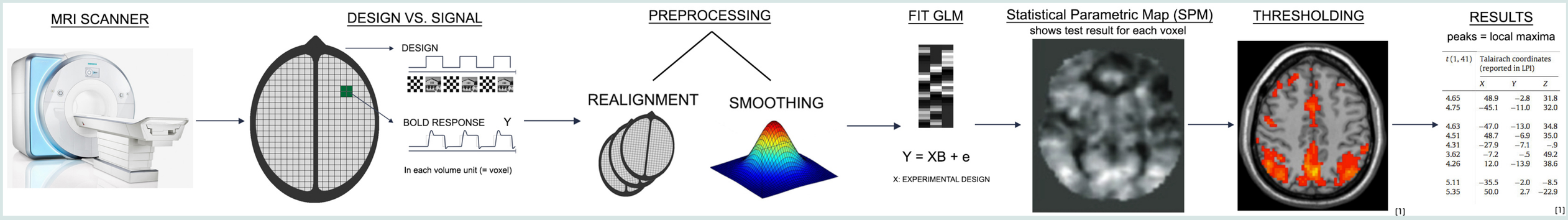
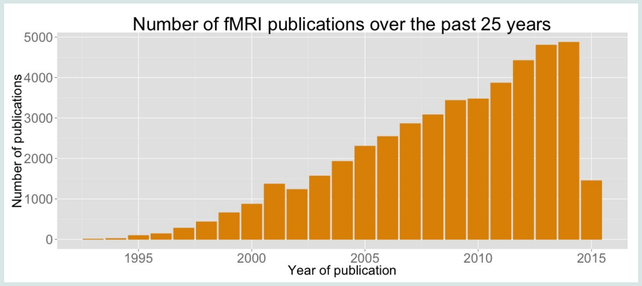
Assessing publication bias in coordinate-based meta-analysis techniques?

Freya Acar[1], Simone Kühn[2], Beatrijs Moerkerke[1], Ruth Seurinck[1]
[1] Ghent University, Department of Data Analysis, [2] Max Planck Institute for Human Development, Berlin

1. Functional Magnetic Resonance Imaging (fMRI): localising brain activation



2. fMRI: methodological problems



- Small sample sizes (due to cost of fMRI research)
- Noisy data (e.g. scanner or physiological artefacts,..)
- Multiple testing problem ($>$ 100,000 voxels are tested simultaneously)
  - explosion of false positives
  - stringent thresholding needed to control false positive rate
  - most studies become underpowered under stringent thresholding

3. Meta-analysis

Advantages meta-analysis of fMRI studies

- Aggregation of results $\Rightarrow$ reproducibility
- Increases power

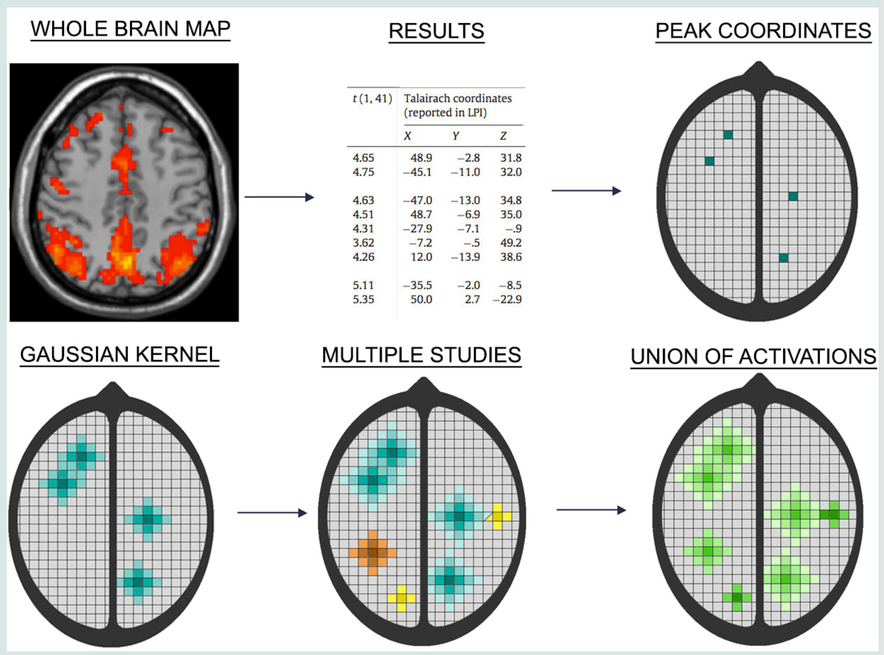
Publication bias

Censored data

- Between studies:
  - studies that fail to show significance in a certain region fail to get published
  - studies apply different thresholds to the data
- Within studies:
  - usually only coordinates of peaks are reported, the entire SPM is not available.

Coordinate-based meta-analysis

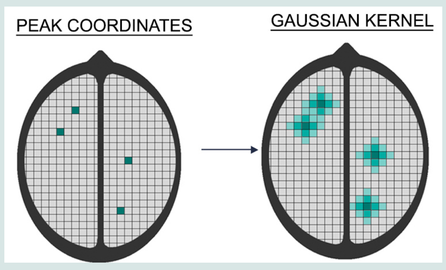
Activation Likelihood Estimation (ALE) [2,3,4]



4. Gaussian smoothing and union of likelihood of activation

Kernel smoothing captures spatial character of true activation

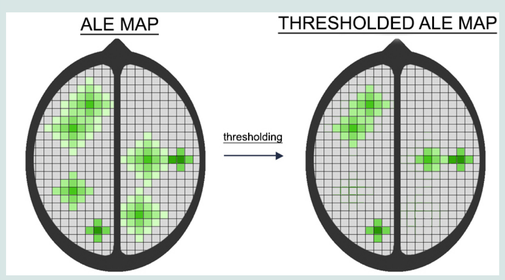
- Result: maps with modeled activation (MA-maps)
  - MA-value: for each voxel the probability of an activation being located at exactly that voxel.
  - Takes into account location and spatial uncertainty of reported coordinates of that study.
  - sample size ↑, spatial uncertainty ↓, FWHM ↓



- In the individual studies (MA maps), all voxels i (i=1,...,V) have a value between 0 and 1.
- A union of these MA-values is computed to construct the summary ALE map.
- Suppose we have K studies (k=1,..,K) than the ALE value in voxel i is equal to

$$ALE_i = 1 - \prod_{k=1}^{K}(1 - MA_{i,k})$$

5. Thresholding



- Uncorrected or
- Account for huge multiple testing problem through control of the False Discovery Rate (FDR).

6. Assessing publication bias?

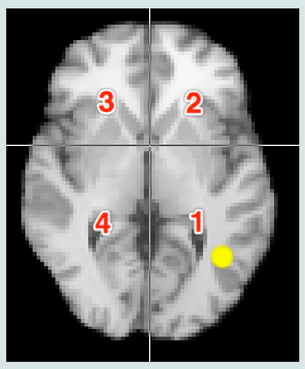
What happens if null studies are added? How many can be added without altering the results?

At which point are the results sufficiently robust?

$\Leftrightarrow$ At which point are the results too lenient? Is the union of activations too liberal?

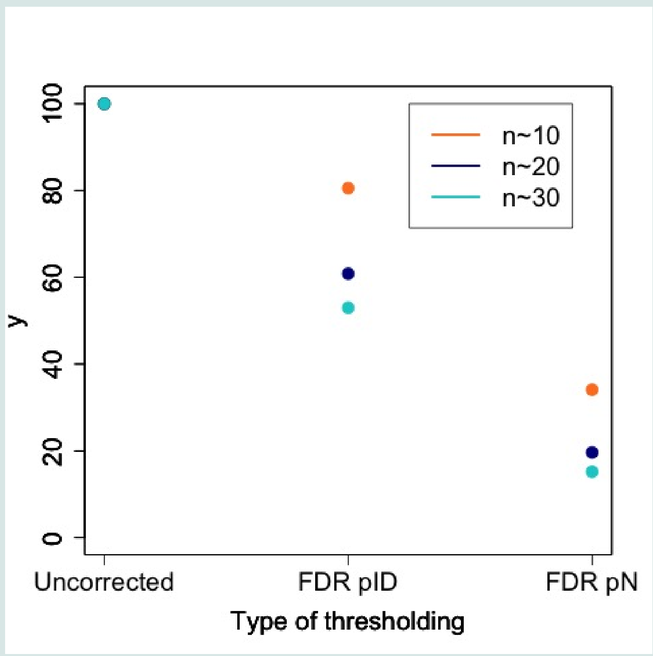
7. Simulation study

- Assumed activation in a specific region (target voxel)
- 3 real studies with peak close to target voxel in quadrant 1
- 100 null studies each with 1 peak in quadrants 2-4
- Effect of sample size
  - small (n$\sim$10), medium (n$\sim$20) or large (n$\sim$30)
- Effect of thresholding
  - uncorrected ($p < 0.001$)
  - FDR pID (assumes independent tests or positive dependence among tests, $p < 0.05$)
  - FDR pN (makes no assumptions about dependencies, more conservative., $p < 0.05$)



8. Results and discussion

maximum number of studies y that can be added without altering the results averaged over 1000 simulations



Robustness versus leniency:

- What is an acceptable number of null studies that can be added without altering the results?
  - Too low? Points at non-robust results. (In spirit of classic Fail-Safe N [5])
  - Too high? One or a small number of studies drives the analysis.
  - Results for sample size: contra-intuitive for robustness but intuitive for leniency.
- Would it be more useful to take an average of MA-values instead of the union (cfr. classic meta-analysis)?
- Employing uncorrected thresholding is unacceptable

9. References

[1] Han et al. (2015) *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 59.
[2] Eickhoff et al., (2009). *Human Brain Mapping*, 30.
[3] Eickhoff et al., (2012). *Neuroimage*, 59.
[4] Turkeltaub et al., (2012). *Human Brain Mapping*, 33.
[5] Rosenthal, (1979). *Human Brain Mapping*, 33.