Biometrika (2025), **112**, 1, asae050
https://doi.org/10.1093/biomet/asae050
Advance Access publication 9 October 2024

A note on *e*-values and multiple testing

BY GUANXUN LI

Department of Statistics, Beijing Normal University at Zhuhai, 18 Jinfeng Road, Zhuhai, Guangdong 519087, China
guanxun@bnu.edu.cn

AND XIANYANG ZHANG

Department of Statistics, Texas A&M University, 400 Bizzell Street, College Station, Texas 77843, U.S.A.
zhangxiany@stat.tamu.edu

SUMMARY

We discover a connection between the Benjamini–Hochberg procedure and the *e*-Benjamini–Hochberg procedure (Wang & Ramdas, 2022) with a suitably defined set of *e*-values. This insight extends to Storey's procedure and generalized versions of the Benjamini–Hochberg procedure and the model-free multiple testing procedure of Barber & Candés (2015) with a general form of rejection rules. We further summarize these findings in a unified form. These connections open up new possibilities for designing multiple testing procedures in various contexts by aggregating *e*-values from different procedures or assembling *e*-values from different data subsets.

Some key words: Benjamini–Hochberg procedure; *E*-value; False discovery rate; Leave-one-out analysis; Multiple testing.

1. INTRODUCTION

When working with high-dimensional data in modern scientific fields, the problem of multiple testing often arises when we explore a vast number of hypotheses with the goal of detecting signals while also controlling some error measures, such as the false discovery rate (FDR). The Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) is perhaps the most widely used FDR-controlling procedure that rejects a hypothesis whenever its *p*-value is less than or equal to an adaptive rejection threshold determined by the whole set of *p*-values. Barber & Candès (2015) proposed a model-free FDR-controlling procedure that estimates the number of false rejections by leveraging the symmetry of *p*-values or test statistics under the null and compares each *p*-value or test statistic with an adaptive threshold.

More recently, there is a growing literature on utilizing *e*-values for statistical inference under different contexts; see, e.g., Shafer (2021), Vovk & Wang (2021), Xu et al. (2021), Dunn et al. (2023), Ignatiadis et al. (2023), Grünwald et al. (2024) and Xu & Ramdas (2024). In particular, Wang & Ramdas (2022) proposed a multiple testing procedure named the *e*-Benjamini–Hochberg procedure by applying the Benjamini–Hochberg procedure to *e*-values, which was shown to control the FDR even when the *e*-values exhibit arbitrary dependence.

In this work, we establish a connection between the Benjamini–Hochberg and *e*-Benjamini–Hochberg procedures with a suitably defined set of *e*-values, proving that they yield identical rejection sets. We next extend this connection to Storey's procedure and generalized versions of the Benjamini–Hochberg and the Barber–Candès procedures, which can have a more general form for the rejection rules. All these connections can be summarized in a unified form. Additionally, these connections provide an effective way of constructing multiple testing procedures in different contexts. Specifically, we propose two new multiple testing procedures by aggregating e-values from different procedures or the same procedure with different tuning quantities, and assembling e-values from different datasets.

2. PRELIMINARIES

2.1. *False discovery rate*

Suppose that we are interested in testing n hypotheses $(H_1, \ldots, H_n)$ simultaneously. Let $\theta = (\theta_1, \ldots, \theta_n) \in \{0, 1\}^n$ indicate the underlying truth of each hypothesis, where $\theta_i = 0$ if H_i is under the null and $\theta_i = 1$ otherwise. Denote by $\delta = (\delta_1, \ldots, \delta_n) \in \{0, 1\}^n$ a decision rule for the n hypotheses, where we reject the ith hypothesis if and only if $\delta_i = 1$. The FDR for the decision rule δ is defined as the expectation of the false discovery proportion (FDP), i.e.,

$$\mathrm{FDR}(\delta) = \mathbb{E}[\mathrm{FDP}(\delta)], \qquad \mathrm{FDP}(\delta) = \frac{\sum_{i=1}^n (1-\theta_i)\delta_i}{1 \vee \sum_{i=1}^n \delta_i},$$

where $a \vee b = \max(a, b)$. The goal of an FDR-controlling procedure is to ensure that the FDR is bounded from above by a prespecified number $\alpha \in (0, 1)$.

2.2. *The Benjamini–Hochberg procedure*

The Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) is perhaps the most widely used FDR-controlling method. To describe the procedure, suppose that we observe a p-value p_i for each H_i. Sort the p-values in ascending order as $p_{(1)} \leqslant \cdots \leqslant p_{(n)}$, and let $\hat{k} = \max\{i : p_{(i)} \leqslant (\alpha i)/n\}$. The Benjamini–Hochberg procedure rejects all hypotheses $H_{(i)}$ with $i \leqslant \hat{k}$, where $H_{(i)}$ is the hypothesis associated with $p_{(i)}$. This procedure is equivalent to rejecting all H_i with $p_i \leqslant T_{\mathrm{BH}}$, where T_{BH} is defined as

$$T_{\mathrm{BH}} = \sup\left\{0 < t \leqslant 1 : \frac{nt}{1 \vee R(t)} \leqslant \alpha\right\} \tag{1}$$

with $R(t) = \sum_{i=1}^n \mathbb{1}\{p_i \leqslant t\}$ the number of rejections given threshold t, and $\mathbb{1}\{A\}$ denoting the indicator function associated with a set A.

Assumption 1. The null p-values are mutually independent, and are independent of the alternative p-values.

We say that a p-value p is superuniform under the null if $P_0(p \leqslant t) \leqslant t$ for each $t \in [0, 1]$, where P_0 denotes the probability measure under the null hypothesis. It is well known that, under Assumption 1 and if the null p-values are superuniform, the Benjamini–Hochberg procedure at level α controls the FDR at level $\alpha n_0/n \leqslant \alpha$, where n_0 is the number of hypotheses under the null (Ferreira & Zwinderman, 2006).

2.3. *Storey's procedure*

Storey's procedure (Storey, 2002; Storey et al., 2004) improves the Benjamini–Hochberg procedure by using the p-values to estimate the null proportion $\pi_0 := n_0/n$. Specifically, we define

$$\pi_0^\lambda := \frac{1 + n - R(\lambda)}{(1-\lambda)n}, \tag{2}$$

where $\lambda \in [0, 1)$ is fixed. Storey's procedure rejects all H_i with $p_i \leqslant T_{\mathrm{ST}}$, where T_{ST} is defined as

$$T_{\mathrm{ST}} = \sup\left\{0 < t \leqslant \lambda : \frac{n\pi_0^\lambda t}{1 \vee R(t)} \leqslant \alpha\right\}. \tag{3}$$

When $\pi_0^\lambda < 1$, Storey's procedure makes more rejections than the Benjamini–Hochberg procedure. If Assumption 1 holds and the null p-values are uniformly distributed on $[0, 1]$, Storey's procedure has finite sample FDR control (Storey et al., 2004).

2.4. *The Barber–Candès procedure*

In a seminal paper by Barber & Candès (2015), the authors proposed a model-free multiple testing procedure that exploits the symmetry of the null p-values or test statistics to estimate the number of false rejections. More precisely, the Barber–Candès procedure specifies a data-dependent threshold, denoted T_{BC}, which is determined as

$$T_{\mathrm{BC}} = \sup\left\{0 < t < 0.5 \colon \frac{1 + \sum_{i=1}^{n} \mathbb{1}\{p_i \geqslant 1 - t\}}{1 \vee R(t)} \leqslant \alpha\right\}, \tag{4}$$

and it rejects all H_i with $p_i \leqslant T_{\mathrm{BC}}$. The Barber–Candès procedure has been shown to provide finite sample FDR control under suitable assumptions (Barber & Candès, 2015).

2.5. *E-values and the e-Benjamini–Hochberg procedure*

A nonnegative random variable e is called an e-value if $\mathbb{E}[e] \leqslant 1$ under the null hypothesis. Suppose that we observe n e-values $e_1, \ldots, e_n$ corresponding to hypotheses $H_1, \ldots, H_n$. The α-level e-Benjamini–Hochberg procedure involves sorting the e-values in decreasing order as $e_{(1)} \geqslant \cdots \geqslant e_{(n)}$ and rejecting the hypotheses associated with the $\hat{k}$ largest e-values, where $\hat{k} := \max\{1 \leqslant i \leqslant n \colon e_{(i)} \geqslant n/(i\alpha)\}$. Note that $P(1/e_i \leqslant t) \leqslant t$ by Markov's inequality, which indicates that $1/e_i$ is superuniform. Thus, the e-Benjamini–Hochberg procedure is simply the Benjamini–Hochberg procedure applied to the p-values $\{1/e_i\}_{i=1}^n$. An advantage of the e-Benjamini–Hochberg procedure is that it controls the FDR at level α even under unknown arbitrary dependence among the e-values.

Proposition 1 (Theorem 2 of Wang & Ramdas, 2022). *Suppose that the nonnegative random variables* $\{e_i\}$ *satisfy*

$$\sum_{i \in \mathcal{H}_0} \mathbb{E}[e_i] \leqslant n, \tag{5}$$

where $\mathcal{H}_0 = \{1 \leqslant i \leqslant n \colon \theta_i = 0\}$. *Then, the* α*-level* e*-Benjamini–Hochberg procedure applied to* $\{e_i\}$ *controls the* FDR *at level* α*, regardless of the dependence structure among* $\{e_i\}$.

In the multiple testing context, the requirement that $\mathbb{E}[e] \leqslant 1$ in the definition of e-values can be relaxed. More precisely, we refer to $\{e_i\}$ as a set of e-values if they satisfy condition (5) throughout the rest of the paper.

3. Connections between the procedures

3.1. *Connections between the Benjamini–Hochberg and e-Benjamini–Hochberg procedures*

We first establish the equivalence between the Benjamini–Hochberg procedure and the corresponding e-Benjamini–Hochberg procedure with a suitably defined set of e-values. This equivalence appears to be a new finding that has not been explicitly stated in the previous literature.

To see the connection between the Benjamini–Hochberg and e-Benjamini–Hochberg procedures, we define the e-value associated with H_i to be

$$e_i = \frac{1}{T_{\mathrm{BH}}} \mathbb{1}\{p_i \leqslant T_{\mathrm{BH}}\}, \tag{6}$$

where T_{BH} is given in (1). The e-value defined in (6) coincides with the e-value defined in equation (1) of Banerjee et al. (2023) when the decision rule therein is specified using the Benjamini–Hochberg procedure. Let $[n] = \{1, 2, \ldots, n\}$ for any positive integer n. Under Assumption 1 and if the null p-values are superuniform, by Lemmas 3–4 of Storey et al. (2004), it is straightforward to show that

$\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i] = n_0$, which implies that the e-values defined by (6) satisfy (5). The detailed derivation is provided in the Supplementary Material. Thus, by Proposition 1, the corresponding e-Benjamini–Hochberg procedure controls the FDR at the desired level. Moreover, we claim that the e-Benjamini–Hochberg procedure based on the e-values defined in (6) is equivalent to the Benjamini–Hochberg procedure in the sense that they produce the same set of rejections; see Theorem 2 below for a precise statement.

3.2. *Connections between the Storey and e-Benjamini–Hochberg procedures*

Define the e-value associated with H_i to be

$$e_i = \frac{1}{\pi_0^{\lambda} T_{\mathrm{ST}}}\mathbb{1}\{p_i \leqslant T_{\mathrm{ST}}\}, \tag{7}$$

where π_0^{λ} is defined in (2) and T_{ST} is given in (3). We have the following result.

THEOREM 1. *Suppose that Assumption 1 holds and that the null p-values follow the uniform distribution on* $[0, 1]$. *Then, the e-values defined in* (7) *satisfy* (5). *Additionally, let* $\mathcal{S}_{\mathrm{ST}}$ *be the set of rejections obtained through Storey's procedure at the* FDR *level* α, *and let* $\mathcal{S}_{\mathrm{eBH}}$ *represent the set of rejections obtained from the e-Benjamini–Hochberg procedure at the same* FDR *level* α, *with the e-values defined in* (7). *Then we have* $\mathcal{S}_{\mathrm{ST}} = \mathcal{S}_{\mathrm{eBH}}$.

3.3. *Connections between the Barber–Candès and e-Benjamini–Hochberg procedures*

As noted in the recent work of Ren & Barber (2024), the Barber–Candès procedure is equivalent to the e-Benjamini–Hochberg procedure based on the e-values

$$e_i = \frac{n\mathbb{1}\{p_i \leqslant T_{\mathrm{BC}}\}}{1 + \sum_{j=1}^{n}\mathbb{1}\{p_j \geqslant 1 - T_{\mathrm{BC}}\}},$$

where T_{BC} is the threshold defined in (4).

4. THE FLEXIBLE BENJAMINI–HOCHBERG AND BARBER–CANDÈS PROCEDURES

4.1. *The flexible Benjamini–Hochberg procedure*

We generalize the Benjamini–Hochberg procedure to allow the rejection rule to take the form $\varphi_i(p_i) \leqslant t$, where φ_i is a strictly increasing function and can differ for each i. This generalization enables the testing procedure to utilize cross-sectional information among the p-values and external structural information for each hypothesis, which often results in a higher multiple testing power. Let $F_i = \varphi_i^{-1}$ represent the inverse function of φ_i, g be some strictly increasing function and g^{-1} be the inverse function of g. Consider the rejection threshold given by

$$T_{\mathrm{FBH}} = \sup\left\{0 < t \leqslant 1 \colon \frac{ng(t)}{1 \vee R(t)} \leqslant \alpha\right\}, \tag{8}$$

where $R(t) = \sum_{i=1}^{n}\mathbb{1}\{\varphi_i(p_i) \leqslant t\}$. The flexible Benjamini–Hochberg procedure rejects H_i whenever $\varphi_i(p_i) \leqslant T_{\mathrm{FBH}}$. Similar to the Benjamini–Hochberg procedure, the flexible Benjamini–Hochberg procedure can be equivalently implemented in the following way. We sort $q_i = \varphi_i(p_i)$ in ascending order, i.e., $q_{(1)} \leqslant \cdots \leqslant q_{(n)}$, and find the largest k, represented as $\hat{k}$, for which $q_{(k)} \leqslant g^{-1}(\alpha k/n)$. We reject $H_{(i)}$ for all $i \leqslant \hat{k}$. The following proposition states that the flexible Benjamini–Hochberg procedure ensures FDR control at a certain level.

PROPOSITION 2. *Suppose that Assumption 1 holds and that the null p-values are superuniform. The flexible Benjamini–Hochberg procedure controls the* FDR *at level* $C\alpha$, *where*

$$C = \sum_{i\in\mathcal{H}_0}\sup_{t\in\mathcal{C}_\alpha}\frac{F_i(t)}{ng(t)}, \qquad \mathcal{C}_\alpha = \{0 < t \leqslant 1 \colon g(t) \leqslant \alpha\}. \tag{9}$$

Additionally, if $g(t) = n^{-1}\sum_{i=1}^{n} F_i(t)$ *and* $F_i(t) = c_i h(t)$*, where* c_i *is some positive constant and* h *is a strictly increasing function of* t*, then the flexible Benjamini–Hochberg procedure controls the* FDR *at level* α.

Proposition 2 broadens and enhances Theorem 7.1 of Peña et al. (2011) in two ways. First, a careful inspection reveals that Theorem 7.1 of Peña et al. (2011) is a specific instance of Proposition 2 with a particular choice of $F_i(t) = \eta_i(t)$ and $g(t) = (1/n)\sum_{i=1}^{n} \eta_i(t)$, where $\{\eta_1(t), \ldots, \eta_n(t)\}$ is the multiple decision size vector defined in Peña et al. (2011). Second, as a consequence of Proposition 2, the flexible Benjamini–Hochberg procedure controls the FDR at level α when $C = \sum_{i\in\mathcal{H}_0} \sup_{t\in C_\alpha}\{F_i(t)/ng(t)\} \leqslant 1$, which is weaker than the condition $n_0 \sup_{i\in\mathcal{H}_0} \sup_{t\in C_\alpha}\{F_i(t)/nng(t)\} \leqslant 1$ required in Theorem 7.1 of Peña et al. (2011).

The following example illustrates that the flexible Benjamini–Hochberg procedure aligns with the weighted Benjamini–Hochberg procedure for particular choices of g and φ_i.

Example 1. Let $g(t) = t$ and $\varphi_i(p) = p/\omega_i$, where ω_i denotes the weight for the ith hypothesis with $\omega_i > 0$ and $\sum_{i=1}^{n} \omega_i = n$. The flexible Benjamini–Hochberg procedure associated with this choice of φ_i and g corresponds to the weighted Benjamini–Hochberg procedure first introduced by Genovese et al. (2006). In this case, the rejection threshold can be expressed as $T_{\mathrm{FBH}} = \sup\{0 < t \leqslant 1 : nt/\{1 \vee R(t)\} \leqslant \alpha\}$, where $R(t) = \sum_{i=1}^{n} \mathbb{1}\{p_i/\omega_i \leqslant t\}$.

4.2. *Connections between the flexible and e-Benjamini–Hochberg procedures*

Analogous to the Benjamini–Hochberg procedure, we show that the flexible Benjamini–Hochberg procedure is equivalent to the e-Benjamini–Hochberg procedure applied to the e-values

$$e_i = \frac{\mathbb{1}\{\varphi_i(p_i) \leqslant T_{\mathrm{FBH}}\}}{g(T_{\mathrm{FBH}})}, \tag{10}$$

where T_{FBH} is defined in (8). By the leave-one-out argument, we prove the following result.

PROPOSITION 3. *Under the assumptions in Proposition* 2*, the e-Benjamini–Hochberg procedure with e-values defined in* (10) *controls the* FDR *at level* $C\alpha$*, where* C *is defined in* (9).

Additionally, we can prove that the e-Benjamini–Hochberg procedure and the flexible Benjamini–Hochberg procedure deliver the same set of rejections.

THEOREM 2. *Let* $\mathcal{S}_{\mathrm{FBH}}$ *be the set of rejections obtained through the flexible Benjamini–Hochberg procedure at the* FDR *level* α*, and let* $\mathcal{S}_{\mathrm{eBH}}$ *represent the set of rejections obtained from the e-Benjamini–Hochberg procedure at the same* FDR *level* α*, with the e-values defined in* (10)*. Then we have* $\mathcal{S}_{\mathrm{FBH}} = \mathcal{S}_{\mathrm{eBH}}$.

The e-value for the Benjamini–Hochberg procedure is a special case of (10) with $\varphi_i(t) = t$ and $g(t) = t$. Consequently, the e-Benjamini–Hochberg procedure based on (6) yields the same rejection set as the Benjamini–Hochberg procedure.

4.3. *The flexible Barber–Candès procedure*

In this section, we generalize the Barber–Candès procedure with the rejection rule given by $\varphi_i(p_i) \leqslant t$. Similar ideas have been pursued in the literature for structure-adaptive multiple testing (Lei & Fithian, 2018; Zhang & Chen, 2022). We assume that the null p-value satisfies the condition

$$P(p_i \leqslant a) \leqslant P(p_i \geqslant 1 - a) = P(1 - p_i \leqslant a) \quad \text{for all } 0 \leqslant a \leqslant 0.5. \tag{11}$$

Condition (11) is weaker than the mirror conservativeness in Lei & Fithian (2018), and it can be shown that superuniformity implies (11). Indeed, $P(1 - p_i \leqslant a) \geqslant 1 - P(p_i \leqslant 1 - a) \geqslant 1 - (1 - a) = a \geqslant P(p_i \leqslant a)$. Assume that φ_i is an increasing and continuous function, and define

$F_i(x) = \sup\{0 \leqslant p \leqslant 1 : \varphi_i(p) \leqslant x\}$. We claim that $P\{\varphi_i(p_i) \leqslant b\} = P\{p_i \leqslant F_i(b)\}$. To see this, consider two cases. If $\varphi_i(p_i) \leqslant b$, by the definition of $F_i(b)$, we have $p_i \leqslant F_i(b)$. On the other hand, if $p_i \leqslant F_i(b)$ then $\varphi_i(p_i) \leqslant \varphi_i\{F_i(b)\} = \lim_{p\uparrow F_i(b)} \varphi_i(p) \leqslant b$, where we use the fact that φ_i is increasing to get the two inequalities, and the equality is due to the continuity of φ_i. Therefore, the above claim together with (11) implies that

$$P\{\varphi_i(p_i) \leqslant b\} = P\{p_i \leqslant F_i(b)\} \leqslant P\{1 - p_i \leqslant F_i(b)\} = P\{\varphi_i(1 - p_i) \leqslant b\}$$

for all $\varphi_i(0) \leqslant b \leqslant \varphi_i(0.5)$. Hence, we have

$$\frac{\sum_{i\in\mathcal{H}_0} \mathbb{1}\{\varphi_i(p_i) \leqslant t\}}{1 \vee \sum_{i=1}^n \mathbb{1}\{\varphi_i(p_i) \leqslant t\}} \lesssim \frac{\sum_{i\in\mathcal{H}_0} \mathbb{1}\{\varphi_i(1-p_i) \leqslant t\}}{1 \vee \sum_{i=1}^n \mathbb{1}\{\varphi_i(p_i) \leqslant t\}} \leqslant \frac{1 + \sum_{i=1}^n \mathbb{1}\{\varphi_i(1-p_i) \leqslant t\}}{1 \vee \sum_{i=1}^n \mathbb{1}\{\varphi_i(p_i) \leqslant t\}},$$

where $\lesssim$ means less than or equal to asymptotically and the last term can be viewed as a conservative estimate of the FDP. Motivated by this observation, we define the threshold for the flexible Barber–Candès procedure as

$$T_{\text{FBC}} = \sup\left\{0 < t \leqslant T_{\text{up}} : \frac{1 + \sum_{i=1}^n \mathbb{1}\{\varphi_i(1-p_i) \leqslant t\}}{1 \vee \sum_{i=1}^n \mathbb{1}\{\varphi_i(p_i) \leqslant t\}} \leqslant \alpha\right\}, \tag{12}$$

which is the largest cut-off such that the FDP estimate is bounded above by α, where T_{up} satisfies $T_{\text{up}} < \min_i \varphi_i(0.5)$. The flexible Barber–Candès procedure rejects H_i whenever $\varphi_i(p_i) \leqslant T_{\text{FBC}}$.

PROPOSITION 4. *Suppose that Assumption 1 holds and that the null p-values satisfy (11). Assuming that φ_i is a monotonic increasing and continuous function for all i, then the flexible Barber–Candès procedure ensures* FDR *control at level α.*

Remark 1. Compared to the flexible Benjamini–Hochberg procedure, the flexible Barber–Candès approach affords us greater flexibility in selecting φ_i, as it no longer requires φ_i to be strictly increasing, and its generalized inverse function does not have to fulfill the condition in Proposition 2 to achieve FDR control.

Example 2. Suppose that the p-value p_i is generated independently from the two-group mixture model: $\pi_i f_0 + (1 - \pi_i) f_{1,i}$, where $\pi_i \in (0, 1)$ is the mixing proportion and f_0 and $f_{1,i}$ denote the p-value distributions under the null and alternative, respectively. The local FDR is defined as $\text{Lfdr}_i(p) = \pi_i f_0(p)/\{\pi_i f_0(p) + (1 - \pi_i) f_{1,i}(p)\}$, which is the posterior probability that the ith hypothesis is under the null given the observed p-value being p. The monotone likelihood ratio assumption (Sun & Cai, 2007) states that $f_{1,i}(p)/f_0(p)$ is decreasing in p. Under this assumption, $\varphi_i(p) = \text{Lfdr}_i(p)$ is monotonically increasing in p and thus fulfills the requirement in Proposition 4. Additionally, it has been shown in the literature that the rejection rule $\varphi_i(p_i) = \text{Lfdr}_i(p_i) \leqslant t$ is optimal in the sense of maximizing the expected number of true positives among the decision rules that control the marginal FDR at level α; see, e.g., Sun & Cai (2007), Lei & Fithian (2018) and Cao et al. (2022).

4.4. *Connections between the flexible Barber–Candès and e-Benjamini–Hochberg procedures*

We show that the flexible Barber–Candès procedure is equivalent to the e-Benjamini–Hochberg procedure with the e-values

$$e_i = \frac{n\mathbb{1}\{\varphi_i(p_i) \leqslant T_{\text{FBC}}\}}{1 + \sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \leqslant T_{\text{FBC}}\}}, \tag{13}$$

where T_{FBC} is defined in (12). By equation (B.1) in the Supplementary Material, we have $\sum_{i\in\mathcal{H}_0} \mathbb{E}[e_i] \leqslant n$, which implies that the corresponding e-Benjamini–Hochberg procedure controls the

Table 1. *The selections of $m(t)$ and $R_i(t)$ for different methods*

Method	$m(t)$	$R_i(t)$
BH	nt	$\mathbb{1}\{p_i \leqslant t\}$
FBH	$ng(t)$	$\mathbb{1}\{\varphi_i(p_i) \leqslant t\}$
ST	$n\pi_0^{\lambda}t$	$\mathbb{1}\{p_i \leqslant t\}$
BC	$1+\sum_{i=1}^{n}\mathbb{1}\{p_i \geqslant 1-t\}$	$\mathbb{1}\{p_i \leqslant t\}$
FBC	$1+\sum_{i=1}^{n}\mathbb{1}\{\varphi_i(1-p_i) \leqslant t\}$	$\mathbb{1}\{\varphi_i(p_i) \leqslant t\}$

BH, the Benjamini–Hochberg procedure; FBH, the flexible Benjamini–Hochberg procedure; ST, Storey's procedure; BC, the Barber–Candès procedure; FBC, the flexible Barber–Candès procedure.

FDR at the desired level. Furthermore, the following theorem shows that the *e*-Benjamini–Hochberg procedure with the *e*-values defined above is equivalent to the flexible Barber–Candès procedure.

Theorem 3. *Let $\mathcal{S}_{\mathrm{FBC}}$ be the set of rejections obtained through the flexible Barber–Candès procedure at the* FDR *level α, and let $\mathcal{S}_{\mathrm{eBH}}$ represent the set of rejections obtained from the e-Benjamini–Hochberg procedure at the same* FDR *level α, with the e-values defined in* (13)*. Then we have $\mathcal{S}_{\mathrm{FBC}} = \mathcal{S}_{\mathrm{eBH}}$.*

4.5. *A unified viewpoint*

The connection between the aforementioned procedures and the *e*-Benjamini–Hochberg procedure can be unified in the following way. Suppose that we reject the ith hypothesis if $R_i(T) = 1$ with

$$T = \sup\left\{t \in \mathcal{D}\colon \frac{m(t)}{1 \vee \sum_{j=1}^{n} R_j(t)} \leqslant \alpha\right\}.$$

Here $\mathcal{D}$ denotes the domain of the threshold, $m(t)$ is an estimate of the number of false discoveries and $\sum_{j=1}^{n} R_j(t)$ is the total number of rejections, with $R_j(t)$ indicating whether or not the jth hypothesis should be rejected at threshold t. The corresponding *e*-Benjamini–Hochberg procedure is defined based on the *e*-values $e_i = nR_i(T)/m(T)$ for $1 \leqslant i \leqslant n$. The selections of $m(t)$ and $R_i(t)$ for different methods are summarized in Table 1.

5. Aggregating and assembling *e*-values

We have shown that the Benjamini–Hochberg and Barber–Candès procedures and their generalized versions are all equivalent to the *e*-Benjamini–Hochberg procedure based on specific forms of *e*-values. This equivalence opens up new possibilities for designing multiple testing procedures by aggregating/combining *e*-values from different procedures, or the same procedure with different tuning quantities, or assembling *e*-values from various subsets of the data. We present the following results for combining and assembling *e*-values, which have not been explicitly stated in the existing literature. We refer the reader to the Supplementary Material for a more detailed illustration. The result in Proposition 5 below is under the case where we have L sets of *e*-values from L procedures, while the result in Proposition 6 below is under the case where we have L sets of *e*-values obtained from L different datasets.

Proposition 5. *Suppose that we have L sets of e-values $\{e_i^l\colon i \in [n]\}_{l=1}^{L}$ from L different procedures, where $\{e_i^l\}_{l=1}^{L}$ are the L e-values associated with H_i and $\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i^l] \leqslant n$. Let $e_i = \sum_{l=1}^{L} w_{l,i}e_i^l$ be the weighted e-value, where $w_{l,i} \geqslant 0$ is the aggregating weight. If $\sum_{l=1}^{L}\max_i w_{l,i} \leqslant 1$, the weighted e-values satisfy* (5)*.*

The condition $\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i^l] \leqslant n$ for all l ensures that each procedure controls the FDR. Proposition 5 suggests that the *e*-Benjamini–Hochberg procedure applied to the weighted *e*-values still controls the FDR. Moreover, when $\mathbb{E}[e_i^l] \leqslant 1$ for all i and l, the condition $\sum_{l=1}^{L} \max_i w_{l,i} \leqslant 1$ can be relaxed to $\sum_{l=1}^{L} \sum_{i=1}^{n} w_{l,i}/n \leqslant 1$.

Proposition 6. *Suppose that we have L sets of e-values $\{e_i^l : i \in \mathcal{G}_l, |\mathcal{G}_l| = n_l\}$ from L different datasets, where $\bigcup_l \mathcal{G}_l = [n]$, $\mathcal{G}_{l_1} \cap \mathcal{G}_{l_2} = \emptyset$ if $l_1 \neq l_2$, e_i^l is associated with hypothesis H_i and $\sum_{i \in \mathcal{G}_l \cap \mathcal{H}_0} \mathbb{E}[e_i^l] \leqslant n_l$. Let $e_i = w_{l,i} e_i^l$ be the weighted e-value, where $w_{l,i} \geqslant 0$ is the assembling weight. If $\sum_{l=1}^{L} n_l \max_{i \in \mathcal{G}_l} w_{l,i} \leqslant n$, the weighted e-values satisfy (5).*

The condition $\sum_{i \in \mathcal{G}_l \cap \mathcal{H}_0} \mathbb{E}[e_i] \leqslant n_l$ ensures that the FDR is controlled within each $\mathcal{G}_l$. Proposition 6 suggests that the e-Benjamini–Hochberg procedure applied to the weighted e-values controls the overall FDR.

Supplementary material

The Supplementary Material includes all proofs.

References

Banerjee, T., Gang, B. & He, J. (2023). Harnessing the collective wisdom: fusion learning using decision sequences from diverse sources. *arXiv:* 2308.11026v1.

Barber, R. F. & Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43**, 2055–85.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B **57**, 289–300.

Cao, H., Chen, J. & Zhang, X. (2022). Optimal false discovery rate control for large scale multiple testing with auxiliary information. *Ann. Statist.* **50**, 807–57.

Dunn, R., Ramdas, A., Balakrishnan, S. & Wasserman, L. (2023). Gaussian universal likelihood ratio testing. *Biometrika* **110**, 319–37.

Ferreira, J. & Zwinderman, A. (2006). On the Benjamini–Hochberg method. *Ann. Statist.* **34**, 1827–49.

Genovese, C. R., Roeder, K. & Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* **93**, 509–24.

Grünwald, P., de Heide, R. & Koolen, W. (2024). Safe testing. *J. R. Statist. Soc.* B, doi: 10.1093/jrsssb/qkae011.

Ignatiadis, N., Wang, R. & Ramdas, A. (2023). *E*-values as unnormalized weights in multiple testing. *Biometrika* **111**, 417–39.

Lei, L. & Fithian, W. (2018). Adapt: an interactive procedure for multiple testing with side information. *J. R. Statist. Soc.* B **80**, 649–79.

Peña, E. A., Habiger, J. D. & Wu, W. (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist.* **39**, 556–83.

Ren, Z. & Barber, R. F. (2024). Derandomised knockoffs: leveraging e-values for false discovery rate control. *J. R. Statist. Soc.* B **86**, 122–54.

Shafer, G. (2021). Testing by betting: a strategy for statistical and scientific communication. *J. R. Statist. Soc.* A **184**, 407–31.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc.* B **64**, 479–98.

Storey, J. D., Taylor, J. E. & Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Statist. Soc.* B **66**, 187–205.

Sun, W. & Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Am. Statist. Assoc.* **102**, 901–12.

Vovk, V. & Wang, R. (2021). E-values: calibration, combination and applications. *Ann. Statist.* **49**, 1736–54.

Wang, R. & Ramdas, A. (2022). False discovery rate control with e-values. *J. R. Statist. Soc.* B **84**, 822–52.

Xu, Z. & Ramdas, A. (2024). More powerful multiple testing under dependence via randomization. *arXiv:* 2305.11126v3.

Xu, Z., Wang, R. & Ramdas, A. (2021). A unified framework for bandit multiple testing. In *Proc. 35th Int. Conf. Neural Info. Proces. Syst.*, pp. 16833–45. Red Hook, NY: Curran Associates.

Zhang, X. & Chen, J. (2022). Covariate adaptive false discovery rate control with applications to omics-wide multiple testing. *J. Am. Statist. Assoc.* **117**, 411–27.

[*Received on* 19 *December* 2023. *Editorial decision on* 30 *September* 2024]