Multiple testing for modern data: structure, curation, and replicability

Eugene Katsevich
Department of Statistics
Stanford University

March 12, 2019
A modern data set

(Image source: Nature)
UK Biobank data

Extensive data on 500,000 individuals, including

- Genotypes
- Diseases (from electronic health records)
- Blood pressure and other clinical diagnostics
- Socioeconomic variables
- Environmental risk factors
- Imaging data
- Diet and exercise questionnaires
- ...
A genotype is an individual’s allele at a given *single nucleotide polymorphism* (SNP).

Genotypes measured at 1,000,000 SNPs.
Genotype data have spatial structure

Nearby SNPs are strongly correlated with each other.
Disease data

Disease codes from hospital episodes, using *International Classification of Diseases* (ICD-10).

ICD-10 is very comprehensive and includes 20K codes.
Disease data

Disease codes from hospital episodes, using *International Classification of Diseases* (ICD-10).

ICD-10 is very comprehensive and includes 20K codes.

(Image source: Google)
Disease data have tree structure

Diseases of the musculoskeletal system and connective tissue

- Inflammatory polyarthropathies
  - Gout
  - Rheumatoid arthritis
- Spondylopathies
  - Ankylosing spondylitis
  - Spondylosis
UK Biobank: a complex multiple testing problem

Type-I error rates like the false discovery rate (FDR) controlled for replicability.
Findings from modern data sets often need curation

**Manual curation (exploration):**
Domain experts search for interesting patterns in the data.

**Automatic curation (filtering):**
Structured hypotheses often lead to redundant findings; filtering is commonly used to reduce redundancy.
Findings from modern data sets often need curation

**Manual curation (exploration):**
Domain experts search for interesting patterns in the data.

**Automatic curation (filtering):**
Structured hypotheses often lead to redundant findings; filtering is commonly used to reduce redundancy.

*Curation may conflict with replicability!*
Phenome-wide association studies (PheWAS)

- Diseases of the musculoskeletal system and connective tissue
  - Inflammatory polyarthropathies
    - Gout
  - Spondyloarthropathies
    - Ankylosing spondylitis
    - Spondylosis

SNP1, SNP2, SNP3, SNP4, SNP5, SNP6, SNP7, SNP8, SNP9
Rejection sets in phenotype space can be redundant

Diseases of the musculoskeletal system and connective tissue

Inflammatory polyarthropathies

- Gout
- Rheumatoid arthritis

Spondylopathies

- Ankylosing spondylitis
- Spondylosis
Redundancy can be fixed by applying the outer nodes filter.
Outer nodes filter may inflate the FDR

- Diseases of the musculoskeletal system and connective tissue
  - Inflammatory polyarthropathies
    - Gout
    - Rheumatoid arthritis
  - Spondylopathies
    - Ankylosing spondylitis
    - Spondylosis

- cyan nodes: non-null; red nodes: null; shaded nodes: rejected.

Yekutieli (2008)
Existing options to control outer nodes FDR are limited

- Yekutieli proposed a procedure and bounded its outer nodes FDR, but only under independence.
- Structured Holm procedure\(^1\) controls FWER on DAGs. It allows arbitrary dependence but is conservative.

\(^1\)Meijer and Goeman (2016)
Similar problems arise in other applications as well

- Genome-wide association studies\(^2\)
- Imaging applications such as fMRI\(^3\)
- Gene Ontology enrichment analysis\(^4\)

---

\(^2\) Siegmund, Zhang, Yakir (2011)
\(^4\) Goeman and Buhlmann (2007), Meijer and Goeman (2016)
A general problem

Filtering may inflate the FDR, and must be accounted for.

Partial solutions exist, but a general-purpose solution is lacking.

Focus of this talk

Reconciling curation with replicability for modern data analysis pipelines.

Goeman and Solari (2011), Berk et al (2013), Taylor and Tibshirani (2015), ...
Part I (automatic curation): For any pre-specified filter, we propose **Focused BH**\(^5\) to control the FDR after filtering.

\[^5\text{K.}, \text{Sabatti, Bogomolov (arXiv, 2019+)}\]

\[^6\text{K. and Ramdas (AOS, in revision, 2019+), K. and Sabatti (AOAS, 2019)}\]
Preview: Reconciling curation with replicability

Part I (automatic curation): For any pre-specified filter, we propose **Focused BH**\(^5\) to control the FDR after filtering.

\[
\text{p-values} \quad \xrightarrow{\text{BH}} \quad \text{Initial rejections} \quad \xrightarrow{\text{Filter}} \quad \text{Filtered rejections}
\]

Part II (manual curation): We propose **simultaneous selective inference**\(^6\) to allow directed exploration while bounding FDP whp.

\(^5\) K., Sabatti, Bogomolov (arXiv, 2019+)

\(^6\) K. and Ramdas (AOS, in revision, 2019+), K. and Sabatti (AOAS, 2019)
Part I: Controlling FDR while filtering
A general definition of a filter

Hypotheses $\mathcal{H} = (H_1, \ldots, H_m)$ and p-values $\mathbf{p} = (p_1, \ldots, p_m)$.

**Definition**

Given $\mathcal{R} \subseteq \mathcal{H}$ and $\mathbf{p} \in [0, 1]^m$, a *filter* $\mathcal{F}$ is any mapping

$$\mathcal{F} : (\mathcal{R}, \mathbf{p}) \mapsto \mathcal{U}, \text{ such that } \mathcal{U} \subseteq \mathcal{R}.$$ 

For example,

- $\mathcal{F}$ is the outer nodes filter;
- $\mathcal{R}$ is the set of rejected nodes;
- $\mathcal{U}$ is the set of outer nodes.
Adjusting the FDR for filtering

The false discovery proportion (FDP) of a set \( \mathcal{U} \subseteq \mathcal{H} \) is

\[
\text{FDP}(\mathcal{U}) = \frac{\left| \mathcal{U} \cap \mathcal{H}_0 \right|}{|\mathcal{U}|},
\]

where \( \mathcal{H}_0 \subseteq \mathcal{H} \) is the set of nulls.

**Definition**

Given a filter \( \mathcal{F} \), the false filtered discovery rate of a testing procedure (mapping \( p \mapsto \mathcal{R}^* \)) is

\[
\text{FDR}_{\mathcal{F}} = \mathbb{E}[\text{FDP}(\mathcal{U}^*)] = \mathbb{E}[\text{FDP}(\mathcal{F}(\mathcal{R}^*, p))].
\]

Given a filter \( \mathcal{F} \) and a pre-specified target FDR level \( q \), our goal is to design a testing procedure for which \( \text{FDR}_{\mathcal{F}} \leq q \).
Adjusting BH to account for filtering

For a p-value cutoff \( t \in [0, 1] \), consider \( \mathcal{R}(t) = \{ j : p_j \leq t \} \).

**BH procedure**

BH employs the FDP estimate (Storey, 2002)

\[
\hat{\text{FDP}}_{BH}(t) = \frac{m \cdot t}{|\mathcal{R}(t)|};
\]

choosing the threshold

\[
t^*_{BH} = \max\{ t \in [0, 1] : \hat{\text{FDP}}_{BH}(t) \leq q \}.
\]

We are interested instead in \( \mathcal{U}(t) = \mathcal{F}(\{ j : p_j \leq t \}, p) \).

BH too optimistic in counting discoveries: \( |\mathcal{R}(t)| \gg |\mathcal{U}(t)| \).
Adjusting BH to account for filtering

Instead of

\[ \hat{FDP}_{BH}(t) = \frac{m \cdot t}{|\mathcal{R}(t)|}, \]

correct the denominator and define

\[ \hat{FDP}(t) = \frac{m \cdot t}{|\mathcal{U}(t)|} = \frac{m \cdot t}{\mathcal{F}(\{j : p_j \leq t\}, \mathcal{P})}. \]

We keep the numerator as is, since

\[ |\mathcal{U}(t) \cap \mathcal{H}_0| \leq |\mathcal{R}(t) \cap \mathcal{H}_0|. \]
Focused BH procedure

**Data:** p-values $p_1, \ldots, p_m$, filter $\mathcal{F}$, target level $q$

**for** $t \in \{0, p_1, \ldots, p_m\}$ **do**

\[
\text{Compute } \hat{\text{FDP}}(t) = \frac{m \cdot t}{|\mathcal{F}(\{j : p_j \leq t\}, p)|};
\]

**end**

Compute $t^* \equiv \max\{t \in \{0, p_1, \ldots, p_m\} : \hat{\text{FDP}}(t) \leq q\}$;

**Result:** $\mathcal{R}^* = \{j : p_j \leq t^*\}$.

- Focused BH is a general-purpose way of dealing with filters; note that $\mathcal{F}$ can be a black box.
- When $\mathcal{F}$ does nothing, Focused BH reduces to BH.
- Procedure can be expanded to filters that prioritize rejections.
Focused BH provably controls FDR$_\mathcal{F}$

A filter $\mathcal{F}$ is **monotonic** if for $\mathcal{R}^1 \supseteq \mathcal{R}^2$ and $p^1 \leq p^2$, we have

$$|\mathcal{F}(\mathcal{R}^1, p^1)| \geq |\mathcal{F}(\mathcal{R}^2, p^2)|.$$

A filter is **simple** if $|\mathcal{F}(\mathcal{R}, p)|$ is independent of $p$.

**Theorem (K., Sabatti, Bogomolov)**

Focused BH controls FDR$_\mathcal{F}$ if either

1. p-values are independent, $\mathcal{F}$ is simple or monotonic.
2. p-values are “positively dependent” (PRDS), $\mathcal{F}$ is monotonic.

- Proof for item 1 inspired by Benjamini and Bogomolov (2014);
- Proof for item 2 inspired by Blanchard and Roquain (2008).

Simulations suggest Focused BH is robust.
Corollary

Focused BH controls the outer nodes FDR on trees if the p-values are positively dependent.

Proof: The outer nodes filter is monotonic on trees.
Specializing to the outer nodes filter

**Corollary**

Focused BH controls the outer nodes FDR on trees if the p-values are positively dependent.

*Proof*: The outer nodes filter is monotonic on trees.

Focused BH is the first procedure provably controlling outer nodes FDR under dependence.
Improving the power of Focused BH

The numerator $m \cdot t$ in

$$\hat{\text{FDP}}(t) = \frac{m \cdot t}{|\mathcal{S}(\{j : p_j \leq t\}, p)|}$$

can be a conservative estimate of $V(t) = |\mathcal{U}(t) \cap \mathcal{H}_0|$. Can improve procedure’s power by tightening FDP estimate, e.g.

$$\hat{V}_{\text{oracle}}(t) = \mathbb{E}[V(t)] \leq m \cdot t.$$
Let $\tilde{p}$ be a “permuted” version of $p$. Then,

$$\mathbb{E}[V(t)] = \mathbb{E}[|\mathcal{F}\{j: p_j \leq t\}, p\} \cap \mathcal{H}_0|]$$
$$\approx \mathbb{E}[|\mathcal{F}\{j: \tilde{p}_j \leq t\}, \tilde{p}\} \cap \mathcal{H}_0|]$$
$$\leq \mathbb{E}[|\mathcal{F}\{j: \tilde{p}_j \leq t\}, \tilde{p}\}].$$

Given permutations $\tilde{p}^1, \ldots, \tilde{p}^B$, define

$$\hat{V}_{\text{perm}}(t) = \frac{1}{B} \sum_{b=1}^{B} |\mathcal{F}\{j: \tilde{p}_j^b \leq t\}, \tilde{p}^b\}|.$$ 

No theoretical results yet, but performs well in simulations.
Simulation: Setup

**Graph structure:** Forest of 20 binary trees of depth 6, with $m = 1260$ total nodes.

**Data generating mechanism:**
- 21 non-null leaves (out of 640), 98 total non-nulls;
- Leaf nodes get independent $p$-values;
- Internal nodes get $p$-values by applying Simes global test to their leaf descendants.

**Filter:** Outer nodes filter.
Simulation: Methods compared

- BH (targeting pre-filter FDR at level $q = 0.1$)
- Structured Holm\textsuperscript{7} (targeting FWER at level $q = 0.1$)
- Yekutieli\textsuperscript{8} (targeting post-filter FDR at level $q = 0.1$)
- Focused BH (targeting post-filter FDR at level $q = 0.1$)
  - Original version
  - Permutation version
  - Oracle version

\textsuperscript{7} Meijer and Goeman (2016)
\textsuperscript{8} Yekutieli (2008)
Simulation: Results

- False Filtered Discovery Rate
- Power

Signal Amplitude

- Focused BH
- Focused BH (permutation)
- Focused BH (oracle)
- BH
- Structured Holm
- Yekutieli
HLA region on chromosome 6 is known to affect many diseases. Conducted PheWAS analysis for the HLA-B*27:05 allele, studied previously by Cortes et al (Nature Genetics, 2017).

Computed p-values testing marginal association between this allele and the \( m = 3265 \) ICD-10 codes that had at least 50 cases.\(^9\)

BH, Structured Holm, Yekutieli, Focused BH applied with \( q = 0.05 \).

\(^9\)This filtering step does not need to be corrected for, since it does not take the response variable into account.
Number of outer node rejections made by each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Outer node rejections</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH</td>
<td>28</td>
</tr>
<tr>
<td><strong>Focused BH</strong></td>
<td><strong>24</strong></td>
</tr>
<tr>
<td>Structured Holm</td>
<td>13</td>
</tr>
<tr>
<td>Yekutieli</td>
<td>1</td>
</tr>
</tbody>
</table>
Focused BH rejects 34 nodes, 24 outer nodes

Key
- Not rejected by Focused BH
- Rejected by Focused BH
FBH rejects 11 outer nodes more than Structured Holm

Key
- □ Not rejected by Focused BH
- □ Rejected by Focused BH
- ■ Outer node rejected by Focused BH but not Structured Holm
Focused BH guarantees Type-I error control when data analysis involves automatic curation via a pre-specified filter.

Filtering framework is general; applies beyond examples presented.
Part II: From automatic to manual curation
Manually curating promising hypotheses

Consider the practice of re-running an FDR procedure with different target levels until one obtains a “good” rejection set.

\[ \mathcal{R}_k = \{ H(1), \ldots, H(k) \} : \text{set corresponding to } k \text{ smallest p-values.} \]

\[ \emptyset = \mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_m \subseteq \mathcal{H} \]
Manually curating promising hypotheses

Consider the practice of re-running an FDR procedure with different target levels until one obtains a “good” rejection set.

\[ \mathcal{R}_k = \{ H_{(1)}, \ldots, H_{(k)} \} \]: set corresponding to \( k \) smallest p-values.

\[ \emptyset = \mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_m \subseteq \mathcal{H} \]
Manually curating promising hypotheses

Consider the practice of re-running an FDR procedure with different target levels until one obtains a “good” rejection set.

\[ \mathcal{R}_k = \{ H_{(1)}, \ldots, H_{(k)} \} : \text{set corresponding to } k \text{ smallest } p\text{-values.} \]

\[ \emptyset = \mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_m \subseteq \mathcal{H} \]
Consider the practice of re-running an FDR procedure with different target levels until one obtains a “good” rejection set.

\[ \mathcal{R}_k = \{H(1), \ldots, H(k)\} \]: set corresponding to \( k \) smallest p-values.

\[ \varnothing = \mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_m \subseteq \mathcal{H} \]
Manually curating promising hypotheses

Consider the practice of re-running an FDR procedure with different target levels until one obtains a “good” rejection set.

\[ R_k = \{ H(1), \ldots, H(k) \} \]: set corresponding to \( k \) smallest p-values.

\[ \emptyset = R_0 \subseteq R_1 \subseteq \cdots \subseteq R_m \subseteq \mathcal{H} \]
Manually curating promising hypotheses

Consider the practice of re-running an FDR procedure with different target levels until one obtains a “good” rejection set.

$\mathcal{R}_k = \{H(1), \ldots, H(k)\}$: set corresponding to $k$ smallest p-values.

$\emptyset = \mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_m \subseteq \mathcal{H}$

Simultaneous inference is one solution (e.g. Goeman and Solari 2011, Berk et al 2013), but can be conservative.
Simultaneous selective inference

Data scientist wants to inspect a “menu” of options

\[ \emptyset = \mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_m \subseteq \mathcal{H}. \]

Idea: provide corresponding upper bounds

\[
\overline{\text{FDP}}(\mathcal{R}_k) = \frac{\log(\alpha^{-1})}{\log(1 + \log(\alpha^{-1}))} \frac{1 + n \cdot p(k)}{|\mathcal{R}_k|}
\]

such that

**Theorem (K. and Ramdas, AOS, in revision, 2019+)**

Under independence of null p-values,

\[
\mathbb{P}[\text{FDP}(\mathcal{R}_k) \leq \overline{\text{FDP}}(\mathcal{R}_k) \text{ for all } k] \geq 1 - \alpha
\]

for all \( n \) and all \( \alpha \leq 0.31 \).

Data scientist can freely choose from menu while maintaining validity of FDP bounds.
Simultaneous selective inference in a toy example

![Graph showing false discovery proportion bounds and hypothesis index]

- **Simultaneous Selective Bound (KR19)**
- **Simultaneous Bound (GS11)**
- **True FDP**
For bounds of the form \( \overline{\text{FDP}}(t) = \frac{a + bt}{R(t)} \), we seek \( a, b \) such that

\[
P[V(t) \leq a + bt \text{ for all } t \in [0, 1]] \geq 1 - \alpha,
\]

where \( V(t) = \sum_{j \in \mathcal{H}_0} I(p_j \leq t) \).

\[ \text{References:} \]
Robbins (1954) and Dvoretsky Kiefer Wolfowitz (1956)
Linear upper bounds for empirical processes

For bounds of the form $\overline{\text{FDP}}(t) = \frac{a + bt}{R(t)}$, we seek $a, b$ such that

$$\mathbb{P}[V(t) \leq a + bt \text{ for all } t \in [0, 1]] \geq 1 - \alpha,$$

where $V(t) = \sum_{j \in \mathcal{H}_0} I(p_j \leq t)$.

Existing finite-sample bounds:\(^{10}

- $\overline{V}(t) = \frac{1}{\alpha} nt$;  
  tight very near 0.
- $\overline{V}(t) = \sqrt{\frac{n}{2} \log \frac{1}{\alpha}} + nt$;  
  tight near 1.

We obtain a new bound by exploiting connection between empirical and Poisson processes.

\(^{10}\text{Robbins (1954) and Dvoretsky Kiefer Wolfowitz (1956)}\)
Comparing to existing bounds \((n = 500, \alpha = 0.05)\)
Simultaneous selective inference with side information

KR19+ bounds can leverage side information to give data scientists a better menu of rejection sets to choose from.

- Hypotheses ordered a priori
  (same menu as accumulation test\textsuperscript{11})
- Hypotheses ordered adaptively
  (same menu as AdaPT or STAR\textsuperscript{12})
- Hypotheses ordered according to variable selection importance
  (same menu as knockoffs\textsuperscript{13})

\textsuperscript{11}Li and Barber (2017)
\textsuperscript{12}Lei and Fithian (2018), Lei, Ramdas, Fithian (2019+)
\textsuperscript{13}Barber and Candes (2015)
Simultaneous selective inference for knockoffs

Knockoffs method (Barber and Candes, 2015) developed for variable selection with FDR control.

**Knockoff statistics** $W_1, \ldots, W_m$ assigned to variables instead of p-values, ordering variables based on

$$W_{(1)} \geq W_{(2)} \geq \cdots \geq W_{(m)}.$$

BR19+ derived uniform FDP bounds for knockoffs as well:

$$\overline{\text{FDP}}(\mathcal{R}_k) = \frac{\log\left(\frac{1}{\alpha}\right)}{\log(2 - \alpha)} \frac{1 + |\{j : W_j \leq -W_{(k)}\}|}{|\mathcal{R}_k|}.$$

Uniform bounds for knockoffs first considered by K. and Sabatti (AOAS, 2019).
Replicability guarantees for modern data analysis pipelines

Different modes of curation require different statistical approaches:

<table>
<thead>
<tr>
<th>Mode of curation</th>
<th>Statistical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Automatic (filtering)</td>
<td>Focused BH</td>
</tr>
</tbody>
</table>

Open questions:
- (Applications) Pairing applications with inferential guarantees;
- (Theory, Methodology) Filling in the spectrum with powerful procedures using realistic assumptions.
Replicability guarantees for modern data analysis pipelines

Different modes of curation require different statistical approaches:

<table>
<thead>
<tr>
<th>Mode of curation</th>
<th>Statistical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Automatic (filtering)</td>
<td>Focused BH</td>
</tr>
</tbody>
</table>

These lie on a spectrum from selective to simultaneous inference:

- Selective inference (1) - Less flexibility, but less conservative guarantees.
- Simultaneous inference (2) - More flexibility, but more conservative guarantees.

Open questions:
- ▶ (Applications) Pairing applications with inferential guarantees;
- ▶ (Theory, Methodology) Filling in the spectrum with powerful procedures using realistic assumptions.
Replicability guarantees for modern data analysis pipelines

Different modes of curation require different statistical approaches:

<table>
<thead>
<tr>
<th>Mode of curation</th>
<th>Statistical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Automatic (filtering)</td>
<td>Focused BH</td>
</tr>
</tbody>
</table>

These lie on a spectrum from selective to simultaneous inference:

<table>
<thead>
<tr>
<th>Selective inference</th>
<th>1</th>
<th>2</th>
<th>Simultaneous inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less flexibility, but less conservative guarantees</td>
<td></td>
<td></td>
<td>More flexibility, but more conservative guarantees.</td>
</tr>
</tbody>
</table>

Open questions:

- (Applications) Pairing applications with inferential guarantees;
- (Theory, Methodology) Filling in the spectrum with powerful procedures using realistic assumptions.
Thank you.

The vector $p$ is PRDS if for any null $j$ and non-decreasing set $D \subseteq [0, 1]^m$, the quantity $\mathbb{P}[p \in D | p_i \leq t]$ is nondecreasing in $t \in (0, 1]$. 

---

**Definition (Benjamini Yekutieli 2001)**

The vector $p$ is PRDS if for any null $j$ and non-decreasing set $D \subseteq [0, 1]^m$, the quantity $\mathbb{P}[p \in D | p_i \leq t]$ is nondecreasing in $t \in (0, 1]$. 

---
Definition of power in the context of filtering

Maximum possible weighted number of non-null rejections is

\[ T_{\text{max}} \equiv \max_{\mathcal{R}, p} \left\{ \sum_{j \in \mathcal{H}_1} U_j \right\} ; \quad U = \mathcal{F}(\mathcal{R}, p), \]

Then, define power via

\[ \pi(U) = \mathbb{E} \left[ \frac{\sum_{j \in \mathcal{H}_1} U_j}{T_{\text{max}}} \right]. \]
Simulation 2: GWAS with clump filtering

- Genome of length 3000, with 100 LD blocks of size 30
- Simulated genotype data with local correlations
- Phenotypes from linear model with 10 nonzero coefficients
- Univariante association p-values generated for each SNP
- For simplicity, filter uses a priori LD blocks as clumps
Simulation 2: Results

![Graphs showing False Filtered Discovery Rate and Power vs Signal Amplitude]

- BH
- Focused BH
- Focused BH (permutation)
- Focused BH (oracle)
Robustness experiment

![Graph showing FDR across signal amplitude for different experiments: non-monotonic, non-monotonic and non-PRDS, and non-PRDS. The x-axis represents signal amplitude, ranging from 0.00 to 1.00, and the y-axis represents FDR, ranging from 0.00 to 0.10. The graph includes error bars indicating variability. The legend is labeled as follows: Experiment, Non-monotonic, Non-monotonic and non-PRDS, and Non-PRDS.]
Outer nodes found by BH but not Focused BH

- Other and unspecified antidepressants [as a cause of death via complication of medical care]
- Urticaria [also known as hives]
- Localisation-related (focal) (partial) symptomatic epilepsy and epileptic syndromes with complex partial seizures
- Meniere’s disease
Outer nodes found by Focused BH but not Structured Holm

- Symptoms, signs and abnormal clinical and laboratory findings
- Other benign neoplasms of connective and other soft tissues
- Meningitis, unspecified
- Other specified polyneuropathies
- Cardiomegaly
- Scrotal varices
- Chronic sinusitis
- Paralysis of vocal cords and larynx
- Cellulitis of other sites
- Rheumatoid arthritis, unspecified (Multiple sites)
- Other synovitis and tenosynovitis
FBH rejects 4 nodes fewer than BH

Key
- Not rejected by Focused BH
- Rejected by Focused BH
- Rejected by BH but not Focused BH
Focusing on diseases of the musculoskeletal system

Diseases of the musculoskeletal system and connective tissue

- Disorders of synovium and tendon
  - Synovitis and tenosynovitis
    - Other synovitis and tenosynovitis
      - Rheumatoid arthritis, unspecified
        - Rheumatoid arthritis, unspecified (Multiple sites)
        - Rheumatoid arthritis, unspecified (Shoulder region)
        - Rheumatoid arthritis, unspecified (Hand)

- Inflammatory polyarthropathies
  - Other rheumatoid arthritis
    - Rheumatoid arthritis, unspecified (Shoulder region)

- Infectious arthropathies
  - Ankylosing spondylitis
    - Ankylosing spondylitis (Site unspecified)

- Spondylopathies

"Focusing on diseases of the musculoskeletal system"
Focusing on diseases of the skin

Diseases of the skin and subcutaneous tissue

- Infections of the skin and subcutaneous tissue
  - Cellulitis
    - Cellulitis of other sites
  - Papulosquamous disorders
    - Psoriasis
      - Arthropathic psoriasis
  - Urticaria and erythema
    - Urticaria (Hives)
Soft outer nodes filter
Multi-filter Focused BH

Given $M$ filters $\mathcal{F}_1, \ldots, \mathcal{F}_M$, suppose one wants $\mathcal{R}^*$ such that

$$\text{FDP}_{\mathcal{F}_k} = \mathbb{E}[\text{FDP}(\mathcal{F}_k(\mathcal{R}^*, p))] \leq q_k \text{ for all } k = 1, \ldots, m.$$  

For a threshold $t$, we can construct $\widehat{\text{FDP}}_k(t)$ as in Focused BH, and then choose

$$t^* = \max\{t \in \{0, p_1, \ldots, p_m\} : \widehat{\text{FDP}}_k(t) \leq q_k \text{ for all } k\}.$$  

This will control FDR for all filtered rejection sets if $p$ is PRDS and all filters are monotonic.
Writing

\[ \hat{m}_0^\lambda = \frac{1 + |\{j : p_j > \lambda\}|}{1 - \lambda}, \]

following Storey, we can define

\[ \hat{\text{FDP}}_{\text{Storey}}(t) = \frac{\hat{m}_0^\lambda \cdot t}{|\tilde{\mathcal{S}}(\mathcal{R}(t, p), p)|}. \]

The corresponding procedure controls FDR under independence for simple filters.