pARI package: valid double-dipping via permutation-based All Resolutions Inference

Angela Andreella[1] Livio Finos[2] Jelle Goeman[3]

[1]Department of Statistical Sciences, University of Padua

[2]Department of Developmental Psychology and Socialisation, University of Padua

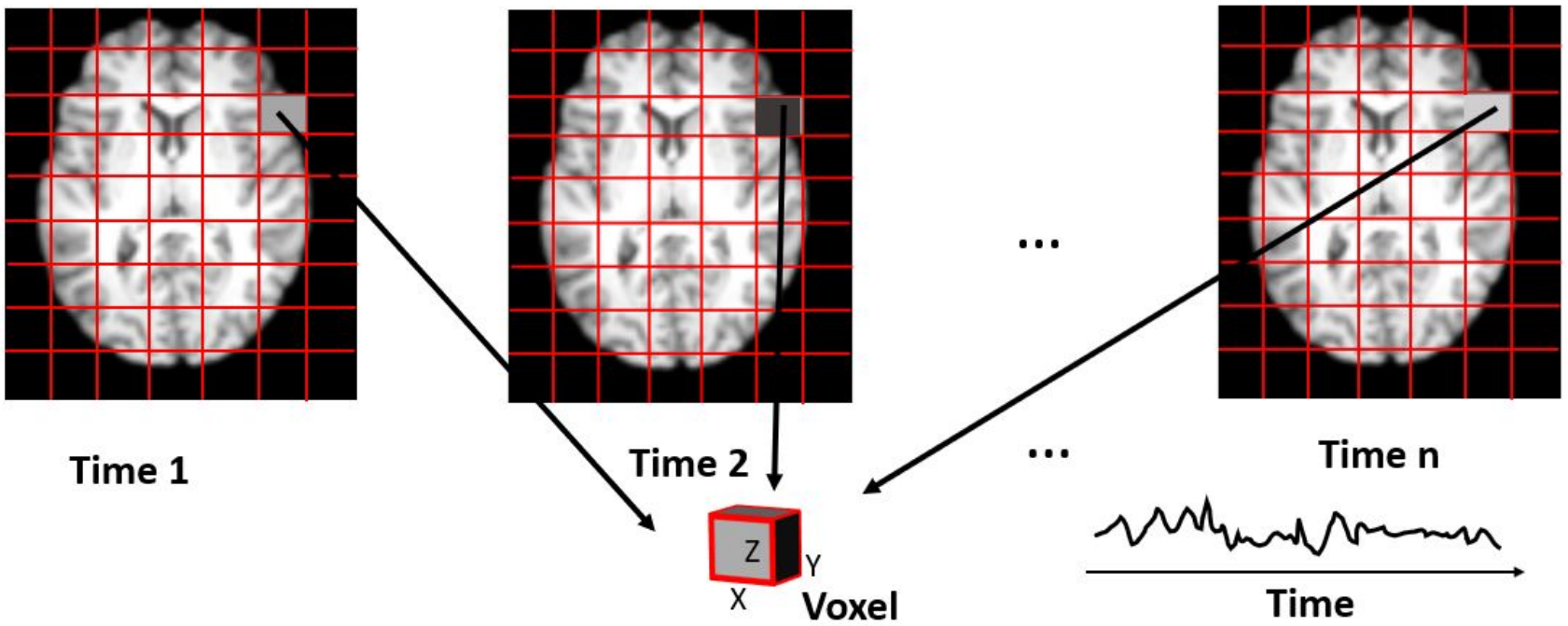
[3]Biomedical Data Sciences, Leiden University Medical Center

EuroBioc - 15 December 2020

Dealing with **multiple testing problem** leads to decide which correction method must be applied to handle potential false positives.

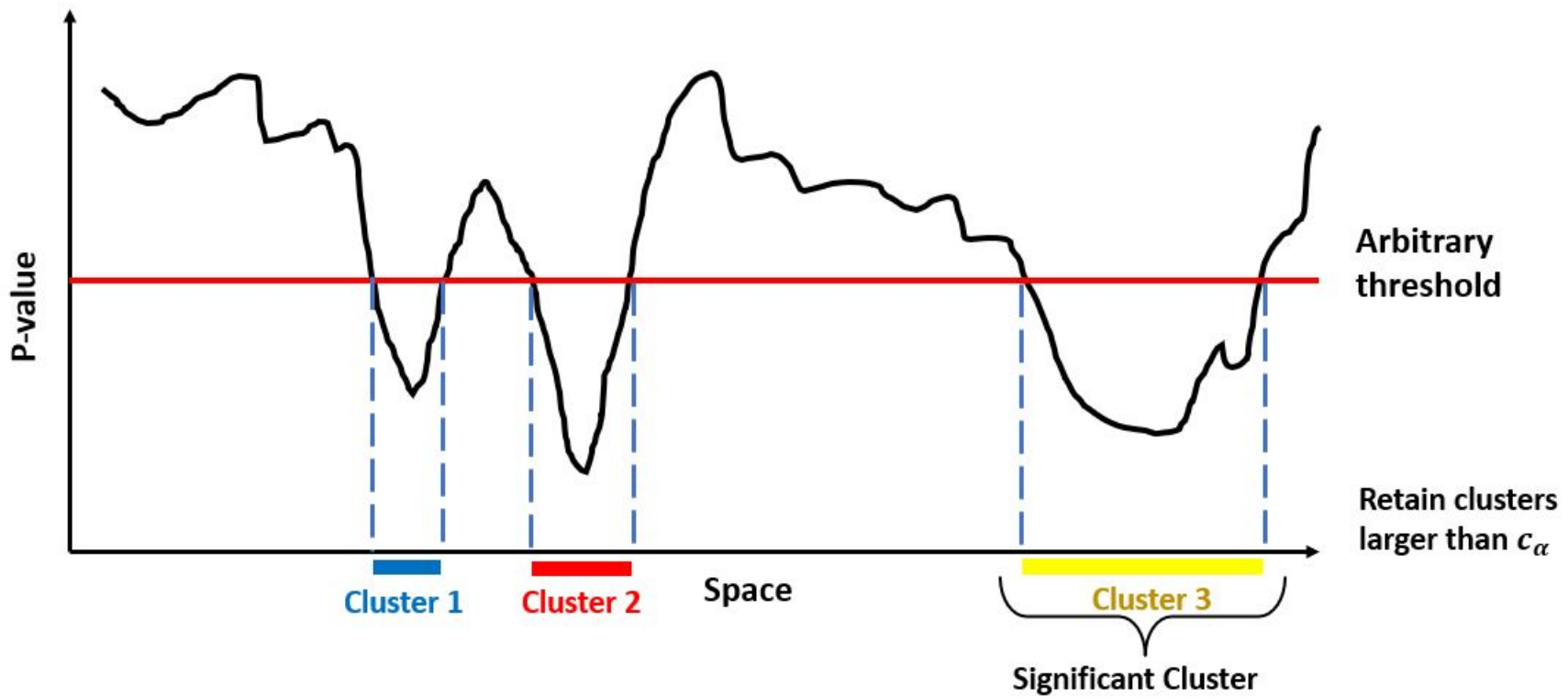
- At level of single variable or set of variables?
- Low **power** or low **information**?

fMRI is the most used technology to study human cognition's neural underpinnings.



We have a **test statistic** ∀ voxel corresponding to the **null hypothesis that the voxel is not active**.

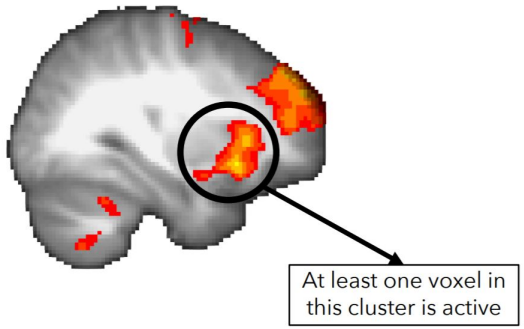
In fMRI data analysis, the most widely used method for locating brain activity is **cluster-extent based thresholding** [1].



[1] Woo, C. et al. (2014).

Problems:

- **Spatial Specificity paradox** [2] → The larger the cluster, the weaker the finding!
- Inferring inside data-driven clusters is problematic!→ **Double-dipping**, **Circular analysis**, **Cherry picking** ...



Solution: All-resolutions inference (ARI)[3] → Inference on the **number of truly active voxels**.

[2] Woo, C. et al. (2014).
[3] Rosenblatt, J. et al. (2018).

Every time that we want to infer inside a **data-driven (and not) cluster** (features set), we can use **ARI**:

- **Cluster fMRI data analysis** [4];
- **Gene expression cluster analysis** [5];
- **Cluster EEG data analysis** [6].

[4] Woo, C. et al. (2014).
[5] Berge, K. et al. (2017).
[6] Maris, E. et al. (2007).

Let consider B collection of m features to test, e.g., voxel activation, gene count, etcetera.

- **Hypothesis**: $H_1, \ldots, H_m$;
- **True Discoveries**: $A \subset \{1, \ldots, m\}$;
- **Cluster**: $S \subset \{1, \ldots, m\}$ selected hypothesis;
- **True Discoveries in S**: $|A \cap S| \subset \{1, \ldots, m\}$.

We want to infer on $a(S) = |A \cap S|$, or equivalently true discovery proportion (TDP), i.e., $\pi(S) = a(S)/|S|$.

Two ingredients:

- **Closed testing procedure** for controlling the familywise error rate[7];
- **Local test** used in the closed testing procedure;

The closed testing principle rejects any H_i, if all possible intersection hypotheses involving H_i are rejected by using valid local level α tests. $\rightarrow 2^m$ hypothesis to tests!!

However...Goeman et al. (2020) finds a good general shortcut by defining the concept of **critical vector**.

[7] Marcus, E. et al. (1976).

Definition

A vector $l_1, \ldots, l_m$ is a critical vector if and only if

$$\Pr(\cap_{i=1}^{|N|}\{q_{(i)} \geq l_i\}) \geq 1 - \alpha, \tag{1}$$

where $N = B \setminus A$ is the set of inactive voxels, and $q_{(i)}$, $1 \leq i \leq |N|$ are their sorted p-values.

Lemma

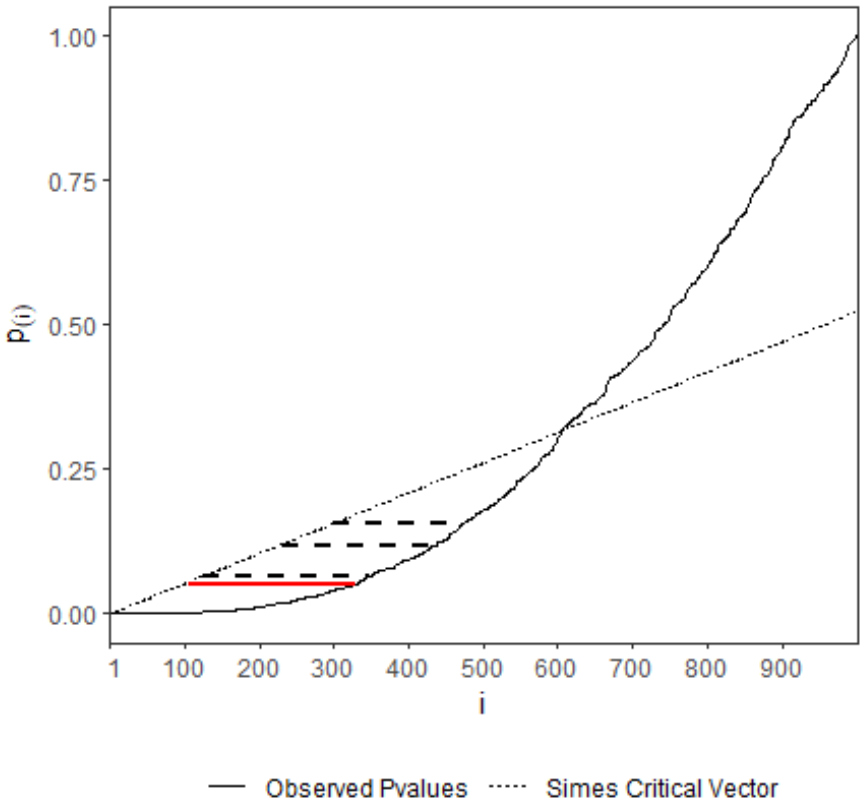
Let l_i satisfy (1). Then for every $\emptyset \neq S \subseteq B$,

$$\bar{a}(S) = \max_{1 \leq u \leq |S|} 1 - u + |\{i \in S : p_i \leq l_u\}|. \tag{2}$$

is a lower $(1-\alpha)$ confidence bound of $a(S)$, simultaneously for all $S \subseteq B$, that is

$$\Pr(\forall S \subseteq B : \bar{a}(S) \leq a(S)) \geq 1 - \alpha. \tag{3}$$

All resolution inference (ARI)



where $z = |\{i \in S : p_i \leq l_u\}|$ of the previous Lemma.

The **parametric** version, i.e., local tests are based on the **Simes tests**, uses:

$$\boxed{l_i = \frac{\alpha\, i}{h}} \tag{4}$$

where h is the largest size of a features set not rejected by the Simes test.

Parametric ARI is not assumption-free → **Simes inequality**, i.e., positive regression dependency on subsets [8], also adopted for the FDR [9] controlling approach.

However... it can be **conservative** in the case of strong **correlation** between tests.

[8] Sarkar, S. K. (2008).
[9] Benjamini, Y. (1995).

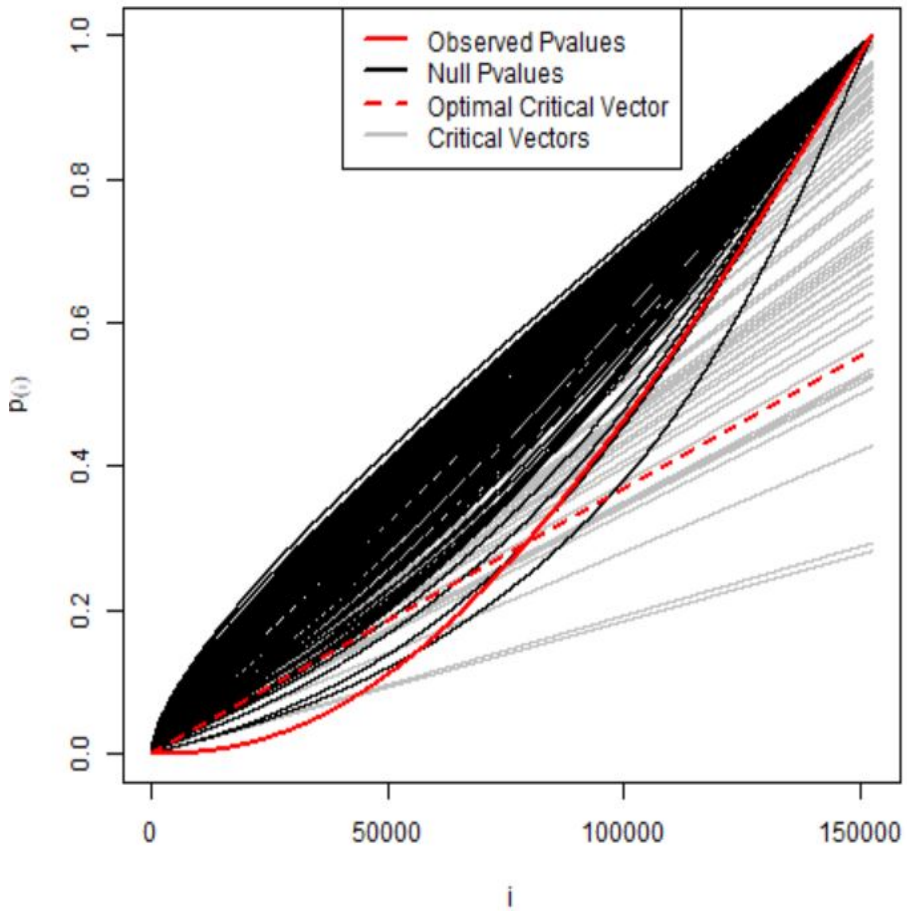
Solution: Permutation approach !

It accounts for the p_i's **dependence structure** → higher critical curve than parametric methods → more **power**.

- Let $p_1^1, \ldots, p_m^1 = p_1, \ldots, p_m$ be the p-values for the real data;
- Let $p_1^j, \ldots, p_m^j$ be the p-values obtained for the j-th random permutation of the data, where w is the total number of random **permutations** or **sign-flipping** transformations;
- Let the **family of candidate critical vectors** be $\mathcal{F} = \{l(\lambda_\alpha) : \lambda_\alpha \in \Lambda\}$, where $\lambda_\alpha \in \Lambda \subseteq \mathbb{R}$.

Permutation-based ARI - λ_α calibration

$$\lambda_\alpha = \sup\{\lambda \in \Lambda : w^{-1}|\{1 \le j \le w : p_i^j \ge l_i(\lambda) \quad \forall i\}| \ge 1-\alpha\}$$



Let be $\delta \in \mathbb{R}^+$:

- **Simes**[10]: $l_i(\lambda_\alpha) = \frac{(i-\delta)\lambda_\alpha}{m-\delta}$, where $\Lambda \subseteq \mathbb{R}_{>0}$;
- **Asymptotically Optimal Rejection Curves (AORC)**[11]: $l_i(\lambda_\alpha) = \frac{(i-\delta)\lambda_\alpha}{(m-\delta)-(i-\delta)(1-\lambda_\alpha)}$, where $\Lambda \subseteq \mathbb{R}_{>0}$;
- **Higher Criticism**[12]: $l_i(\lambda_\alpha) = \frac{2i+\lambda_\alpha^2-\sqrt{(2i+\lambda_\alpha^2)^2-4i^2(m+\lambda_\alpha^2)/m)}}{2(m+\lambda_\alpha^2)}$, where $\Lambda \subseteq \mathbb{R}$;
- **Beta**[13]: $l_i(\lambda_\alpha) = \inf\{x : \lambda_\alpha \leq F_i(x)\}$, where $\Lambda = [0,1]$ and $F_i(X)$ is the cdf of Beta$(i, m+1-i)$.

[10] Simes (1986).
[11] Finner et al. (2009).
[12] Donoho and Jin (2004).
[13] Hemerik et al. (2019).

Theorem

The vector $l(\lambda_\alpha)$ is a critical vector, i.e., it satisfies (1).

The parametric ARI equals the permutation-based ARI if:

- Simes-based $\mathcal{F}$;
- $\lambda_\alpha = \alpha$ and $\delta = 0$.

```
devtools::install_github(angeella/pARI)
library(pARI)
```

- fMRI framework:

```
pARIbrain(copes, thr, mask, alpha, ...)
```

- General framework:

```
pARI(data, ix, alpha, test.type, ...)
```

where `ix` is the **features set** of interest. It can be a vector of indices or a vector with length equals the number of features where different values indicate the different sets.

We analyzed the **Auditory data** collected by Pernet et al. (2015), i.e, people listening vocal and non-vocal sounds.

Group analysis on 140 subjects of the Vocal > Non-vocal **contrast** by the one sample t-test flipping the sign of 140 voxel-wise contrasts maps.

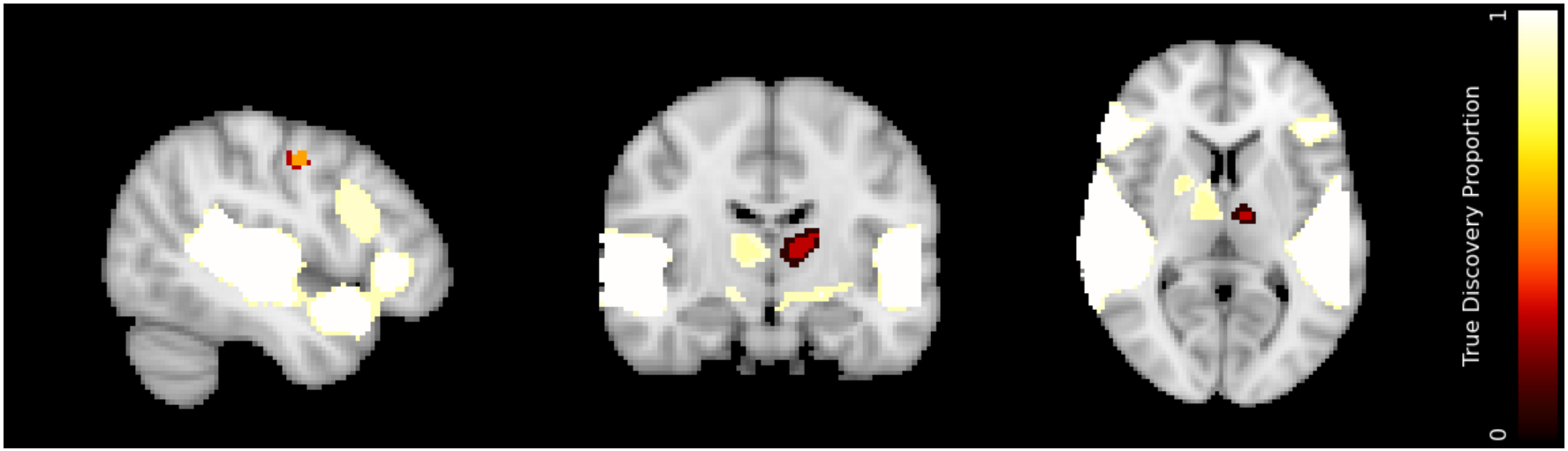
First, let download the data from the `fMRIdata` package:

```
devtools::install_github(angeella/fMRIdata)
library(fMRIdata)
data(Auditory_clusterTH3_2)
data(Auditory_copes)
data(Auditory_mask)
```

```
pARIbrain(copes = Auditory_copes,
          cluster = Auditory_clusterTH3_2,
          mask = Auditory_mask,
          alpha = 0.05)
```

Cluster S	**Threshold** t	**Size** $\lvert S \rvert$	**% active** $\bar{\pi}(S)$ Perm. Simes	Param. Simes	**P-Values** p_{FWER}
Right STG/PT HG/IFG/T	3.2	11683	92.36%	84.98%	< 0.0001
Right STG/PT HG/IFG/T	4	8875	99.54%	98.5%	–
Right IFG	4	422	91.47%	83.18%	–
Right T	4	292	85.96%	64.04%	–
Right T	4	15	13.33%	0%	–

`map_TDP`: Create **true discovery proportion map** in nifti format.



Let consider a simple example using Bottomly et. al (2011) data, i.e., a comparative RNA-seq analysis of different mouse strains:

- 1049 genes and 21 samples, where 10 C57BL/6J strain and 11 DBA/2J strain;
- After pre-processing steps, we perform a **two-samples t-tests** for each gene;
- We define as **sets of interest** the ones computed by `hclust`.

```
pARI(X = bottomly.eset_preprocessed,
     ix = my.clusters,
     test.type = "two_samples",
     label = label,
     alternative = "two.sided")
```

Gene expression data application - Results

Cluster S	**Size** $\lvert S\rvert$	**% active** $\bar{\pi}(S)$	
		Perm. Simes	Param. Simes
1	3183	27.66%	3.5%
2	3150	34.27%	5.58%
3	2313	22.17%	2.7%
4	2243	12.77%	0%

Key Points

The **permutation-based ARI**:

- **flexible**, **mild** and **post-hoc** → You can choose the features set as many time as you want!
- resolves the **spatial specificity paradox** of cluster fMRI data analysis;
- is general, allowing different families of **confidence bounds**;
- accounts for the **dependence** structure of the data;
- makes **no assumptions** on the null distribution of the p-values → only **exchangeability**;
- the **power** does not depend on the number of features sets tested.

Thanks to the amazing group that worked on the paper *Permutation-based true discovery proportions for fMRI cluster analysis*! (in arXiv)

- Livio Finos University of Padua;
- Jelle Goeman Leiden University Medical Centre;
- Jesse Hemerik Wageningen University;
- Wouter Weeda Leiden University;

and thanks for your attention!

 angeella

 @aangeella

 angela.andreella@stat.unipd.it