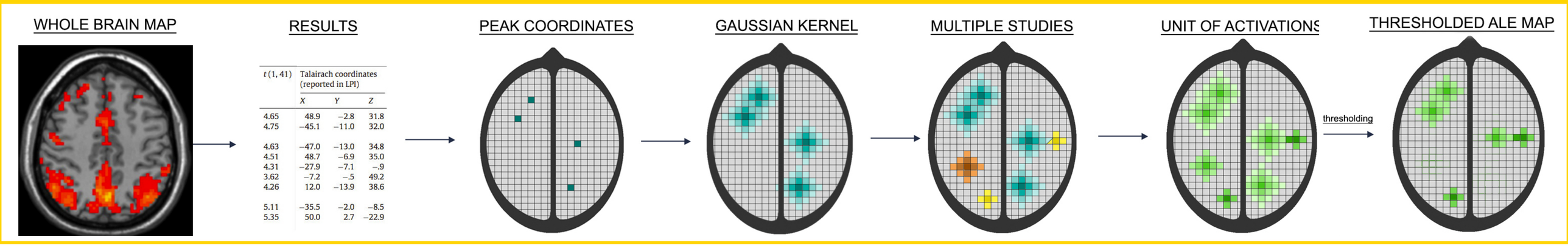
Assessing cluster validity in coordinate-based meta-analysis for fMRI

Freya Acar[1]*, Ruth Seurinck[1], Simon B. Eickhoff[2, 3] and Beatrijs Moerkerke[1]

[1]Data Analysis, Ghent University, Belgium
[2]Institute for Systems Neuroscience, Medical Faculty, Heinrich-Heine University Düsseldorf, Germany
[3]Institute for Neuroscience and Medicine (INM-7), Research Center Jülich, Germany

ALE[1] meta-analysis of fMRI studies



t (1, 41)	Talairach coordinates (reported in LPI)		
	X	Y	Z
4.65	48.9	−2.8	31.8
4.75	−45.1	−11.0	32.0
4.63	−47.0	−13.0	34.8
4.51	48.7	−6.9	35.0
4.31	−27.9	−7.1	−.9
3.62	−7.2	−.5	49.2
4.26	12.0	−13.9	38.6
5.11	−35.5	−2.0	−8.5
5.35	50.0	2.7	−22.9

Assessment of publication bias

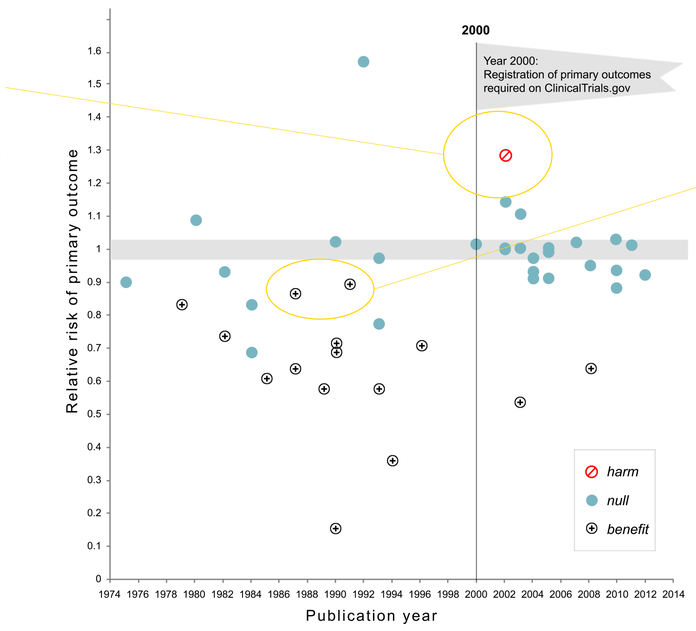
Publication bias occurs when the results of published and unpublished studies differ significantly. There are two different forms, within- and between study publication bias. In the graph below we see the results of clinical trials before and after preregistration became mandatory in 2000[2].

Between-study: File-drawer problem

Studies that fail to show statistically significant results or show results that are not in line with the research hypothesis, remain in the file drawer.

Fail-Safe N[3]

The Fail-Safe N (FSN) quantifies the amount of null studies necessary before a previously statistically significant cluster is no longer statistically significant. This shows the possible influence of studies that remain in the file drawer and is a measure for robustness of the cluster against random noise.



Within-study: p-hacking

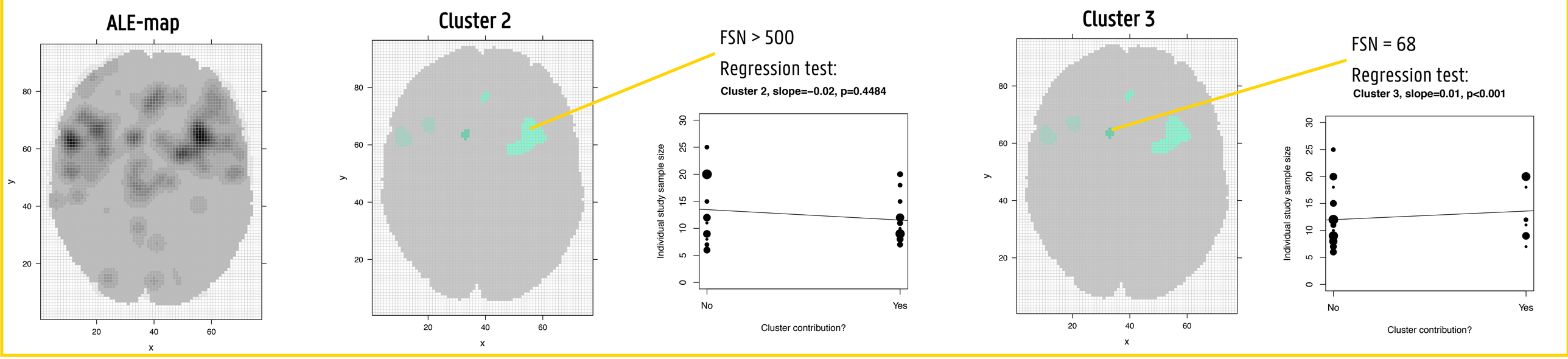
The procedure of a study is altered to obtain statistically significant results (e.g. adding or removing participants, more lenient thresholding)

Regression test[4]

Because they are underpowered, small studies tend to employ more lenient thresholds, resulting in and explosion of false positives (small sample bias). Is the statistically significant cluster driven by studies with small sample sizes and therefore vulnerable to a small sample bias?

Example: Meta-analysis of fMRI studies on taste

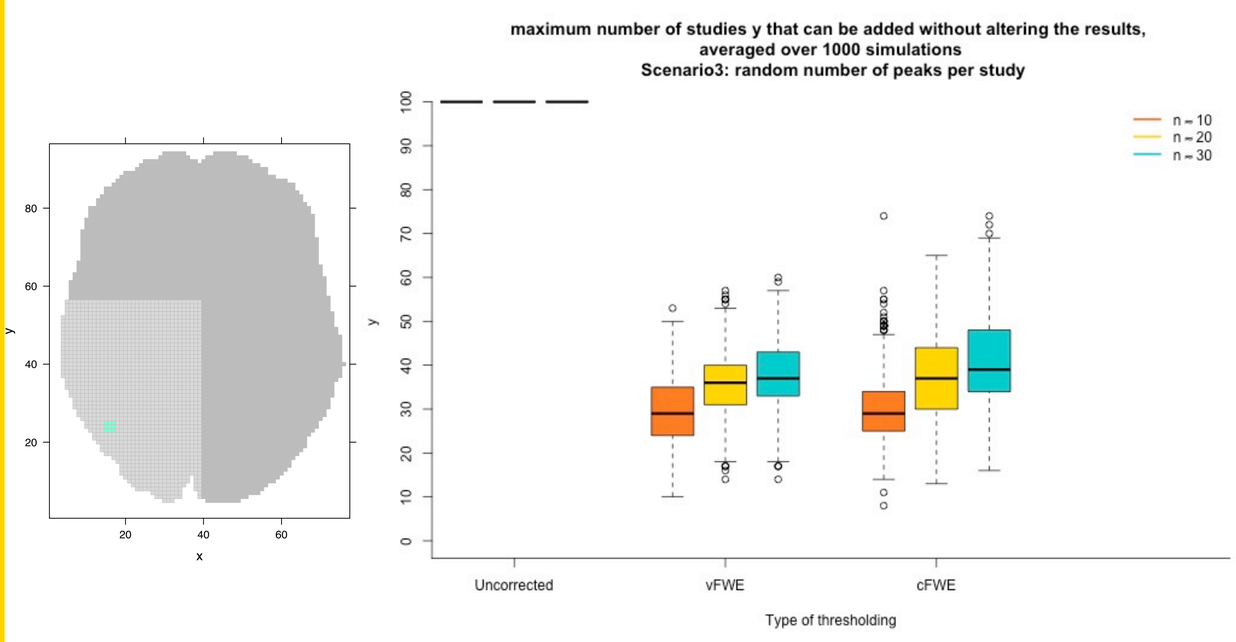
Hypothetical meta-analysis on the "taste" paradigm, studies are collected from the BrainMap database through Sleuth. 86 experiments, 521 foci, 1075 participants, cluster-level FWE $p < 0.05$, cluster-forming threshold $p < 0.001$ uncorrected. On the left we see the resulting ALE-map and on the right the statistically significant clusters after thresholding.



Simulations

Fail-Safe N

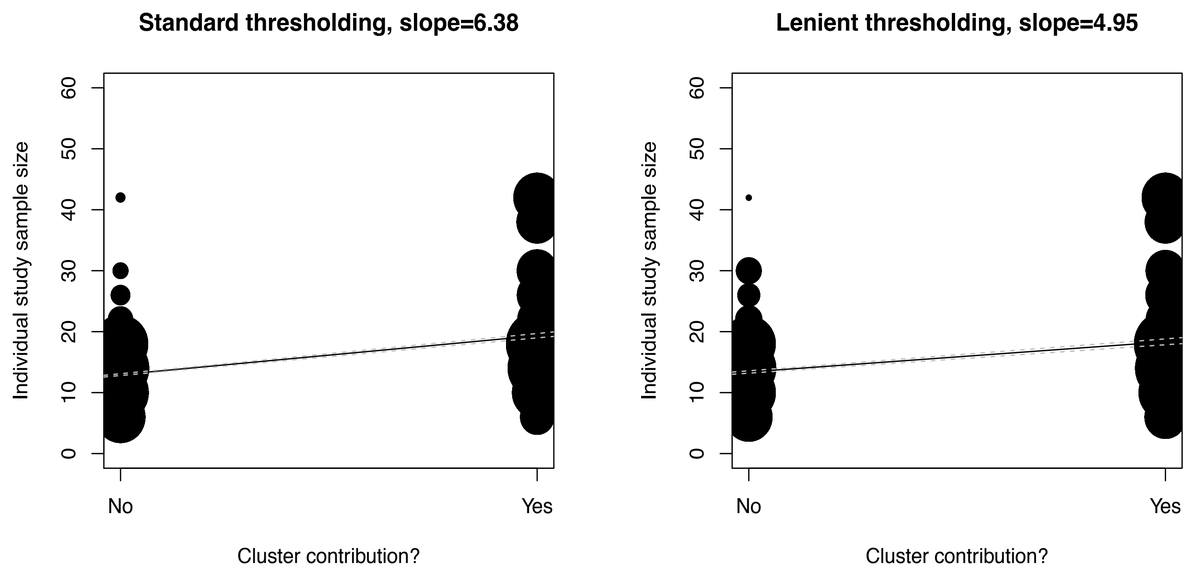
- 3 real studies with activation in target region
- Up to 100 null studies with activation in other quadrants than target region
- Look at effect of
  - Individual study sample size (n ~ 10, n ~ 20 or n ~ 30)
  - Thresholding method: uncorrected, voxel- or cluster-level FWE



Results: effect of number of peaks, sample size and thresholding method

Regression test

- Select 1 t-map from a meta-analysis & compute average effect size in ROI
- Compute power in ROI for different sample sizes with standard (FDR $q < 0.01$) & lenient thresholding ($p < 0.05$)
- Simulate cluster contribution based on power (x 100)
  => Depends on sample size, effect size and thresholding method



Results: slope changes significantly if lenient thresholding is applied

[1] Eickhoff et al., (2009, 2012). *Human Brain Mapping*, 30; *Neuroimage*, 59; Turkeltaub et al., (2012). *Human Brain Mapping*, 33.
[2] Kaplan & Irvin (2015). *PLoS ONE*, 10(8). [3] Rosenthal, (1979). *Psychological Bulletin*, 86(3).
[4] Egger & Smith, (1997). *BMJ*, 315.

freya.acar@ugent.be