# A Note on E-values and Multiple Testing

Guanxun Li[1], Xianyang Zhang[2,†]
[1] Department of Statistics, Beijing Normal University at Zhuhai
[2] Department of Statistics, Texas A&M University

**Abstract**

We discover a connection between the Benjamini-Hochberg (BH) procedure and the e-BH procedure [Wang and 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 in Barber and Candès [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.

*Keywords:* Benjamini-Hochberg procedure; E-values; 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 (BH) procedure [Benjamini and 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 and Candès [2015] proposed a model-free FDR-controlling (BC) 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., Grünwald et al. [2020], Shafer [2021], Vovk and Wang [2021], Xu et al. [2021], Ignatiadis et al. [2023], Dunn et al. [2023], Xu and Ramdas [2023]. In particular, Wang and Ramdas [2022] proposed a multiple testing procedure named e-BH procedure by applying the BH 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 BH and e-BH 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 BH and BC procedures, which can have a more general form for the rejection rules. All these connections can be summarized in

[†]Corresponding author: zhangxiany@stat.tamu.edu

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 data sets.

# 2 Preliminaries

## 2.1 False discovery rate (FDR)

Suppose we are interested in testing $n$ hypotheses $(H_1, \dots, H_n)$ simultaneously. Let $\theta = (\theta_1, \dots, \theta_n) \in \{0,1\}^n$ 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 $\delta = (\delta_1, \dots, \delta_n) \in \{0,1\}^n$ a decision rule for the $n$ hypotheses, where we reject the $i$th hypothesis if and only if $\delta_i = 1$. The FDR for the decision rule $\delta$ is defined as the expectation of the false discovery proportion (FDP), i.e.,

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

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 pre-specified number $\alpha \in (0,1)$.

## 2.2 The Benjamini-Hochberg procedure

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

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

with $R(t) = \sum_{i=1}^n \mathbb{1}\{p_i \le t\}$ being the number of rejections given the 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 super-uniform under the null if $P_0(p \le t) \le t$ 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 super-uniform, the BH procedure at level $\alpha$ controls the FDR at the level $\alpha n_0/n \le \alpha$, where $n_0$ is the number of hypotheses under the null [Ferreira and Zwinderman, 2006].

## 2.3 Storey's procedure

Storey's procedure [Storey, 2002, Storey et al., 2004] improves the BH procedure by using the p-values to estimate the null proportion $\pi_0 := n_0/n$. Specifically, we define

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

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

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

When $\pi_0^\lambda < 1$, Storey's procedure makes more rejections than the BH 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 and Candès procedure

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

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

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

### 2.5 E-values and e-BH procedure

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

**Proposition 1** (Theorem 2 of Wang and Ramdas [2022])**.** *Suppose the non-negative random variables $\{e_i\}$ satisfy*

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

*where $\mathcal{H}_0 = \{1 \le i \le n : \theta_i = 0\}$. Then, the $\alpha$-level e-BH procedure applied to $\{e_i\}$ controls the FDR at the level $\alpha$, regardless of the dependence structure among $\{e_i\}$.*

*Proof.* See Section B in the Appendices. □

In the multiple testing context, the requirement that $\mathbb{E}[e] \le 1$ in the definition of e-values can be relaxed. More precisely, we will 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 BH/Storey/BC and e-BH procedures

## 3.1 Connection between the BH and e-BH procedures

We first establish the equivalence between the BH procedure and the corresponding e-BH 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 BH and e-BH procedures, we define the e-value associated with $H_i$ to be

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

where $T_{\mathrm{BH}}$ 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 BH procedure. Denote $[n] = \{1, 2, \cdots, n\}$ for any positive integer $n$. Under Assumption 1 and if the null p-values are super-uniform, by Lemmas 3-4 in 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 Section B.1. Thus, by Proposition 1, the corresponding e-BH procedure controls the FDR at the desired level. Moreover, we claim that the e-BH procedure based on the e-values defined in (6) is equivalent to the BH procedure in the sense that they produce the same set of rejections; see Theorem 2 for a precise statement.

## 3.2 Connection between the Storey's and e-BH procedures

Define the e-value associated with $H_i$ to be

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

where $\pi_0^\lambda$ is defined in (2) and $T_{\mathrm{ST}}$ is given in (3). We have the following results.

**Theorem 1.** *Suppose Assumption 1 holds and 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* $\alpha$*, and let* $\mathcal{S}_{\mathrm{eBH}}$ *represent the set of rejections obtained from the e-BH procedure at the same FDR level* $\alpha$*, with the e-values defined in* (7)*. Then we have* $\mathcal{S}_{\mathrm{ST}} = \mathcal{S}_{\mathrm{eBH}}$*.*

*Proof.* See Section A in the Appendices. □

## 3.3 Connection between the BC and e-BH procedures

As noted in the recent work by Ren and Barber [2024], the BC procedure is equivalent to the e-BH procedure based on the following e-values:

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

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

# 4 The Flexible BH and BC procedures

## 4.1 Flexible BH procedure

We generalize the BH procedure to allow the rejection rule to take the form of $\varphi_i(p_i) \le t$, where $\varphi_i$ 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 $\varphi_i$, $g$ be some strictly increasing function and $g^{-1}$ be the inverse function of $g$. Consider the rejection threshold given by

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

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

**Proposition 2.** *Suppose Assumption 1 holds and the null p-values are super-uniform. The* FBH *procedure controls the FDR at the level $C\alpha$, where*

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

*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* FBH *procedure controls the FDR at level $\alpha$.*

*Proof.* See Section B in the Appendices. □

Proposition 2 broadens and enhances Theorem 7.1 from 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 $g(t) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(t)$, 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 FBH procedure controls the FDR at level $\alpha$ when $C = \sum_{i \in \mathcal{H}_0} \sup_{t \in C_\alpha} \frac{F_i(t)}{ng(t)} \le 1$, which is weaker than the condition $n_0 \sup_{i \in \mathcal{H}_0} \sup_{t \in C_\alpha} \frac{F_i(t)}{ng(t)} \le 1$ required in Theorem 7.1 of Peña et al. [2011]. Please refer to Section C for a more comprehensive comparison between FBH and other related works.

The following example illustrates that the FBH procedure aligns with the weighted BH procedure for particular choices of $g$ and $\varphi_i$.

**Example 1.** Let $g(t) = t$ and $\varphi_i(p) = p/\omega_i$, where $\omega_i$ denotes the weight for the $i$th hypothesis with $\omega_i > 0$ and $\sum_{i=1}^{n} \omega_i = n$. The FBH procedure associated with this choice of $\varphi_i$ and $g$ corresponds to the weighted BH procedure first introduced by Genovese et al. [2006]. In this case, the rejection threshold can be expressed as

$$T_{\mathrm{FBH}} = \sup\left\{0 < t \le 1 \colon \frac{nt}{1 \vee R(t)} \le \alpha\right\},$$

where $R(t) = \sum_{i=1}^{n} \mathbb{1}\{p_i/w_i \le t\}$.

### 4.2 Connection between the FBH and e-BH procedures

Analogous to the BH procedure, we show that the FBH procedure is equivalent to the e-BH procedure applied to the following e-values:

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

where $T_{\mathrm{FBH}}$ 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-BH procedure with e-values defined in* (10) *controls the FDR at the level* $C\alpha$*, where* $C$ *is defined in* (9).

*Proof.* See Section B in the Appendices. □

Additionally, we can prove that the e-BH procedure and the FBH procedure deliver the same set of rejections.

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

*Proof.* See Section A in the Appendices. □

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

### 4.3 Flexible BC procedure

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

$$P(p_i \leq a) \leq P(p_i \geq 1-a) = P(1-p_i \leq a), \qquad \text{for all } 0 \leq a \leq 0.5. \tag{11}$$

Condition (11) is weaker than the mirror conservativeness in Lei and Fithian [2018], and it can be shown that super-uniformity implies (11). Indeed, $P(1-p_i \leq a) \geq 1 - P(p_i \leq 1-a) \geq 1-(1-a) = a \geq P(p_i \leq a)$. Assume that $\varphi_i$ is an increasing and continuous function, and define $F_i(x) = \sup\{0 \leq p \leq 1\colon \varphi_i(p) \leq x\}$. We claim that $P(\varphi_i(p_i) \leq b) = P(p_i \leq F_i(b))$. To see this, consider two cases. If $\varphi_i(p_i) \leq b$, by the definition of $F_i(b)$, we have $p_i \leq F_i(b)$. On the other hand, if $p_i \leq F_i(b)$, then $\varphi_i(p_i) \leq \varphi_i(F_i(b)) = \lim_{p\uparrow F_i(b)} \varphi_i(p) \leq b$, where we use the fact that $\varphi_i$ is increasing to get the two inequalities, and the equality is due to the continuity of $\varphi_i$. Therefore, the above claim together with equation (11) implies that $P(\varphi_i(p_i) \leq b) = P(p_i \leq F_i(b)) \leq P(1-p_i \leq F_i(b)) = P(\varphi_i(1-p_i) \leq b)$, for all $\varphi_i(0) \leq b \leq \varphi_i(0.5)$. Hence, we have

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

where "$\lesssim$" means "less 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 FBC procedure as

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

which is the largest cutoff such that the FDP estimate is bounded above by $\alpha$, where $T_{\mathrm{up}}$ satisfies $T_{\mathrm{up}} < \min_i \varphi_i(0.5)$. The FBC procedure rejects $H_i$ whenever $\varphi_i(p_i) \leq T_{\mathrm{FBC}}$.

**Proposition 4.** *Suppose that Assumption 1 holds and the null p-values satisfy Condition* (11)*. Assuming $\varphi_i$ is a monotonic increasing and continuous function for all $i$, then the* FBC *procedure ensures FDR control at level $\alpha$.*

*Proof.* See Section B in the Appendices. □

**Remark 1.** Compared to the FBH procedure, the FBC approach affords us greater flexibility in selecting $\varphi_i$, as it no longer requires $\varphi_i$ 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 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 $\mathrm{Lfdr}_i(p) = \pi_i f_0(p)/\{\pi_i f_0(p) + (1-\pi_i) f_{1,i}(p)\}$, which is the posterior probability that the $i$th hypothesis is under the null given the observed p-value being $p$. The monotone likelihood ratio assumption [Sun and Cai, 2007] states that $f_{1,i}(p)/f_0(p)$ is decreasing in $p$. Under this assumption, $\varphi_i(p) = \mathrm{Lfdr}_i(p)$ 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 $\varphi_i(p_i) = \mathrm{Lfdr}_i(p_i) \leq t$ is optimal in the sense of maximizing the expected number of true positives among the decision rules that control the marginal FDR at level $\alpha$, see, e.g., Sun and Cai [2007], Lei and Fithian [2018], Cao et al. [2022].

## 4.4 Connection between the FBC and e-BH procedures

We show that the FBC procedure is equivalent to the e-BH procedure with the e-values:

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

where $T_{\mathrm{FBC}}$ is defined in (12). By equation (B.1) in the proof of Proposition 4, we have $\sum_{i\in\mathcal{H}_0} \mathbb{E}[e_i] \leq n$, which implies that the corresponding e-BH procedure controls the FDR at the desired level. Furthermore, the following theorem shows that the e-BH procedure with the e-values defined above is equivalent to the FBC procedure.

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

*Proof.* See Section A in the Appendices. □

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

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

### 4.5 A unified viewpoint

The connection between the aforementioned procedures and the e-BH procedure can be unified in the following way. Suppose we reject the $i$th hypothesis if $R_i(T) = 1$ with

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

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 $R_j(t)$ being the indicator function that indicates whether the $j$th hypothesis should be rejected or not at the threshold $t$. The corresponding e-BH procedure is defined based on the $e$-values $e_i = nR_i(T)/m(T)$ for $1 \le i \le n$. 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 BH and BC procedures and their generalized versions are all equivalent to the e-BH 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 readers to Section D in Appendices for a more detailed illustration. The result in Proposition 5 is under the case where we have $L$ sets of e-values from $L$ procedures, while the result in Proposition 6 is under the case where we have $L$ sets of e-values obtained from $L$ different datasets.

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

*Proof.* See Section D in the Appendices. □

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

**Proposition 6.** *Suppose 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 $\cup_l \mathcal{G}_l = [n]$, $\mathcal{G}_{l_1} \cap \mathcal{G}_{l_2} = \emptyset$ if $l_1 \ne l_2$, $e_i^l$ is associated with the hypothesis $H_i$ and*

$\sum_{i\in\mathcal{G}_l\cap\mathcal{H}_0}\mathbb{E}[e_i^l] \leq n_l$. *Let* $e_i = w_{l,i}e_i^l$ *be the weighted e-value, where* $w_{l,i} \geq 0$ *is the assembling weight. If* $\sum_{l=1}^L n_l \max_{i\in\mathcal{G}_l} w_{l,i} \leq n$*, the weighted e-values satisfy Condition* (5).

*Proof.* See Section D in the Appendices. □

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

# Appendices

## A Proofs of the main results

We first state the following propositions whose proofs are deferred to Appendix B. These results will be used frequently in the subsequent proofs.

**Proposition A.1** (Lemma 6 of Barber et al. [2020])**.** *Let* $T_{\mathrm{BC},i}$ *be the threshold for the BC method when* $p_i$ *is replaced with* $\min\{p_i, 1-p_i\}$. *For any* $i$, $j$, *if* $\min(p_i,p_j) \geq 1 - \max\{T_{\mathrm{BC},i}, T_{\mathrm{BC},j}\}$, *then we have* $T_{\mathrm{BC},i} = T_{\mathrm{BC},j}$.

**Proposition A.2.** *Suppose the assumptions in Proposition 4 hold. Let* $T_i$ *be the threshold for the FBC procedure when* $p_i$ *is replaced with* $\min\{p_i, 1-p_i\}$. *For any* $i$, $j$, *if* $\max\{\varphi_i(1-p_i), \varphi_j(1-p_j)\} \leq \max\{T_i, T_j\}$, *then we have* $T_i = T_j$.

### A.1 Proof of Theorem 1

*Proof.* Let us first prove that the e-values defined by (7) satisfy (5). Under Assumption 1, by Lemmas 3-4 in Storey et al. [2004], $\sum_{i\in\mathcal{H}_0}\mathbb{1}\{p_i \leq t\}/t$ for $0 < t \leq 1$ is a martingale with time running backwards with respect to the filtration $\mathcal{F}_t = \sigma(\mathbb{1}\{p_i \leq s\} : t \leq s \leq 1, i = 1, 2, \ldots, n)$, and $T_{\mathrm{ST}}$ is a stopping time with respect to $\mathcal{F}_t$, where $\mathcal{F}_t$ is the sigma field generated by $\mathbb{1}\{p_i \leq s\}$ for $t \leq s \leq 1$. By the optional stopping theorem, we have

$$\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i] = \mathbb{E}\left[\frac{\sum_{i\in\mathcal{H}_0}\mathbb{1}\{p_i \leq T_{\mathrm{ST}}\}}{\pi_0^\lambda T_{\mathrm{ST}}}\right] = \mathbb{E}\left[\frac{n(1-\lambda)}{1+n-R(\lambda)}\sum_{i\in\mathcal{H}_0}\frac{\mathbb{1}\{p_i \leq \lambda\}}{\lambda}\right].$$

Denote $V(\lambda) = \sum_{i\in\mathcal{H}_0}\mathbb{1}\{p_i \leq \lambda\}$. Since

$$1 + n - R(\lambda) = 1 + (n_0 - V(\lambda)) + \{(n - n_0) - (R(\lambda) - V(\lambda))\} \geq 1 + n_0 - V(\lambda),$$

we have

$$\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i] \leq \frac{n(1-\lambda)}{\lambda}\mathbb{E}\left[\frac{V(\lambda)}{1+n_0-V(\lambda)}\right].$$

Because the p-values follow the uniform distribution on $[0, 1]$ under the null, we have $V(\lambda) \sim \mathrm{Bin}(n_0, \lambda)$, which implies

$$
\begin{aligned}
\mathbb{E}\left[\frac{V(\lambda)}{1+n_0-V(\lambda)}\right] &= \sum_{i=1}^{n_0} P\big(V(\lambda)=i\big)\frac{i}{1+n_0-i} \\
&= \sum_{i=1}^{n_0} \binom{n_0}{i} \lambda^i (1-\lambda)^{n_0-i} \frac{i}{1+n_0-i} \\
&= \sum_{i=1}^{n_0} \lambda^i (1-\lambda)^{n_0-i} \frac{n_0! \times i}{(1+n_0-i) \times (n_0-i)! \times i!} \\
&= \sum_{i=1}^{n_0} \lambda^i (1-\lambda)^{n_0-i} \frac{n_0!}{(1+n_0-i)!(i-1)!} \\
&= \sum_{j=0}^{n_0-1} \lambda^{j+1} (1-\lambda)^{n_0-j-1} \frac{n_0!}{(n_0-j)!j!} \\
&= \frac{\lambda}{1-\lambda} \sum_{j=0}^{n_0-1} \lambda^{j} (1-\lambda)^{n_0-j} \frac{n_0!}{(n_0-j)!j!} \\
&= \frac{\lambda}{1-\lambda}\big((\lambda+1-\lambda)^{n_0} - \lambda^{n_0}\big) \\
&= \frac{\lambda(1-\lambda^{n_0})}{1-\lambda}.
\end{aligned}
$$

Hence,

$$
\sum_{i \in \mathcal{H}_0} \mathbb{E}[e_i] \le n(1-\lambda^{n_0}) \le n.
$$

The last part is to prove that $\mathcal{S}_{\mathrm{ST}} = \mathcal{S}_{\mathrm{eBH}}$. Let $\hat{k}$ be the cardinality of $\mathcal{S}_{\mathrm{ST}}$. If $i \notin \mathcal{S}_{\mathrm{ST}}$, then $p_i > T_{\mathrm{ST}}$ and thus $e_i = 0$. This implies that the $i$th hypothesis is not rejected by the e-BH procedure, and hence $\mathcal{S}_{\mathrm{eBH}} \subset \mathcal{S}_{\mathrm{ST}}$. Conversely, if $i \in \mathcal{S}_{\mathrm{eBH}}$, we have $p_i \le T_{\mathrm{ST}} \le \frac{\alpha\hat{k}}{n\pi_0^\lambda}$, leading to

$$
e_i = \frac{1}{\pi_0^\lambda T_{\mathrm{ST}}} \ge \frac{n}{\alpha\hat{k}}.
$$

Define the ordered e-values $e_{(1)} \ge e_{(2)} \ge \cdots \ge e_{(n)}$, we have $e_{(\hat{k})} \ge n/(\alpha\hat{k})$, which indicates that $|\mathcal{S}_{\mathrm{eBH}}| \ge \hat{k}$. Because $\mathcal{S}_{\mathrm{eBH}} \subset \mathcal{S}_{\mathrm{ST}}$, it is clear that $\mathcal{S}_{\mathrm{eBH}} = \mathcal{S}_{\mathrm{ST}}$. □

### A.2 Proof of Theorem 2

*Proof.* Let $\hat{k}$ be the cardinality of $\mathcal{S}_{\mathrm{FBH}}$. If $i \notin \mathcal{S}_{\mathrm{FBH}}$, then $\varphi_i(p_i) > T_{\mathrm{FBH}}$ and thus $e_i = 0$. This implies that the $i$th hypothesis is not rejected by the e-BH procedure, and hence $\mathcal{S}_{\mathrm{eBH}} \subset \mathcal{S}_{\mathrm{FBH}}$. Conversely, if $i \in \mathcal{S}_{\mathrm{FBH}}$, we have $\varphi_i(p_i) \le T_{\mathrm{FBH}} \le g^{-1}\big(\alpha\hat{k}/n\big)$, leading to

$$
e_i = \frac{1}{g(T_{\mathrm{FBH}})} \ge \frac{n}{\alpha\hat{k}}
$$

as $g$ is strictly increasing. Define $e_{(i)} = \mathbb{1}\{q_{(i)} \leq T_{\text{FBH}}\}/g(T_{\text{FBH}})$ for $1 \leq i \leq n$, where $q_{(1)} \leq \cdots \leq q_{(n)}$ are the order statistics of $\{q_i = \varphi_i(p_i)\}_{i=1}^n$. Let $\hat{k}$ represent the maximum $i$ for which $q_{(i)} \leq T_{\text{FBH}}$. We get

$$e_{(\hat{k})} \geq \frac{n}{\alpha\hat{k}}, \tag{A.1}$$

which indicates that $|\mathcal{S}_{\text{eBH}}| \geq \hat{k}$. Because $\mathcal{S}_{\text{eBH}} \subset \mathcal{S}_{\text{FBH}}$, it is clear that $\mathcal{S}_{\text{eBH}} = \mathcal{S}_{\text{FBH}}$. □

### A.3 Proof of Theorem 3

*Proof.* Let $\hat{k}$ be the cardinality of $\mathcal{S}_{\text{FBC}}$. If $i \notin \mathcal{S}_{\text{FBC}}$, then $\varphi_i(p_i) > T_{\text{FBC}}$ and thus $e_i = 0$. Hence the $i$th hypothesis is not rejected by the e-BH procedure, which implies that $\mathcal{S}_{\text{eBH}} \subset \mathcal{S}_{\text{FBC}}$. For the other direction, note that if $i \in \mathcal{S}_{\text{FBC}}$, then $\varphi_i(p_i) \leq T_{\text{FBC}}$ and $1+\sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \leq T_{\text{FBC}}\} \leq \hat{k}\alpha$. Hence, we have

$$e_i \geq \frac{n}{\hat{k}\alpha}.$$

We sort the e-values in descending order as $e_{(1)} \geq \cdots \geq e_{(n)}$. It is clear that $e_{(\hat{k})} \geq n/(\hat{k}\alpha)$. Thus, $|\mathcal{S}_{\text{eBH}}| \geq \hat{k}$, which implies that $\mathcal{S}_{\text{eBH}} = \mathcal{S}_{\text{FBC}}$. □

## B Additional proofs

### B.1 Proof of the Result in Section 3.1

Under Assumption 1 and if the null p-values are super-uniform, by Lemmas 3-4 in Storey et al. [2004], $\sum_{i\in\mathcal{H}_0} \mathbb{1}\{p_i \leq t\}/t$ for $0 < t \leq 1$ is a martingale with time running backwards with respect to the filtration $\mathcal{F}_t = \sigma(\mathbb{1}\{p_i \leq s\} : t \leq s \leq 1, i \in [n])$, and $T_{\text{BH}}$ is a stopping time with respect to $\mathcal{F}_t$, where $\mathcal{F}_t$ is the sigma field generated by $\mathbb{1}\{p_i \leq s\}$ for $t \leq s \leq 1$. By the optional stopping theorem, we have

$$\sum_{i\in\mathcal{H}_0} \mathbb{E}[e_i] = \mathbb{E}\left[\frac{\sum_{i\in\mathcal{H}_0} \mathbb{1}\{p_i \leq T_{\text{BH}}\}}{T_{\text{BH}}}\right] = \mathbb{E}\left[\sum_{i\in\mathcal{H}_0} \mathbb{1}\{p_i \leq 1\}\right] = n_0.$$

### B.2 Proof of Proposition 1

*Proof.* Note that

$$\begin{aligned}
\text{FDP} &= \sum_{i=1}^n \frac{\mathbb{1}\{ie_{(i)} \geq n/\alpha, H_{(i)} \text{ is under the null}\}}{1 \vee \hat{k}} \\
&\leq \sum_{i=1}^n \frac{\mathbb{1}\{ie_{(i)} \geq n/\alpha, H_{(i)} \text{ is under the null}\}}{1 \vee i} \\
&\leq \sum_{i=1}^n \mathbb{1}\{H_{(i)} \text{ is under the null}\} \frac{\alpha e_{(i)}}{n} = \frac{\alpha}{n} \sum_{i\in\mathcal{H}_0} e_i.
\end{aligned}$$

Under Condition (5), we have

$$\text{FDR} = \mathbb{E}[\text{FDP}] \leq \alpha.$$

□

### B.3 Proof of Proposition 2

*Proof.* Let $V_i = \mathbb{1}\{H_i \text{ is rejected}\}$. We have

$$\mathrm{FDP}(T_{\mathrm{FBH}}) = \sum_{i\in\mathcal{H}_0} \frac{V_i}{R(T_{\mathrm{FBH}}) \vee 1} = \sum_{i\in\mathcal{H}_0} \frac{V_i}{ng(T_{\mathrm{FBH}})} \frac{ng(T_{\mathrm{FBH}})}{R(T_{\mathrm{FBH}}) \vee 1} \leq \alpha \sum_{i\in\mathcal{H}_0} \frac{V_i}{ng(T_{\mathrm{FBH}})}.$$

Therefore, we need to bound

$$\mathbb{E}\left[\sum_{i\in\mathcal{H}_0} \frac{V_i}{ng(T_{\mathrm{FBH}})}\right]$$

from above. Observing that, for a given $R$, $T_{\mathrm{FBH}} = T_{\mathrm{FBH}}(R)$ is a deterministic function of $R$, we have:

$$\mathbb{E}\left[\sum_{i\in\mathcal{H}_0} \frac{V_i}{ng(T_{\mathrm{FBH}})}\right] = \sum_{i\in H_0}\sum_{k=1}^{n} \mathbb{E}\left[\frac{V_i \mathbb{1}\{R=k\}}{ng\big(T_{\mathrm{FBH}}(k)\big)}\right] = \sum_{i\in H_0}\sum_{k=1}^{n} \mathbb{E}\left[\frac{V_i \mathbb{1}\{R(p_i\to 0)=k\}}{ng\big(T_{\mathrm{FBH}}(k)\big)}\right],$$

where $R(p_i \to 0)$ is the number of rejections obtained by replacing the p-value $p_i$ with 0. To clarify the second equality, note that if $V_i = 0$, the equation is trivially true. When $V_i = 1$, setting $p_i$ to 0 does not change the number of rejections. By direct calculation,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in\mathcal{H}_0} \frac{V_i}{ng(T_{\mathrm{FBH}})}\right] &= \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \frac{1}{ng\big(T_{\mathrm{FBH}}(k)\big)} \mathbb{E}[\mathbb{1}\{\varphi_i(p_i) \leq T_{\mathrm{FBH}}(k)\}]\mathbb{E}[\mathbb{1}\{R(p_i\to 0)=k\}] \\
&\leq \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \frac{F_i(T_{\mathrm{FBH}}(k))}{ng\big(T_{\mathrm{FBH}}(k)\big)} \mathbb{E}[\mathbb{1}\{R(p_i\to 0)=k\}] \\
&\leq \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \sup_{t\in\mathcal{C}_\alpha} \frac{F_i(t)}{ng(t)} \mathbb{E}[\mathbb{1}\{R(p_i\to 0)=k\}] \\
&\leq C.
\end{aligned}$$

From (8), we know the domain of $T_{\mathrm{FBH}}$ is $C_\alpha$. Therefore, we only need to take the supremum over $\mathcal{C}_\alpha$ in the second inequality. Hence, the proposed method controls the FDR at level $C\alpha$. Additionally, if $g(t) = n^{-1}\sum_{i=1}^{n} F_i(t)$ and $F_i(t) = c_i h(t)$, we obtain

$$C \leq \sum_{i=1}^{n} \sup_{t\in\mathcal{C}_\alpha} \frac{F_i(t)}{\sum_{j=1}^{n} F_j(t)} = \sum_{i=1}^{n} \frac{c_i}{\sum_{j=1}^{n} c_j} = 1.$$

Hence, the proposed method controls the FDR at level $\alpha$. □

### B.4 Proof of Proposition 3

*Proof.* Recall from the proof of Proposition 2 that $T_{\mathrm{FBH}} = T_{\mathrm{FBH}}(R_{\mathrm{FBH}})$ is a deterministic function of $R_{\mathrm{FBH}}$. Additionally, for any $i \in \mathcal{H}_0$ and $\varphi_i(p_i) \leq T_{\mathrm{FBH}}$, replacing the p-value $p_i$ with 0 does not change the number of rejections, i.e., $R_{\mathrm{FBH}}(p_i \leftarrow 0) = R_{\mathrm{FBH}}$. Therefore, we can infer that

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i\in\mathcal{H}_0} e_i\right] &= \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \frac{\mathbb{E}[\mathbb{1}\{\varphi_i(p_i) \le T_{\mathrm{FBH}}(k), R_{\mathrm{FBH}} = k\}]}{g\big(T_{\mathrm{FBH}}(k)\big)} \\
&= \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \frac{\mathbb{E}[\mathbb{1}\{\varphi_i(p_i) \le T_{\mathrm{FBH}}(k), R_{\mathrm{FBH}}(p_i \leftarrow 0) = k\}]}{g\big(T_{\mathrm{FBH}}(k)\big)} \\
&\le \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \frac{F_i\big(T_{\mathrm{FBH}}(k)\big)\mathbb{E}[\mathbb{1}\{R_{\mathrm{FBH}}(p_i \leftarrow 0) = k\}]}{g\big(T_{\mathrm{FBH}}(k)\big)} \\
&\le n \sum_{i\in\mathcal{H}_0}\sum_{k=1}^{n} \sup_{t\in\mathcal{C}_\alpha} \frac{F_i(t)}{ng(t)} \mathbb{E}[\mathbb{1}\{R_{\mathrm{FBH}}(p_i \to 0) = k\}] \\
&\le nC.
\end{aligned}
$$

The final result can be obtained by going through the proof of Proposition 1. □

### B.5 Proof of Proposition 4

*Proof.* Write $T = T_{\mathrm{FBC}}$ for the ease of notation. First note that

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\mathcal{H}_0} \frac{\mathbb{1}\{\varphi_i(p_i) \le T\}}{1 \vee \sum_{j=1}^n \mathbb{1}\{\varphi_j(p_j) \le T\}}\right] \\
=& \sum_{i\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \le T\}}{1 \vee \sum_{j=1}^n \mathbb{1}\{\varphi_j(p_j) \le T\}} \frac{1 + \sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \le T\}}{1 + \sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \le T\}}\right] \\
\le& \alpha \sum_{i\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \le T\}}{1 + \sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \le T\}}\right].
\end{aligned}
$$

Hence, we only need to show that

$$
\sum_{i\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \le T\}}{1 + \sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \le T\}}\right] \le 1.
$$

One approach to prove FDR control is through the construction of a super-martingale and the use of the optional stopping theorem. Here, we employ an alternative argument based on the leave-one-out technique. Let $\tilde{p}_i = \min\{p_i, 1-p_i\}$ and $p_{-i} = (p_1, \ldots, p_{i-1}, \tilde{p}_i, p_{i+1}, \ldots, p_n)$. Define $T_i = T(p_{-i})$, where we view $T$ as a function of the p-values. Notice that if $\varphi_i(p_i) \le T$, then we have

$$
\varphi_i(p_i) \le T \le T_{\mathrm{up}} < \varphi_i(0.5),
$$

which implies that $p_i < 0.5$ since $\varphi_i$ is increasing. Hence, if the $i$th hypothesis is rejected, then $p_i = \tilde{p}_i$. Thus, $\mathbb{1}\{\varphi_i(p_i) \le T\} = \mathbb{1}\{\varphi_i(p_i) \le T_i\}$, which further implies that

$$
\mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \le T\}}{1 + \sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \le T\}}\right] = \mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \le T_i\}}{1 + \sum_{j\ne i} \mathbb{1}\{\varphi_j(1-p_j) \le T_i\}}\right],
$$

where we use the fact that if $p_i < 0.5$, then $\varphi_i(1-p_i) \geq \varphi_i(0.5) > T_{\text{up}} \geq T_i$. Let $\mathcal{F}_i$ be the sigma algebra generated by $p_{-i}$. For $i \in \mathcal{H}_0$, we have

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \leq T\}}{1+\sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \leq T\}}\right] =& \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \leq T_i\}}{1+\sum_{j\neq i} \mathbb{1}\{\varphi_j(1-p_j) \leq T_i\}}\middle|\mathcal{F}_i\right]\right] \\
=& \mathbb{E}\left[\frac{1}{1+\sum_{j\neq i} \mathbb{1}\{\varphi_j(1-p_j) \leq T_i\}}\mathbb{E}\left[\mathbb{1}\{\varphi_i(p_i) \leq T_i\}|\mathcal{F}_i\right]\right] \\
\leq& \mathbb{E}\left[\frac{1}{1+\sum_{j\neq i} \mathbb{1}\{\varphi_j(1-p_j) \leq T_i\}}\mathbb{E}\left[\mathbb{1}\{\varphi_i(1-p_i) \leq T_i\}|\mathcal{F}_i\right]\right] \\
=& \mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(1-p_i) \leq T_i\}}{1+\sum_{j\neq i} \mathbb{1}\{\varphi_j(1-p_j) \leq T_i\}}\right],
\end{aligned}$$

where we use the assumption that $p_i$ satisfies Condition (11) to get the inequality. By Proposition A.2, we have

$$\frac{\mathbb{1}\{\varphi_i(1-p_i) \leq T_i\}}{1+\sum_{j\neq i} \mathbb{1}\{\varphi_j(1-p_j) \leq T_i\}} = \frac{\mathbb{1}\{\varphi_i(1-p_i) \leq T_i\}}{1+\sum_{j\neq i} \mathbb{1}\{\varphi_j(1-p_j) \leq T_j\}} = \frac{\mathbb{1}\{\varphi_i(1-p_i) \leq T_i\}}{\sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \leq T_j\}}.$$

If $\varphi_i(1-p_i) > T_i$, both sides are equal to 0. If $\varphi_i(1-p_i) \leq T_i$, we claim that $\mathbb{1}\{\varphi_j(1-p_j) \leq T_i\} = \mathbb{1}\{\varphi_j(1-p_j) \leq T_j\}$. Indeed, if $\varphi_j(1-p_j) > T_i$ but $\varphi_j(1-p_j) \leq T_j$, then we have $T_i < T_j$. Hence, $\varphi_i(1-p_i) \leq T_i < T_j$. By proposition A.2, we have $T_i = T_j$, which contradicts with the assumption $T_i < T_j$. The other direction can be proved similarly.

Hence,

$$\sum_{i\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbb{1}\{\varphi_i(p_i) \leq T\}}{1+\sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \leq T\}}\right] \leq \mathbb{E}\left[\frac{\sum_{i\in\mathcal{H}_0} \mathbb{1}\{\varphi_i(1-p_i) \leq T_i\}}{\sum_{j=1}^n \mathbb{1}\{\varphi_j(1-p_j) \leq T_j\}}\right] \leq 1, \tag{B.1}$$

which finishes the proof. □

### B.6 Proof of Proposition A.1

*Proof.* Proposition A.1 is a special case of Proposition A.2 by choosing $\varphi_i$ as the identity function for all $1 \leq i \leq n$. □

### B.7 Proof of Proposition A.2

*Proof.* Write $T = T_{\text{FBC}}$ for the ease of notation. First, given a p-value vector $p = (p_1, \cdots, p_n)$, recall that the threshold $T$ is defined as

$$T = \max\left\{0 < t \leq T_{\text{up}} \colon \underbrace{\frac{1+\sum_{l=1}^n \mathbb{1}\{\varphi_l(1-p_l) \leq t\}}{\sum_{l=1}^n \mathbb{1}\{\varphi_l(p_l) \leq t\}}}_{:=g(p,t)} \leq \alpha\right\},$$

where $T_{\text{up}}$ satisfies $T_{\text{up}} < \varphi_l(0.5)$ for all $l$.

Without loss of generality, let us assume $T_i \geq T_j$. By the assumption that $\max\{\varphi_i(1-p_i), \varphi_j(1-p_j)\} \leq \max\{T_i, T_j\}$, we have $\varphi_i(1-p_i) \leq T_i$ and $\varphi_j(1-p_j) \leq T_i$. Since $\varphi_i$ is an increasing function, we have $\varphi_i(1-p_i) \leq T_{\text{up}} < \varphi_i(0.5)$, which implies $1-p_i < 0.5$. Thus $\varphi_i(p_i) \geq \varphi_i(0.5) > T_{\text{up}} \geq T_i$. The same discussion for $p_j$ leads to $\varphi_j(p_j) > T_i$.

Denote $\tilde{p}_i = \min\{p_i, 1-p_i\}$ and $p_{-i} = (p_1, \cdots, p_{i-1}, \tilde{p}_i, p_{i+1}, \cdots, p_n)$ for all $i$. Consider the function

$$g(p_{-j}, T_i) = \frac{1 + \sum_{l=1}^n \mathbb{1}\{\varphi_l(1-p_{-j,l}) \leq T_i\}}{\sum_{l=1}^n \mathbb{1}\{\varphi_l(p_{-j,l}) \leq T_i\}},$$

where $p_{-j,l}$ is the $l$th entry of $p_{-j}$. For the denominator, we have

$$\begin{aligned}
&\sum_{l=1}^n \mathbb{1}\{\varphi_l(p_{-j,l}) \leq T_i\} \\
=&\sum_{l=1}^n \mathbb{1}\{\varphi_l(p_{-i,l}) \leq T_i\} + \underbrace{\mathbb{1}\{\varphi_j(p_{-j,j}) \leq T_i\}}_{=1} + \underbrace{\mathbb{1}\{\varphi_i(p_{-j,i}) \leq T_i\}}_{=0} \\
&\quad - \underbrace{\mathbb{1}\{\varphi_j(p_{-i,j}) \leq T_i\}}_{=0} - \underbrace{\mathbb{1}\{\varphi_i(p_{-i,i}) \leq T_i\}}_{=1} \\
=&\sum_{l=1}^n \mathbb{1}\{\varphi_l(p_{-i,l}) \leq T_i\}.
\end{aligned}$$

Similarly, for the numerator, we have

$$\begin{aligned}
&\sum_{l=1}^n \mathbb{1}\{\varphi_l(1-p_{-j,l}) \leq T_i\} \\
=&\sum_{l=1}^n \mathbb{1}\{\varphi_l(1-p_{-i,l}) \leq T_i\} + \underbrace{\mathbb{1}\{\varphi_j(1-p_{-j,j}) \leq T_i\}}_{=0} \\
&\quad + \underbrace{\mathbb{1}\{\varphi_i(1-p_{-j,i}) \leq T_i\}}_{=1} - \underbrace{\mathbb{1}\{\varphi_j(1-p_{-i,j}) \leq T_i\}}_{=1} - \underbrace{\mathbb{1}\{\varphi_i(1-p_{-i,i}) \leq T_i\}}_{0} \\
=&\sum_{l=1}^n \mathbb{1}\{\varphi_l(1-p_{-i,l}) \leq T_i\}.
\end{aligned}$$

Hence, $g(p_{-j}, T_i) = g(p_{-i}, T_i) \leq \alpha$. By the definition of $T_j$, we must have $T_i \leq T_j$. Similarly, we get $T_j \leq T_i$ and hence $T_i = T_j$. $\square$

# C Additional discussions about the FBH procedure

**Example C.1.** Consider the scenario where $g(t) = n^{-1}\sum_{i=1}^n F_i(t) \coloneqq \bar{F}(t)$ and $F_i(t) = c_i h(t)$. In this case, our approach reduces to the weighted BH procedure in Genovese et al. [2006]. To see this, note that when $F_i(t) = c_i h(t)$, we have $\varphi_i(x) = F_i^{-1}(x) = h^{-1}(x/c_i)$, $\bar{F}(t) = \bar{c}h(t)$, and $\bar{F}^{-1}(x) = h^{-1}(x/\bar{c})$, where $\bar{c} = \sum_{i=1}^n c_i/n$ and $\bar{F}^{-1}$ is the inverse function of $\bar{F}$. Denote $q_i = \varphi_i(p_i) = h^{-1}(p_i/c_i)$, and sort $q_i$ in an ascending order, i.e., $q_{(1)} \leq \cdots \leq q_{(n)}$, where $(i)$ denotes the index of the $i$th smallest value in the set $\{q_i\}_{i=1}^n$. For any index $(i)$, we have $q_{(i)} = h^{-1}(p_{(i)}/c_{(i)})$. Since $h$ is strictly increasing, so is $h^{-1}$. Therefore, $p_{(1)}/c_{(1)} \leq \cdots \leq p_{(n)}/c_{(n)}$.

For the FBH procedure, we reject $H_{(i)}$ for all $i \le \hat{k}$, where

$$\begin{aligned}\hat{k} &= \max_i \left\{ i\colon q_{(i)} \le \bar{F}^{-1}(\alpha i/n) = h^{-1}\left(\frac{\alpha i}{n\bar{c}}\right) \right\} \\ &= \max_i \left\{ i\colon h^{-1}(p_{(i)}/c_{(i)}) \le h^{-1}\left(\frac{\alpha i}{n\bar{c}}\right) \right\} \\ &= \max_i \left\{ i\colon \frac{p_{(i)}}{c_{(i)}/\bar{c}} \le \frac{\alpha i}{n} \right\},\end{aligned}$$

which coincides with the weighted BH procedure with the weights being $c_i/\bar{c}$.

**Remark C.1.** When $g(t) = n^{-1}\sum_{i=1}^n F_i(t)$ and $F_i(t) = c_i h(t)$, the FBH method satisfies the two sufficient conditions for controlling FDR, as proposed by Blanchard and Roquain [2008]. Therefore, the FDR control, in this case, can also be proven using Proposition 2.7 in Blanchard and Roquain [2008].

**Remark C.2.** The p-testing procedure in Section 6.5 of Wang and Ramdas [2022] is a specific case of our FBH procedure. In the p-testing procedure, a strictly decreasing and continuous function $\psi\colon [0,1] \to [0,\infty]$ is used to decide the set of rejections. Specifically, the p-testing procedure rejects $k_\psi^*$ hypotheses with the smallest p-values, where $k_\psi^* = \max\{i\colon \psi(p_{(i)}) \ge n/i\}$. Wang and Ramdas [2022] obtained an upper bound for the FDR of the p-testing procedure. In the FBH procedure, the rejection rule depends on the strictly increasing functions $\{\varphi_i\}$ that can differ for each $H_i$, which allows us to incorporate external structural information for each hypothesis. If we choose $g$ as the identity function and $\varphi_i(x) = \alpha/\psi(x)$ for all $i$, then the FBH procedure reduces to the p-testing procedure in Wang and Ramdas [2022].

**Remark C.3.** The weighted BH procedure proposed in Remark 1 of Sarkar [2023] is a specific case of our FBH procedure. In Sarkar [2023], the author proposed a transformation of the p-values using the formula $\tilde{q}_i := F_0\big(w_i^{-1}F_0^{-1}(p_i)\big)$, where $w_i$ represents the weight for the $i$th hypothesis. He then applied the BH procedure to $\{\tilde{q}_i\}$, i.e., compared them with the set of thresholds $\{i\alpha/n\colon i = 1,2,\cdots,n\}$. In contrast, in the FBH procedure, $F_0\big(w_i^{-1}F_0^{-1}\big)$ can be viewed as a special case of $\varphi_i$, where we compare $\varphi_i(p_i)$ with the threshold $g^{-1}(i\alpha/n)\colon i = 1,2,\cdots,n$. If we choose $g$ as the identity function, the FBH procedure simplifies to the weighted BH procedure proposed in Sarkar [2023].

# D Additional discussions about aggregating and assembling e-values

We have shown that the BH and BC procedures and their generalized versions are all equivalent to the e-BH procedure based on specific forms of e-values. This equivalence opens up new possibilities for designing multiple testing procedures in different contexts by combining e-values from different procedures (or the same procedure with different tuning quantities) or assembling e-values from various subsets of the data. See Figure 1 below for an illustration of aggregating and assembling e-values.

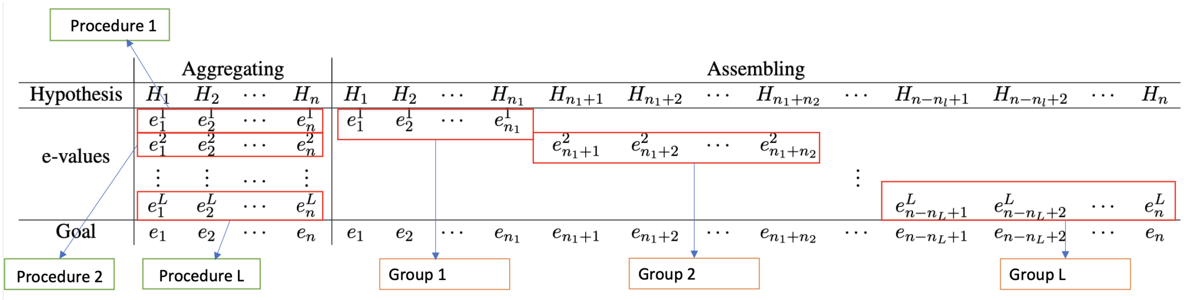


Figure 1: An illustration of aggregating e-values from different procedures and assembling e-values from different datasets.

## D.1 Aggregating e-values from different procedures

Suppose we have $L$ sets of e-values $\{e_i^l : i \in [n]\}_{l=1}^L$ (possibly) from $L$ different multiple testing procedures, where $\{e_i^l\}_{l=1}^L$ are the $L$ e-values associated with $H_i$ and $\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i^l] \le n$. In other words, for each hypothesis $H_i$, we have $L$ e-values $\{e_i^l\}_{l=1}^L$. In this scenario, our goal is to aggregate the $L$ sets of e-values $\{e_i^l : i \in [n]\}_{l=1}^L$ into a single e-value vector $[e_1, \cdots, e_n]$, satisfying Condition (5). In Proposition 5, we propose to define the weighted e-value through $e_i = \sum_{l=1}^L w_{l,i} e_i^l$, where $w_{l,i} \ge 0$ is the aggregating weight. The condition $\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i^l] \le n$ for all $l$ ensures that each procedure controls the FDR. Proposition 5 suggests that the e-BH procedure applied to the weighted e-values still controls the FDR. Below is the proof of Proposition 5.

*Proof of Proposition 5.*

$$\begin{aligned}
\sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i] &= \sum_{i\in\mathcal{H}_0}\sum_{l=1}^L w_{l,i}\mathbb{E}[e_i^l] \\
&\le \sum_{i\in\mathcal{H}_0}\sum_{l=1}^L \max_i w_{l,i}\mathbb{E}[e_i^l] \\
&= \sum_{l=1}^L \max_i w_{l,i} \sum_{i\in\mathcal{H}_0}\mathbb{E}[e_i^l] \\
&\le \sum_{l=1}^L \max_i w_{l,i} n \\
&\le n.
\end{aligned}$$

□

When $n = 1$, this problem is the same as the e-value merging problem considered in Vovk and Wang [2021]. In their Proposition 3.1, the authors proved that the arithmetic mean dominates any symmetric aggregation function. For a detailed explanation of "domination" and how to prove this result, please refer to Vovk and Wang [2021]. Let $f\colon [0,1]^n \to [0,1]$ be an aggregation function. We say that $f$ is a symmetric aggregation function if it is invariant with respect to any permutation of its arguments, i.e., $f(e_1, \cdots, e_n) = f(e_{i_1}, \cdots, e_{i_n})$, where $[i_1, \cdots, i_n]$ is a permutation of $[1, \cdots, n]$.

Notice that our proposed method is a weighted mean, where the weight can be different for each $e_i^l$. Hence, our aggregation function is not symmetric.

Banerjee et al. [2023] considered a similar aggregation problem. In their scenario, only a portion of $\{e_i^l\}_{l=1}^L$ is observable for each hypothesis $H_i$. Their method relies on arithmetic mean and does not cover the case of weighted averages.

## D.2 Assembling e-values from different datasets

Suppose 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 $\cup_l \mathcal{G}_l = [n]$, $\mathcal{G}_{l_1} \cap \mathcal{G}_{l_2} = \emptyset$ if $l_1 \neq l_2$, $e_i$ is associated with the hypothesis $H_i$ and $\sum_{i\in\mathcal{G}_l\cap\mathcal{H}_0} \mathbb{E}[e_i^l] \leq n_l$. In this scenario, our goal is to assemble $L$ sets of e-values $\{e_i^l : i \in \mathcal{G}_l, |\mathcal{G}_l| = n_l\}$ into a single e-value vector $[e_1, \cdots, e_n]$, satisfying Condition (5). To our knowledge, this problem seems less explored in the existing e-value literature.

In Proposition 6, we propose to define the weighted e-value through $e_i = w_{l,i} e_i^l$, where $w_{l,i} \geq 0$ is the assembling weight. The most direct method is to set all weights equal to 1. However, in practice, we may have some extra information, which allows different weights to make our approach more flexible. The condition $\sum_{i\in\mathcal{G}_l\cap\mathcal{H}_0} \mathbb{E}[e_i] \leq n_l$ for all $l$ ensures that the FDR is controlled within each group $\mathcal{G}_l$. Proposition 6 suggests that the e-BH procedure applied to the weighted e-values controls the overall FDR. Following is the proof of Proposition 6.

*Proof of Proposition 6.*

$$
\begin{aligned}
\sum_{i\in\mathcal{H}_0} \mathbb{E}[e_i] &= \sum_{l=1}^{L} \sum_{i\in\mathcal{G}_l\cap\mathcal{H}_0} w_{l,i}\mathbb{E}[e_i^l] \\
&\leq \sum_{l=1}^{L} \sum_{i\in\mathcal{G}_l\cap\mathcal{H}_0} \max_{i\in\mathcal{G}_l} w_{l,i}\mathbb{E}[e_i^l] \\
&= \sum_{l=1}^{L} \max_{i\in\mathcal{G}_l} w_{l,i} \sum_{i\in\mathcal{G}_l\cap\mathcal{H}_0} \mathbb{E}[e_i^l] \\
&\leq \sum_{l=1}^{L} \max_{i\in\mathcal{G}_l} w_{l,i} n_l \\
&\leq n.
\end{aligned}
$$

□

# References


Trambak Banerjee, Bowen Gang, and Jianliang He. Harnessing the collective wisdom: Fusion learning using decision sequences from diverse sources. *arXiv preprint arXiv:2308.11026*, 2023.

Rina Foygel Barber and Emmanuel J Candès. Controlling the false discovery rate via knockoffs. *Annals of Statistics*, 2015.

Rina Foygel Barber, Emmanuel J Candès, and Richard J Samworth. Robust inference with knockoffs. *Annals of Statistics*, 2020.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

Gilles Blanchard and Etienne Roquain. Two simple sufficient conditions for fdr control. *Electronic Journal of Statistics*, 2008.

Hongyuan Cao, Jun Chen, and Xianyang Zhang. Optimal false discovery rate control for large scale multiple testing with auxiliary information. *Annals of statistics*, 50(2):807, 2022.

Robin Dunn, Aaditya Ramdas, Sivaraman Balakrishnan, and Larry Wasserman. Gaussian universal likelihood ratio testing. *Biometrika*, 110(2):319–337, 2023.

JA Ferreira and AH Zwinderman. On the benjamini–hochberg method. *Annals of Statistics*, 2006.

Christopher R Genovese, Kathryn Roeder, and Larry Wasserman. False discovery control with p-value weighting. *Biometrika*, 93(3):509–524, 2006.

Peter Grünwald, Rianne de Heide, and Wouter M Koolen. Safe testing. In *2020 Information Theory and Applications Workshop (ITA)*, pages 1–54. IEEE, 2020.

Nikolaos Ignatiadis, Ruodu Wang, and Aaditya Ramdas. E-values as unnormalized weights in multiple testing. *Biometrika*, page asad057, 09 2023. ISSN 1464-3510.

Lihua Lei and William Fithian. Adapt: an interactive procedure for multiple testing with side information. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(4): 649–679, 2018.

Edsel A Peña, Joshua D Habiger, and Wensong Wu. Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Annals of statistics*, 39(1):556, 2011.

Zhimei Ren and Rina Foygel Barber. Derandomised knockoffs: leveraging e-values for false discovery rate control. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):122–154, 2024.

Sanat K Sarkar. On controlling the false discovery rate in multiple testing of the means of correlated normals against two-sided alternatives. *arXiv preprint arXiv:2304.05261*, 2023.

Glenn Shafer. Testing by betting: A strategy for statistical and scientific communication. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(2):407–431, 2021.

John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):479–498, 2002.

John D Storey, Jonathan E Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(1):187–205, 2004.

Wenguang Sun and T Tony Cai. Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, 102(479):901–912, 2007.

Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination and applications. *Annals of Statistics*, 49(3):1736–1754, 2021.

Ruodu Wang and Aaditya Ramdas. False discovery rate control with e-values. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):822–852, 2022.

Ziyu Xu and Aaditya Ramdas. More powerful multiple testing under dependence via randomization. *arXiv preprint arXiv:2305.11126*, 2023.

Ziyu Xu, Ruodu Wang, and Aaditya Ramdas. A unified framework for bandit multiple testing. *Advances in Neural Information Processing Systems*, 34:16833–16845, 2021.

Xianyang Zhang and Jun Chen. Covariate adaptive false discovery rate control with applications to omics-wide multiple testing. *Journal of the American Statistical Association*, 117(537):411–427, 2022.