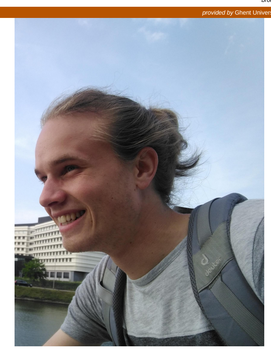
The interplay between sample size and replicability of results in fMRI studies

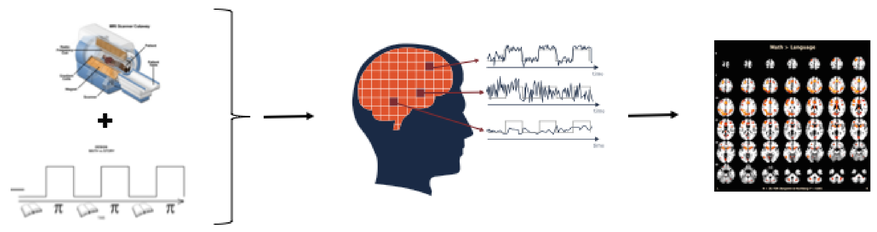
Han Bossier[1], Ruth Seurinck[1], Sanne Roels[1], Simone Kühn[2] and Beatrijs Moerkerke[1]

Department of Data Analysis, Ghent University, Belgium[1]
Department of Psychiatry and Psychotherapy, University Clinic, Germany[2]



[1] INTRODUCTION

A popular tool to localise cognitive abilities in the brain is **functional Magnetic Resonance Imaging** (fMRI).



- Brain: $>$ 100,000 artificially created volume units (voxels).
- Measures a time series of the blood oxygenated level dependent response.
- Relate this to experimental design.
- Localize cognitive tasks.
- Limited **replicability** (Poldrack et al., 2017).
- Slow adoption of power and sample size calculations.

$\rightarrow$ At what sample size do we achieve a reasonable replicability?

REPLICABILITY

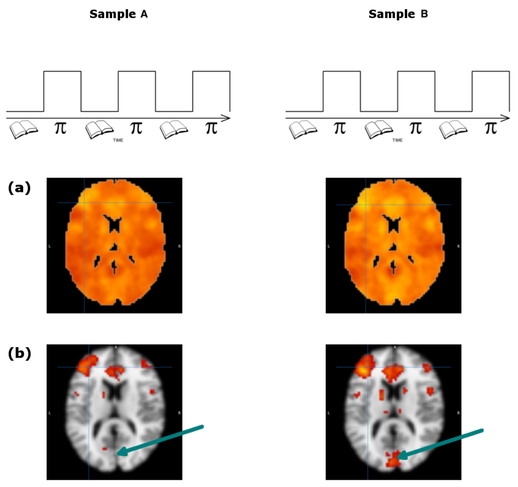
The ability of obtaining identical spatial results with the same experimental procedure + data analysis using a new (independent) dataset.

[2] MEASURES OF REPLICABILITY

1] **Test-retest reliability** for (un)conditional (i.e. non-thresholded) statistical parametric maps (SPM).
$\rightarrow V_{A,B}$ = intersection of activated voxels in image A and B.
$\rightarrow V_A = V_B$ = number of activated voxels in respectively image A and B.

2] Assume sum of thresholded & binarised images (A, B, ...) is realisation of a mixture of two binomial distributions (H_0 and H_a).
$\rightarrow$ Good separation of the parameters in the mixture = **coherence**.
$\rightarrow$ Need to correct for chance of correct classification. See table where p_0 = voxels correctly classified by mixture model and p_c = those correctly classified by chance.

3] **Stability** of fMRI results.



Measurement	Calculation	Feature
Unconditional test-retest reliability	Pearson's Product - Moment Correlation Coefficient (ρ)	Voxel: SPM
Conditional test-retest reliability (overlap)	$\omega_{A,B} = \frac{V_{A,B}}{V_A + V_B - V_{A,B}}$	Voxel: thresholded SPM*
Coherence	$\kappa = \frac{p_0 - p_c}{1 - p_c}$	Voxel: thresholded SPM*
Stability	Number of (unique) clusters	Clusters: thresholded SPM*
	Variability of cluster size	Clusters: thresholded SPM*

* thresholded SPM = here binary data

PDF + Contact Information

https://github.com/NeuroStat
@HBossier
Han.Bossier@Ugent.be

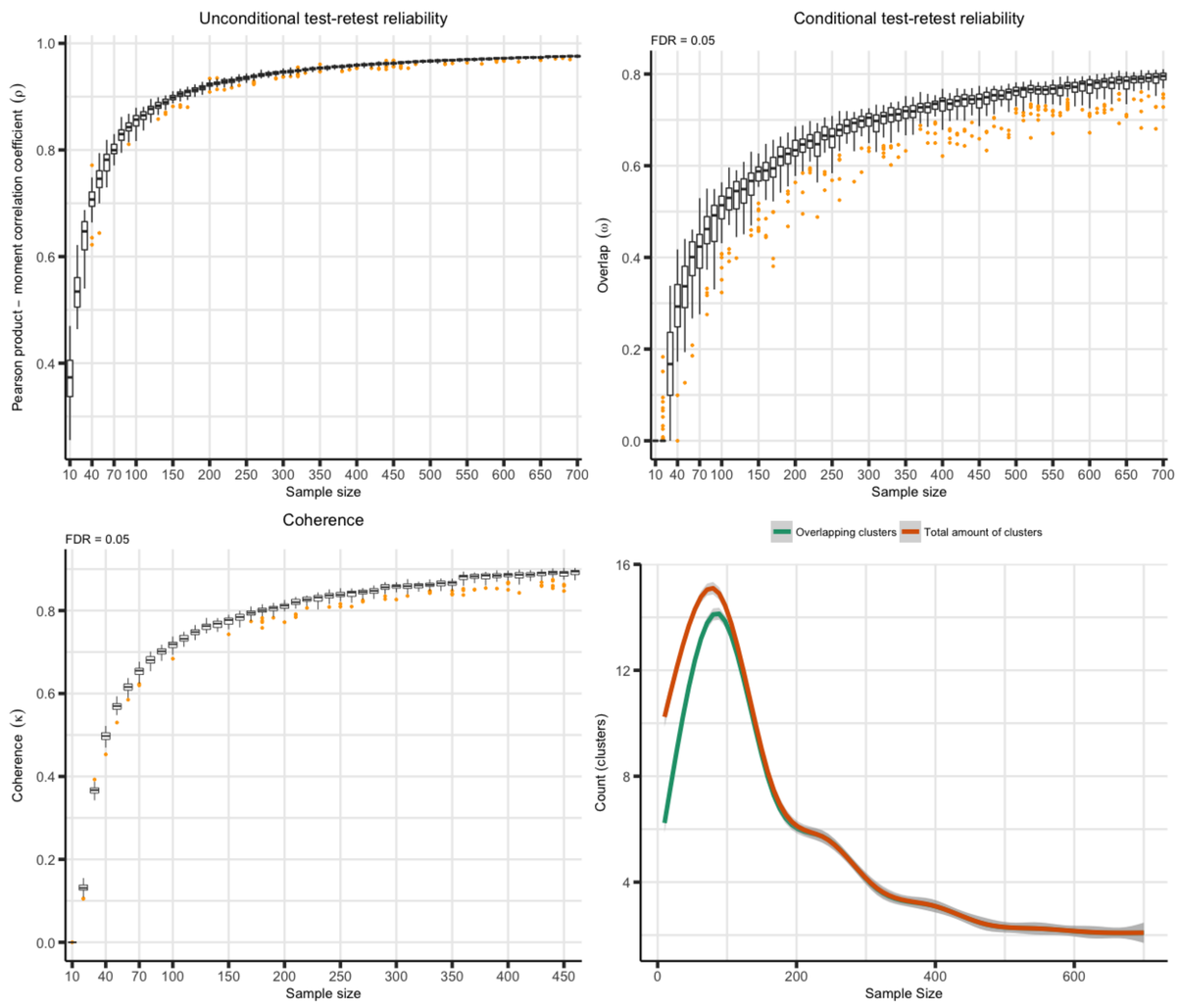
OBJECTIVE

- Assess the empirical replicability of fMRI results in function of sample size (N).
- We will use a similar approach as in Thirion et al. (2007, max(N) = 40) and Turner et al. (2017, max(N) = 120) using a database with max(N) = 700.

[3] METHOD

- IMAGEN database (Schumann et al. 2010, N = 1400).
- Cognitive task: MATH $>$ LANGUAGE.
- Resampling procedure:
  - START
  - Sample $\mathbf{N_A}$ subjects and fit GLM for task in every voxel.
  - Sample $\mathbf{N_B}$ subjects and fit GLM for task in every voxel.
  - Compare results analysis A and B.
  - REPEAT for 50 times.
  - Increase $\mathbf{N_A}$ and $\mathbf{N_B}$ with 10: go to START.
  - Maximum $\mathbf{N_A}$ and $\mathbf{N_B}$ = 700.

[4] RESULTS



[5] DISCUSSION

- Unconditional test-retest reliability $\Rightarrow$ good (! decoding).
- Voxelwise overlap between images $\Rightarrow$ poor performance at low N (need at least 200 in this dataset).
- Voxelwise coherence $\Rightarrow$ better performance.
- Overlap at cluster level $\Rightarrow$ reasonable replicability.

Acknowledgements

Illustrations in the introduction were provided by Joke Durnez. Data is provided by the IMAGEN consortium. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government - department EWI.

References

- Schumann, G. et al. (2010): The imagen study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry, 15(12):1128-1139.
- Thirion, B. et al. (2007): Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. Neuroimage, 35(1), 105-120.
- Turner, B. O. et al. (2017): How Sample Size Influences The Replicability Of Task-Based fMRI. BioRxiv, 1-22.
- Poldrack, R. A. et al. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. Nature Reviews Neuroscience, 18(2), 115-126.