Received: 16 October 2018 | Revised: 22 October 2019 | Accepted: 22 January 2020

DOI: 10.1111/sjos.12453

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Post hoc false positive control for structured hypotheses

Guillermo Durand[1] | **Gilles Blanchard**[2] | **Pierre Neuvial**[3] | **Etienne Roquain**[1]

[1]Laboratoire de probabilités Statistique et Modélisation, LPSM, Sorbonne Université, France

[2]Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS, Université Paris-Saclay, France

[3]Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS, UPS IMT, France

Correspondence
Roquain Etienne, Sorbonne Université, LPSM, Campus Pierre et Marie Curie Case courrier 158, 4 Place Jussieu 75252 Paris Cedex 05, France.
Email: etienne.roquain@upmc.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-16-CE40-0019, Grant/Award Number: ANR-17-CE40-0001; Deutsche Forschungsgemeinschaft, Grant/Award Number: FOR-1735, Grant/Award Number: SFB-1294; GDR ISIS, Grant/Award Number: TASTY

Abstract

In a high-dimensional multiple testing framework, we present new confidence bounds on the false positives contained in subsets S of selected null hypotheses. These bounds are post hoc in the sense that the coverage probability holds simultaneously over all S, possibly chosen depending on the data. This article focuses on the common case of structured null hypotheses, for example, along a tree, a hierarchy, or geometrically (spatially or temporally). Following recent advances in post hoc inference, we build confidence bounds for some prespecified forest-structured subsets and deduce a bound for any subset S by interpolation. The proposed bounds are shown to improve substantially previous ones when the signal is locally structured. Our findings are supported both by theoretical results and numerical experiments. Moreover, our bounds can be obtained by an algorithm (with complexity bilinear in the sizes of the reference hierarchy and of the selected subset) that is implemented in the open-source `R` package `sansSouci` available from https://github.com/pneuvial/sanssouci, making our approach operational.

KEYWORDS

Dvoretzky Kiefer Wolfowitz inequality, forest structure, multiple testing, post hoc bound, selective inference

1 INTRODUCTION

1.1 Background

Modern statistical data analysis often involves asking many questions of interest simultaneously, possibly using the data repeatedly, as long as the user feels that this could provide additional information. To avoid selection bias due to various forms of data snooping, specific strategies can be proposed to take into account the procedure as a whole and be investigated as to the statistical guarantees they provide. This problem is often referred to as selective (or post hoc) inference, a long standing research field, with a recent renewal of interest. To name only a few studies, inference on the observed selection set can be done by controlling the false coverage rate control (Benjamini & Bogomolov, 2014; Benjamini & Yekutieli, 2005; Weinstein & Ramdas, 2019) or by controlling a criterion conditional to a specific initial selection step (see the series of works Choi, Taylor, & Tibshirani, 2017; Fithian, Sun, & Taylor, 2017; Taylor & Tibshirani, 2015, 2018; Tibshirani, Taylor, Lockhart, & Tibshirani, 2016). In other studies, the selection step is based on sample splitting (see Bühlmann & Mandozzi, 2014; Cox, 1975; Dezeure, Bühlmann, Meier, & Meinshausen, 2015), which is another way to tackle selective inference by explicitly avoiding data reuse. Another strategy is to provide a statistical guarantee uniformly over all selected sets. A historical reference is the work of Scheffé (1953) (see also Scheffé, 1959, p. 69; also Bachoc, Blanchard, & Neuvial, 2018; Bachoc, Leeb, & Pötscher, 2019; Berk, Brown, Buja, Zhang, & Zhao, 2013 for recent developments on this issue).

We follow in this article the aim of establishing confidence bounds on the number of false positives specifically in the multiple testing framework, simultaneously over all possible sets of selected hypotheses, so that the user can resort to arbitrary use of the data before selecting a set of hypotheses, consider several concurrent selections, change their mind, and so on. Such a post hoc setting has been considered in particular by Genovese and Wasserman (2006), Goeman and Solari (2011), and Blanchard, Neuvial, and Roquain (in press).

We are more specifically interested in the common situation where the set of null hypotheses has a geometrical structure, such as when hypotheses are organized along a tree or a hierarchy or are indexed by an underlying metric space (for instance spatially or temporally). We expect correspondingly the false hypotheses to be clustered at some contiguous positions in this structure, possibly at different scales, and the set of selected hypotheses to be at least partially dictated by the structure; if this is the case, we aim at deriving confidence bounds that will be sharper than unstructured approaches.

In the context of multiple testing methods, hierarchically structured null hypotheses have appeared in various application settings and statistical models. In bioinformatics applications, there often exists prior knowledge about the units under analysis (such as genes, SNPs, proteins), summarized as an ontology (which can take several forms) and it is common to apply a hierarchical clustering algorithm resulting in a binary hierarchy. This structure can be known from prior knowledge or data (Meinshausen, 2008), from independent data (Goeman & Mansmann, 2008), or be data driven from the same data (Kim, Roquain, & van de Wiel, 2010). In the situation where the units have follow spatial, temporal, or other topological structure, the use of hierarchies is also common to get a suitable multiscale representation of this structure (Blanchard & Geman, 2005; Ehm, Kornmeier, & Heinrich, 2010). The goal of controlling the false discovery rate (FDR) in such a hierarchically structured context, and for a specific selection procedure (thus not post hoc) has been considered by Yekutieli (2008), Guo, Lynch, and Romano (2018), Ramdas, Chen, Wainwright, and Jordan (2019), and among others.

Concerning the incorporation of structural information into post hoc bounds, this line of research has been followed in Meijer and Goeman (2015) and Meijer, Krebs, and Goeman (2015) by using the sequential rejection principle of Goeman and Solari (2010). In particular, the method in Meijer et al. (2015), whose goal inspired the present work, deals with geometrically structured null hypotheses along space or time and shows that incorporating such an external information can substantially improve the detection of signal and thus can increase the accuracy of the resulting post hoc bound.

1.2 | Confidence bounds and the selection effect

Let us motivate our work by the following simple example. Suppose we have at hand m independent variables $X_i \sim \mathcal{N}(\mu_i, 1)$, for some unknown "signal" vector $\mu \in \mathbb{R}_+^m$ and we denote $\mathcal{H}_0 = \{i : \mu_i = 0\}$ the set of coordinates where the signal is zero. The user's goal is to find subset of coordinates with nonzero signal. For a subset of selected coordinates $S \subseteq \{1, \dots, m\}$, the number of false positives in S is given by $|S \cap \mathcal{H}_0|$. In practice, it is crucial to give an upper bound of this quantity in order to quantify the relevance of the elements of S. In the case where the set S is *fixed in advance*, we can prove that the following quantity

$$V(S) = \left\lfloor \left(\left(2 \sum_{i \in S} \mathbf{1}\{X_i \leq 0\} + \frac{\log(1/\alpha)}{2} \right)^{1/2} + \left(\frac{\log(1/\alpha)}{2} \right)^{1/2} \right)^2 \right\rfloor \wedge |S| \tag{1}$$

is a $(1-\alpha)$-upper bound on $|S \cap \mathcal{H}_0|$, that is, $\mathbb{P}(|S \cap \mathcal{H}_0| \leq V(S)) \geq 1 - \alpha$ (in (1), $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x). This bound will be justified further on (see Section 4). The rationale is that since a $\mathcal{N}(0,1)$ distribution is symmetric, the quantity $2\sum_{i \in S} \mathbf{1}\{X_i \leq 0\}$ is an upper estimate of $|S \cap \mathcal{H}_0|$, the other terms being deviation terms depending on α.

Figure 1 reports the value of this bound on one particular simulation run with $m = 100$ and $S = 11 : 35 (= \{11, \dots, 35\})$. Here, the obtained bound is trivial, that is, $V(S) = 25$. Starting from this, let us seek for another set S with more "signal," that is, with a smaller bound. In view of (1), choosing the set S consisting of the 25 highest realizations will give a much smaller bound $V(S) = 4$, which means that S should contain at least 21 significant elements. In this example, despite the fact that $\mu = 0$ (meaning that there is no signal at all in the data), the event $V(S) = 4$ occurs with probability very close to 1. This illustrates that the bound (1) does not account for the selection effect, that is, is only valid for set S that are deterministic, not driven by the data as the top 25 realizations.

1.3 | Post hoc bounds

To provide guarantee for any data-driven set S that can stem from any procedure followed by the user, one should provide a confidence bound uniformly valid over all $S \subseteq \mathbb{N}_m = \{1, \dots, m\}$, which is called a post hoc bound.

The general framework is as follows : let us observe a random variable $X \sim P$, P belonging to some model $\mathcal{P}$, for which m null hypotheses $H_{0,i} \subseteq \mathcal{P}$, $i \in \mathbb{N}_m$ are under investigation for

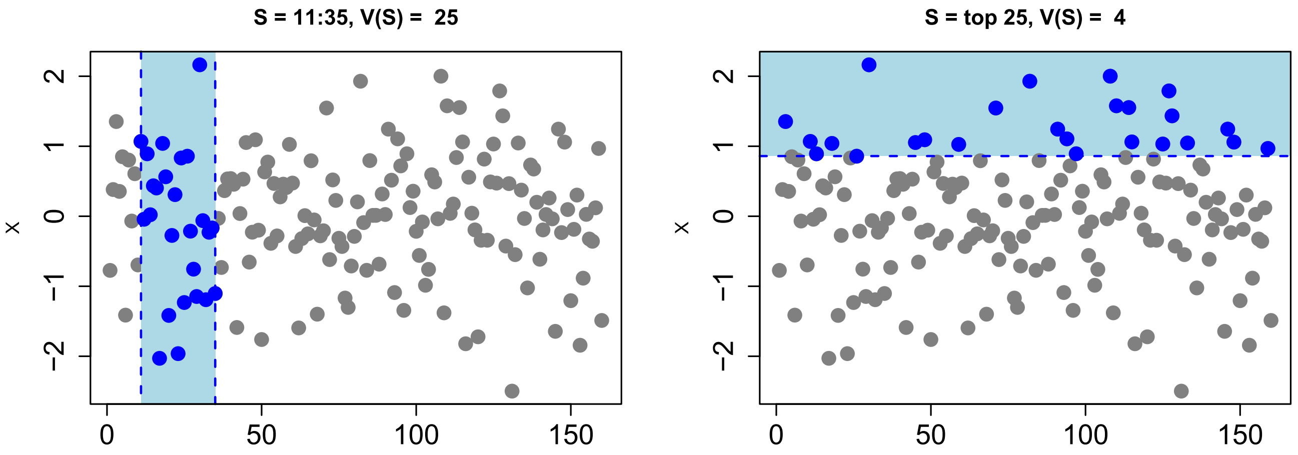

FIGURE 1 Confidence bound $V(S)$ (not post hoc) given by (1) for S in two cases ($\alpha = 0.1$). Left: valid use of this bound. Right : invalid "data snooped" use of this bound. In both pictures, X has been generated as $m = 160$ i.i.d. $\mathcal{N}(0, 1)$ [Color figure can be viewed at wileyonlinelibrary.com]

P, the aim is to build a function $V(X, \cdot) : S \subseteq \mathbb{N}_m \mapsto V(X, S) \in \mathbb{N}$ (denoted by $V(S)$ for short) satisfying

$$\forall P \in \mathcal{P}, \quad \mathbb{P}_{X\sim P}\left(\forall S \subseteq \mathbb{N}_m, |S \cap \mathcal{H}_0(P)| \leq V(S)\right) \geq 1 - \alpha, \tag{2}$$

where $\mathcal{H}_0(P) = \{i \in \mathbb{N}_m \ : \ P \text{ satisfies } H_{0,i}\}$ is the set of true null hypotheses. The bound $V(\cdot)$ will be referred to as a post hoc bound throughout this article.

In (2), the number $|S \cap \mathcal{H}_0(P)|$ thus corresponds to the number of false positives of the procedure rejecting the nulls $H_{0,i}$ with $i \in S$. Note that (2) directly implies that $V(S)/|S|$ is a bound for the celebrated false discovery proportion $|S \cap \mathcal{H}_0(P)|/|S|$ of this procedure (still holding uniformly over the sets $S \subseteq \mathbb{N}_m, S \neq \emptyset$). Obviously, since the trivial bound $V(S) = |S|, S \subseteq \mathbb{N}_m$, always satisfies (2), we implicitly aim at finding post hoc bounds $V(S)$ as small as possible, and possibly for "many interesting S."

The problem of constructing post hoc bounds has been first tackled specifically in the case where the selection sets S are of the form of p-value level sets: $\{i : p_i(X) \leq t\}, t \in [0, 1]$, where each $p_i(X)$ is a p-value for the null hypothesis $H_{0,i}$, $1 \leq i \leq m$. The resulting bounds are often referred to as confidence envelopes (see Genovese & Wasserman, 2004; Meinshausen, 2006). Later, Genovese and Wasserman (2006) and Goeman and Solari (2011) proposed to extend this approach to arbitrary subsets S, by using a methodology based on performing $2^m - 1$ local tests (one for each intersection hypothesis), with a possible complexity reduction by using shortcuts. In particular, the approach of Goeman and Solari (2011) extensively relies on the closed testing principle, which was introduced by Marcus, Eric, and Gabriel (1976). More recently, Blanchard et al. (in press) (BNR below) have proposed a flexible methodology that adjusts the complexity of the bound by way of a reference family: the post hoc bound is based on a family $\mathfrak{R} = ((R_k(X), \zeta_k(X))_{k\in\mathcal{K}}$ (R_k, ζ_k for short), where $\mathcal{K}$ is a finite index set, $R_k \subseteq \mathbb{N}_m$ (and $R_k \neq R_{k'}$ if $k \neq k'$), $\zeta_k \in \mathbb{N}$, that satisfies the following joint error rate (JER) control:

$$\forall P \in \mathcal{P}, \quad \mathbb{P}_{X\sim P}\left(\forall k \in \mathcal{K}, |R_k \cap \mathcal{H}_0(P)| \leq \zeta_k\right) \geq 1 - \alpha. \tag{3}$$

In this article, we will focus on the case where the sets R_k are deterministic (not depending on X), while the bounds $\zeta_k(X)$ are random, and chosen so that (3) holds. This approach is sensible when the set of null hypotheses has some (a priori known) structure, which is reflected in the fixed choice of the regions R_k: they will typically be clusters of hypotheses that are “close” in the underlying structure; possibly at different scales. This rationale will be discussed in more detail in the next sections. An important difference between (2) and (3) is that S in (2) is let arbitrary and typically chosen by the user, whereas R_k, ζ_k in (3) is part of the methodology and is chosen by the statistician to make (3) hold. Once the reference family is fixed, a post hoc bound is obtained from (3) simply by interpolation, by exploiting the constraints that the event appearing in (3) imposes to the unknown set $\mathcal{H}_0(P)$, namely, that it is a subset A satisfying the property “$\forall k \in \mathcal{K}, |R_k \cap A| \leq \zeta_k$”:

$$V^*_{\mathfrak{R}}(S) = \max_{A \in \mathcal{A}(\mathfrak{R})} \{|S \cap A|\}, \quad \mathcal{A}(\mathfrak{R}) = \{A \subseteq \mathbb{N}_m \mid \forall k \in \mathcal{K} : |R_k \cap A| \leq \zeta_k\}, \quad S \subseteq \mathbb{N}_m. \tag{4}$$

Hence, if (3) holds, then $V = V^*_{\mathfrak{R}}$ satisfies (2). This post hoc bound will be referred to as the *optimal bound* (relative to a given reference family).

1.4 | Toy example

Let us first mention that the numerical results of this section can be reproduced online at `https://pneuvial.shinyapps.io/posthoc-bounds_ordered-hypotheses/` (see Appendix C.1 for more details).

To explain the rationale behind our approach, let us detail how it behaves in the simple example mentioned in Section 1.2. Again, consider m independent variables $X_i \sim \mathcal{N}(\mu_i, 1)$, but this time assume that there is some signal present, that is, $\mu \neq 0$. One classical post hoc bound is given by the Simes bound V_{Simes} (see (9) below), which is a common benchmark that has the correct coverage in that case, that is, satisfies (3). From an intuitive point of view, this bound will be accurate for data driven S that are level sets, such as the k_0 largest X_i. However, we are interested in this article in the common situation where the unknown signal μ (and thus the corresponding null hypotheses) has some underlying structure, for example, that the nonzero entries are clustered at some contiguous positions in $\mathbb{N}_m$. In that case, our approach consists in using some fixed, deterministic, reference sets R_k that are also contiguous and are typically defined in either of the two following manners :

- Partition reference family : $R^{\text{part}}_{\text{k}} = \{(k-1)s + 1, \dots, ks\}, k \in \mathcal{K}^{\text{part}} = \{1, \dots, m/s\}, \zeta^{\text{part}}_{\text{k}}$ as in (1) with α divided by $K^{\text{part}} = |\mathcal{K}^{\text{part}}| = m/s$;
- Tree reference family : assuming that $m = 2^i s$ for some $i \in \mathbb{N}$: the sets $R^{\text{tree}}_{\text{k}}$ are the nodes of a perfect binary tree whose leaves are elements of the above partition, and $\zeta^{\text{tree}}_{\text{k}}$ as in (1) with $S = R^{\text{part}}_{\text{k}}$ and α divided by $K^{\text{tree}} = |\mathcal{K}^{\text{tree}}| = 2^{i+1} - 1$. One possible indexation is $\mathcal{K}^{\text{tree}} = \{1, \dots, K^{\text{tree}}\}$.

In both reference families, the scale parameter s (we assume s divides m) has to be fixed in advance. By using a union bound argument, (3) is satisfied by both reference families, which, according to (4), gives rise to the post hoc bounds V_{part} and V_{tree}, respectively. We will prove (see Corollary 1 below) that $V_{\text{part}}(S) = \sum_{k=1}^{m/s} \zeta^{\text{part}}_k \wedge |S \cap R_k|$ and that $V_{\text{tree}}(S)$ can be computed as follows:

$$V_{\text{tree}}(S) = \min_{\substack{Q\subseteq \mathcal{K}^{\text{tree}}\\ \text{s.t. } \left\{R_k^{\text{tree}}, k\in Q\right\}\\ \text{is a partition of } \mathbb{N}_m}} \left\{ \sum_{k\in Q} \zeta_k^{\text{tree}} \wedge |S\cap R_k^{\text{tree}}| \right\}. \tag{5}$$

Since the number of partitions to be explored is large, formula (5) might be time-consuming in practice and we propose a algorithm of complexity $O(|S|\,|K^{\text{tree}}|)$ to compute $V_{\text{tree}}(S)$ (see Algorithm 1 below). The idea is to compute the bound recursively by selecting the "best" partition, possibly using nodes at different depths in the tree. Note that our algorithm is not only valid for a perfect binary tree structure, but for any reference sets that have a forest structure (see Definition 1 below).

The bounds $V_{\text{part}}(S)$ and $V_{\text{tree}}(S)$ for $s = 20$ are illustrated in Figures 2 and 3, and the data are given in Figure 4, for $S = \{1, \dots, 45\}$, for some realization generated in the case where $\mu_i = \overline{\mu}$ for $i \in \{1, \dots, 40\}$ and $\mu_i = 0$ otherwise, so that $\mathcal{H}_0 = \{41, \dots, 180\}$. We easily read in Figure 2 that $V_{\text{part}}(S) = 21 = 8 + 8 + 5 = \zeta_1^{\text{part}} + \zeta_2^{\text{part}} + |S \cap R_3^{\text{part}}|$. As for $V_{\text{tree}}(S)$, according to (5), we should optimize over all possible partitions formed by the nodes. Since the only informative $\zeta_{\text{k}}^{\text{tree}}$'s are ζ_4^{tree}, ζ_8^{tree}, and ζ_9^{tree}, the result does not depend on the chosen partition covering the set $\{41, \dots, 160\}$ so that we should only choose a partition covering the set $\{1, \dots, 40\}$, that is, either $\{R_4^{\text{tree}}\}$, or $\{R_8^{\text{tree}}, R_9^{\text{tree}}\}$. The first partition gives a bound $15 = 10 + 5$ and the second gives $25 = 10 + 10 + 5$. Hence, $V_{\text{tree}}(S) = 15$.

This simple example already shows that taking into account the structure of the signal leads to an improvement of the bounds. This is further supported by numerical experiments in Section 5 and by a theoretical study in Section 4.2.

1.5 | Contributions and relation to the literature

While our aim is similar in spirit to Meijer et al. (2015), our method is markedly different, as it relies on the general strategy laid down by BNR, with a specifically structured reference family $R_k, k \in \mathcal{K}$ (see Section 6.1 for a comparison between our approach and the one of Meijer et al., 2015). In addition, the way the method is built here is different than the one proposed in sections 3-6 of BNR: the main focus in BNR is the case of (random) reference sets $R_k = R_k(X)$ that are designed in order to satisfy (3) with $\zeta_k = k - 1$ (thus corresponding to a "joint k-family-wise error rate"). By contrast, in the present work the reference sets R_k are fixed in advance, and the (random) bounds on the number false positives $\zeta_k = \zeta_k(X)$ are designed to satisfy the constraint (3). Again, the rationale behind this approach is that the reference sets R_k can be chosen arbitrarily by the statistician, so that it can accommodate any prespecified structure (reflecting some prior knowledge on the considered problem). Since we are interested in structured signal, we focus on a reference family enjoying a forest structure, meaning that two reference sets are either disjoint or nested.

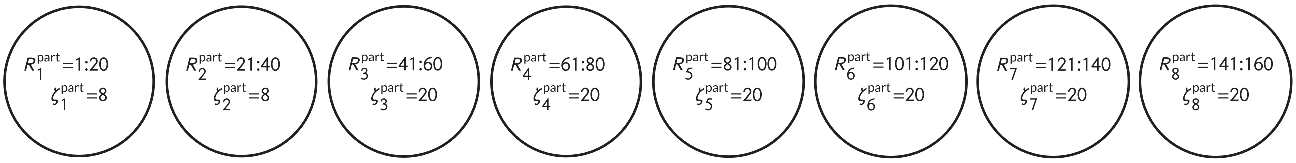

FIGURE 2 Partition reference family with some regular structure (see text). The obtained values of the $\zeta_{\text{k}}^{\text{part}}$ are obtained for the data of Figure 4

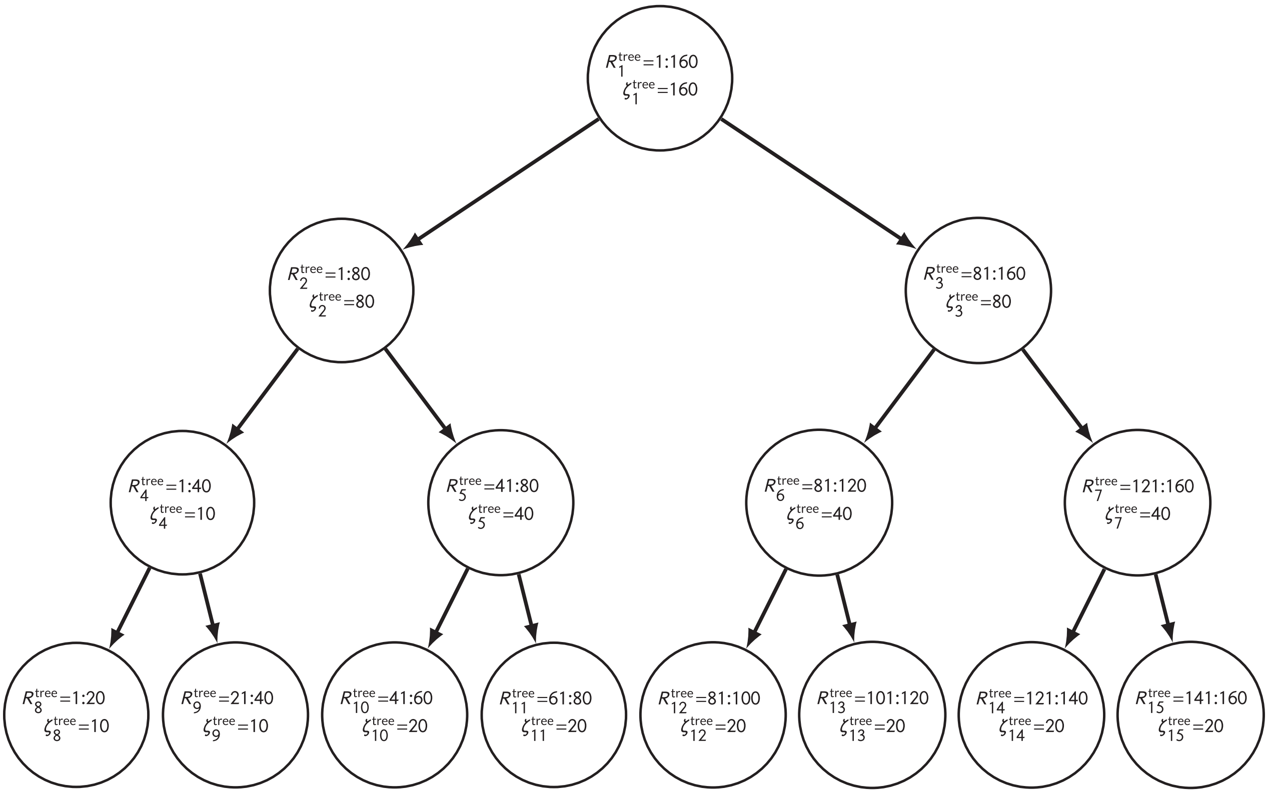

FIGURE 3 Tree reference family for some particular perfect binary structure (see text). The obtained values of the $\zeta_{\text{k}}^{\text{tree}}$ are obtained for the data of Figure 4

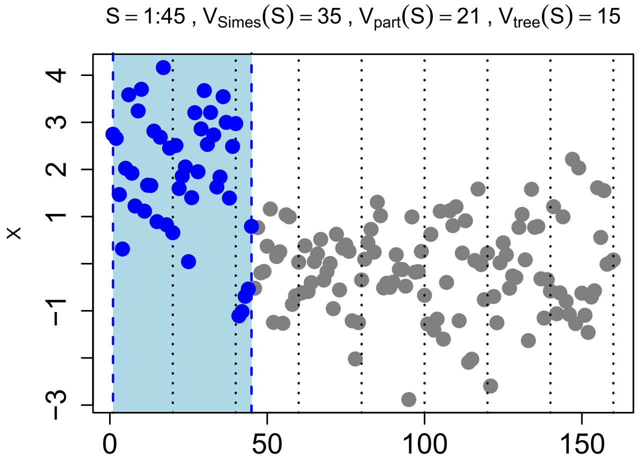

FIGURE 4 Data generated as $m = 160$ i.i.d. $\mathcal{N}(\mu, 1)$, with $\overline{\mu} = 2$, and associated values for the post hoc bounds $V_{\text{Simes}}(S)$ $V_{\text{part}}(S)$ and $V_{\text{tree}}(S)$ (see text). The elements of the partition $R_k^{\text{part}} = \{(k-1)s+1, \dots, ks\}, 1 \le k \le K^{\text{part}}$ for $s = 20$ are separated by dotted vertical lines [Color figure can be viewed at wileyonlinelibrary.com]

The second ingredient of our method is the local bounds $\zeta_k(X)$, that should estimate $|R_k \cap \mathcal{H}_0(P)|$ with a suitable deviation term. While any deviation inequality can be used, we have chosen to focus on the DKW inequality (Dvoretzky, Kiefer, & Wolfowitz, 1956), that has the advantage to be sub-Gaussian. Hence, the uniformity over the range $k \in \mathcal{K}$ can be obtained by a simple union bound without being too conservative.

Using the DKW inequality to obtain a confidence bound for the proportion of null hypotheses is not new (see Genovese & Wasserman, 2004, equation (16) therein; Farcomeni & Pacillo, 2011; Meinshausen, 2006). While our bound is a uniform improvement of the existing version (see Remark 2 for more details), our main innovation is to use the DKW bound in a local manner and to appropriately combine these local bounds to derive an overall post hoc bound. The improvement can be substantial, as illustrated in our numerical experiments.

Let us finally mention the work of Katsevich and Ramdas (in press), which provides FDP bounds along a path driven by a specific multiple testing procedure. They deduce from these particular bounds a fully valid post hoc bounds also by interpolation (see Section 6.2 for more details on this work).

1.6 | Organization of the article

The article is organized as follows: Precise setup and notation are introduced in Section 2. For any reference family with a forest structure, the optimal post hoc bound is computed in Section 3. The calibration of the local bounds ζ_k and of the overall reference family is done in Section 4. This section also includes a theoretical comparison with previous methods, which quantifies formally the amplitude of the improvement induced by the new method. The latter is supported by numerical experiments in Section 5, where a hybrid approach is also introduced to mimic the best between the new approach and the existing Simes bound (the latter being defined in (9)). A discussion is given in Section 6 and the proofs are provided in Appendix A. Additional technical details are postponed to Appendix B, and information to reproduce the numerical experiments of the article are given in Appendix C.

2 | PRELIMINARIES

2.1 | Assumptions

We focus on the common situation where a test statistic $T_i(X)$ is available for each null hypothesis $H_{0,i}$. For $i \in \mathbb{N}_m$, each statistic $T_i(X)$ is transformed into a p-value $p_i(X)$, satisfying the following assumptions:

$$\forall P \in \mathcal{P}, \quad \forall i \in \mathcal{H}_0(P), \quad \forall t \in [0,1], \quad \mathbb{P}_{X\sim P}(p_i(X) \le t) \le t; \qquad \text{(Superunif)}$$

$$\forall P \in \mathcal{P}, \quad \{p_i(X)\}_{i\in\mathcal{H}_0(P)} \text{ is a family of independent } p\text{-values and is independent of } \{p_i(X)\}_{i\in\mathcal{H}_1(P)}, \qquad \text{(Indep)}$$

where $\mathcal{H}_1(P) = \mathbb{N}_m \setminus \mathcal{H}_0(P)$ denotes the indices of the false null hypotheses. Extending our results to the case where (Indep) fails is possible (see the discussion in Section 6.3).

2.2 | Classical post hoc bounds

As argued in BNR, computing the optimal post hoc bound (4) relative to a given reference family $(R_k, \zeta_k)_{k\in\mathcal{K}}$ can be NP-hard, and simpler, more conservative versions can be provided, that is, bounds V such that for all $S \subseteq \mathbb{N}_m$, $V^*_{\mathfrak{R}}(R) \le V(R)$. A simple upper bound for $V^*_{\mathfrak{R}}$ is given by

$$\overline{V}_{\mathfrak{R}}(S) = |S| \wedge \min_{k\in\mathcal{K}} \{\zeta_k + |S \setminus R_k|\}, \quad S \subseteq \mathbb{N}_m. \tag{6}$$

It is straightforward to check that

$$V^*_{\mathfrak{R}}(S) \le \overline{V}_{\mathfrak{R}}(S), \quad S \subseteq \mathbb{N}_m. \tag{7}$$

While this inequality is strict in general, BNR established that it is an equality if the reference family is nested, that is,

$$\mathcal{K} = \{1, \dots, K\} \quad \text{and} \quad R_k \subseteq R_{k+1} \quad \text{for} \;\; 1 \le k \le K-1. \tag{Nested}$$

Condition (Nested) is mild when the sequence ζ_k is nondecreasing, for example, $\zeta_k = k-1$.

A consequence of (7) is that $\overline{V}_{\mathfrak{R}}$ is a post hoc bound in the sense of (2) as soon as the reference family $\mathfrak{R}$ is such that (3) holds. A simple union bound under (Superunif) yields that (3) holds with $\mathfrak{R} = \{(R_1, \zeta_1)\}$, $R_1 = \{i \in \mathbb{N}_m : p_i \le \alpha/m\}$, $\zeta_1 = 0$. This leads to the Bonferroni post hoc bound

$$V_{\mathrm{Bonf}}(S) = \sum_{i \in S} \mathbf{1}\{p_i(X) > \alpha/m\}, \quad S \subseteq \mathbb{N}_m. \tag{8}$$

The more subtle Simes inequality (Simes, 1986), valid under (Superunif) and (Indep), ensures that (3) holds with $\mathfrak{R} = \{(R_k, \zeta_k), 1 \le k \le m\}$, $R_k = \{i \in \mathbb{N}_m : p_i \le \alpha k/m\}$, $\zeta_k = k-1$. This leads to the Simes post hoc bound

$$V_{\mathrm{Simes}}(S) = \min_{1 \le k \le m} \left\{ \sum_{i \in S} \mathbf{1}\{p_i(X) > \alpha k/m\} + k - 1 \right\}, \quad S \subseteq \mathbb{N}_m. \tag{9}$$

As noted in BNR, this bound is identical to the post hoc bound of Goeman and Solari (2011), which will be used as a benchmark in this article.

2.3 | Improved interpolation bound

When the reference family is not nested, inequality (7) can be far too conservative. We introduce the following extension: for a reference family $\mathfrak{R} = (R_k(X), \zeta_k(X))_{k \in \mathcal{K}}$ of cardinality $K = |\mathcal{K}|$,

$$\tilde{V}^q_{\mathfrak{R}}(S) = \min_{Q \subseteq \mathcal{K}, |Q| \le q} \left(\sum_{k \in Q} \zeta_k \wedge |S \cap R_k| + |S \setminus \bigcup_{k \in Q} R_k| \right), \quad 1 \le q \le K, \; S \subseteq \mathbb{N}_m; \tag{10}$$

$$\tilde{V}_{\mathfrak{R}}(S) = \tilde{V}^K_{\mathfrak{R}}(S), \quad S \subseteq \mathbb{N}_m. \tag{11}$$

Obviously, we have $\tilde{V}^1_{\mathfrak{R}} = \overline{V}_{\mathfrak{R}}$ and $\tilde{V}^q_{\mathfrak{R}}$ is nonincreasing in q. The following result shows that these bounds are all conservative versions of $V^*_{\mathfrak{R}}$.

Lemma 1. *For any reference family $\mathfrak{R}$, we have*

$$V^*_{\mathfrak{R}}(S) \le \tilde{V}_{\mathfrak{R}}(S) \le \tilde{V}^q_{\mathfrak{R}}(S) \le \overline{V}_{\mathfrak{R}}(S), \quad 1 \le q \le K, \quad S \subseteq \mathbb{N}_m. \tag{12}$$

In particular, if $\mathfrak{R}$ is such that (3) holds, then $\tilde{V}_{\mathfrak{R}}$ is a post hoc bound in the sense of (2).

Lemma 1 is proved in Section A.1. The inequality $V^*_{\mathfrak{R}}(S) \le \tilde{V}_{\mathfrak{R}}(S)$ in (12) is strict in general (see Example 1). As we will show in the next section, this relation is nevertheless an equality when $\mathfrak{R}$ has a specific forest structure, which makes $\tilde{V}_{\mathfrak{R}}$ a particularly interesting bound.

Example 1. Let $m = 4$, $K = 3$, $R_1 = \{1, 2, 4\}$, $R_2 = \{2, 3, 4\}$, $R_3 = \{1, 3, 4\}$. Consider the event where $\zeta_1(X) = \zeta_2(X) = \zeta_3(X) = 1$. For $S = \mathbb{N}_4$, we easily check that $V^*_{\mathfrak{R}}(S) = 1$ and $\tilde{V}_{\mathfrak{R}}(S) = 2$.

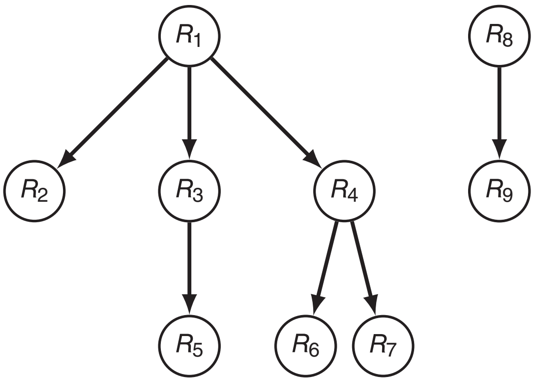

FIGURE 5 Graph corresponding to the reference family given in Example 2

3 | POST HOC BOUND FOR FOREST STRUCTURED REFERENCE FAMILY

3.1 | Forest structure

Definition 1. A reference family $\mathfrak{R} = (R_k, \zeta_k)_{k\in\mathcal{K}}$ is said to have a forest structure if the following property is satisfied:

$$\forall k, k' \in \mathcal{K}, \quad R_k \cap R_{k'} \in \{R_k, R_{k'}, \emptyset\}, \tag{Forest}$$

that is, two elements of $\{R_k\}_{k\in\mathcal{K}}$ are either disjoint or nested.

The forest structure is general enough to cover a wide range of different situations, as for instance the disjoint case

$$\forall k, k' \in \mathcal{K}, \quad k \neq k' \Rightarrow R_k \cap R_{k'} = \emptyset. \tag{Disjoint}$$

and the case (Nested). In general, if each R_k is considered as a node and an oriented edge $R_k \leftarrow R_{k'}$ is depicted between two different sets R_k and $R_{k'}$ if and only if $R_k \subseteq R_{k'}$ and there is no $R_{k''}$ such that $R_k \subsetneq R_{k''} \subsetneq R_{k'}$; the obtained graph corresponds to a (directed) forest in the classical graph theory sense (see Kolaczyk, 2009). An illustration is given in Figure 5. The positions of the nodes in this picture rely on the depth of $\mathfrak{R}$, which can be defined as the function

$$\phi : \begin{cases} \mathcal{K} \to \mathbb{N}^* \\ k \;\mapsto 1 + |\{k' \in \mathcal{K} : R_{k'} \supsetneq R_k\}| . \end{cases} \tag{13}$$

For instance, under (Disjoint), $\phi(k) = 1$ for all $k \in \mathcal{K}$, while under (Nested), $\phi(k) = K + 1 - k$ for all $1 \leq k \leq K$.

Example 2. Let $m = 25$, $R_1 = \{1, \dots, 20\}$, $R_2 = \{1, 2\}$, $R_3 = \{3, \dots, 10\}$, $R_4 = \{11, \dots, 20\}$, $R_5 = \{5, \dots, 10\}$, $R_6 = \{11, \dots, 16\}$, $R_7 = \{17, \dots, 20\}$, $R_8 = \{21, 22\}$, $R_9 = \{22\}$. Then the corresponding reference family $\mathfrak{R} = (R_k, \zeta_k)_{1\leq k\leq 9}$ satisfies (Forest). The sets R_1, R_8 are of depth 1; the sets R_2, R_3, R_4, R_9 are of depth 2; and the sets R_5, R_6, R_7 are of depth 3.

A useful characterization of a forest-structure reference family is given in the next lemma.

Lemma 2. *For any reference family $\mathfrak{R} = (R_k, \zeta_k)_{k\in\mathcal{K}}$ having the structure (Forest), there exists a partition $(P_n)_{1\leq n\leq N}$ of $\mathbb{N}_m$ such that for each $k \in \mathcal{K}$, there exists some (i,j) with $1 \leq i \leq j \leq N$ and $R_k = P_{i:j}$, where we denote*

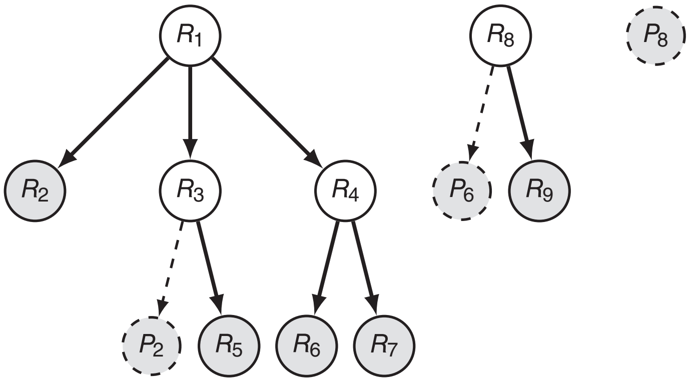

FIGURE 6 Graph corresponding to the reference family given by Example 2, with the associated partition (atoms) $\{P_n, 1 \leq n \leq N\}$, displayed by light gray nodes and given in Example 3. The nodes that correspond to atoms that are not in the reference family are depicted with a dashed circle

$$P_{i:j} = \underset{i \leq n \leq j}{\cup} P_n, \quad 1 \leq i \leq j \leq N. \tag{14}$$

Conversely, for some partition $(P_n)_{1 \leq n \leq N}$ *of* $\mathbb{N}_m$*, consider any reference family of the form* $\mathfrak{R} = (P_{i:j}, \zeta_{i,j})_{(i,j) \in C}$ *with* $C \subseteq \{(i,j) \in \mathbb{N}_N^2 \, : \, i \leq j\}$ *such that for* $(i,j), (i',j') \in C$*, we have*

$$[\![i,j]\!] \cap [\![i',j']\!] = \emptyset; \quad \text{or} \quad [\![i,j]\!] \subseteq [\![i',j']\!]; \quad \text{or} \quad [\![i',j']\!] \subseteq [\![i,j]\!],$$

where $[\![i,j]\!]$ *denotes the set of all integers between i and j. Then* $\mathfrak{R}$ *has the structure (*Forest*).*

For the ease of notation, the set C will be identified to $\mathcal{K}$ throughout the article, which leads to the following slight abuse: denoting indifferently $k \in \mathcal{K}$ or $(i,j) \in \mathcal{K}$, and

$$\mathfrak{R} = (R_k, \zeta_k)_{k \in \mathcal{K}} \quad \text{or} \quad \mathfrak{R} = (P_{i:j}, \zeta_{i,j})_{(i,j) \in \mathcal{K}}. \tag{15}$$

We call “atoms” the elements of the underlying partition $(P_n)_{1 \leq n \leq N}$ because they have the thinnest granularity in the structure and because any subset R_k of the family can be expressed as a combination of these atoms. Note however that this partition is not unique. A simple algorithm to compute $(P_n)_n$ and the proof of Lemma 2 are provided in Appendix B.2. An example of such a partition is given in Example 3 and Figure 6.

Example 3. For the reference family given in Example 2, a partition as in Lemma 2 is given by $P_1 = R_2, P_2 = R_3 \setminus R_5, P_3 = R_5, P_4 = R_6, P_5 = R_7, P_6 = R_8 \setminus R_9, P_7 = R_9, P_8 = \mathbb{N}_m \setminus \{R_1 \cup R_8\}$.

An important particular case in our analysis is the case where the forest structure includes all atoms, that is

$$\forall n \in \{1, \ldots, N\}, \quad P_n \in \{R_k, k \in \mathcal{K}\}. \tag{Complete}$$

When (Complete) does not hold (as in Example 3), we can impose this condition by adding missing atoms (together with a trivial bound) to the structure, building in this way the completed reference family:

Definition 2. Consider any reference family $\mathfrak{R} = (P_{i:j}, \zeta_{i,j})_{(i,j) \in \mathcal{K}}$ satisfying (Forest) and associated with atoms $(P_n)_{1 \leq n \leq N}$ by (15). Let $\mathcal{K}^+ = \{(i,i), 1 \leq i \leq N \, : \, (i,i) \notin \mathcal{K}\}$, $\zeta_{i,i} = |P_{i:i}| = |P_i|$ for all $(i,i) \in \mathcal{K}^+$, and $\mathcal{K}^\oplus = \mathcal{K} \cup \mathcal{K}^+$. Then the completed version of $\mathfrak{R}$ is given by $\mathfrak{R}^\oplus = (P_{i:j}, \zeta_{i,j})_{(i,j) \in \mathcal{K}^\oplus}$.

For the reference family $\mathfrak{R}$ given by Example 2, the completed version $\mathfrak{R}^\oplus$ is depicted in Figure 7.

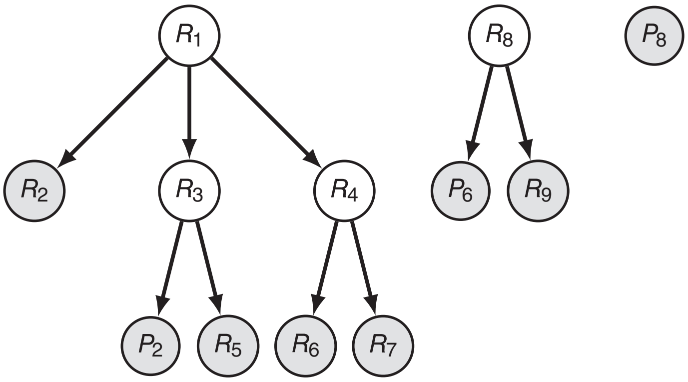

FIGURE 7 Graph corresponding to the completed version $\mathfrak{R}^{\oplus}$ of the reference family $\mathfrak{R}$ given by Example 2 with the atoms given in Example 3

3.2 | Deriving the optimal post hoc bound

The next result shows that the expression of the optimal post hoc bound $V^*_{\mathfrak{R}}$ can be simplified when $\mathfrak{R}$ satisfies (Forest).

Theorem 1. *Let $\mathfrak{R}$ be a reference family having the structure* (Forest)*. Then the optimal bound $V^*_{\mathfrak{R}}$* (4) *can be derived from the bounds $\tilde{V}^q_{\mathfrak{R}}$* (10) *and $\tilde{V}_{\mathfrak{R}}$* (11) *in the following way:*

$$V^*_{\mathfrak{R}}(S) = \tilde{V}_{\mathfrak{R}}(S), \quad S \subseteq \mathbb{N}_m; \tag{16}$$

$$V^*_{\mathfrak{R}}(S) = \tilde{V}^d_{\mathfrak{R}}(S), \quad S \subseteq \mathbb{N}_m, \tag{17}$$

where d is the maximum number of disjoint sets that can be found in the reference family, that is,

$$d = \max\{|Q|, Q \subseteq \mathcal{K} : \forall k, k' \in Q, k \neq k' \Rightarrow R_k \cap R_{k'} = \emptyset\}.$$

A byproduct of Theorem 1 is that, if (Nested) holds, $V^*_{\mathfrak{R}} = \tilde{V}^1_{\mathfrak{R}}(S) = \overline{V}_{\mathfrak{R}}$ and we recover proposition 2.5 of BNR. Another interesting case is the structure (Disjoint), where $\tilde{V}_{\mathfrak{R}}$ has a simpler form. Finally, a third particular case of interest allowing for some simplification is when (Forest) and (Complete) hold, that is, when the forest structure already contains all atoms. We summarize these particular cases in the following result.

Corollary 1. *Let $\mathfrak{R} = (R_k, \zeta_k)_{k \in \mathcal{K}}$ be a reference family.*

(i) if $\mathfrak{R}$ satisfies (Nested)*, then $V^*_{\mathfrak{R}} = \overline{V}_{\mathfrak{R}}$.*
(ii) if $\mathfrak{R}$ satisfies (Disjoint)*, then $V^*_{\mathfrak{R}}(S) = \sum_{k \in \mathcal{K}} \zeta_k \wedge |S \cap R_k| + |S \setminus \cup_{k \in \mathcal{K}} R_k|, \ S \subseteq \mathbb{N}_m$.*
(iii) if $\mathfrak{R}$ satisfies (Forest) *and* (Complete)*, then*

$$V^*_{\mathfrak{R}}(S) = \min_{\substack{Q \subseteq \mathcal{K} \\ \text{s.t. } \{R_k, k \in Q\} \\ \text{is a partition of } \mathbb{N}_m}} \left\{ \sum_{k \in Q} \zeta_k \wedge |S \cap R_k| \right\}. \tag{18}$$

Note that point (iii) justifies the formula (5) used for the motivating example in the Introduction.

Theorem 1 and Corollary 1 are, respectively, proved in Sections A.2 and A.3.

The proof of Theorem 1 being constructive, it provides an algorithm to compute easily $V^*_{\mathfrak{R}}(S)$, that we now describe. Let us first introduce an additional piece of notation. For some reference family $\mathfrak{R} = (P_{i:j}, \zeta_{i,j})_{(i,j)\in\mathcal{K}}$ of depth function ϕ (see (13)), we denote

$$\mathcal{K}^h = \{(i,j) \in \mathcal{K} : \phi(i,j) = h \text{ or } (i = j \text{ and } \phi(i,i) \le h)\}, \quad h \ge 1.$$

Hence, each $\mathcal{K}^h$ contains the indices of the sets of depth h and also the atoms with an inferior depth. Figure 8 displays some $\mathcal{K}^h$ for the reference family of Example 2.

Algorithm 1 gives the steps to compute $V^*_{\mathfrak{R}}(S)$: first, complete the family $\mathfrak{R}$ by adding all the members of the partition, as explained in Definition 2, in order to get $\mathfrak{R}^{\oplus}$. By Lemma 6, we have $V^*_{\mathfrak{R}^{\oplus}}(S) = V^*_{\mathfrak{R}}(S)$, so that this operation does not change the targeted quantity. In particular,

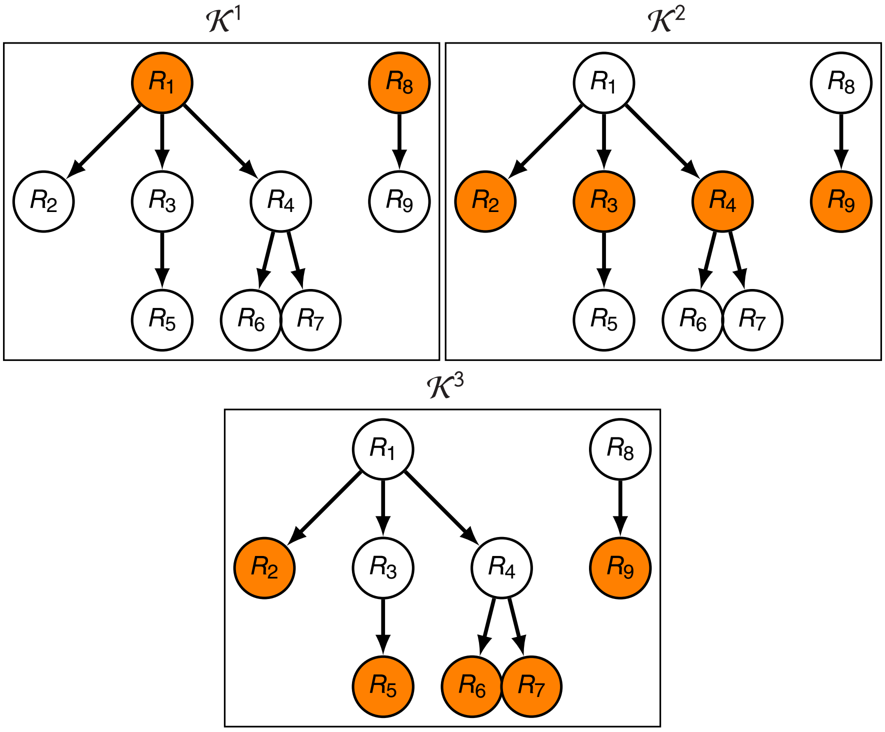

FIGURE 8 Display of the nodes corresponding to $\mathcal{K}^1$, $\mathcal{K}^2$, $\mathcal{K}^3$ (in orange) for the reference family given in Example 2 [Color figure can be viewed at wileyonlinelibrary.com]

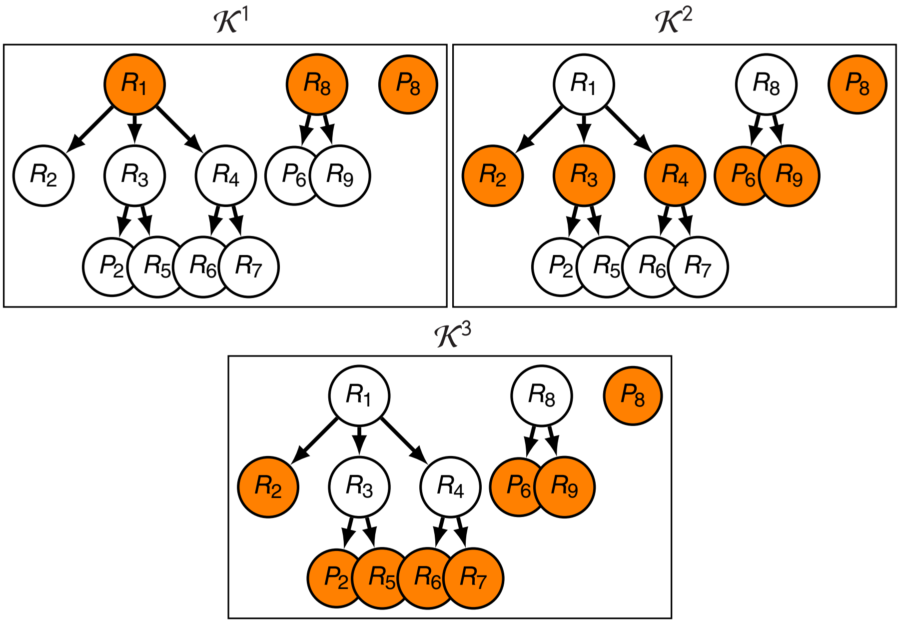

FIGURE 9 Same as Figure 8 but for the completed version [Color figure can be viewed at wileyonlinelibrary.com]

(Complete) holds after this step. Second, the algorithm uses a reverse loop, which successively updates a vector V whose components correspond to active nodes; the current value of the bound is equal to the sum of the components of V. Each step of the loop will update the value of V to make the bound possibly smaller, to obtain at the end $V^*_{\mathfrak{R}}(S)$. It is possible to write equivalently the algorithm by applying the main computation formula (*new* V_{k} line) in succession to all regions R_{k} sorted by decreasing order of depth (another equivalent formulation is as a recursive formula starting at the lowest depth nodes). If we assume that the (completed) hierarchy structure has been computed beforehand (in particular, that the list of direct descendents of a given region is directly available), as well as the bounds ζ_{k}, the time complexity of Algorithm 1 for a given S is $O(|S|K)$.Indeed, the completed forest structure is at most twice as large as the original; each node and edge of the structure is visited once by the algorithm; and the computation of the set intersection with S costs at most $O(|S|)$.

Let us describe the loop in more detail by using the particular situation of Figure 9. Initialization: $H = 3$ and $\mathcal{K}^H = \mathcal{K}^3$, which corresponds to the orange nodes in the bottom graph. Hence, V is equal to the vector of values $\zeta_k \wedge |S \cap R_k|$ among these nodes. First step: $h = 2$ hence $\mathcal{K}^h = \mathcal{K}^2$, for which the active nodes are displayed in orange in the top-right graph. Each of these nodes $k \in \mathcal{K}^2$, gives a bound $\zeta_k \wedge |S \cap R_k|$ that should be compared with the one of the previous step, that is, $\sum_{k' \in \mathrm{Succ}_k} V_{k'}$, where Succ_k denotes the offspring of R_k. The vector V is defined by the best choice among these two. Second (and final) step: $h = 1$ hence $\mathcal{K}^h = \mathcal{K}^1$ (top-left graph) that only contains the roots of the forest and where V is updated following the same process. The algorithm then returns $V^*_{\mathfrak{R}}(S) = \sum_{k \in \mathcal{K}^1} V_k$.

Algorithm 1. Computation of $V^*_{\mathfrak{R}}(S)$

```
Data: $\mathfrak{R} = (P_{i:j}, \zeta_{i,j})_{(i,j)\in\mathcal{K}}$ and $S \subseteq \mathbb{N}_m$.
Result: $V^*_{\mathfrak{R}}(S)$.
$\mathfrak{R} \longleftarrow \mathfrak{R}^{\oplus}$; $\mathcal{K} \longleftarrow \mathcal{K}^{\oplus}$ (completion, see Definition 7);
$H \longleftarrow \max_{k\in\mathcal{K}} \phi(k)$, see (13);
$V \longleftarrow (\zeta_k \wedge |S \cap R_k|)_{k\in\mathcal{K}^H}$;
for $h \in \{H-1, \dots, 1\}$ do
    $newV \longleftarrow (0)_{k\in\mathcal{K}^h}$;
    for $k \in \mathcal{K}^h$ do
        $Succ_k \longleftarrow \{k' \in \mathcal{K}^{h+1} : R_{k'} \subseteq R_k\}$;
        $newV_k \longleftarrow \min\left(\zeta_k \wedge |S \cap R_k|, \sum_{k'\in Succ_k} V_{k'}\right)$;
    end
    $V \longleftarrow newV$;
end
return $\sum_{k\in\mathcal{K}^1} V_k$.
```

4 | LOCAL CALIBRATION OF THE REFERENCE FAMILY

In this section, we explain how to build a reference family $\mathfrak{R}$ such that (3) holds. The results presented in this section hold for any deterministic $(R_k)_k$ and the calibration concerns only $(\zeta_k)_k$ here.

4.1 | Calibration of ζ_k by DKW inequality

In this section, we estimate $|S \cap \mathcal{H}_0|$ by using an approach close in spirit to the so-called Storey estimator (Storey, 2002). The latter depends on a parameter, denoted by t here, that has to be chosen appropriately (see Blanchard & Roquain, 2009 for a discussion on this issue). To avoid this caveat while improving accuracy, we can derive an estimator uniform on t by using the DKW inequality (Dvoretzky et al., 1956), with the optimal constant of Massart (1990).

For any deterministic subsets $R_k \subseteq \mathbb{N}_m$, $k \in \mathcal{K}$, $K = |\mathcal{K}|$, let

$$\zeta_k(X) = |R_k| \wedge \min_{t\in[0,1)} \left\lfloor \left(\frac{C}{2(1-t)} + \left(\frac{C^2}{4(1-t)^2} + \frac{\sum_{i\in R_k} \mathbf{1}\{p_i(X) > t\}}{1-t} \right)^{1/2} \right)^2 \right\rfloor, \quad k \in \mathcal{K}, \tag{19}$$

where $C = \sqrt{\frac{1}{2}\log\left(\frac{K}{\alpha}\right)}$ and $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x.

Proposition 1. *Consider any deterministic (different) subsets $R_k \subseteq \mathbb{N}_m$, $k \in \mathcal{K}$ ($K = |\mathcal{K}|$) and assume $\alpha/K < 1/2$. Assume that for all $k \in \mathcal{K}$, the p-value family $\{p_i(X), i \in R_k\}$ satisfies (*Superunif*) and (*Indep*). Then the JER control (*3*) holds for the reference family $\mathfrak{R} = (R_k, \zeta_k(X))_{k\in\mathcal{K}}$, for which the local bounds ζ_k are given by (*19*).*

Combining Proposition 1 with Lemma 1, we obtain that, under the assumptions of Proposition 1, the bound

$$V_{\mathrm{DKW}} = \tilde{V}_{\mathfrak{R}} \text{ given by (11) with } \mathfrak{R} = (R_k, \zeta_k(X))_{k\in\mathcal{K}} \quad \text{and} \quad \zeta_k(X) \text{ given by (19)}, \tag{20}$$

satisfies (2) and thus is a valid post hoc bound.

Proposition 1 is proved in Section A0.4. Note that $\zeta_k(X) \geq \lfloor \log(K/\alpha)/2 \rfloor \geq 1$ as soon as $\alpha \leq e^{-2}K$. Hence, this contrasts with previous approaches Blanchard et al., (in press); Goeman & Solari, 2011, for which $\zeta_k = 0$ was included in the reference family. This means that using this reference family induces a minimum cost. In the next section, we will see that this cost is generally compensated by the accuracy of the joint estimation of $|R_k \cap \mathcal{H}_0|$, $k \in \mathcal{K}$.

Remark 1. In practice, $\zeta_k(X)$ in (19) can be computed as

$$\zeta_k(X) = s \wedge \min_{0\leq \ell \leq s} \left\lfloor \left(\frac{C}{2(1-p_{(\ell)})} + \left(\frac{C^2}{4(1-p_{(\ell)})^2} + \frac{s-\ell}{1-p_{(\ell)}} \right)^{1/2} \right)^2 \right\rfloor,$$

where $s = |R_k|$ and $0 = p_{(0)} \leq p_{(1)} \leq \cdots \leq p_{(s)}$ are the ordered p-values of $\{p_i(X), i \in R_k\}$.

Remark 2. With our notation, the previous $(1-\alpha)$-confidence bound of Genovese and Wasserman (2004, equation (16) therein) corresponds to taking

$$\zeta_k^{\mathrm{GW}}(X) = |R_k| \wedge \min_{t\in[0,1)} \left\lfloor \frac{\sum_{i\in R_k} \mathbf{1}\{p_i(X) > t\} + |R_k|^{1/2} C}{1-t} \right\rfloor.$$

By using (B1) in Lemma 3 with $a = 1 - t$, $b = C$, $c = \sum_{i\in R_k} \mathbf{1}\{p_i(X) > t\}$, and $d = |R_k|$, we can see that the quantity $\zeta_k^{\mathrm{GW}}(X)$ is always larger than the $\zeta_k(X)$ given by (19). Hence our result is a uniform improvement over Genovese and Wasserman (2004).

Remark 3. The local bounds ζ_k in (19) depend on the target level α only through C, where $2C^2 = \log(K/\alpha)$. Therefore, the post hoc bounds derived from Proposition 1 are expected to depend only weakly on α. This important point is illustrated in our numerical experiments (Section 5), where this property is used to propose a hybrid post hoc bound taking the best of both the Simes and the DKW-based bounds.

4.2 | Comparison to existing post hoc bounds

To explore the benefit of the new reference family when the signal is localized, let us consider a stylized model where the signal is localized according to a regular partition

$$R_k = \{1 + (k-1)s, \dots, ks\}, \quad 1 \le k \le K, \tag{21}$$

composed of K regions of equal size s. In particular, this reference family satisfies (Disjoint). Among the regions R_k, only R_1 contains false nulls, and $r \in (0, 1)$ denotes the proportion of signal in R_1, that is,

$$r = |R_1 \cap \mathcal{H}_1|/|R_1|. \tag{22}$$

The remaining regions contain no signal, that is, $|R_k \cap \mathcal{H}_1| = 0$, for $k \ge 2$.

In addition, we consider an independent Gaussian one-sided setting where the false nulls have mean $\mu > 0$, that is, we assume that $X_i \sim \mathcal{N}(0, 1)$ if $i \in \mathcal{H}_0$ and $X_i \sim \mathcal{N}(\mu, 1)$ if $i \in \mathcal{H}_1$, and the p-values are derived as $p_i(X) = \overline{\Phi}(X_i)$, $i \in \mathbb{N}_m$, where $\overline{\Phi}$ denotes the upper tail of the standard normal distribution.

Proposition 2. *Let us consider the post hoc bounds* V_{Bonf} (8); V_{Simes} (9) *and the new post hoc bound* V_{DKW} *given by* (20) *and associated with the reference regions* R_k *defined above. In the setting defined above, we have*

$$\frac{\mathbb{E}(V_{\mathrm{DKW}}(R_1))}{|R_1|} \le 1 \wedge \left(1 - r + 2r\overline{\Phi}(\mu) + \frac{4C}{\sqrt{s}}\left(1 + \frac{C}{\sqrt{s}}\right)\right), \tag{23}$$

$$\frac{\mathbb{E}(V_{\mathrm{Simes}}(R_1))}{|R_1|} \ge (1 - r)(1 - \alpha s/m) + r\overline{\Phi}(\mu - \overline{\Phi}^{-1}(\alpha s/m)); \tag{24}$$

$$\frac{\mathbb{E}(V_{\mathrm{Bonf}}(R_1))}{|R_1|} = (1 - r)(1 - \alpha/m) + r\overline{\Phi}(\mu - \overline{\Phi}^{-1}(\alpha/m)). \tag{25}$$

This proposition is proved in Section A.5. In particular, combining (23) and (24) yields

$$\frac{\mathbb{E}(V_{\mathrm{DKW}}(R_1))}{\mathbb{E}(V_{\mathrm{Simes}}(R_1))} \le \frac{1 \wedge \left(1 - r + 2r\overline{\Phi}(\mu) + \frac{4C}{\sqrt{s}}\left(1 + \frac{C}{\sqrt{s}}\right)\right)}{(1 - r)(1 - \alpha s/m) + r\overline{\Phi}(\mu - \overline{\Phi}^{-1}(\alpha s/m))}. \tag{26}$$

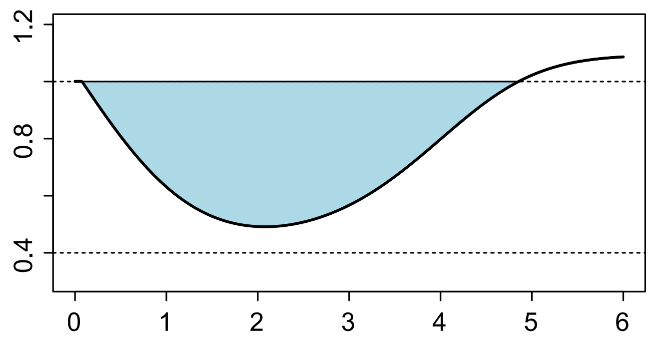

FIGURE 10 Y-axis: upper bound of the ratio between the new bound and the Simes bound (see (26)). X-axis: effect size μ. $m = 10^7$, $s = m^{2/3}$, $K = m/s$, $r = 3/5$, $\alpha = 0.1$ [Color figure can be viewed at wileyonlinelibrary.com]

This ratio is displayed in Figure 10 for a choice of model parameters. The new bound can substantially improve the Simes bound over a wide range of effect sizes.

This improvement can also be highlighted by an asymptotic approach.

Corollary 2. *Let us consider the framework of Proposition* 2. *In the asymptotic setting in m where s tends to infinity with* $s \gg \log K$ *and* μ *tends to infinity with* $\mu - \overline{\Phi}^{-1}(\alpha/m) \to -\infty$, *we have*

$$\limsup_m \left\{ \frac{\mathbb{E}(V_{\mathrm{DKW}}(R_1))}{|R_1|} \right\} \leq 1 - r \quad \text{and} \quad \limsup_m \left\{ \frac{\mathbb{E}(V_{\mathrm{Bonf}}(R_1))}{|R_1|} \right\} = 1.$$

If moreover $s \ll m$ *(i.e.,* $K \to \infty$*) and* $\mu - \overline{\Phi}^{-1}(\alpha s/m) \to -\infty$, *we have*

$$\limsup_m \left\{ \frac{\mathbb{E}(V_{\mathrm{DKW}}(R_1))}{|R_1|} \right\} \leq 1 - r \quad \text{and} \quad \limsup_m \left\{ \frac{\mathbb{E}(V_{\mathrm{Simes}}(R_1))}{|R_1|} \right\} = 1.$$

In particular, this corollary establishes that the order of the new bound can improve the Simes bound by a factor $1 - r$.

5 | NUMERICAL EXPERIMENTS

The open-source R package `sansSouci` (Blanchard, Durand, Neuvial, & Roquain, 2019) implements the bounds proposed in this article and provides R code to reproduce the numerical experiments reported in this section (see Appendix C.2). All these experiments have been carried out using the R language (R Core Team, 2019).

5.1 | Setting

In this section, we perform numerical experiments to compare our new post hoc bound V_{DKW} (20) with the Simes post hoc bound (9). Let q be some fixed integer, say larger than 1. We consider two versions of our new bound:

- The first version of our post hoc bound, denoted V_{part}, is defined by (20) in which the reference family $\mathfrak{R}^{\mathrm{part}}$ is the regular partition of $\mathbb{N}_m$ given by (21) for $K^{\mathrm{part}} = 2^q$ ($s = m/2^q$ being assumed to be an integer).
- The second version of our post hoc bound, denoted V_{tree}, is defined similarly by (20), but the reference family $\mathfrak{R}^{\mathrm{tree}}$ is given this time by the perfect binary tree whose leaves are the elements of $\mathfrak{R}^{\mathrm{part}}$. Hence, by using the notation of Lemma 2, this means $P_k = \{1 + (k-1)s, \ldots, ks\}$, $1 \leq k \leq 2^q$. The cardinal of the reference family is thus $K^{\mathrm{tree}} = 2^{q+1} - 1$.

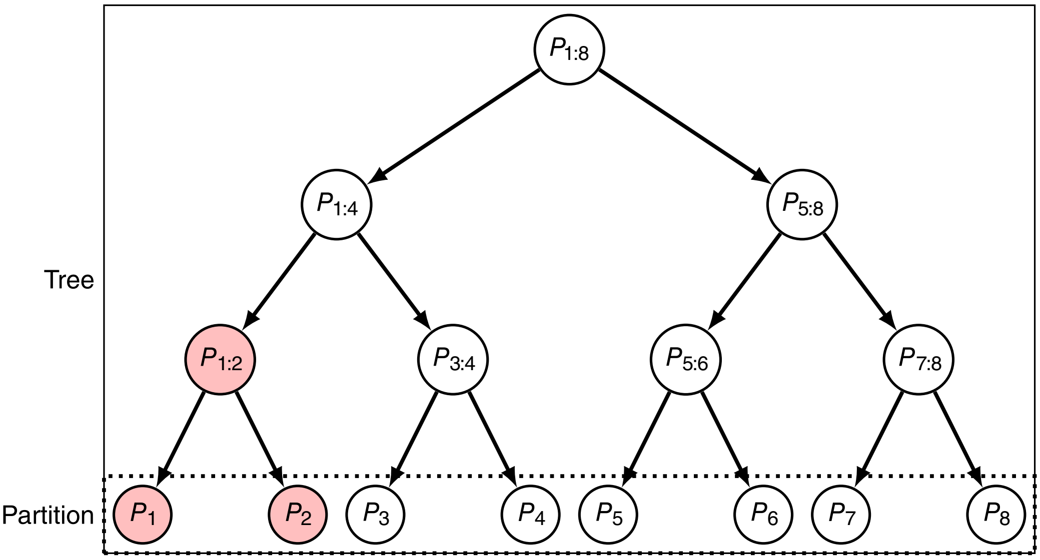

FIGURE 11 Partition and perfect binary tree structures used in simulations, here with $q = 3$ and $K_1 = 2$ ($K^{\text{part}} = 8$ and $K^{\text{tree}} = 15$). The pink nodes are those containing some signal [Color figure can be viewed at wileyonlinelibrary.com]

Note that, compared with the bounds used in the toy example of the Introduction, V_{part} and V_{tree} here use slightly improved values of the ζ_k's (see (20)).

The true/false null hypothesis configuration is as follows: The false null hypotheses are contained in P_k for $1 \leq k \leq K_1$, for some fixed value of K_1. The quantity r is defined similarly as in (22), as the fraction of false null hypotheses in those P_k, and is set to $r \in \{0.5, 0.75, 0.9, 1\}$. All of the other partition pieces only contain true null hypotheses. *The* p-values are one-sided and computed from Gaussian observations. The true null observations are distributed as i.i.d. $\mathcal{N}(0, 1)$, and false null observations are distributed as i.i.d. $\mathcal{N}(\overline{\mu}, 1)$, where $\overline{\mu}$ is a fixed value in $\{2, 3, 4\}$. This construction is illustrated in Figure 11 for $q = 3$ (leading to $K^{\text{part}} = 8$ and $K^{\text{tree}} = 15$) and $K_1 = 2$. In our experiments, we have chosen $q = 7$ and $s = 100$ (corresponding to $K^{\text{part}} = 128$ and $K^{\text{tree}} = 255$ and $m = 12,800$), and $K_1 = 8$.

We also performed numerical experiments with $s \in \{10, 20, 50\}$ and $K_1 \in \{1, 4, 16\}$, and with Poisson- and Gaussian-distributed $\overline{\mu}$. Because the results are qualitatively similar, we only report the above-described setting.

5.2 | Comparing confidence envelopes

One possible way to evaluate the performance of post hoc bounds is to consider the associated confidence envelopes on the number of true discoveries among the most significant hypotheses. Formally, for $k = 1, \dots, m$, we let $S_k = \{i_1, \dots, i_k\}$, where i_j is the index of the jth smallest p-value. Note that focusing on such sets is a priori favorable to the Simes bound, for which the elements of the reference family are among the S_k. In Figure 12, each panel corresponds to a particular choice of the model parameters r (in rows) and $\overline{\mu}$ (in columns). Each panel compares the actual number of true positives $(k - |S_k \cap \mathcal{H}_0|)$, $k = 1, \dots, m$ (labeled "Oracle") to post hoc bounds of the form $(k - V(S_k))$, $k = 1, \dots, m$, where V is V_{Simes}, V_{part}, or V_{tree}. In this figure, the confidence level is set to $1 - \alpha = 95\%$.

The chosen model parameters span a wide range of situations between very low and very high signal. For very low signal ($\overline{\mu} = 2, r = 0.75$, top-left panel), all the bounds are trivial, that is, output

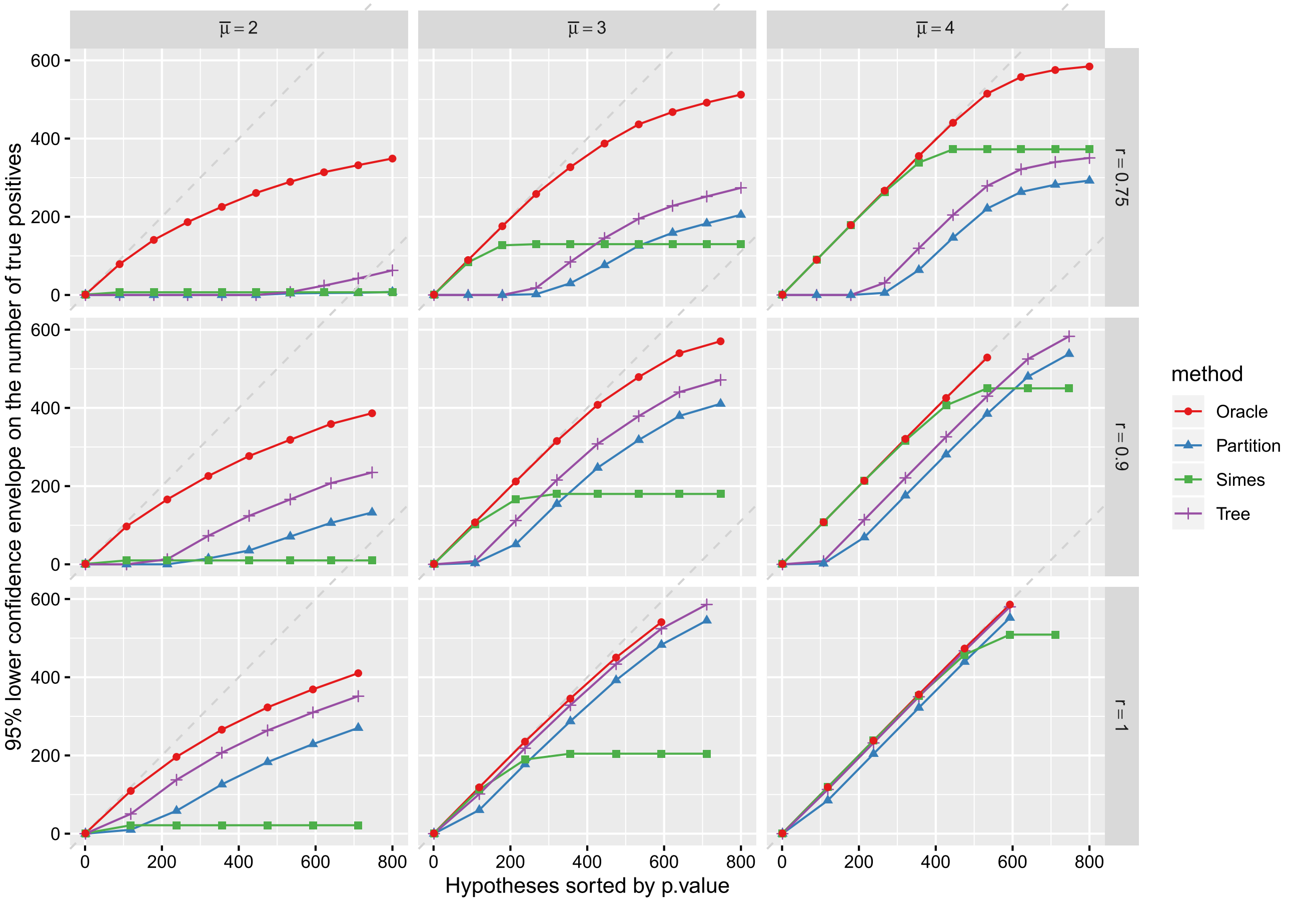

FIGURE 12 95% lower confidence envelopes on the number of true positives obtained from Simes inequality and from the proposed methods are compared with the actual (Oracle) number of true positives [Color figure can be viewed at wileyonlinelibrary.com]

a value $V(S_k)$ close to $|S_k|$ $(= k)$. As expected, all the bounds get sharper as the signal to noise ratio increases, that is, as $\overline{\mu}$ or r increase, and for very high signal ($\overline{\mu} = 4$, $r = 1$, bottom-right panel), all the bounds are very close to the actual number of true positives. The tree-based bound dominates the partition-based bound, which is expected because in this particular experiment, the regions P_k containing signal are adjacent (see Figure 11), and the multiscale nature of the tree-based bound allows it to take advantage of large-scale clusters. When the signal regions are not adjacent, these two bounds are very close (additional numerical experiments not shown). Our proposed bounds are more sensitive to the proportion of signal in each active region, while the Simes bound is more sensitive to the strength of the signal in those regions. As a result, none of the Simes and the "tree" bound is uniformly better than the other one. The Simes bound is typically sharper than the "tree" bound for small values of k, but becomes more conservative for larger values of k. This is expected, because the "tree" bound is based on *estimating the proportion of* nonnull items, while the Simes bound is based on *pinpointing* nonnull items.

5.3 | Hybrid approach

An interesting question raised in Section 4.1 (Remark 3) is how these bounds are influenced by the target confidence level, which is fixed to $1 - \alpha = 95\%$ in Figure 12. In Figure 13, we compare

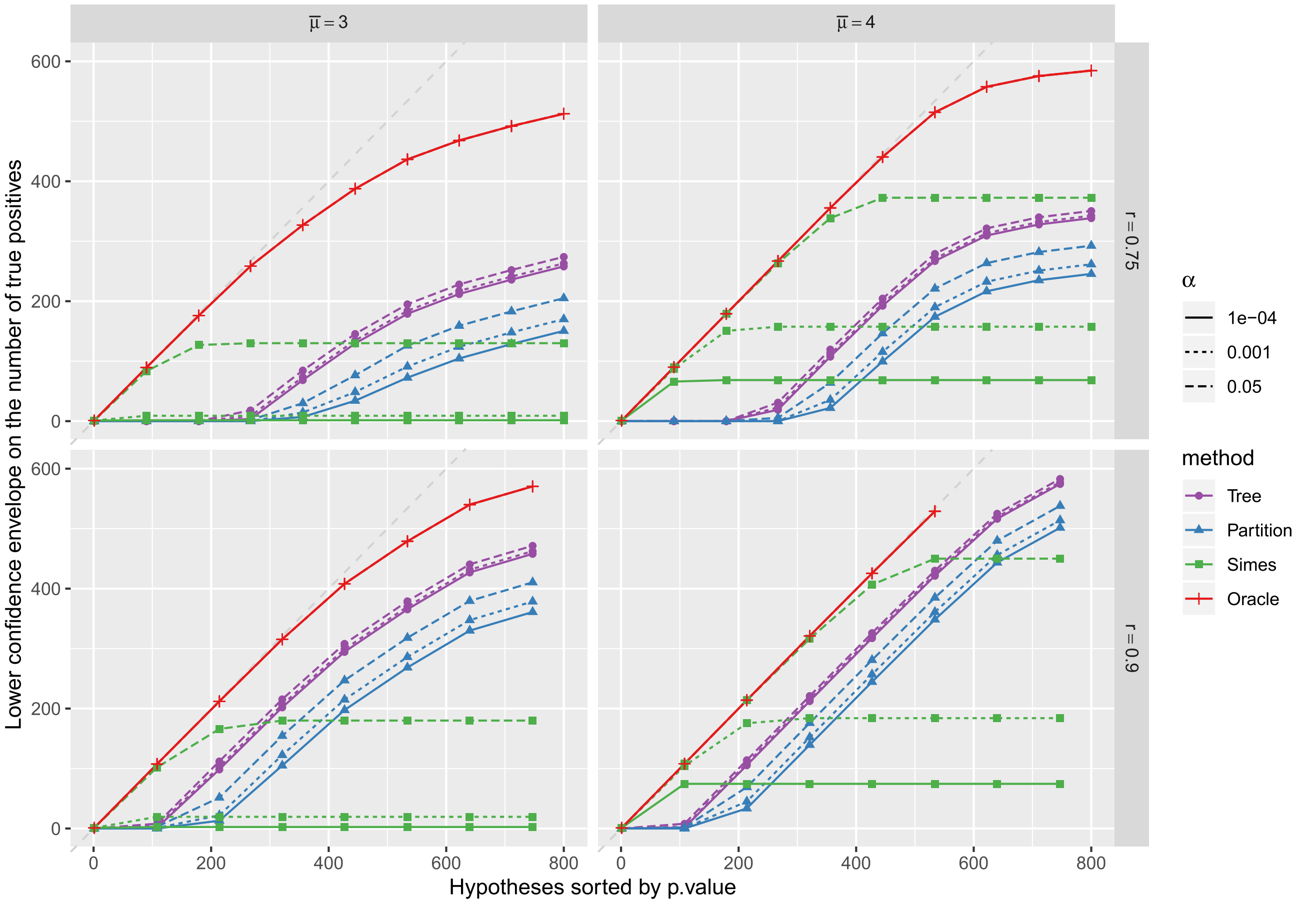

FIGURE 13 Influence of the target level parameter α on upper confidence envelopes on the number of true positives [Color figure can be viewed at wileyonlinelibrary.com]

the bounds obtained across values of α (corresponding to different line types) for $\overline{\mu} \in \{3, 4\}$ and $r \in \{0.75, 0.9\}$. The influence of α on the Simes bound is quite substantial. This is consistent with the shape of the bound (9); the p-values are directly compared with α. The influence of α on the bounds derived from (19) is much weaker, as expected from Remark 3. In particular, the envelopes derived from the "tree" method are very close to each other when α varies from .001 to .05. These striking differences suggest to introduce hybrid confidence envelopes that could take advantage of the superiority of the Simes bound on sets S_k for small k with that of the DKW-tree-based bound on sets S_k for larger k. For a fixed $\gamma \in [0, 1]$, let us define the bound $V^{\gamma}_{\text{hybrid}}$ as follows. For $S \subseteq \mathbb{N}_m$,

$$V^{\gamma}_{\text{hybrid}}(\alpha, S) = \min\left(V_{\text{Simes}}((1-\gamma)\alpha, S), V_{\text{tree}}(\gamma\alpha, S)\right),$$

where the notation in the bounds explicitly acknowledges the dependence of the bounds in the target level α. By an union bound, $V^{\gamma}_{\text{hybrid}}(\alpha, \cdot)$ is a $(1-\alpha)$-level post hoc bound. Figure 14 gives an illustration with $\alpha = 0.05$ and $\gamma = 0.02$. In this case, the hybrid envelope is the minimum of the Simes envelope at level $(1-\gamma)\alpha = 0.049$ and the DKW-tree-based envelope at level 0.001. Because $(1-\gamma)\alpha$ is very close to α, the confidence envelope $V^{0.02}_{\text{hybrid}}$ is essentially equivalent to the Simes-based confidence envelope for small k; for larger values of k, $V^{0.02}_{\text{hybrid}}$ is only slightly worse than the DKW-tree-based confidence envelope at level $\gamma\alpha = 0.001$.

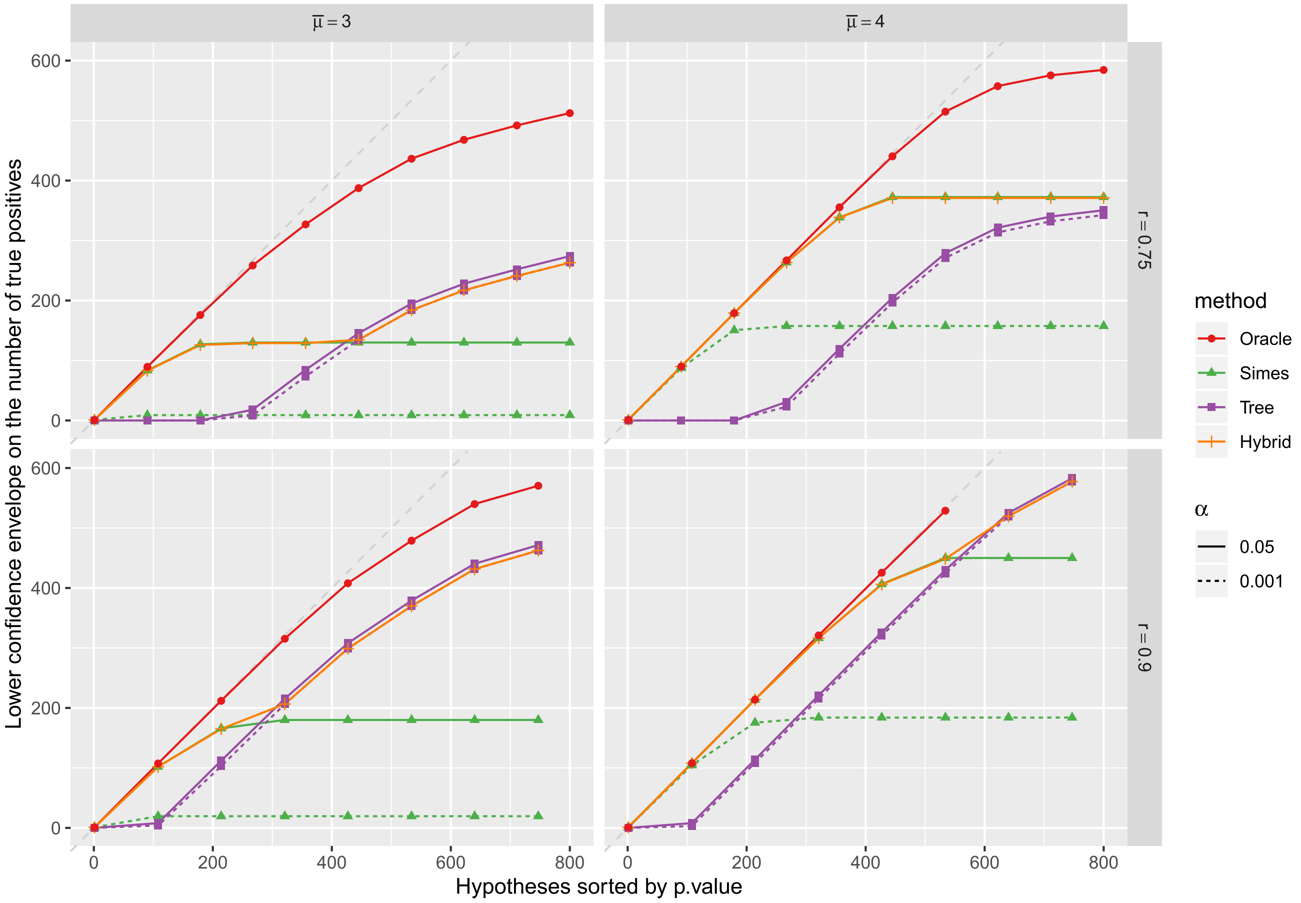

FIGURE 14 Combining Simes and tree-based confidence envelopes on the number of true positives into a hybrid confidence envelope [Color figure can be viewed at wileyonlinelibrary.com]

6 | DISCUSSION

6.1 | Comparison with Meijer et al. (2015)

Since our aim is similar to the one of Meijer et al. (2015) (denoted MKG for short), let us make a short qualitative comparison between MKG and our study. First, while both approaches are based on graph-structured subsets $\{R_k, k \in \mathcal{K}\}$, the geometrical shapes of the nodes R_k are different: the nodes in MKG correspond to all possible consecutive intervals, possibly overlapping, while our reference sets are based on recursive partitioning of the hypothesis space at different resolutions. Our approach avoids redundancies of the tests, but is suitable when the signal is structured according to the prespecified partition structure and may lead to a less accurate bound otherwise. This in turn impacts the way the local pieces of information are combined. The MKG approach uses a sequential, top-down algorithm, with an α-recycling method (that allows, for instance, to spend the same nominal level α both for a parent and its child). By contrast, our approach uses a bottom-up algorithm, with an overall nominal level adjusted by a simple union bound, which is generally conservative but seems fair here as the nodes are disjoint (at each resolution).

Second, the criteria used are different: MKG focus on simultaneous FWER control of local tests of intersections of null hypotheses $\cap_{i \in R_k} H_{0,i}$, $k \in \mathcal{K}$, while our statistical criterion ensures with high probability $|R_k \cap \mathcal{H}_0| \leq \zeta_k$, for all $k \in \mathcal{K}$, for some bounds ζ_k. As already noted in BNR (see the supplementary file therein), the two approaches coincide when $\zeta_k = |R_k| - 1$ because $|R_k \cap \mathcal{H}_0| > |R_k| - 1$ is equivalent to the fact that $\cap_{i \in R_k} H_{0,i}$ is true. Hence, a family $\{R_k, k \in \mathcal{K}\}$

violating $|R_k \cap \mathcal{H}_0| > |R_k| - 1$ for some k will also wrongly reject $\cap_{i \in R_k} H_{0,i}$ for some k. However, when using another form of bound ζ_k, such as the DKW device used here, such a connection is not valid anymore and the two criteria does not incorporate the local structure of the nodes in the same way. Here, using bounds ζ_ks based on classical estimators will in principle lead to better post hoc bounds.

Third, within each node, the local statistics used are not of the same nature: in MKG, the local tests are based on a multivariate χ^2-type test (see Goeman, Van der Geer, De Kort, & Van Houwelingen, 2004). Here, we use an estimator relying on individual p-values that exploits the independence structure. This means that the assumptions made in MKG are much weaker, since it is valid under arbitrary dependence. Our approach can in principle also accommodate such a distributional setting, but this needs additional investigations (see the discussion in Section 6.3).

Finally, let us mention a setting for which the two methods can be fairly compared. First take the MKG method with Bonferroni local tests. As proved in MKG, the resulting FWER controlling procedure (reject the $H_{0,i}$ for which $V(\{i\}) = 0$) then reduces to the Holm procedure (Holm, 1979). By contrast, if we consider ζ_k equals to the number of accepted null hypotheses by the Holm procedure restricted to R_k (satisfying (Disjoint)), our methodology induces another overall FWER controlling procedure: Simply the one rejecting all the null hypotheses rejected by the local Holm procedures. Both FWER controlling procedures are valid under arbitrary dependence. Interestingly, if the signal is sparse but localized in one of the prespecified R_k, the new procedure will dominate the global Holm procedure (this is supported by a numerical experiment and a theoretical study, not reported here for short). This illustrates, once again, that our methodology can improve the state of the art, even in a very elementary framework.

6.2 | Comparison with Katsevich and Ramdas (in press)

The approach followed by Katsevich and Ramdas (in press) is similar to that of BNR in the sense that it builds on a family of reference regions R_k, satisfying a JER control of the form (3), from which a post hoc bound (4) can be interpolated. The main focus of Katsevich and Ramdas is on how to construct suitable bounds ζ_k, under several settings, using powerful martingale techniques. In order to apply those, all settings considered there have in common the assumption (Indep) and the fact that the regions R_k are nested; in fact, these regions are of the form $R_k = \{\pi(i); 1 \leq i \leq k\}$ for different orderings π of the hypotheses. The three orderings considered by Katsevich and Ramdas are (a) ordered p-values; (b) a priori ordering available from side information; and (c) “interactive” ordering. In all three cases, the rationale is that the obtained bounds will have better performance if the false hypotheses tend to be ordered before true null hypotheses.

As we have discussed earlier, ordering p-values by their magnitude without using any structural information is not appropriate for the kind of problem we consider in the present work, and the hierarchical tree structure we focused on does not correspond naturally to an a priori ordering of the hypotheses: the structure we consider will be well suited to a situation where the false hypotheses are clustered in a specific part of the tree, but we do not know which one.

Finally, the *interactive* ordering considered by Katsevich and Ramdas (in press) uses the principle of “p-value masking” introduced by Lei and Fithian (2018) and Lei, Ramdas, and Fithian (2017). In that setting, the ordering is constructed iteratively, where the next element $\pi(k)$ in the order can be decided arbitrarily based on the information $((g(p_i))_{i \in \mathbb{N}_m}, (p_i)_{i<k})$, $g : [0,1] \to [0,1]$ being a “masking function” (one simple example is $g(p) = \min(p, 1-p)$). Thus, while the

constructed reference regions R_k are still nested, the iterative nature of the order construction, by progressive unmasking of the p-values, can orient the procedure. For instance, it is conceivable that, in the tree-structured setting considered here, the user would choose to "unmask" first the masked p-values that appear to cluster in a same region of the tree. A more in-depth comparison to our approach would however require to specify a precise "unmasking policy" depending on the data and on the tree structure, which falls outside of the scope of the present discussion. The approach considered in the present article does not require to specify (or ask interactively from the user) such an "unmasking" policy. We also note that, although we focused on the (Indep) assumption in order to obtain explicit bounds (using the DKW device), our approach is potentially valid under weaker dependence assumption, provided the number of false positives in a fixed region can be estimated in a suitable way by a confidence bound (see Section 6.3).

6.3 | Extension to general local confidence bounds

In this work, the local bounds ζ_k have been designed by using the DKW inequality. This can be straightforwardly extended to the case where the bound (19) is replaced by $\zeta_k(X) = L_k(\alpha/K)$, for which the function $L_k(\cdot)$ is a local bound satisfying the condition

$$\forall \lambda \in (0,1), \quad \forall k \in \mathcal{K}, \quad \forall P \in \mathcal{P}, \quad \mathbb{P}_{X \sim P}\left(|R_k \cap \mathcal{H}_0(P)| \leq L_k(\lambda)\right) \leq \lambda. \tag{27}$$

The properties of the final post hoc bound will obviously depend on the choice of L_k.

In the present work, the validity of our post hoc bounds relies on (Indep), which is a strong assumption. The latter is only used to make the DKW inequality valid. If this assumption is violated, we should use another local bound L_k, that satisfies condition (27) under the specific dependence setting of the data. For instance, when the dependence of the individual test statistics is known or satisfies a randomization hypothesis (see Hemerik & Goeman, 2018), such a local bound can be constructed by applying the λ-calibration methodology of BNR (e.g., the one corresponding to the balanced template therein).

Such extension would be particularly relevant in the case of linear regression models, where the test statistics are correlated in a complex way. Certainly, testing null regression coefficients in a linear regression model is a theme of major interest throughout the literature; furthermore, whenever correlation is present between the covariates, it is different from the (simpler) goal of testing multiple associations between the target and the each covariate (see, for instance, the introductory discussion of Meinshausen, 2008). Beyond the presence of correlations, this setting raises important additional issues, in particular how to handle the high-dimensional situation where the number of covariates is much larger than the number of observations. In a general configuration, this is a thorny issue, not to say a major hindrance. Under some specific assumptions used in high-dimensional statistics such as the restricted isometry property, one can hope that false positive inference on arbitrary selected sets of bounded size is feasible, though such an assumption might be deemed unrealistic for practice. In any case, future work should concentrate on obtaining local bounds of the form (27) for regression models under assumptions as general as possible.

ACKNOWLEDGEMENTS

The authors would like to thank the editor, associate editor, and three referees for their very helpful comments. This work has been supported by ANR-16-CE40-0019 (SansSouci), ANR-17-CE40-0001 (BASICS) and by the GDR ISIS through the “projets exploratoires” program (project TASTY). G.B. acknowledges the support from the german DFG, under the Research Unit FOR-1735 “Structural Inference in Statistics - Adaptation and Efficiency” and under the Collaborative Research Center SFB-1294 “Data Assimilation.”

ORCID

Etienne Roquain https://orcid.org/0000-0003-0214-5185

REFERENCES

Bachoc, F., Blanchard, G., & Neuvial, P. (2018). On the post selection inference constant under restricted isometry properties. *Electronic Journal of Statistics*, *12*, 3736–3757 Retrieved from https://hal.archives-ouvertes.fr/hal-01772538

Bachoc, F., Leeb, H., & Pötscher, B. M. (2019). Valid confidence intervals for post-model-selection predictors. *Annals of Statistics*, *47*, 1475–1504.

Benjamini, Y., & Bogomolov, M. (2014). Selective inference on multiple families of hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*, 297–318.

Benjamini, Y., & Yekutieli, D. (2005). False discovery rate–adjusted multiple confidence intervals for selected parameters. *J Am Stat Assoc*, *100*, 71–81.

Berk, R., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, *41*, 802–837.

Blanchard, G., Durand, G., Neuvial, P., & Roquain, E. (2019) *Sanssouci: Post hoc inference via joint error rate control. R package version 0.8.1*. Retrieved from https://github.com/pneuvial/sanssouci.

Blanchard, G., & Geman, D. (2005). Hierarchical testing designs for pattern recognition. *The Annals of Statistics*, *33*, 1155–1202.

Blanchard, G., Neuvial, P., & Roquain, E. (in press). Post hoc confidence bounds on false positives using reference families. *The Annals of Statistics* Retrieved from https://hal.archives-ouvertes.fr/hal-01483585

Blanchard, G., & Roquain, É. (2009). Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research*, *10*, 2837–2871.

Bühlmann, P., & Mandozzi, J. (2014). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Computational Statistics*, *29*, 407–430.

Chang, W., Cheng, J., Allaire, J., Xie, Y., &McPherson, J. (2019) Shiny: Web application framework for R. R package version 1.4.0. Retrieved from https://CRAN.R-project.org/package=shiny.

Choi, Y., Taylor, J., & Tibshirani, R. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *The Annals of Statistics*, *45*, 2590–2617.

Cox, D. R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika*, *62*, 441–444.

Dezeure, R., Bühlmann, P., Meier, L., & Meinshausen, N. (2015). High-dimensional inference: Confidence intervals, p-values and r-software hdi. *Statistical Science*, *30*, 533–558.

Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, *27*, 642–669.

Ehm, W., Kornmeier, J., & Heinrich, S. P. (2010). Multiple testing along a tree. *Electronic Journal of Statistics*, *4*, 461–471.

Farcomeni, A., & Pacillo, S. (2011). A conservative estimator for the proportion of false nulls based on Dvoretzky, Kiefer and Wolfowitz inequality. *Statistics & Probability Letters*, *81*, 1867–1870.

Fithian, W., Sun, D., & Taylor, J. (2017). Optimal inference after model selection. arXiv preprint arXiv:1410.2597.

Genovese, C., & Wasserman, L. (2004). A stochastic process approach to false discovery control. *The Annals of Statistics*, *32*(3), 1035–1061.

Genovese, C. R., & Wasserman, L. (2006). Exceedance control of the false discovery proportion. *Journal of the American Statistical Association*, *101*, 1408–1417.

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

Goeman, J. J., & Solari, A. (2010). The sequential rejection principle of familywise error control. *The Annals of Statistics*, *38*, 3782–3810.

Goeman, J. J., & Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, *26*, 584–597.

Goeman, J. J., Van der Geer, S. A., De Kort, F., & Van Houwelingen, H. C. (2004). A global test for groups of genes: Testing association with a clinical outcome. *Bioinformatics*, *20*, 93–99.

Guo, W., Lynch, G., & Romano, J. P. (2018) A new approach for large scale multiple testing with application to FDR control for graphically structured hypotheses. arXiv preprint arXiv:1812.00258.

Hemerik, J., & Goeman, J. J. (2018). False discovery proportion estimation by permutations: Confidence for significance analysis of microarrays. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*, 137–155.

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

Katsevich, E., & Ramdas, A. (in press). Simultaneous high-probability bounds on the false discovery proportion in structured, regression, and online settings. *Annals of Statistics* Retrieved from https://arxiv.org/abs/1803.06790

Kim, K. I., Roquain, E., & van de Wiel, M. A. (2010). Spatial clustering of array CGH features in combination with hierarchical multiple testing. *Statistical Applications in Genetics and Molecular Biology*, *9*, 40.

Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. New York, NY: Springer Science & Business Media.

Lei, L., & Fithian, W. (2018). AdaPT: An interactive procedure for multiple testing with side information. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*, 649–679.

Lei, L., Ramdas, A., & Fithian, W. (2017) STAR: A general interactive framework for FDR control under structural constraints. arXiv preprint arXiv:1710.02776.

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

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, *18*, 1269–1283.

Meijer, R. J., & Goeman, J. J. (2015). A multiple testing method for hypotheses structured in a directed acyclic graph. *Biometrical Journal*, *57*, 123–143.

Meijer, R. J., Krebs, T. J., & Goeman, J. J. (2015). A region-based multiple testing method for hypotheses ordered in space or time. *Statistical Applications in Genetics and Molecular Biology*, *14*, 1–19.

Meinshausen, N. (2006). False discovery control for multiple tests of association under general dependence. *Scandinavian Journal of Statistics*, *33*, 227–237.

Meinshausen, N. (2008). Hierarchical testing of variable importance. *Biometrika*, *95*, 265–278.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.R-project.org/

Ramdas, A., Chen, J., Wainwright, M. J., & Jordan, M. I. (2019). A sequential algorithm for false discovery rate control on directed acyclic graphs. *Biometrika*, *106*, 69–86.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, *40*, 87–110.

Scheffé, H. (1959). *The analysis of variance*. New York, NY / London, UK: John Wiley & Sons, Inc / Chapman & Hall, Ltd.

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

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

Taylor, J., & Tibshirani, R. (2018). Post-selection inference for-penalized likelihood models. *The Canadian Journal of Statistics*, *46*, 41–61.

Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 7629–7634.

Tibshirani, R. J., Taylor, J., Lockhart, R., & Tibshirani, R. (2016). Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association*, *111*, 600–620.

Weinstein, A., & Ramdas, A. (2019) Online control of the false coverage rate and false sign rate. arXiv preprint arXiv:1905.01059.

Yekutieli, D. (2008). Hierarchical false discovery rate-controlling methodology. *Journal of the American Statistical Association*, *103*, 309–316.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Durand G, Blanchard G, Neuvial P, Roquain E. Post hoc false positive control for structured hypotheses. *Scand J Statist*. 2020;1–35. https://doi.org/10.1111/sjos.12453

APPENDIX A. PROOFS

A.1 Proof of Lemma 1

The second and third inequalities in (12) are straightforward from the fact that $\tilde{V}^q_{\mathfrak{R}}$ is nonincreasing in q and $\tilde{V}^1_{\mathfrak{R}} = \overline{V}_{\mathfrak{R}}$. For the first inequality, let $S \subseteq \mathbb{N}_m$ and consider $A \subseteq \mathbb{N}_m$ such that $\forall k \in \mathcal{K}$, $|R_k \cap A| \le \zeta_k$. For any $Q \subseteq \mathcal{K}$, we get

$$\begin{aligned}|S\cap A| &\le \sum_{k\in Q} |S\cap A\cap R_k| + |S\cap A\cap (\underset{k\in Q}{\cup} R_k)^{\mathrm{c}}| \\ &\le \sum_{k\in Q} \zeta_k \wedge |S\cap R_k| + |S\setminus \underset{k\in Q}{\cup} R_k|,\end{aligned}$$

which implies the result.

A.2 Proof of Theorem 1

In this proof, we fix $S \subseteq \mathbb{N}_m$. Also, we let

$$\mathcal{A}(\mathfrak{R}) = \{A \subseteq \mathbb{N}_m \,:\, \forall k \in \mathcal{K}, |R_k \cap A| \le \zeta_k\}\,, \tag{A1}$$

so that $V^*_{\mathfrak{R}}(S) = \max_{A\in\mathcal{A}(\mathfrak{R})}|S\cap A|$. Also note that (10) and (11) can be rewritten as

$$\tilde{V}_{\mathfrak{R}}(S) = \min_{\mathcal{K}'\subseteq\mathcal{K}}\left(\sum_{k\in\mathcal{K}'} \zeta_k \wedge |S\cap R_k| + \left|S\setminus \underset{k\in\mathcal{K}'}{\cup} R_k\right|\right). \tag{A2}$$

A.2.2 Proof of (16)

First, by Lemma 6, it is sufficient to prove (16) for $\mathfrak{R}^{\oplus}$. Hence, we can focus without generality on the case where (Complete) holds. Recall that this means that $(i,i) \in \mathcal{K}$ for all $1 \le i \le N$. Now, to prove that $\tilde{V}_{\mathfrak{R}}(S) = V^*_{\mathfrak{R}}(S)$, it suffices to build $A \subseteq S$ such that $A \in \mathcal{A}(\mathfrak{R})$ and $|A| = \tilde{V}_{\mathfrak{R}}(S)$. The key point is that for any h, A is the disjoint union of the $A \cap R_k$, $k \in \mathcal{K}^h$, because the R_k, $k \in \mathcal{K}^h$, form a partition of $\mathbb{N}_m$ (by Lemma 4). Let $H = \max_{k \in \mathcal{K}} \phi(k)$ be the greater depth of the Forest structure, we will construct A with a decreasing recursion over $h \in \{1, \dots, H\}$. To this end, we need some additional notation: first, for any $k \in \mathcal{K}$, let $\mathcal{K}_k = \{k' \in \mathcal{K} : R_{k'} \subseteq R_k\}$ be the set of indexes of elements that are subsets of R_k. Then, for any h, let $\mathcal{K}^{\geq h} = \cup_{h \leq h' \leq H} \mathcal{K}^{h'}$. Note that $\mathcal{K}^{\geq 1} = \mathcal{K}$. Finally let

$$\mathfrak{P}^h = \{\mathcal{P} \subseteq \mathcal{K}^{\geq h} : \text{ the } R_k,\ k \in \mathcal{P}, \text{ form a partition of } \mathbb{N}_m\},$$

and note that the result of Lemma 5 (i.e., Equation (B2)) can be rewritten in

$$\tilde{V}_{\mathfrak{R}}(S) = \min_{\mathcal{P} \in \mathfrak{P}^1} \sum_{k \in \mathcal{P}} \zeta_k \wedge |S \cap R_k|. \tag{A3}$$

The decreasing recursion starts like this: noting that $\mathcal{K}^H$ is the set of all the (i, i)'s, $1 \leq i \leq N$, we define A^H by choosing (arbitrarily) $\zeta_{i,i} \wedge |S \cap P_{i:i}|$ distinct elements of $S \cap P_{i:i}$ for each $1 \leq i \leq N$. Note that we have both

$$\forall k \in \mathcal{K}^{\geq H},\quad |A^H \cap R_k| \leq \zeta_k,$$

and

$$|A^H| = \sum_{k \in \mathcal{K}^H} \zeta_k \wedge |S \cap R_k| = \min_{\mathcal{P} \in \mathfrak{P}^H} \sum_{k \in \mathcal{P}} \zeta_k \wedge |S \cap R_k|,$$

since $\mathfrak{P}^H = \left\{\mathcal{K}^H\right\}$.

Now let h be given and assume we have constructed an $A^{h+1} \subseteq S$ such that both

$$\forall k \in \mathcal{K}^{\geq h+1},\quad |A^{h+1} \cap R_k| \leq \zeta_k,$$

and

$$\begin{aligned} |A^{h+1}| &= \min_{\mathcal{P} \in \mathfrak{P}^{h+1}} \sum_{k \in \mathcal{P}} \zeta_k \wedge |S \cap R_k| \\ &= \sum_{k \in \mathcal{P}^{h+1}} \zeta_k \wedge |S \cap R_k|, \end{aligned} \tag{A4}$$

for a given $\mathcal{P}^{h+1} \in \mathfrak{P}^{h+1}$. Using that $|A^{h+1}| = \sum_{k \in \mathcal{P}^{h+1}} |A^{h+1} \cap R_k|$ and that $|A^{h+1} \cap R_k| \leq \zeta_k \wedge |S \cap R_k|$ for all $k \in \mathcal{P}^{h+1}$, we deduce that $|A^{h+1} \cap R_k| = \zeta_k \wedge |S \cap R_k|$ for all $k \in \mathcal{P}^{h+1}$.

Now we want to construct A^h by defining all the $A^h \cap R_k$, $k \in \mathcal{K}^h$. By writing that $R_k = \cup_{k' \in \mathcal{P}^{h+1} \cap \mathcal{K}_k} R_{k'}$, the union being disjoint, we have first that, for all $k \in \mathcal{K}^h$,

$$\begin{aligned} |A^{h+1} \cap R_k| &= \sum_{k' \in \mathcal{P}^{h+1} \cap \mathcal{K}_k} |A^{h+1} \cap R_{k'}| \\ &= \sum_{k' \in \mathcal{P}^{h+1} \cap \mathcal{K}_k} \zeta_{k'} \wedge |S \cap R_{k'}|. \end{aligned}$$

Second, we have that:

$$\min_{\mathcal{P}\in\mathfrak{P}^h}\sum_{k\in\mathcal{P}}\zeta_k\wedge|S\cap R_k| = \sum_{k\in\mathcal{K}^h}\min_{\mathcal{P}\in\mathfrak{P}^h}\left(\sum_{k'\in\mathcal{P}\cap\mathcal{K}_k}\zeta_{k'}\wedge|S\cap R_{k'}|\right) \tag{A5}$$

$$= \sum_{k\in\mathcal{K}^h}\min(\zeta_k\wedge|S\cap R_k|, \min_{\mathcal{P}\in\mathfrak{P}^{h+1}}\left(\sum_{k'\in\mathcal{P}\cap\mathcal{K}_k}\zeta_{k'}\wedge|S\cap R_{k'}|\right)) \tag{A6}$$

$$\begin{aligned} &= \sum_{k\in\mathcal{K}^h}\min(\zeta_k\wedge|S\cap R_k|, \sum_{k'\in\mathcal{P}^{h+1}\cap\mathcal{K}_k}\zeta_{k'}\wedge|S\cap R_{k'}|) \\ &= \sum_{k\in\mathcal{K}^h}\min\left(\zeta_k\wedge|S\cap R_k|, |A^{h+1}\cap R_k|\right). \end{aligned} \tag{A7}$$

In the above, (A5) holds by additivity and because for every $\mathcal{P}\in\mathfrak{P}^h$, any element of $\mathcal{P}$ is also an element of one of the $\mathcal{P}\cap\mathcal{K}_k$, $k\in\mathcal{K}^h$. Moreover, for every $\mathcal{P}\in\mathfrak{P}^h$ and $k\in\mathcal{K}^h$, $\mathcal{P}\cap\mathcal{K}_k$ is either $\{k\}$, either a set of elements of depth $\geq h+1$, hence (A6). Finally, (A7) holds because all the minima in (A6) are realized in $\mathcal{P}^{h+1}$, otherwise the minimality of $\mathcal{P}^{h+1}$ in (A4) would be contradicted.

We finally construct all the $A^h\cap R_k$, $k\in\mathcal{K}^h$, in the following way: if $|A^{h+1}\cap R_k|\leq\zeta_k\wedge|S\cap R_k|$, we let $A^h\cap R_k = A^{h+1}\cap R_k$, else we let $A^h\cap R_k$ be a subset of $\zeta_k\wedge|S\cap R_k|$ distinct elements of $A^{h+1}\cap R_k$. This both ensures that

$$|A^h| = \min_{\mathcal{P}\in\mathfrak{P}^h}\sum_{k\in\mathcal{P}}\zeta_k\wedge|S\cap R_k|,$$

and that

$$\forall k\in\mathcal{K}^{\geq h},\quad |A^h\cap R_k|\leq\zeta_k,$$

because $\mathcal{K}^{\geq h}=\mathcal{K}^h\cup\mathcal{K}^{\geq h+1}$ and $A^h\subseteq A^{h+1}$, which ends the recursion.

Now letting $A=A^1$, we have found an $A\subseteq S$ such that $A\in\mathcal{A}(\mathfrak{R})$ and $|A|=\tilde{V}_{\mathfrak{R}}(S)$ (by (A3)).

A.2.3 Proof of (17)

By (16) and Lemmas 5 and 6, we have

$$V^*_{\mathfrak{R}}(S) = V^*_{\mathfrak{R}^\oplus}(S) = \tilde{V}_{\mathfrak{R}^\oplus}(S) = \sum_{k\in\overline{\mathcal{K}}}\zeta_k\wedge|S\cap R_k|,$$

for some $\overline{\mathcal{K}}\subseteq\mathcal{K}^\oplus$ such that the R_k, $k\in\overline{\mathcal{K}}$, form a partition of $\mathbb{N}_m$. Hence,

$$\begin{aligned} V^*_{\mathfrak{R}}(S) &= \sum_{k\in\mathcal{K}\cap\overline{\mathcal{K}}}\zeta_k\wedge|S\cap R_k| + \sum_{k\in\overline{\mathcal{K}}\setminus\mathcal{K}}\zeta_k\wedge|S\cap R_k| \\ &= \sum_{k\in\mathcal{K}\cap\overline{\mathcal{K}}}\zeta_k\wedge|S\cap R_k| + \sum_{k\in\overline{\mathcal{K}}\setminus\mathcal{K}}|S\cap R_k| \\ &= \sum_{k\in\mathcal{K}\cap\overline{\mathcal{K}}}\zeta_k\wedge|S\cap R_k| + \left|S\setminus\bigcup_{k\in\mathcal{K}\cap\overline{\mathcal{K}}}R_k\right| \end{aligned}$$

because the $R_k, k \in \overline{\mathcal{K}} \setminus \mathcal{K}$ are all disjoint. Now, $|\mathcal{K} \cap \overline{\mathcal{K}}| \le d$ by definition of d, which means that the latter display is larger than or equal to $\tilde{V}^{d}_{\mathfrak{R}}(S)$, which proves the result.

A.3 Proof of Corollary 1

A.3.2 Proof of (i)

This is a direct byproduct of Theorem 1 because if (Nested) holds, then $d = 1$ and thus $V^*_{\mathfrak{R}} = \tilde{V}^{d}_{\mathfrak{R}} = \tilde{V}^{1}_{\mathfrak{R}} = \overline{V}_{\mathfrak{R}}$.

A.3.3 Proof of (ii)

By Theorem 1, $V^*_{\mathfrak{R}} = \tilde{V}_{\mathfrak{R}} = \tilde{V}^{K}_{\mathfrak{R}}$ defined by (10) and (11). Now, for any $S \subseteq \mathbb{N}_m$, for any $Q \subseteq \mathcal{K}$ with $|Q| \le K - 1$, by denoting k_0 any element not in Q, we have

$$R_{k_0} \cap \left(\underset{k \in Q}{\cup} R_k \right) = \emptyset,$$

by (Disjoint), and

$$\begin{aligned}
\sum_{k \in Q} \zeta_k \wedge |S \cap R_k| + \left| S \setminus \underset{k \in Q}{\cup} R_k \right| &= |S \cap R_{k_0}| + \sum_{k \in Q} \zeta_k \wedge |S \cap R_k| + \left| S \setminus \left(\underset{k \in Q}{\cup} R_k \cup R_{k_0} \right) \right| \\
&\ge \zeta_{k_0} \wedge |S \cap R_{k_0}| + \sum_{k \in Q} \zeta_k \wedge |S \cap R_k| + \left| S \setminus \left(\underset{k \in Q}{\cup} R_k \cup R_{k_0} \right) \right| \\
&= \sum_{k \in Q \cup \{k_0\}} \zeta_k \wedge |S \cap R_k| + \left| S \setminus \underset{k \in Q \cup \{k_0\}}{\cup} R_k \right| .
\end{aligned}$$

Hence, the minimum in (10) within the $\tilde{V}^{K}_{\mathfrak{R}}$ expression is attained for $Q = \mathcal{K}$ and the result is proved.

A.3.3 Proof of (iii)

This is established in Lemma 5.

A.4 Proof of Proposition 1

Let us show that for all $\lambda \in (0, 1/2)$, for any $S \subseteq \mathbb{N}_m$ with cardinal $s = |S|$, we have with probability at least $1 - \lambda$ that

$$|S \cap \mathcal{H}_0| \le \min_{t \in [0,1)} \left(\frac{\sqrt{\log(1/\lambda)/2}}{2(1-t)} + \left\{ \frac{\log(1/\lambda)/2}{4(1-t)^2} + \frac{N_t(S)}{1-t} \right\}^{1/2} \right)^2, \tag{A8}$$

for $N_t(S) = \sum_{i \in S} \mathbf{1}\{p_i(X) > t\}$. Let $v = |S \cap \mathcal{H}_0|$ (assumed to be positive without loss of generality) and $U_1, \ldots, U_v$ being v i.i.d. uniform random variables. The DKW inequality (with the optimal constant of Massart, 1990) ensures that, with probability at least $1 - \lambda$, for all $t \in [0, 1]$, we have

$$v^{-1}\sum_{i=1}^{v}\mathbf{1}\{U_i > t\} - (1-t) \geq -\sqrt{\log(1/\lambda)/(2v)}.$$

Now using Lemma 3 with $x = v^{1/2}$, $a = 1-t$, $b = \sqrt{\log(1/\lambda)/2}$ and $c = \sum_{i=1}^{v}\mathbf{1}\{U_i > t\}$ provides (A8) but with $N_t(S)$ replaced by c. Since $p_i(X)$ stochastically dominates U_i, by independence $N_t(S)$ also dominates c, which yields

$$\forall k \in \mathcal{K}, \mathbb{P}\left(|R_k \cap \mathcal{H}_0| > \zeta_k(X)\right) \leq \frac{\alpha}{K},$$

by choosing $\lambda = \frac{\alpha}{K}$. Then (3) follows by a classical union bound argument.

A.5 Proof of Proposition 2

We have for any $t \in [0,1)$,

$$\begin{aligned}\frac{\mathbb{E}(V_{\text{Bonf}}(R_1))}{|R_1|} &= s^{-1}\sum_{i\in R_1\cap\mathcal{H}_0}\mathbb{P}(p_i(X) > \alpha/m) + s^{-1}\sum_{i\in R_1\cap\mathcal{H}_1}\mathbb{P}(p_i(X) > \alpha/m)\\ &= (1-r)(1-\alpha/m) + r\left(1-\overline{\Phi}(\overline{\Phi}^{-1}(\alpha/m)-\mu)\right),\end{aligned}$$

which gives (25). Next,

$$\begin{aligned}V_{\text{Simes}}(R_1) &= \min_{1\leq k\leq s}\left\{\sum_{i\in R_1}\mathbf{1}\left\{p_i(X) > \alpha k/m\right\} + k - 1\right\}\\ &\geq \sum_{i\in R_1}\mathbf{1}\left\{p_i(X) > \alpha s/m\right\},\end{aligned}$$

which gives (24). Finally, for all $t \in [0,1)$, by denoting $N = \sum_{i\in R_1}\mathbf{1}\{p_i(X) > t\}$, we have

$$\begin{aligned}\mathbb{E}(V_{\text{DKW}}(R_1)) &\leq \mathbb{E}\left[\left(\frac{C}{2(1-t)} + \left\{\frac{C^2}{4(1-t)^2} + \frac{N}{1-t}\right\}^{1/2}\right)^2\right]\\ &\leq \mathbb{E}\left[\left(\frac{C}{1-t} + \left\{\frac{N}{1-t}\right\}^{1/2}\right)^2\right]\\ &\leq \frac{C^2}{(1-t)^2} + \frac{\mathbb{E}N}{1-t} + \frac{2C}{(1-t)^{3/2}}\mathbb{E}\left[N^{1/2}\right]\\ &\leq \frac{C^2}{(1-t)^2} + \frac{\mathbb{E}N}{1-t} + \frac{2C}{1-t}\left(\frac{\mathbb{E}N}{1-t}\right)^{1/2},\end{aligned}$$

where we used $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for all $x, y \geq 0$ and that $x \mapsto x^{1/2}$ is concave. Since

$$\mathbb{E}\left[N/|R_1|\right] = (1-r)(1-t) + r\left(1-\overline{\Phi}(\overline{\Phi}^{-1}(t)-\mu)\right),$$

and $\mathbb{E}[N] \leq s(1-t)$, this provides

$$\frac{\mathbb{E}(V_{\mathrm{DKW}}(R_1))}{|R_1|} \leq \min_t \left\{ s^{-1}\frac{C^2}{(1-t)^2} + 1 - r + r\frac{\overline{\Phi}(\mu - \overline{\Phi}^{-1}(t))}{1-t} + s^{-1/2}\frac{2C}{1-t} \right\}.$$

Taking $t = 1/2$ gives (23).

APPENDIX B. AUXILIARY RESULTS

B.1 Auxiliary lemmas

The following lemma holds.

Lemma 3. *For all $a > 0$ and $b, c, x \geq 0$, the two following assertions are equivalent*

(i) $c - ax^2 \geq -bx$;
(ii) $x \leq \frac{b}{2a} + \sqrt{\frac{b^2}{4a^2} + \frac{c}{a}}$.

In particular, we have for all $d \geq 0$,

$$d \wedge \left(\frac{b}{2a} + \sqrt{\frac{b^2}{4a^2} + \frac{c}{a}}\right)^2 \leq d \wedge \left(\frac{c + d^{1/2}b}{a}\right). \tag{B1}$$

Proof. The equivalence between (i) and (ii) is obvious. For $d \geq 0$, if we have the inequality $\left(b/(2a) + \sqrt{b^2/(4a^2) + c/a}\right)^2 \geq d$, then (ii) is satisfied with $x = d^{1/2}$, which entails $c - ad \geq -bd^{1/2}$ and gives $d \leq (c + d^{1/2}b)/a$. If, on the contrary, $\left(b/(2a) + \sqrt{b^2/(4a^2) + c/a}\right)^2 \leq d$, then

$$\begin{aligned}\left(\frac{b}{2a} + \sqrt{\frac{b^2}{4a^2} + \frac{c}{a}}\right)^2 &= \frac{b^2}{2a^2} + \frac{c}{a} + \frac{b}{a}\sqrt{b^2/(4a^2) + c/a} \\ &= \frac{c}{a} + \frac{b}{a}\left(b/(2a) + \sqrt{b^2/(4a^2) + c/a}\right) \leq \frac{c}{a} + \frac{b}{a}d^{1/2}.\end{aligned}$$

This entails the result. ∎

The two following lemmas are used in the proof of Theorem 1, in the case where condition (Complete) holds.

Lemma 4. *For a reference family that has a Forest structure, if (Complete) holds, then for any $h \geq 1$, the $P_{i:j}$, $(i,j) \in \mathcal{K}^h$, form a partition of $\mathbb{N}_m$.*

Proof. Let $h \geq 1$. Let $(i,j), (i',j') \in \mathcal{K}^h$ such that $(i,j) \neq (i',j')$. By (Forest), either $P_{i:j}$ and $P_{i':j'}$ are disjoint, or, without loss of generality, $P_{i:j} \subseteq P_{i':j'}$. If $\phi(i',j') = h$ then the latter is not possible because that would mean that $\phi(i,j) \geq h + 1$. If $i' = j'$, then $P_{i:j} \subseteq P_{i':j'}$ would imply that $P_i \cup \ldots \cup P_j \subseteq P_{i'}$ which in turn implies $i = j = i' = j'$ which is also impossible. Hence, $P_{i:j}$ and $P_{i':j'}$ are disjoint.

Now take any $e \in \mathbb{N}_m$. $(P_n)_{1 \leq n \leq N}$ is a partition so there exists some $n \leq N$ such that $e \in P_n$. If $\phi(n,n) \leq h$ then $(n,n) \in \mathcal{K}^h$. If $\phi(n,n) > h$, then $\{k \in \mathcal{K} : P_n \subsetneq R_k\}$ has at least h elements.

Furthermore, those elements are nested by (Forest), so there exists $k \in \mathcal{K}$ such that $P_n \subsetneq R_k$ and $\phi(k) = h$, hence $e \in R_k$ with $k \in \mathcal{K}^h$. Finally in both cases $e \in \cup_{k \in \mathcal{K}^h} R_k$ so $\mathbb{N}_m = \cup_{k \in \mathcal{K}^h} R_k$, which concludes the proof. ■

Lemma 5. *For a reference family that satisfies (Forest) and (*Complete*), we have*

$$\tilde{V}_{\mathfrak{R}}(S) = \min_{\substack{\overline{\mathcal{K}} \subseteq \mathcal{K} \\ \text{s.t. } \left\{R_k, k \in \overline{\mathcal{K}}\right\} \\ \text{is a partition of } \mathbb{N}_m}} \left\{ \sum_{k \in \overline{\mathcal{K}}} \zeta_k \wedge |S \cap R_k| \right\}. \tag{B2}$$

*that is, the minimum in (*A2*) is always achieved on a partition of* $\mathbb{N}_m$.

Proof. Let any $\mathcal{K}' \subseteq \mathcal{K}$. Let $\mathcal{K}'_1 \subseteq \mathcal{K}'$ be the indices of sets that are maximal for inclusion in the family $\left\{R_k, k \in \mathcal{K}'\right\}$. Because of property (Forest), the R_k, $k \in \mathcal{K}'_1$, are pairwise disjoint, and

$$\forall k \in \mathcal{K}', \exists k' \in \mathcal{K}'_1, R_k \subseteq R_{k'}.$$

Note that this implies that $\cup_{k \in \mathcal{K}'_1} R_k = \cup_{k \in \mathcal{K}'} R_k$. Likewise, because $\mathcal{K}$ includes all the (i, i), $1 \leq i \leq N$, there exists $\mathcal{K}'_2 \subseteq \mathcal{K}$ such that the R_k, $k \in \mathcal{K}'_2$, are pairwise disjoint, and

$$\mathbb{N}_m \setminus \bigcup_{k \in \mathcal{K}'_1} R_k = \bigcup_{k \in \mathcal{K}'_2} R_k.$$

Let $\overline{\mathcal{K}} = \mathcal{K}'_1 \cup \mathcal{K}'_2$ and note that the R_k, $k \in \overline{\mathcal{K}}$, form a partition of $\mathbb{N}_m$. To conclude the proof of (B2), we write that

$$\begin{aligned}
\sum_{k \in \mathcal{K}'} \zeta_k \wedge |S \cap R_k| + \left| S \setminus \bigcup_{k \in \mathcal{K}'} R_k \right| &= \sum_{k \in \mathcal{K}'} \zeta_k \wedge |S \cap R_k| + \left| S \cap \left(\mathbb{N}_m \setminus \bigcup_{k \in \mathcal{K}'_1} R_k \right) \right| \\
&\geq \sum_{k \in \mathcal{K}'_1} \zeta_k \wedge |S \cap R_k| + \sum_{k \in \mathcal{K}'_2} |S \cap R_k| \\
&\geq \sum_{k \in \mathcal{K}'_1} \zeta_k \wedge |S \cap R_k| + \sum_{k \in \mathcal{K}'_2} \zeta_k \wedge |S \cap R_k| \\
&= \sum_{k \in \overline{\mathcal{K}}} \zeta_k \wedge |S \cap R_k|.
\end{aligned}$$

■

The last lemma is useful for the general case where (Complete) no longer holds, by making use of the completed Forest structure introduced in Definition 2.

Lemma 6. *For a reference family* $\mathfrak{R} = (R_k, \zeta_k)_{k \in \mathcal{K}}$ *that has a Forest structure, and* $\mathcal{K}^+$, $\mathcal{K}^\oplus$, $\mathfrak{R}^\oplus$ *as in Definition 2, we have for all* $S \subseteq \mathbb{N}_m$, $V^*_{\mathfrak{R}^\oplus}(S) = V^*_{\mathfrak{R}}(S)$, *and* $\tilde{V}_{\mathfrak{R}^\oplus}(S) = \tilde{V}_{\mathfrak{R}}(S)$.

Proof. It is trivial that $\mathcal{A}(\mathfrak{R}) = \mathcal{A}\left(\mathfrak{R}^\oplus\right)$ (see (A1)) because $\zeta_k = |R_k|$ for $k \in \mathcal{K}^+$, hence $V^*_{\mathfrak{R}^\oplus}(S) = V^*_{\mathfrak{R}}(S)$. It is also obvious that $\tilde{V}_{\mathfrak{R}}(S) \geq \tilde{V}_{\mathfrak{R}^\oplus}(S)$ by (A2) and since $\mathcal{K} \subseteq \mathcal{K}^\oplus$. Now let any $\mathcal{K}' \subseteq \mathcal{K}^\oplus$. Let $\mathcal{K}'_1 = \mathcal{K}' \cap \mathcal{K}$ and $\mathcal{K}'_2 = \mathcal{K}' \cap \mathcal{K}^+$. Note that $\mathcal{K}'$ is the disjoint union of $\mathcal{K}'_1$ and $\mathcal{K}'_2$. Then,

$$\begin{aligned}\sum_{k\in\mathcal{K}'}\zeta_k\wedge|S\cap R_k|+\left|S\setminus\underset{k\in\mathcal{K}'}{\cup}R_k\right| &= \sum_{k\in\mathcal{K}_1'}\zeta_k\wedge|S\cap R_k|+\sum_{k\in\mathcal{K}_2'}|S\cap R_k|+\left|S\setminus\underset{k\in\mathcal{K}'}{\cup}R_k\right|\\ &\geq \sum_{k\in\mathcal{K}_1'}\zeta_k\wedge|S\cap R_k|+|S\setminus\underset{k\in\mathcal{K}_1'}{\cup}R_k|\\ &\geq \tilde{V}_{\mathfrak{R}}(S)\end{aligned}$$

because $\zeta_k=|R_k|$ for $k\in\mathcal{K}_2'$. Hence $\tilde{V}_{\mathfrak{R}\oplus}(S)\geq\tilde{V}_{\mathfrak{R}}(S)$, which concludes the proof. ■

B.2 Material for Lemma 2

Algorithm 2 builds (P_n) and follows directly from the proof.

Proof of Lemma 2. Let $H=\max_{k\in\mathcal{K}}\phi(k)$, where ϕ is the depth function defined by (13). We use a recursion to build, for each $1\leq h\leq H$, an integer $N^h\geq 1$ and a partition $P^h=(P_n^h)_{1\leq n\leq N^h}$, which satisfy the following three properties:

$$P^h \text{ is a partition of } \mathbb{N}_m, \tag{P1 h }$$

$$\forall k\in\mathcal{K}\text{such that } \phi(k)<h, \exists(i,j)\in\left\{1,\ldots,N^h\right\}^2 : R_k=\underset{i\leq n\leq j}{\cup}P_n^h, \tag{P 2h}$$

$$\forall k\in\mathcal{K}\text{such that } \phi(k)=h, \exists n\in\left\{1,\ldots,N^h\right\} : R_k=P_n^h. \tag{P 3h}$$

We start the recursion with $h=1$. Let $\mathrm{Succ}_1=\{k\in\mathcal{K} : \phi(k)=1\}$,

$$\mathrm{New}_1=\{R_k : k\in\mathrm{Succ}_1\}\cup\left\{\mathbb{N}_m\setminus\underset{k\in\mathrm{Succ}_1}{\cup}R_k\right\}\setminus\{\emptyset\},$$

and $N^1=|\mathrm{New}_1|$. We let P^1 be the family of elements of New_1. $(\mathscr{P}_1{}^1)$ is true because, by (Forest), for $k,k'\in\mathrm{Succ}_1$, $k\neq k'$, R_k and $R_{k'}$ are disjoint (otherwise they cannot have same depth). $(\mathscr{P}_2{}^1)$ and $(\mathscr{P}_3{}^1)$ are trivially true.

Now let $h\in\{2,\ldots,H\}$ and assume that there exists N^{h-1} and P^{h-1} satisfying $(\mathscr{P}_1{}^{h-1})$, $(\mathscr{P}_2{}^{h-1})$, and $(\mathscr{P}_3{}^{h-1})$. For all $n\in\{1,\ldots,N^{h-1}\}$, let

$$\mathrm{Succ}_{h,n}=\left\{k\in\mathcal{K} : \phi(k)=h \text{ and } R_k\subseteq P_n^{h-1}\right\},$$

$$\mathrm{New}_{h,n}=\left\{R_k : k\in\mathrm{Succ}_{h,n}\right\}\cup\left\{P_n^{h-1}\setminus\underset{k\in\mathrm{Succ}_{h,n}}{\cup}R_k\right\}\setminus\{\emptyset\},$$

$\mathbb{S}_n^h=\sum_{n'=0}^{n}\left|\mathrm{New}_{h,n'}\right|$ (with $\left|\mathrm{New}_{h,0}\right|=0$ by convention), and $\left(P^h_{\mathbb{S}^h_{n-1}+1},\ldots,P^h_{\mathbb{S}^h_n}\right)$ be the family of the elements of $\mathrm{New}_{h,n}$. Then let $N^h=\mathbb{S}^h_{N^{h-1}}$ and $P^h=(P_1^h,\ldots,P^h_{N^h})$. Note that for each $1\leq n\leq N^{h-1}$, P_n^{h-1} is the disjoint union of $P^h_{\mathbb{S}^h_{n-1}+1},\ldots,P^h_{\mathbb{S}^h_n}$, because by (Forest), for $k,k'\in Succ_{h,n}$, $k\neq k'$, R_k and $R_{k'}$ are disjoint (otherwise they can't have same depth). This and $(\mathscr{P}_1{}^{h-1})$ imply $(\mathscr{P}_1{}^h)$. Let $k\in\mathcal{K}$ such that $\phi(k)<h$, then $(\mathscr{P}_2{}^{h-1})$ and $(\mathscr{P}_3{}^{h-1})$ imply that there exists $(i,j)\in\{1,\ldots,N^{h-1}\}^2$ such that $R_k=\cup_{i\leq n\leq j}P_n^{h-1}$. Hence

$$R_k=\underset{\mathbb{S}^{h-1}_{i-1}+1\leq n\leq\mathbb{S}^{h-1}_j}{\cup}P_n^h,$$

and we get $(\mathscr{P}_2^h)$. Finally, let $k \in \mathcal{K}$ such that $\phi(k) = h$. Let k' be the unique element of $\mathcal{K}$ such that $\phi(k') = h - 1$ and $R_k \subsetneq R_{k'}$. By $(\mathscr{P}_3^{h-1})$, there exists $n \in \{1, \dots, N^{h-1}\}$ such that $R_{k'} = P_n^{h-1}$. Hence $k \in \mathrm{Succ}_{h,n}$ and R_k is equal to one of the elements of $\mathrm{New}_{h,n}$, which gives us $(\mathscr{P}_3^h)$.

Now that the recursion has ended, properties $(\mathscr{P}_1^H)$, $(\mathscr{P}_2^H)$, and $(\mathscr{P}_3^H)$ imply the existence of the desired partition. The proof of the converse statement is straightforward from (14). ■

For the purpose of Algorithm 2, we let `len` and `con` be the concatenation and length functions such that, for all $S_1, \dots, S_n, S_{n+1} \subseteq \mathbb{N}_m$ and $S = (S_1, \dots, S_n)$, `len`$(S)= n$, `con`$(S, S_{n+1}) = (S_1, \dots, S_n, S_{n+1})$ if $S_{n+1} \neq \emptyset$ and `con`$(S, \emptyset) = S$.

Algorithm 2. Computation of $(P_n)_{1 \le n \le N}$

```
Data: $\mathfrak{R} = (R_k, \zeta_k)_{k \in \mathcal{K}}$ satisfying (Forest).
Result: $P = (P_n)_{1 \le n \le N}$ such that for each $k \in \mathcal{K}$, there exists some $(i,j)$ such that
        $R_k = \bigcup_{i \le n \le j} P_n$.
$P \longleftarrow (\mathbb{N}_m)$;
$N \longleftarrow 1$;
$H \longleftarrow \max_{k \in \mathcal{K}} \phi(k)$;
for $h \in (1, \dots, H)$ do
    $newP \longleftarrow ()$;
    for $n \in \{1, \dots, N\}$ do
        $Succ_{h,n} \longleftarrow \{k \in \mathcal{K} : R_k \subseteq P_n, \phi(k) = h\}$;
        for $k \in Succ_{h,n}$ do
            $newP \longleftarrow$ con$(newP, R_k)$;
        end
        $newP \longleftarrow$ con$\left(P_n \setminus \bigcup_{k \in Succ_{h,n}} R_k, newP\right)$;
    end
    $P \longleftarrow newP$;
    $N \longleftarrow$ len$(P)$;
end
return $P$
```

APPENDIX C. REPRODUCIBILITY

C.1 Interactive application to reproduce the introductory example

An interactive application to reproduce the introductory example is available at: `https://pneuvial.shinyapps.io/posthoc-bounds_ordered-hypotheses/`. This application has been created with the R package `shiny` Chang, Cheng, Allaire, Xie, & McPherson, 2019. The code for this application is itself distributed with the R package `sansSouci` at `https://github.com/pneuvial/sanssouci/tree/develop/inst/shiny-examples/ordered-hypotheses`.

C.2 Code and vignette to reproduce numerical experiments

The code used to perform the numerical experiments described in Section 5 is distributed with the R package sansSouci (see inst/DBNR/envelopes).

A more user-friendly pdf vignette to reproduce the middle panel of Figure 12 is also available as supplementary material (pdf format). The source code for this vignette is distributed with the R package sansSouci (see the file: vignettes/confidenceEnvelopes_localized.Rmd). It is also available directly from R via:

```
> remotes::install_github("pneuvial/sanssouci@develop", build_vignettes = TRUE)
> browseVignettes(package = "sansSouci")
```