Bringing Flexibility to the Benjamini-Hochberg Procedure

Aurele Mingam[1] Rovanos Tsafack Nzanguim[2] Rianne de Heide[2] Jelle Goeman[1]

[1] Leiden University Medical Center (LUMC) [2] University of Twente

Classical Procedure: Benjamini and Hochberg 1995 (BH)

- m hypotheses $\mathcal{H}_1, \dots, \mathcal{H}_m$, $\mathcal{N} \subset [m]$ null indices
- m ordered PRDS p-variables $p = (p_1 \leq \dots \leq p_m)$
- PRDS: $\forall D \subset [0,1]^m$ increasing set, $\forall i \in \mathcal{N}$ the function $x \mapsto \mathbb{P}(p \in D | p_i \leq x)$ is increasing
- BH rejects the k^{BH} smallest p-values (■ in Figure 2) controlling the **False Discovery Rate** at level α:

$$FDR([k^{BH}]) = \mathbb{E}\left[FDP_{\mathcal{N}}([k^{BH}])\right] = \mathbb{E}\left[\frac{|[k^{BH}] \cap \mathcal{N}|}{k^{BH}}\right] \leq \alpha$$

$$\text{Where} \quad k^{BH} := \sup\left\{k \in [m] \mid p_k \leq \alpha \times \frac{k}{m}\right\} \tag{1}$$

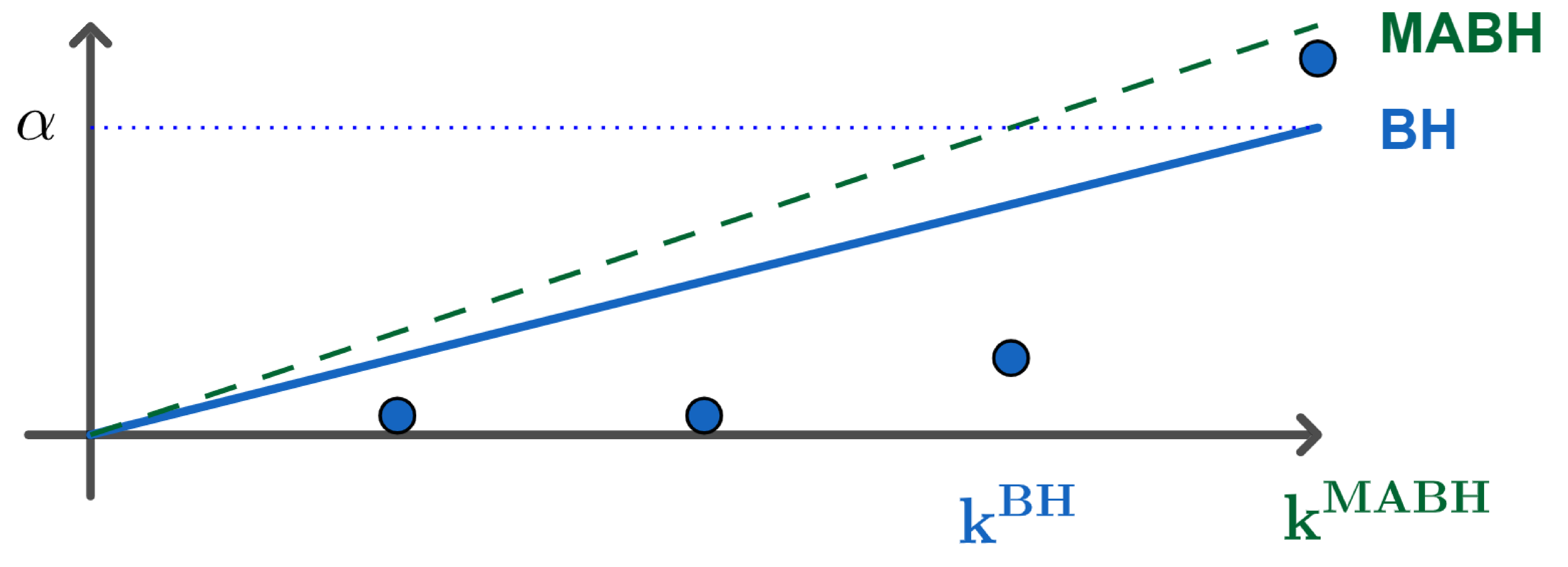

Figure 1: BH and MABH comparison

GOALS:

- **Rejecting multiple sets with $[k^{BH}]$, all controlling FDR simultaneously:** Guaranteed for FWER and FDP tail probability control, possible for some FDR procedures (BY,SU) but not yet for BH.
- **Enabling flexibility between FDR and FWER:** Choosing between FWER and FDP after having seen the data is always possible, when is FDR compatible remains an open question.

Expressing BH with an e-Collection

With the e-Closure principle we can **reformulate BH via the e-Closure** by defining:

$$e_S^{BH} = \frac{|[k^{BH}] \cap S|}{k^{BH}\alpha} \tag{2}$$

When plugging this e-Collection in the e-Closure procedure we **recover exactly the BH rejection set**, there is **no** flexibility:

$$\mathcal{R}(e^{BH}) = \{[k^{BH}], \emptyset\}$$

An Expectation Gap to Exploit

BH controls FDR at the stringent level $\alpha \times |\mathcal{N}|/m$ therefore:

$$\mathbb{E}\left[e_{\mathcal{N}}^{BH}\right] = \mathbb{E}\left[\frac{|[k^{BH}] \cap \mathcal{N}|}{k^{BH}\alpha}\right] \leq \alpha \times \frac{|\mathcal{N}|}{\alpha m} \leq 1 \tag{3}$$

The last inequality reveals a **gap in expectation** since in practice there are non-null hypotheses.

State of the Art: Restricted Flexibility

Xu et al. proposed multiplying the BH e-Variable (2) to fill the expectation gap (3):

$$e_S^{\times BH} = \frac{|[k^{BH}] \cap S|}{k^{BH}\alpha} \times \frac{m}{|n|}$$

$$\mathbb{E}\left[e_{\mathcal{N}}^{\times BH}\right] \leq \alpha \times \frac{|\mathcal{N}|}{\alpha m} \times \frac{m}{|\mathcal{N}|} = 1$$

As acknowledge by the authors, the result is unsatisfactory:

Theorem: (Xu et al.):

- If $k^{BH} < m$: $\mathcal{R}(e^{\times BH}) = \{[k^{BH}], \emptyset\}$
- If $k^{BH} = m$: $\mathcal{R}(e^{\times BH}) = 2^{[m]}$

This yields no flexibility unless everything is rejected. More, in this case we control the restrictive FWER. This shows that the **additional expectation is not exploited** efficiently by the procedure.

Main Tool: Reformulating Multiple Testing via e-Variable

(Xu et al. 2026) frames multiple testing via **e-Collections** $(e_S)_{S \in 2^{[m]}}$: a collection of **non-negative random variables satisfying** $\mathbb{E}[e_{\mathcal{N}}] \leq 1$. The output of their e-Closure procedure is the collection of all simultaneously rejectable sets:

$$\mathcal{R}(e) = \left\{R \in 2^{[m]} \;\middle|\; \forall S \in 2^{[m]} : \alpha e_S \geq \frac{|S \cap R|}{|R|}\right\}. \tag{4}$$

Any set $R \in \mathcal{R}(e)$ can be rejected post-hoc, the **e-Closure enables flexibility for FDR**. If the e-values reach $1/\alpha$, **e-Closure will also simultaneously control FWER**.

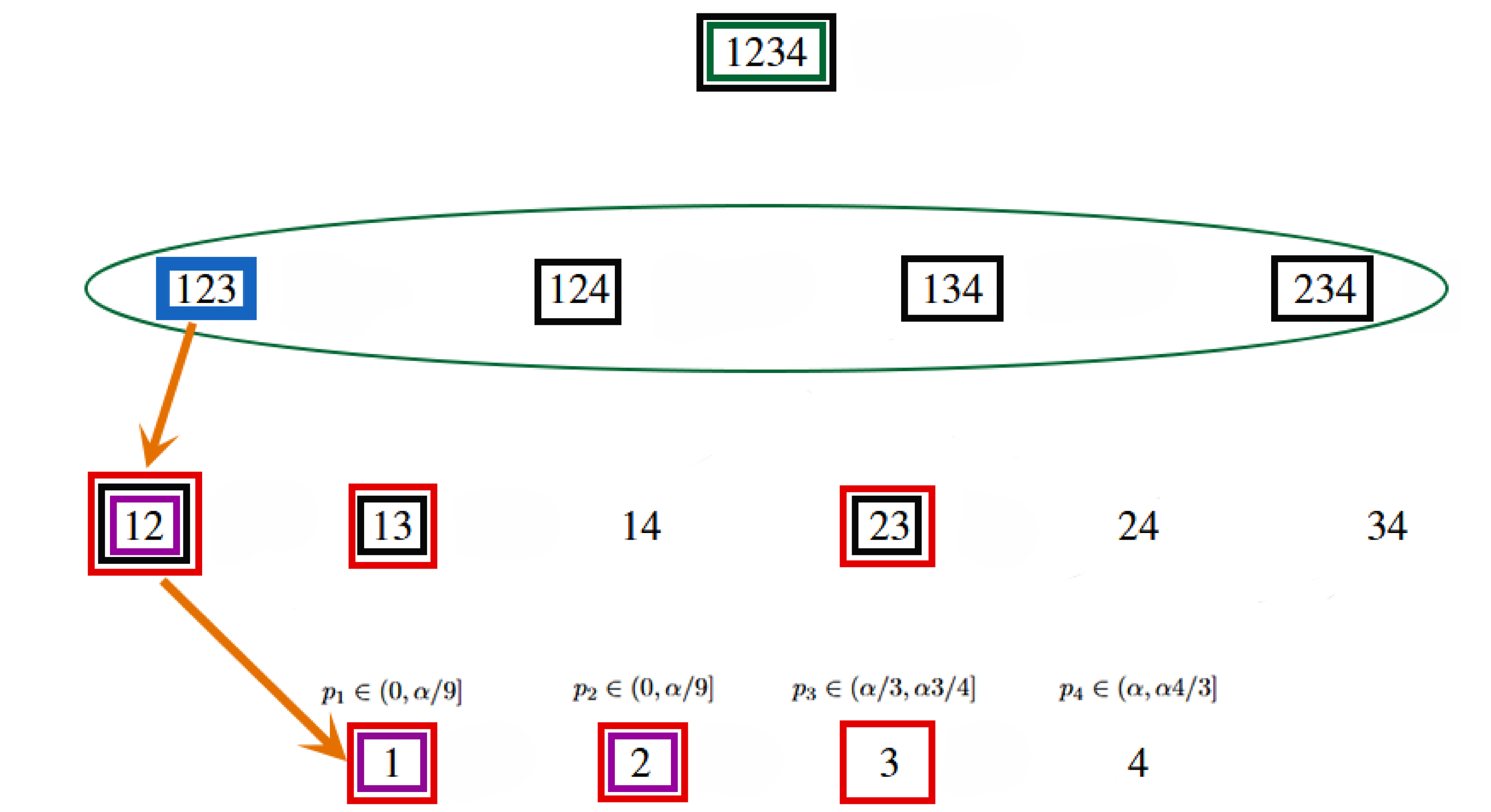

Figure 2: Rejection sets produced by BH(1), MABH (10), adaBH(8), pBH(9), cstBH(6) for four hypotheses and given p-values. All methods reject the BH set.

The e-Closure Principle

Theorem (Xu et al.):
$\mathcal{R}(e)$ controls FDR at level α whenever $(e_S)_{S \in 2^{[m]}}$ is an e-Collection: $\mathbb{E}\left[\sup_{R \in \mathcal{R}(e)} FDP_{\mathcal{N}}(R)\right] \leq \alpha$

The practitioner can **choose any set in this collection** after having seen the data and **retains FDR control** of her final rejection set.

Theorem (Xu et al.): $\forall R$ random rejection set controlling FDR at level α, $\exists (e_S)_{S \in 2^{[m]}}$ an e-Collection such that $\mathcal{R}(e) = \{R, \emptyset\}$.

Any existing valid procedure can be expressed through the e-Closure, this is what we do in (2) and (11).

Key Idea: Filling e-Variable

To fill the expectation gap we add a second e-Collection and consider **filling** and **filled** e-Collections:

$$e_S^{\text{filled}} := \underbrace{\frac{|S \cap [k^{BH}]|}{\alpha k^{BH}}}_{\text{BH e-Variable}} + \frac{m - |S|}{m} \times \underbrace{e_S^*}_{\text{filling e-Variable}}. \tag{5}$$

Proposition:
$\left(e^{\text{filled}}{}_S\right)_{S \in 2^{[m]}}$ is an e-Collection if $(e^*{}_S)_{S \in 2^{[m]}}$ is one.

Since $e_{\mathcal{N}}^*$ has expectation at most 1, we remain valid while increasing our e-Collection, enabling more rejection sets and error rate flexibility.

Useful e-Variable for Multiple Testing in the Literature

We focused on filling e-Variable of the form:

$$e_S^{\tau BH} = \mathbb{1}_{\{p_S \leq \tau_S(p)\}} \frac{1}{\tau_S(p)} \quad \text{with} \quad p_S = \min_{i \in [|S|]} \frac{|S|}{i} p_{i:S} \quad \text{the Simes' } p\text{-value} \tag{6}$$

Theorem (Blanchard and Roquain):
If $\forall S \in 2^{[m]}$ τ_S is non-increasing then $\left(e^{\tau BH}{}_S\right)_{S \in 2^{[m]}}$ is a valid e-Collection

Those e-Variable help us to target the precise and known thresholds given by the e-Closure principle.

Examples of Thresholds for Filling e-Collections

- Constant Threshold (■ in Figure 2)

$$\tau_S^{cstBH}(P) := \alpha \frac{m - |S|}{m - 1} \tag{7}$$

This simplistic example is the largest constant threshold for which e_S^{filled} **attains the value** $1/\alpha$ **whenever** $p_S \leq \tau$, $1/\alpha$ being necessary for FWER control. Though it seems reasonable, it **often exceeds** $1/\alpha$, leading to useless expectation attributions, so few rejection sets (see Figure 2).

- Adaptive Flexibility Improvement (■ in Figure 2)

$$\tau_S^{adaBH}(P) := \alpha \frac{m - |S|}{m\left(1 - \inf_{k \in [m]} FDP_S([k])\right)} \tag{8}$$

To target $1/\alpha$ more accurately we adopt an **adversarial approach**: the infimum targets the lowest value the BH e-Variable e_S^{BH} can take for S depending on the number of rejection k^{BH}. The rest is to deal with the coefficient $(m - |S|)/m$ and indeed reach $1/\alpha$ when summing.

- Random Power Improvement (■ in Figure 2)

$$\tau_S^{pBH} := \alpha \frac{m - |S|}{m} |R_{BH}| \tag{9}$$

Enlarging this threshold we reach a power+flexibility improvement. If the p-values are small enough, **rejecting one additional hypothesis incurs little FDR cost**. *pBH* leverages this by adding 1 to the numerator of e_S^{BH} when possible. Nevertheless, we lost all FWER flexibility except when $k^{BH} = 2$.

CONJECTURE: An Optimized Filling e-Collection

A natural question is what would be the most flexible filling e-Collection? To define a flexibility criteria we focus on a specific path: knowing we reject $\{[k^{BH}], \emptyset\}$ can we reject $\{[k^{BH}], [k^{BH} - 1], \emptyset\}$? and so on until $\{[k^{BH}], [k^{BH} - 1], \dots, [1], \emptyset\}$ (see Figure 2 in ■). For such path we designed the following threshold:

$$\tau_S^{optiBH} := \alpha \frac{(m - |S|)}{m \times \left(\max_{l' \in [\![l^*, k^{BH}]\!]} FDP_S([l']) - FDP_S(k^{BH})\right)}$$

where

$$l^* := \inf\left\{l \in [k^{BH}] \mid \forall S \; p_S \leq \frac{\alpha(m - |S|)}{f_S(l, k^{BH}) m}\right\}$$

τ_S^{optiBH} targets exactly the thresholds needed for FDR control of rejected sets. There will be no rejections outside the chosen path. l^* is a "stopping-time" stopping at the smallest set we can reject. Only the sets between $[k^{BH} - l^*]$ and $[k^{BH}]$ will be rejected. We focus the power on the path. We conjecture that this sequence $\left(e_S^{optiBH}\right)_{S \in 2^{[m]}}$ is an e-Collection but we were unable to prove it due to non-monotonicity of FDP.

Minimally Adaptive BH: No gap to Exploit

MABH uniformly improves upon BH (Solari and Goeman), under PRDS it rejects $[k^{BH}]$ (■ in Figure 2):

$$\text{Where} \quad k^{MABH} := \mathbb{1}_{\{k^{BH} > 0\}} \times \sup\left\{k \in [m] \mid p_k \leq \alpha \times \frac{k}{m - 1}\right\} \tag{10}$$

The e-Collection for MABH is:

$$e_S^{MABH} = \frac{|[k^{MABH}] \cap S|}{k^{MABH}\alpha} \quad \text{with expectation} \quad \mathbb{E}\left[e_{\mathcal{N}}^{MABH}\right] \leq \alpha \frac{|\mathcal{N}|}{(m-1) \vee |\mathcal{N}|} \tag{11}$$

There is no expectation gap on the first line (— in Figure 2), this prevents any flexibility improvement.

Theorem:
$\forall$ e-Collection e, if $R^{MABH} \overset{a.s.}{\in} \mathcal{R}(e)$, then $\mathcal{R}(e) = \{R^{MABH}, \emptyset\}$

Relation to the State of the Art in Multiple Testing

Goeman et al. (2020) showed (but did not state) that one can choose between FWER and FDP tail probability control in a post-hoc way for free. Xu et al. (2026) extended this error rate flexibility to FDR while bringing post-hoc set selection to it. We showed here that this flexibility can be added in many different way to a preserved BH procedure, and this, again, for free. We achieve this by interpolating between e-Collection targeting the FDR thresholds and others targeting the FWER threshold.

References: Benjamini and Hochberg (1995) Xu et al. (2026) Blanchard & Roquain (2008) Goeman et al. (2021)

email: ajf.mingam@lumc.nl