\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}_2) \in D|p_{1i} = x\right] = \int_{\boldsymbol{p} \in [0,1]^{n_2}} \mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}) \in D|p_{1i} = x\right] f_{\boldsymbol{p}_2}(\boldsymbol{p}) d\boldsymbol{p}, \;\; (\text{A.4})$$

where $f_{\boldsymbol{p}_2}$ is the density function of $\boldsymbol{p}_2$. As shown above, for each $\boldsymbol{p} \in [0,1]^{n_2}$, $\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}) \in D \,|\, p_{1i} = x\right]$ is a non-decreasing function of x. Therefore, based on (A.4) we obtain that $\mathbb{E}\left[\mathbb{1}(\boldsymbol{p}_1, \boldsymbol{p}_2) \in D \,|\, p_{1i} = x\right]$ is a non-decreasing function of x, which completes the proof. $\square$

A.5. Proof of Theorem 5.2

The proof is based on the following lemma.

Lemma A.3. *The procedure in Section 5.3 with shape function $\beta(\cdot)$ guarantees*

$$\mathbb{E}\left[\frac{\sum_{i \in \mathcal{S}} v_i \mathbb{1}(\hat{k}(i) > k(i))}{|\mathcal{S}|_{\boldsymbol{v}}}\right] \leq q \tag{A.5}$$

if either of the following conditions is satisfied:

1. *For each $i \in \{1, \ldots, n\}$ such that $k(i) < n$, the pair $(P_i^{(k(i)+1)/n}, |\mathcal{S}|_{\boldsymbol{v}})$ satisfies the dependency control condition with shape function β.*
2. *The selection rule $\mathcal{S}$ is stable (as defined in item C2 in Section 5.3), and for each $i \in \{1, \ldots, n\}$ such that $k(i) < n$, the pair $(P_i^{(k(i)+1)/n}, |\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}})$ satisfies the dependency control condition with shape function β.*

The proof of this lemma is given in the end of this section. Let us first prove items 1(a) and 1(b) of Theorem 5.2. For each $i \in \{1, \ldots, m\}$ such that $k(i) < n$, $H_i^{(k(i)+1)/n}$ is a true null hypothesis, because the number of false null hypotheses in row i, $k(i)$, is smaller than $k(i) + 1$. According to item 1 of Lemma 6.2, if the assumptions of item 1(a) or item 1(b) of Theorem 5.2 hold, then for each $i \in \{1, \ldots, m\}$ such that $k(i) < n$, the pair $(P_i^{(k(i)+1)/n}, |\mathcal{S}|_{\boldsymbol{v}})$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$. Therefore, according to item 1 of Lemma A.3, the sequential procedure in Section 5 with identity shape function guarantees (A.5). Thus we have proved items 1(a) and 1(b) of Theorem 5.2. Item 1(c) follows similarly, based on item 2 of Lemma 6.2 and item 2 of Lemma A.3.

Let us now prove item 2 of Theorem 5.2. Under the assumptions of this item, the p-values within each row are independent, and the partial conjunction p-values are valid. The p-values across the rows are also independent, so for each $i \in \{1, \ldots, m\}$, $P_i^{(k(i)+1)/n}$ is independent of $\boldsymbol{p}_{(-i)\cdot}$, because $P_i^{(k(i)+1)/n}$ is a combination of p-values in row i. Let us show that in this case for each $i \in \{1, \ldots, n\}$ such that $k(i) < n$, the pair $(P_i^{(k(i)+1)/n}, |\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}})$ satisfies the dependency control condition with the shape function $\beta(x) = x$, yielding that item 2 of Lemma A.3 with $\beta(x) = x$ holds. Let $i \in \{1, \ldots, n\}$ such that $k(i) < n$. Then $P_i^{(k(i)+1)/n}$ is a p-value of a true null hypothesis, and it is stochastically lower bounded by a uniform variable on $[0, 1]$. Since $P_i^{(k(i)+1)/n}$ is independent of $\boldsymbol{p}_{(-i)\cdot}$, the conditional distribution of $P_i^{(k(i)+1)/n}$ given $\boldsymbol{p}_{(-i)\cdot}$ is the same as its marginal distribution, and therefore is stochastically lower bounded by a uniform distribution. In addition, for a given $\boldsymbol{p}_{(-i)\cdot}$, $\mathcal{S}^{(-i)\cdot}$ is fixed. Hence we obtain for any $c > 0$:

$$\mathbb{E}\left[\frac{\mathbb{1}(P_i^{(k(i)+1)/n} \leq c|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}})}{|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}} \mid \boldsymbol{p}_{(-i)\cdot}\right] \leq c,$$

therefore

$$\mathbb{E}\left[\frac{\mathbb{1}(P_i^{(k(i)+1)/n}) \leq c|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}}{|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}}\right] =$$

$$\mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{1}(P_i^{(k(i)+1)/n}) \leq c|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}}{|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}} \mid \boldsymbol{p}_{(-i)\cdot}\right]\right] \leq c.$$

Thus we have shown that item 2 of Lemma A.3 with $\beta(x) = x$ holds, therefore the result of item 2 of Theorem 5.2 follows. Finally, item 3 of Theorem 5.2 follows from item (iii) of Lemma 3.2 of [12], showing that the condition of item 1 of Lemma A.3 holds with respect to the shape function β of form (3.5).

Proof of Lemma A.3. Let us first prove item 1 of Lemma A.3. Define $I = \{i \in \{1, \ldots, m\} : k(i) < n\}$. Note that the event $\{i \in \mathcal{S}, \hat{k}(i) > k(i)\}$ is equivalent to the event $\{i \in I \cap \mathcal{S}, \hat{k}(i) > k(i)\}$, because $\hat{k}(i) \leq n$ for each $i \in \mathcal{S}$. Therefore,

$$\mathbb{E}\left[\frac{\sum_{i \in \mathcal{S}} v_i \mathbb{1}(\hat{k}(i) > k(i))}{|\mathcal{S}|_{\boldsymbol{v}}}\right] = \mathbb{E}\left[\frac{\sum_{i \in I} v_i \mathbb{1}(i \in \mathcal{S}, \hat{k}(i) > k(i))}{|\mathcal{S}|_{\boldsymbol{v}}}\right]$$
$$= \sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, \hat{k}(i) > k(i))}{|\mathcal{S}|_{\boldsymbol{v}}}\right] \qquad (A.6)$$

For the sequential procedure with identity shape function, given in Section 5.3, the event $\{i \in \mathcal{S}, \hat{k}(i) > k(i)\}$ implies the event $\{i \in \mathcal{S}, P_i^{(k(i)+1)/n} \leq w_i|\mathcal{S}|_{\boldsymbol{v}}q/m\}$. Indeed, by the definition of $\hat{k}(i)$, for each selected feature i and for each $u \leq \hat{k}(i)$, $P_i^{u/n} \leq w_i|\mathcal{S}|_{\boldsymbol{v}}q/m$. If $\hat{k}(i) > k(i)$, then $P^{(k(i)+1)/n} \leq w_i|\mathcal{S}|_{\boldsymbol{v}}q/m$. Therefore, continuing from (A.6), we obtain

$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, \hat{k}(i) > k(i))}{|\mathcal{S}|_{\boldsymbol{v}}}\right] \leq \qquad (A.7)$$
$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, P_i^{(k(i)+1)/n} \leq w_i\beta(|\mathcal{S}|_{\boldsymbol{v}})q/m}{|\mathcal{S}|_{\boldsymbol{v}}}\right] \leq$$
$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(P_i^{(k(i)+1)/n} \leq w_i\beta(|\mathcal{S}|_{\boldsymbol{v}})q/m}{|\mathcal{S}|_{\boldsymbol{v}}}\right] \leq \qquad (A.8)$$
$$\frac{q}{m}\sum_{i=1}^{m} v_i w_i = q$$

The inequality in (A.8) follows from the fact that for each $i \in \{1, \ldots, m\}$ such that $k(i) < n$, $H_i^{(k(i)+1)/n}$ is a true null hypothesis, and the pair

$$(P_i^{(k(i)+1)/n}, |\mathcal{S}|_{\boldsymbol{v}})$$

satisfies the dependency control condition with shape function β. This completes the proof of item 1 of Lemma A.3.

Let us prove item 2 of Lemma A.3. In this case $\mathcal{S}$ is stable, therefore $\mathcal{S}^{(-i)\cdot}$ is defined, and

$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, P_i^{(k(i)+1)/n} \leq w_i\beta(|\mathcal{S}|_{\boldsymbol{v}})q/m}{|\mathcal{S}|_{\boldsymbol{v}}}\right] =$$

$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, P_i^{(k(i)+1)/n} \le w_i \beta(|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}})q/m)}{|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}}\right].$$

Therefore, using (A.6) and (A.7), we obtain

$$\mathbb{E}\left[\frac{\sum_{i \in \mathcal{S}} v_i \mathbb{1}(\hat{k}(i) > k(i))}{|\mathcal{S}|_{\boldsymbol{v}}}\right] \le$$

$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, P_i^{(k(i)+1)/n} \le w_i \beta(|\mathcal{S}|_{\boldsymbol{v}})q/m)}{|\mathcal{S}|_{\boldsymbol{v}}}\right] =$$

$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(i \in \mathcal{S}, P_i^{(k(i)+1)/n} \le w_i \beta(|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}})q/m)}{|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}}\right] \le$$

$$\sum_{i \in I} v_i \mathbb{E}\left[\frac{\mathbb{1}(P_i^{(k(i)+1)/n} \le w_i \beta(|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}})q/m)}{|\mathcal{S}^{(-i)\cdot}|_{\boldsymbol{v}}}\right] \le \qquad (A.9)$$

$$\frac{q}{m}\sum_{i=1}^{m} v_i w_i = q,$$

where the inequality in (A.9) follows from the fact that for each $i \in \{1, \ldots, m\}$ such that $k(i) < n$, the pair $(P_i^{(k(i)+1)/n}, |\mathcal{S}^{(-i)}|_{\boldsymbol{v}})$ satisfies the dependency control condition with shape function β. $\qquad \square$

A.6. Proof of Lemma 6.1

The proof is based on the following auxiliary lemma.

Lemma A.4. *Consider the setting of Section 3.1. Let $g \in \mathcal{G}_0$, and assume that $P_g^{u_g/n_g}$ is of form (4.1):*

$$P_g^{u_g/n_g} = \max\{P^{1/(n_g - u_g + 1)}[A] : A \subseteq A_g, |A| = n_g - u_g + 1\},$$

where for each null group $A \subseteq A_g$ such that $|A| = n_g - u_g + 1$, $P^{1/(n_g - u_g + 1)}[A]$ is a valid global null p-value for group A. Let U be a non-negative random variable, $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing shape function satisfying $\beta(0) = 0$. Assume that for each null group $A \subseteq A_g$ such that $|A| = n_g - u_g + 1$, the pair $(P^{1/(n_g - u_g + 1)}[A], U)$ satisfies the dependency control condition with respect to the shape function β. Then the pair $(P_g^{u_g/n_g}, U)$ satisfies the dependency control condition with respect to the shape function β.

The proof of this lemma is given in the end of this section. Note that the condition $\beta(0) = 0$ holds for the shape functions considered in Lemmas 6.1 and 6.2, i.e. for the identity shape function and for the shape function of form (3.5). The proof of Lemma 6.1 is based on the techniques of the proof of the group-level super-uniformity lemma of [48].

Proof of item 1. Let c be a positive constant, and $g \in \mathcal{G}_0$. According to the assumptions of item 1,

$$P_g^{u_g/n_g} = \max\{P^{1/(n_g-u_g+1)}[A] : A \subseteq A_g, |A_g| = n_g - u_g + 1\},$$

where $P^{1/(n_g-u_g+1)}[A]$ is the minimum adjusted p-value based on the multiple testing procedure $\mathcal{M}_g$ applied on the p-values of hypotheses with indices in A, and $\mathcal{M}_g$ is non-increasing and is self-consistent with respect to thresholds of form $\Delta(r) = r\alpha/m$. Since it is assumed that the p-values satisfy the PRDS property on the subset of true null hypotheses, we obtain from Propositions 2.7 and 3.6 in [12] that $\mathcal{M}_g$ controls the FDR when applied on any subset of p-values. Therefore, according to Corollary 2.1, for any null group $A \subseteq A_g$ such that $|A_g| = n_g - u_g + 1$, $P^{1/(n_g-u_g+1)}[A]$ is a valid p-value for testing the global null for group A, i.e. its distribution is uniform or is stochastically larger than uniform.

Let $f : [0,1]^M \to [0,\infty)$ be a non-increasing function. According to Lemma A.4, it is enough to prove that for any null group $A \subseteq A_g$ such that $|A_g| = n_g - u_g + 1$, the pair

$$(P^{1/(n_g-u_g+1)}[A], f(\boldsymbol{p}))$$

satisfies the dependency control condition with the shape function $\beta(x) = x$. Let $A \subseteq A_g$ be a null group such that $|A_g| = n_g - u_g + 1$. Note that such group exists, because $g \in \mathcal{G}_0$, yielding that the number of true null hypotheses in group g is not smaller than $n_g - u_g + 1$. Let c be a positive constant. We need to prove that

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g-u_g+1)}[A] \leq cf(\boldsymbol{p}))}{f(\boldsymbol{p})}\right] \leq c.$$

Note that the event $\{P^{1/(n_g-u_g+1)}[A] \leq cf(\boldsymbol{p})\}$ is equivalent to the event in which $\mathcal{M}_g$ makes at least one rejection when it is applied on the p-values belonging to group A at level $cf(\boldsymbol{p})$. Let $\mathcal{R}_g^A$ be the set of indices of hypotheses rejected by $\mathcal{M}_g$ when it is applied on the p-values of group A at level $cf(\boldsymbol{p})$, and let $|\mathcal{R}_g^A|$ be their number. Then

$$\mathbb{1}(\{P^{1/(n_g-u_g+1)}[A] \leq cf(P)\}) = \mathbb{1}(|\mathcal{R}_g^A| > 0) = \frac{|\mathcal{R}_g^A|}{|\mathcal{R}_g^A|}. \tag{A.10}$$

The last equality follows from the fact that we define $0/0 = 0$, so the fraction on the right hand side is equal to 1 if $|\mathcal{R}_g^A| > 0$, and is equal to 0 otherwise. Since $\mathcal{M}_g$ is self-consistent with respect to thresholds of form (3.2) with unity weights and identity shape function, we obtain that

$$\mathcal{R}_g^A \subseteq \{j \in A : p_j \leq |\mathcal{R}_g^A|cf(\boldsymbol{p})/|A|\},$$

yielding that

$$|\mathcal{R}_g^A| \leq \sum_{j \in A} \mathbb{1}(p_j \leq |\mathcal{R}_g^A|cf(\boldsymbol{p})/|A|). \tag{A.11}$$

Note that if $|\mathcal{R}_g^A| = 0$, then $\sum_{j \in A} \mathbb{1}(p_j \leq |\mathcal{R}_g^A| cf(\boldsymbol{p})/|A|) = 0$ almost surely, because for each $j \in A$, p_j is uniform or is stochastically larger than uniform, so $\mathbb{P}[p_j \leq 0] = 0$. Therefore, according to Remark 1.1, if $|\mathcal{R}_g^A| = 0$, the fraction

$$\frac{\sum_{j \in A} \mathbb{1}(p_j \leq |\mathcal{R}_g^A| cf(\boldsymbol{p})/|A|)}{|\mathcal{R}_g^A|}$$

is defined as 0. It follows that we can divide both sides of (A.11) by $|R_g^A|$, and the inequality will be preserved. Therefore, combining (A.10) and (A.11), we obtain

$$\mathbb{1}(P^{1/(n_g - u_g + 1)}[A] \leq cf(\boldsymbol{p})) \leq \frac{\sum_{j \in A} \mathbb{1}(p_j \leq |\mathcal{R}_g^A| cf(\boldsymbol{p})/|A|)}{|\mathcal{R}_g^A|} \qquad \text{(A.12)}$$

Using the fact that $P^{1/(n_g - u_g + 1)}[A]$ is uniform or is stochastically larger than uniform, and using the same argument as above, it is legitimate to divide both expressions in (A.12) by $f(\boldsymbol{p})$, preserving the inequality. Taking expectations on both sides, we obtain:

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g - u_g + 1)}[A] \leq cf(\boldsymbol{p}))}{f(\boldsymbol{p})}\right] \leq \mathbb{E}\left[\frac{\sum_{j \in A} \mathbb{1}(p_j \leq |\mathcal{R}_g^A| cf(\boldsymbol{p})/|A|)}{|\mathcal{R}_g^A| f(\boldsymbol{p})}\right] \qquad \text{(A.13)}$$

Since $\mathcal{M}_g$ is non-increasing, $|\mathcal{R}_g^A|$ is non-increasing in each p-value in A. In addition, $f(\boldsymbol{p})$ is non-increasing in each p-value, therefore the function $g(\boldsymbol{p}) \equiv c|\mathcal{R}_g^A| f(\boldsymbol{p})/|A|$ is non-increasing in each p-value. Since the p-values satisfy the PRDS property on the subset of true null hypotheses, and A is a null group, we obtain from the super-uniformity lemma of [48], item 1, for each $j \in A$,

$$\mathbb{E}\left[\frac{\mathbb{1}(p_j \leq g(\boldsymbol{p}))}{g(\boldsymbol{p})}\right] \leq 1.$$

Combining this inequality with (A.13), we obtain

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g - u_g + 1)}[A] \leq cf(\boldsymbol{p}))}{f(\boldsymbol{p})}\right] &\leq \mathbb{E}\left[\frac{\sum_{j \in A} \mathbb{1}(p_j \leq |\mathcal{R}_g^A| cf(\boldsymbol{p})/|A|)}{|\mathcal{R}_g^A| f(\boldsymbol{p})}\right] \\
&= \sum_{j \in A} \mathbb{E}\left[\frac{\mathbb{1}(p_j \leq |\mathcal{R}_g^A| cf(\boldsymbol{p})/|A|)}{|\mathcal{R}_g^A| f(\boldsymbol{p})}\right] \qquad \text{(A.14)} \\
&\leq \sum_{j \in A} \frac{c}{|A|} \mathbb{E}\left[\frac{\mathbb{1}(p_j \leq g(\boldsymbol{p}))}{g(\boldsymbol{p})}\right] \\
&\leq c
\end{aligned}$$

This completes the proof. For the last step of the proof, we could rely on the results of [12] rather on the superuniformity lemma of [48]. Specifically, using the arguments in the proof of Proposition 3.6 in [12], one can show that for

each $j \in A$, the pair $(p_j, |\mathcal{R}_g^A| f(\boldsymbol{p}))$ satisfies the dependency control condition with respect to the identity shape function $\beta(x) = x$, which yields that for each $j \in A$,

$$\mathbb{E}\left[\frac{\mathbb{1}(p_j \leq |\mathcal{R}_g^A| c f(\boldsymbol{p})/|A|)}{|\mathcal{R}_g^A| f(\boldsymbol{p})}\right] \leq \frac{c}{|A|}.$$

Combining this inequality with (A.14), the result follows.

Proof of item 2. Let $g \in \mathcal{G}_0$ and let $h : [0,1]^{M-n_g} \to [0, \infty)$ be a non-increasing function. We need to prove that the pair $(P_g^{u_g/n_g}, h(\boldsymbol{p}^{-g}))$ satisfies the dependency control condition with respect to the shape function $\beta(x) = x$. Based on Lemma A.4, it is enough to prove that for any null group $A \subseteq A_g$ such that $|A| = n_g - u_g + 1$, the pair $(P^{1/(n_g - u_g + 1)}[A], h(\boldsymbol{p}^{-g}))$ satisfies the dependency control condition. Let $A \subseteq A_g$ be a null group such that $|A| = n_g - u_g + 1$. Such group exists since $g \in \mathcal{G}_0$, as explained in the proof of item 1 of Lemma 6.1. Let c be a positive constant. According to the assumptions, $P^{1/(n_g - u_g + 1)}[A]$ is the minimum adjusted p-value of a multiple testing procedure $\mathcal{M}_g$ applied on the p-values of A, and $\mathcal{M}_g$ is non-increasing and self-consistent with respect to thresholds of form $\Delta(i, r) = r\alpha/(\hat{\pi}_0(\boldsymbol{p}^A)|A|)$, where $\boldsymbol{p}^A$ is the vector of p-values for the hypotheses in group A, and $\hat{\pi}_0(\boldsymbol{p}^A)$ is an estimator for the proportion of true null hypotheses in group A, π_0^A. It is assumed that the function $\hat{\pi}_0$ satisfies Condition 3.1. Since the p-values are assumed to be independent, according to Theorem 11 of [13] we obtain that $\mathcal{M}_g$ controls the FDR when applied on the p-values of A, therefore $P^{1/(n_g - u_g + 1)}[A]$ is a valid global null p-value for group A. Let $\mathcal{R}_g^A$ be the set of rejections made by the multiple testing procedure $\mathcal{M}_g$ when it is applied on the p-values of group A at level $ch(\boldsymbol{p}^{-g})$, and let $|\mathcal{R}_g^A|$ be their number. Using similar arguments to those leading to (A.13), we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g - u_g + 1)}[A] \leq ch(\boldsymbol{p}^{-g}))}{h(\boldsymbol{p}^{-g})}\right] \leq$$
$$\sum_{i \in A} \mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A| ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}^A)/|A|)}{h(\boldsymbol{p}^{-g})|\mathcal{R}_g^A|}\right]. \tag{A.15}$$

Let $i \in A$ be arbitrary fixed. Let $\boldsymbol{p}_{(0,i)}^A$ be the vector of p-values with indices in A, where p_i is replaced by 0. Since the function $\hat{\pi}_0^{-1}$ is non-increasing in each p-value, $\hat{\pi}_0^{-1}(\boldsymbol{p}^A) \leq \hat{\pi}^{-1}(\boldsymbol{p}_{(0,i)}^A)$. Therefore,

$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A| ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_A)/|A|)}{h(\boldsymbol{p}^{-g})|\mathcal{R}_g^A|}\right] \leq$$
$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A| ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|)}{h(\boldsymbol{p}^{-g})|\mathcal{R}_g^A|}\right]. \tag{A.16}$$

Recall that $\boldsymbol{p}_g^{-i}$ is the vector of the p-values in group g, excluding p_i. For fixed $\boldsymbol{p}_g^{-i}$, the p-values for the hypotheses in $A \setminus \{i\}$ are fixed, so $|\mathcal{R}_g^A|$ is a non-increasing function of p_i, and $\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)$ is fixed. In addition, $h(\boldsymbol{p}^{-g})$ is non-increasing in each entry of the vector $\boldsymbol{p}^{-g}$. This yields that for fixed $\boldsymbol{p}_g^{-i}$, the function $c|\mathcal{R}_g^A|h(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|$ is non-increasing in each coordinate of $\boldsymbol{p}$. Since there is independence within each group, p_i is independent of $\boldsymbol{p}_g^{-i}$, therefore the conditional distribution of p_i given $\boldsymbol{p}_g^{-i}$ is uniform or is stochastically larger than uniform. Due to conditional positive dependence across groups, as defined in item D3 in Section 4, the vector $\boldsymbol{p}$ is PRDS with respect to p_i, conditionally on $\boldsymbol{p}_g^{-i}$. Therefore, using item (b) of super-uniformity lemma of [48], we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|)}{|\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|} \mid \boldsymbol{p}_g^{-i}\right] \leq 1,$$

which yields

$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|)}{h(\boldsymbol{p}^{-g})|\mathcal{R}_g^A|} \mid \boldsymbol{p}_g^{-i}\right] =$$

$$\frac{c\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)}{|A|}\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|)}{|\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|} \mid \boldsymbol{p}_g^{-i}\right] \leq \tag{A.17}$$

$$\frac{c\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)}{|A|} \tag{A.18}$$

Using (A.18) we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|)}{h(\boldsymbol{p}^{-g})|\mathcal{R}_g^A|}\right] = \tag{A.19}$$

$$\mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq |\mathcal{R}_g^A|ch(\boldsymbol{p}^{-g})\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)/|A|)}{h(\boldsymbol{p}^{-g})|\mathcal{R}_g^A|} \mid \boldsymbol{p}_g^{-i}\right]\right] \leq$$

$$\frac{c}{|A|}\mathbb{E}\left[\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)\right] \tag{A.20}$$

Combining (A.20) with (A.15) and (A.16), we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g-u_g+1)} \leq ch(\boldsymbol{p}^{-g}))}{h(\boldsymbol{p}^{-g})}\right] \leq \frac{c}{|A|}\sum_{i \in A}\mathbb{E}\left[\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)\right].$$

Using the assumption that the estimator $\hat{\pi}_0$ satisfies Condition 3.1, we obtain that for any $i \in A$, $\mathbb{E}\left[1/\hat{\pi}_0(\boldsymbol{p}_{(0,i)}^A)\right] \leq 1/\pi_0^A$. Since A is a group consisting only of true null hypotheses, $\pi_0^A = 1$, therefore for any $i \in A$, $\mathbb{E}\left[\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)\right] \leq 1$.

Therefore, we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g-u_g+1)}[A] \leq ch(\boldsymbol{p}^{-g}))}{h(\boldsymbol{p}^{-g})}\right] \leq \frac{c}{|A|}\sum_{i\in A}\mathbb{E}\left[\hat{\pi}_0^{-1}(\boldsymbol{p}_{(0,i)}^A)\right] \leq c.$$

This completes the proof.

Proof of item 3. Let $g \in \mathcal{G}_0$ and $h : [0,1]^{M-n_g} \to [0,\infty)$. Let $A \subseteq A_g$ be a null group such that $|A| = n_g - u_g + 1$ (which exists, because $g \in \mathcal{G}_0$). Based on Lemma A.4, it is enough prove that that the pair $(P^{1/(n_g-u_g+1)}[A],\, h(\boldsymbol{p}^{-g}))$ satisfies the dependency control condition with respect to shape function $\beta(x) = x$. Let c be a positive constant. Since $P^{1/(n_g-u_g+1)}[A]$ is based on Bonferroni, $P^{1/(n_g-u_g+1)}[A] = |A|\min\{p_i : i \in A\}$. Therefore,

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g-u_g+1)} \leq ch(\boldsymbol{p}^{-g}))}{h(\boldsymbol{p}^{-g})}\right] \leq \mathbb{E}\left[\frac{\mathbb{1}(\min\{p_i : i \in A\} \leq ch(\boldsymbol{p}^{-g})/|A|)}{h(\boldsymbol{p}^{-g})}\right]$$

$$\leq \sum_{i\in A}\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq ch(\boldsymbol{p}^{-g})/|A|)}{h(\boldsymbol{p}^{-g})}\right] \quad\quad \text{(A.21)}$$

$$= \sum_{i\in A}\frac{c}{|A|}\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq ch(\boldsymbol{p}^{-g})/|A|)}{ch(\boldsymbol{p}^{-g})/|A|}\right] \quad \text{(A.22)}$$

Let $i \in A$ be arbitrary fixed. Since A consists only of true null hypotheses, p_i corresponds to a true null hypothesis for each $i \in A$. Therefore, the vector $(p_i, \boldsymbol{p}^{-g})$ is PRDS with respect to p_i, according to the assumption of positive dependence across groups, as defined in item D4 in Section 4. Using the fact that $ch(\boldsymbol{p}^{-g})/|A|$ is a coordinate-wise non-increasing function of $\boldsymbol{p}^{-g}$, we obtain from item (b) of the super-uniformity lemma of [48] that

$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq ch(\boldsymbol{p}^{-g})/|A|)}{ch(\boldsymbol{p}^{-g})/|A|}\right] \leq 1.$$

Combining this inequality with (A.22), we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}(P^{1/(n_g-u_g+1)}[A] \leq ch(\boldsymbol{p}^{-g}))}{h(\boldsymbol{p}^{-g})}\right] \leq \sum_{i\in A}\frac{c}{|A|} = c, \quad\quad \text{(A.23)}$$

which completes the proof. As in item 1, we could use the results of [12] rather than those of [48] for the last step of the proof. Specifically, using similar arguments to those used in the proof of Proposition 3.6 of [12], we could show that for each $i \in A$, $(p_i, h(\boldsymbol{p}^{-g}))$ satisfies the dependency control condition with respect to shape function $\beta(x) = x$, which yields that

$$\mathbb{E}\left[\frac{\mathbb{1}(p_i \leq ch(\boldsymbol{p}^{-g})/|A|)}{h(\boldsymbol{p}^{-g})}\right] \leq \frac{c}{|A|}.$$

Using this result, we obtain (A.23) from (A.21).

Proof of Lemma A.4. Let c be a positive constant and let $g \in \mathcal{G}_0$. We need to show that

$$\mathbb{E}\left[\frac{\mathbb{1}\{P_g^{u_g/n_g} \le c\beta(U)\}}{U}\right] \le c.$$

By assumption,

$$P_g^{u_g/n_g} = \max\left\{P^{1/(n_g-u_g+1)} : A \subseteq A_g, |A| = n_g - u_g + 1\right\}. \qquad (A.24)$$

Let $A_g^0 \subseteq A_g$ be the set of indices of true null hypotheses belonging to group g. Since $g \in \mathcal{G}_0$, $H_0^{u_g/n_g}$ is a true null hypothesis, therefore $|A_g^0| \ge n_g - u_g + 1$, and there exists a set $A \subseteq A_g^0$ such that $|A| = n_g - u_g + 1$. Fix an arbitrary set $A' \subseteq A_g^0$ satisfying $|A'| = n_g - u_g + 1$. Then we obtain

$$\mathbb{1}\left[\max\left\{P^{1/(n_g-u_g+1)}[A] : A \subseteq A_g, |A| = n_g - u_g + 1\right\} \le c\beta(U)\right] \le$$
$$\mathbb{1}\left[P^{1/(n_g-u_g+1)}[A'] \le c\beta(U)\right] \qquad (A.25)$$

According to our assumptions, for each null group $A \subseteq A_g$ such that $|A| = n_g - u_g + 1$, $P^{1/(n_g-u_g+1)}[A]$ is a valid global null p-value for group A. Therefore, both $P_g^{u_g/n_g}$ and $P^{1/(n_g-u_g+1)}[A']$ are either uniform or are stochastically larger than uniform, which yields that $\mathbb{P}\left[P_g^{u_g/n_g} \le 0\right] = \mathbb{P}\left[P^{1/(n_g-u_g+1)}[A'] \le 0\right] = 0$. On the other hand, it is assumed that $\beta(0) = 0$. Therefore, according to Remark 1.1, we can divide both expressions in (A.25) by U while preserving the inequality. Therefore, combining (A.24) and (A.25), we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}\left\{P_g^{u_g/n_g} \le c\beta(U)\right\}}{U}\right] \le \mathbb{E}\left[\frac{\mathbb{1}\left\{P^{1/(n_g-u_g+1)}[A'] \le c\beta(U)\right\}}{U}\right] \qquad (A.26)$$

Since it is assumed that for each null group $A \subseteq A_g$ such that $|A| = n_g - u_g + 1$, the pair $(P^{1/(n_g-u_g+1)}[A], U)$ satisfies the dependency control condition with respect to the shape function β, we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}\left\{P^{1/(n_g-u_g+1)}[A'] \le c\beta(U)\right\}}{U}\right] \le c \qquad (A.27)$$

Combining (A.26) with (A.27), we obtain

$$\mathbb{E}\left[\frac{\mathbb{1}\{P_g^{u_g/n_g} \le c\beta(U)\}}{U}\right] \le c,$$

which completes the proof. $\qquad \square$

A.7. Proof of Lemma 6.2

Proof of item 1. Let us first assume that $P_i^{u/n}$ is connected to a multiple testing procedure in the sense of (4.1), which satisfies condition (c) of item 1 of Theorem 4.1. The meta-analysis setting can be viewed as a group setting of Section 3.1, where the n hypotheses for each feature constitute a group. According to Lemma A.1, the p-values satisfy the overall positive dependence condition, as defined in item D2 in Section 4. Therefore, according to item 1 of Lemma 6.1, the pair $(P_i^{u/n}, f(\boldsymbol{p}))$ satisfies the dependency control condition with respect to the identity shape function $\beta(x) = x$.

Let us now assume that $P_i^{u/n}$ is based on Fisher's or on Stouffer's methods. In this case the proof is similar to the proof of Theorem 3 in [3]. If $P_i^{u/n}$ is based on Fisher's method, then

$$P_i^{u/n} = \max\{P_i^{1/(n-u+1)}[A] : A \subseteq \{1, \ldots, n\}, |A| = n - u + 1\}, \qquad \text{(A.28)}$$

where

$$P_i^{1/(n-u+1)}[A] = \mathbb{P}\left[\chi^2_{2(n-u+1)} \geq -2 \sum_{j \in A} \log p_{ij}\right]. \qquad \text{(A.29)}$$

If $P_i^{u/n}$ is based on Stouffer's method, then $P_i^{u/n}$ is given by (A.28), where

$$P_i^{1/(n-u+1)}[A] = \Phi\left\{\frac{\sum_{j \in A} \Phi^{-1}(p_{ij})}{\sqrt{n-u+1}}\right\}. \qquad \text{(A.30)}$$

According to Lemma A.4, it is enough to show that for any subset of indices $A \subseteq \{1, \ldots, n\}$ such that $|A| = n - u + 1$, and H_{ij} is a true null hypothesis for each $j \in A$, the pair $(P_i^{1/(n-u+1)}[A], f(\boldsymbol{p}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$, when $P_i^{1/(n-u+1)}[A]$ is given by (A.29) or (A.30). Let $A \subseteq \{1, \ldots, n\}$ be such a subset (which exists, because $P_i^{u/n}$ is a true null hypothesis). We shall rely on item 2 of Lemma 3.2 of [12], which we give below for completeness.

Lemma A.5 ([12]). *Let* (U, V) *be a couple of non-negative random variables such that* U *is stochastically lower bounded by a uniform random variable on* $[0, 1]$, *i.e.* $\forall t \in [0, 1]$, $\mathbb{P}[U \leq t] \leq t$. *Then the pair* (U, V) *satisfies the dependency control condition with the identity shape function* $\beta(x) = x$ *if for any* $r \geq 0$, *the function* $u \mapsto \mathbb{P}[V < r \mid U \leq u]$ *is non-decreasing.*

The set A consists only of true null hypotheses, and according to our assumptions, their p-values are independent and uniform on $[0, 1]$. Therefore, $P_i^{1/(n-u+1)}[A]$ based on Fisher's or Stouffer's methods (given in (A.29) and (A.30), respectively) is a a uniform random variable on $[0, 1]$. According to Lemma A.5 it is enough to show that for any $r \geq 0$, the function

$$u \mapsto \mathbb{P}\left[f(\boldsymbol{p}) < r \mid P_i^{1/(n-u+1)}[A] \leq u\right]$$

is non-decreasing, where $P_i^{1/(n-u+1)}[A]$ is based on either Fisher's or Stouffer's methods. As shown by [8] and [12], the above result follows if for any $r \geq 0$, the function $u \mapsto \mathbb{P}\left[f(\boldsymbol{p}) < r \mid P_i^{1/(n-u+1)}[A] = u\right]$ is non-decreasing, so it is enough to prove the latter. Let $r \geq 0$ be arbitrary fixed. Define $W : [0,1]^{n-u+1} \to [0,1]$ as follows:

$$W(x_1, \ldots, x_{n-u+1}) = \mathbb{P}\left[f(\boldsymbol{p}) < r \mid \forall k \in A : p_{ik} = x_k\right].$$

Note that the set $D = \{\boldsymbol{p} : f(\boldsymbol{p}) < r\}$ is a non-decreasing set, since f is a non-increasing function. Therefore, using the result of item 3 of Lemma A.1 and the fact that the p-values within each row are independent, we obtain that $W(x_1, \ldots, x_{n-u+1})$ is non-decreasing in each of its arguments. Note that for any $u \in [0,1]$,

$$\mathbb{P}\left[f(\boldsymbol{p}) < r \mid P_i^{1/(n-u+1)}[A] = u\right] = \mathbb{E}\left[W(\boldsymbol{p}_i^A) \mid P_i^{1/(n-u+1)}[A] = u\right], \quad \text{(A.31)}$$

where $\boldsymbol{p}_i^A$ is the vector of p-values for the hypotheses in the set $\{H_{ij}, j \in A\}$. We shall now use the following theorem of [20], addressing random variables with densities which are Polya frequency functions of order 2 (PF_2):

Theorem A1 ([20])**.** *Let $X_1, \ldots, X_n$ be n independent random variables with PF_2 densities $r_1(x), \ldots, r_n(x)$ respectively, and let $H(x_1, \ldots, x_n)$ be a real measurable function on Euclidean n-space which is non-decreasing in each of its arguments. Then $\mathbb{E}\left[H(X_1, \ldots, X_n) \mid \sum_{i=1}^n X_i = y\right]$ is a non-decreasing function of y.*

Note that when $P_i^{1/(n-u+1)}[A]$ is based on either Fisher's or Stouffer's methods, $P_i^{1/(n-u+1)}[A] = G(\sum_{j \in A} g(p_{ij}))$, where $G(\cdot)$ and $g(\cdot)$ are certain strictly increasing functions. For Fisher's method, $G(x) = \mathbb{P}\left[\chi^2_{2(n-u+1)} \geq -2x\right]$ and $g(x) = \log(x)$, while for Stouffer's method, $G(x) = \Phi(x/\sqrt{n-u+1})$ and $g(x) = \Phi^{-1}(x)$. Recall that A consists only of true null hypotheses, therefore, according to our assumptions, the p-values in the set $\{p_{ij}, j \in A\}$ are independent and uniform on $[0,1]$. Let us consider the random variables in the set $\{g(p_{ij}), j \in A\}$. These random variables are independent. For each $j \in A$, $g(p_{ij})$ is distributed as $\log(U)$ for Fisher's method, and $g(p_{ij})$ is distributed as $\Phi^{-1}(U)$ for Stouffer's method, where $U \sim U[0,1]$. Note that $-\log(U)$ has an exponential distribution, while $\Phi^{-1}(U)$ has a standard normal distribution. Therefore, for each $j \in A$, $g(p_{ij})$ has a PF_2 density (see [20]). We shall consider the general case, where

$$P_i^{1/(n-u+1)}[A] = G\left[\sum_{j \in A} g(p_{ij})\right], \quad \text{(A.32)}$$

where both $G(\cdot)$ and $g(\cdot)$ are strictly increasing functions, and the variables in the set $\{g(p_{ij}), j \in A\}$ are independent variables with PF_2 densities. As shown

above, this case covers both Fisher's and Stouffer's combining functions. In this case the inverse functions $G^{-1}(\cdot)$ and $g^{-1}(\cdot)$ exist, and they are also strictly increasing. We define $H(x_1, \ldots, x_n) = W(g^{-1}(x_1), \ldots, g^{-1}(x_n))$. Obviously, $H(g(x_1), \ldots, g(x_n)) = W(x_1, \ldots, x_n)$. Since $W(x_1, \ldots, x_n)$ is non-decreasing in each of its arguments, and g^{-1} is strictly increasing, $H(x_1, \ldots, x_n)$ is also non-decreasing in each of its arguments. Let $g(\boldsymbol{p}_i^A)$ be the vector obtained by applying the function g on each of its entries. Based on (A.31) and (A.32), we obtain

$$\mathbb{P}\left[f(\boldsymbol{p}) < r \,|\, P_i^{1/(n-u+1)}[A] = u\right] =$$

$$\mathbb{E}\left[W(\boldsymbol{p}_i^A) \,|\, G\left\{\sum_{j \in A} g(p_{ij})\right\} = u\right] =$$

$$\mathbb{E}\left[H(g(\boldsymbol{p}_i^A)) \,|\, \sum_{j \in A} g(p_{ij}) = G^{-1}(u)\right]. \tag{A.33}$$

According to Theorem A1, the expression in (A.33) is non-decreasing in $G^{-1}(u)$, and since G^{-1} is an increasing function, this expression is non-decreasing in u. Thus we have proved that

$$\mathbb{P}\left[f(\boldsymbol{p}) < r \,|\, P_i^{1/(n-u+1)}[A] = u\right]$$

is non-decreasing in u. As shown above, this completes the proof.

Proof of item 2. Assume that $P_i^{u/n}$ is connected to a multiple testing procedure (in the sense of (4.1)), which satisfies condition (c) of item 2 of Theorem 4.1. The meta-analysis setting can be viewed as a group setting of Section 3.1, where the n hypotheses for each feature constitute a group. According to Lemma A.1, the p-values are conditionally positively dependent across groups, and are independent within each group, in the sense of items D3 and D1 in Section 4. Therefore, according to item 2 of Lemma 6.1, the pair $(P_i^{u/n}, h(\boldsymbol{p}_{(-i)\cdot}))$ satisfies the dependency control condition with the identity shape function $\beta(x) = x$.

Appendix B: Example addressing the PRDS property

The following example was presented by Prof. Abba Krieger in his invited talk in a conference in December 2018 in honor of Prof. Yoav Benjamini's 70th birthday. Let X_1, X_2 be independent standard uniform random variables. Let X_3 be a random variable which depends on (X_1, X_2) in the following way. The conditional distribution of X_3 given that $X_1 = x_1, X_2 = x_2$ is defined as follows, according to the colored areas in Figure B.1:

$$\begin{cases} U(0, \frac{1}{2}) & \text{if } (x_1, x_2) \text{ belongs to the red area} \\ U(\frac{1}{2}, 1) & \text{if } (x_1, x_2) \text{ belongs to the green area} \\ U(0, 1) & \text{otherwise} \end{cases}$$

It is easy to see that the vector (X_1, X_2, X_3) satisfies the PRDS property on the subset $\{1, 2\}$. However, the vector $(X_3, f_{\text{Simes}}(X_1, X_2))$, which can be written as $(f_{\text{Simes}}(X_3), f_{\text{Simes}}(X_1, X_2))$, does not satisfy the PRDS property with respect to $f_{\text{Simes}}(X_1, X_2)$, because

$$\mathbb{P}\left[X_3 \geq 0.5 \mid f_{\text{Simes}}(X_1, X_2) = 0.6\right] = 0.5,$$

while

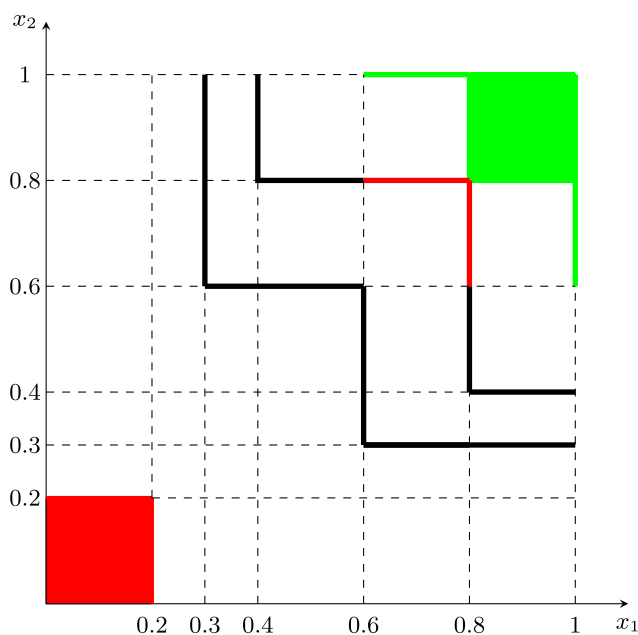$$\mathbb{P}\left[X_3 \geq 0.5 \mid f_{\text{Simes}}(X_1, X_2) = 0.8\right] = \frac{1}{3}.$$

FIG B.1. *An auxiliary figure for the example in Appendix B. The top curve corresponds to the values of (x_1, x_2) for which $f_{Simes}(x_1, x_2) = 0.8$, while the bottom curve corresponds to the values of (x_1, x_2) for which $f_{Simes}(x_1, x_2) = 0.6$.*

Acknowledgment

References

[1] BENJAMINI, Y. and BOGOMOLOV, M. (2014). Selective inference on multiple families of hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 297–318.

[2] BENJAMINI, Y. and HELLER, R. (2007). False discovery rates for spatial signals. *J. Amer. Statist. Assoc.* **102** 1272–1281.

[3] BENJAMINI, Y. and HELLER, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222.

[4] BENJAMINI, Y., HELLER, R. and YEKUTIELI, D. (2009). Selective inference in complex research. *Philos Transact A Math Phys Eng Sci* **367** 4255-4271.

[5] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57** 289–300.

[6] BENJAMINI, Y. and HOCHBERG, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Statist.* **24** 407–418.

[7] BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507.

[8] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29** 1165–1188.

[9] BENJAMINI, Y. and YEKUTIELI, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.* **100** 71–93.

[10] BIRNBAUM, A. (1954). Combining independent tests of significance. *Journal of the American Statistical Association* **49** 559–574.

[11] BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2020). Post hoc confidence bounds on false positives using reference families. *Annals of Statistics* **48** 1281–1303.

[12] BLANCHARD, G. and ROQUAIN, E. (2008). Two simple sufficient conditions for FDR control. *Electronic Journal of Statistics* **2** 963–992.

[13] BLANCHARD, G. and ROQUAIN, E. (2009). Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research* **10** 2837–2871.

[14] BOGOMOLOV, M. and HELLER, R. (2013). Discovering findings that replicate from a primary study of high dimension to a follow-up study. *Journal of the American Statistical Association* **108** 1480-1492.

[15] BOGOMOLOV, M. and HELLER, R. (2022). Replicability across multiple studies. *arXiv preprint arXiv:2210.00522*.

[16] BOGOMOLOV, M., PETERSON, C. B., BENJAMINI, Y. and SABATTI, C. (2020). Hypotheses on a tree: new error rates and testing strategies. *Biometrika*.

[17] BRZYSKI, D., PETERSON, C. B., SOBCZYK, P., CANDES, E. J., BOGDAN, M. and SABATTI, C. (2017). Controlling the rate of GWAS false

discoveries. *Genetics* **205** 61–75.

[18] DICKHAUS, T., HELLER, R. and HOANG, A.-T. (2021). Multiple testing of partial conjunction null hypotheses with conditional *p*-values based on combination test statistics. *arXiv preprint arXiv:2110.06692*.

[19] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics* **32** 962-994.

[20] EFRON, B. (1965). Increasing properties of Polya frequency function. *The Annals of Mathematical Statistics* 272–279.

[21] FINNER, H., ROTERS, M. and STRASSBURGER, K. (2017). On the Simes test under dependence. *Statistical Papers* **58** 775–789.

[22] FISHER, R. A. (1934). Statistical methods for research workers. *Statistical methods for research workers.* **5th Ed**.

[23] FOYGEL BARBER, R. and RAMDAS, A. (2016). The p-filter: multi-layer false discovery rate control for grouped hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* MR3689317

[24] GENOVESE, C. R., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p-value weighting. *Biometrika* **93** 509–524.

[25] GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *Journal of the American Statistical Association* **101** 1408–1417.

[26] GOEMAN, J. J., HEMERIK, J. and SOLARI, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *The Annals of Statistics* **49** 1218–1238.

[27] GOEMAN, J. J., MEIJER, R. J., KREBS, T. J. and SOLARI, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika* **106** 841–856.

[28] GOEMAN, J. J. and SOLARI, A. (2011). Multiple testing for exploratory research. *Statistical Science* **26** 584–597.

[29] GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. M. (2020). Safe testing. In *2020 Information Theory and Applications Workshop (ITA)* 1–54. IEEE.

[30] HELLER, R., BOGOMOLOV, M. and BENJAMINI, Y. (2014). Deciding whether follow-up studies have replicated findings in a preliminary large-scale omics study. *Proceedings of the National Academy of Sciences* **111** 16262–16267.

[31] HELLER, R., CHATTERJEE, N., KRIEGER, A. and SHI, J. (2018). Postselection inference following aggregate level hypothesis testing in large-scale genomic data. *Journal of the American Statistical Association* **113** 1770–1783. MR3902245

[32] HELLER, R. and YEKUTIELI, D. (2014). Replicability analysis for genome-wide association studies. *The Annals of Applied Statistics* **8** 481–498.

[33] HOANG, A.-T. and DICKHAUS, T. (2021). Combining independent p-values in replicability analysis: A comparative study. *arXiv preprint arXiv:2104.13081*.

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

[35] HOCHBERG, Y. and LIBERMAN, U. (1994). An extended Simes' test. *Statis-*

tics & Probability Letters **21** 101–105.

[36] HOMMEL, G. (1983). Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical Journal* **25** 423–430.

[37] HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75** 383–386.

[38] KATSEVICH, E. and RAMDAS, A. (2020). Simultaneous high-probability bounds on the false discovery proportion in structured, regression and online settings. *Annals of Statistics* **48** 3465–3487.

[39] KATSEVICH, E., SABATTI, C. and BOGOMOLOV, M. (2021). Filtering the rejection set while preserving false discovery rate control. *Journal of the American Statistical Association* **0** 1-27.

[40] LOUGHIN, T. M. (2004). A systematic comparison of methods for combining p-values from independent tests. *Computational statistics & data analysis* **47** 467–485.

[41] MARCUS, R., ERIC, P. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660.

[42] MEIJER, R. J., KREBS, T. J. and GOEMAN, J. J. (2019). Hommel's procedure in linear time. *Biometrical Journal* **61** 73–82.

[43] OWEN, A. B. (2009). Karl Pearson's meta-analysis revisited. *The Annals of Statistics* **37** 3867–3892.

[44] PEARSON, K. (1934). On a New Method of Determining "Goodness of Fit". *Biometrika* **26** 425–442.

[45] PENNY, W. and FRISTON, K. (2003). Mixtures of general linear models for functional neuroimaging. *IEEE Transactions on Medical Imaging* **22** 504-514.

[46] PETERSON, C. B., BOGOMOLOV, M., BENJAMINI, Y. and SABATTI, C. (2016). Many phenotypes without many false discoveries: error controlling strategies for multitrait association studies. *Genet. Epidemiol.* **40** 45–56.

[47] RAMDAS, A., CHEN, J., WAINWRIGHT, M. J. and JORDAN, M. I. (2019). A sequential algorithm for false discovery rate control on directed acyclic graphs. *Biometrika* **106** 69–86.

[48] RAMDAS, A. K., BARBER, R. F., WAINWRIGHT, M. J. and JORDAN, M. I. (2019). A unified treatment of multiple testing with prior knowledge using the p-filter. *The Annals of Statistics* **47** 2790–2821.

[49] SARKAR, S. K. (1998). Some probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture. *The Annals of Statistics* 494–504.

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

[51] STOREY, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** 479–498.

[52] STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical*

Society: Series B (Statistical Methodology) **66** 187–205.

[53] Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. and Williams Jr, R. M. (1949). *The American Soldier: Adjustment During Army Life, Vol. 1. Studies in Social Psychology in World War II.* Princeton Univ. Press.

[54] Tamhane, A. C., Liu, W. and Dunnett, C. W. (1998). A generalized step-up-down multiple test procedure. *Canadian Journal of Statistics* **26** 353–363.

[55] Tippett, L. H. C. (1931). The methods of statistics. *The Methods of Statistics.*

[56] Vovk, V. and Wang, R. (2019). E-values: Calibration, combination, and applications. *Forthcoming in the Annals of Statistics.*

[57] Vovk, V. and Wang, R. (2020). Combining p-values via averaging. *Biometrika* **107** 791–808.

[58] Wang, J., Gui, L., Su, W. J., Sabatti, C. and Owen, A. B. (2022). Detecting multiple replicating signals using adaptive filtering procedures. *The Annals of Statistics* **50** 1890–1909. MR4474476

[59] Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences* **116** 1195–1200.

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

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

[62] Yekutieli, D., Reiner-Benaim, A., Benjamini, Y., Elmer, G. I., Kafkafi, N., Letwin, N. E. and Lee, N. H. (2006). Approaches to multiplicity issues in complex research in microarray analysis. *Statist. Neerlandica* **60** 414–437.