Reliability of Effect Sizes and Spatial Localization with Population-Level Sample Sizes

Patrick Sadil, Martin Lindquist

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Background

- Consortium studies allow assessment of analysis methods
- Assessments have previously revealed that brain-wide association studies may require thousands of participants[1]
- Goal: survey several effect types in MRI to help interpret small studies and calibrate expectations for large studies

Discussion

- Studies with thousands of participants provide reliable estimates of several kinds of effects
- For effects like correlation between brain volume and fluid intelligence, variability in published values can be accounted for by sampling
- Peak location and voxel-wise effect size maps[4] can be estimated precisely with hundreds of participants

Methods

- Split UK Biobank[2]: 8k gold standard, 10k study set
- Sample from study set, varying sample size
- Estimate statistics (brain-behavior correlation, cluster peak location, voxel-wise effect size)
- Compare study set distributions to gold standard

Variability in effect size estimates accounted for by sampling

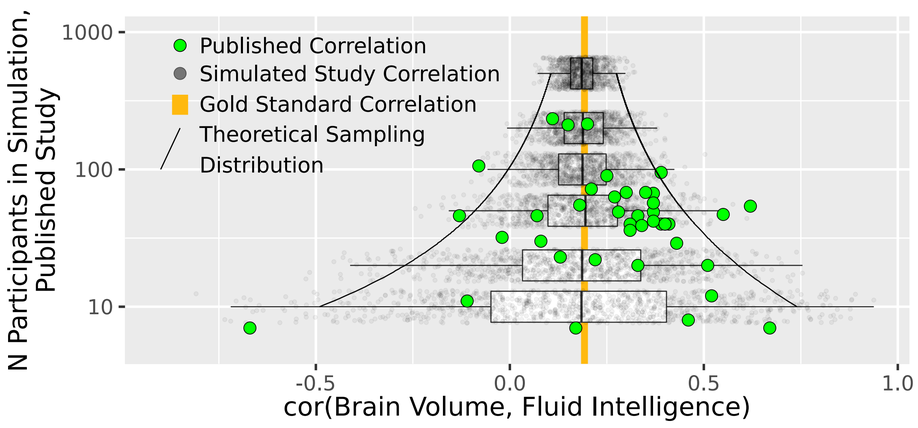


Figure shows reliability of estimated (rank) correlations between whole brain volume and fluid intelligence score. The gold standard correlation is 0.19. For this value, the 2.5% and 97.5% quantiles of a sampling distribution are marked with solid lines. Black dots indicate estimates of this correlation from the simulated studies and are summarized with box plots. Green dots are from published studies and meta analyses. Although a range of correlations have been reported, the variability matches what would be expected from the sampling distribution.

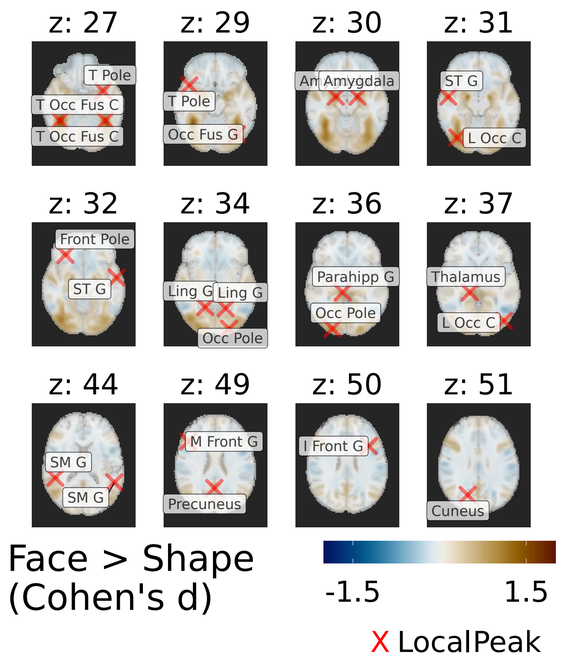
Gold Standard Peaks

Voxel-wise effect sizes were measured with Cohen's d. In the gold standard, they were calculated from the average, μ, and standard deviation, σ, of the faces > shape contrast[3]. In simulated studies, values were estimated with Hedge's correction, $C(N)$.

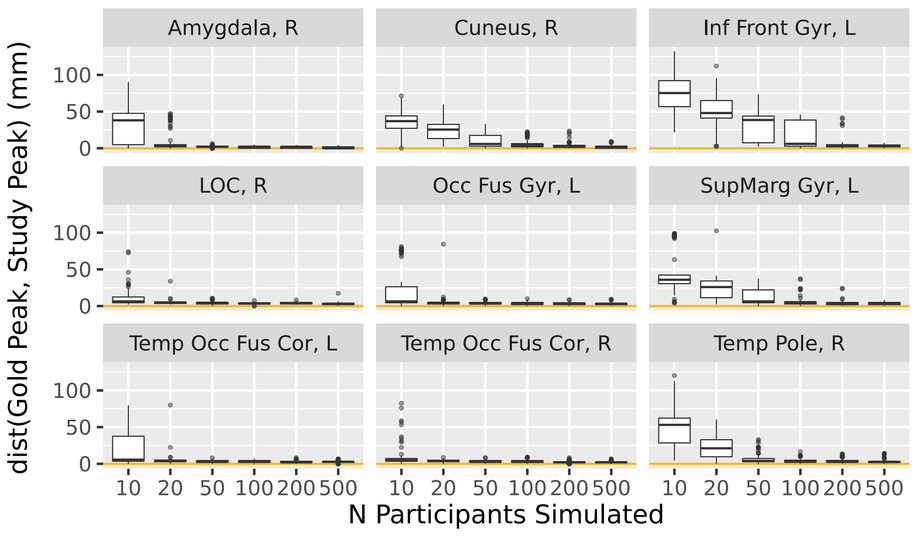
$$d = \frac{\mu}{\sigma}$$

$$g = \frac{\hat{\mu}}{\hat{\sigma}} C(N)$$

For display, the figure shows only the 24 most active local peaks, sorted by (signed) effect size.

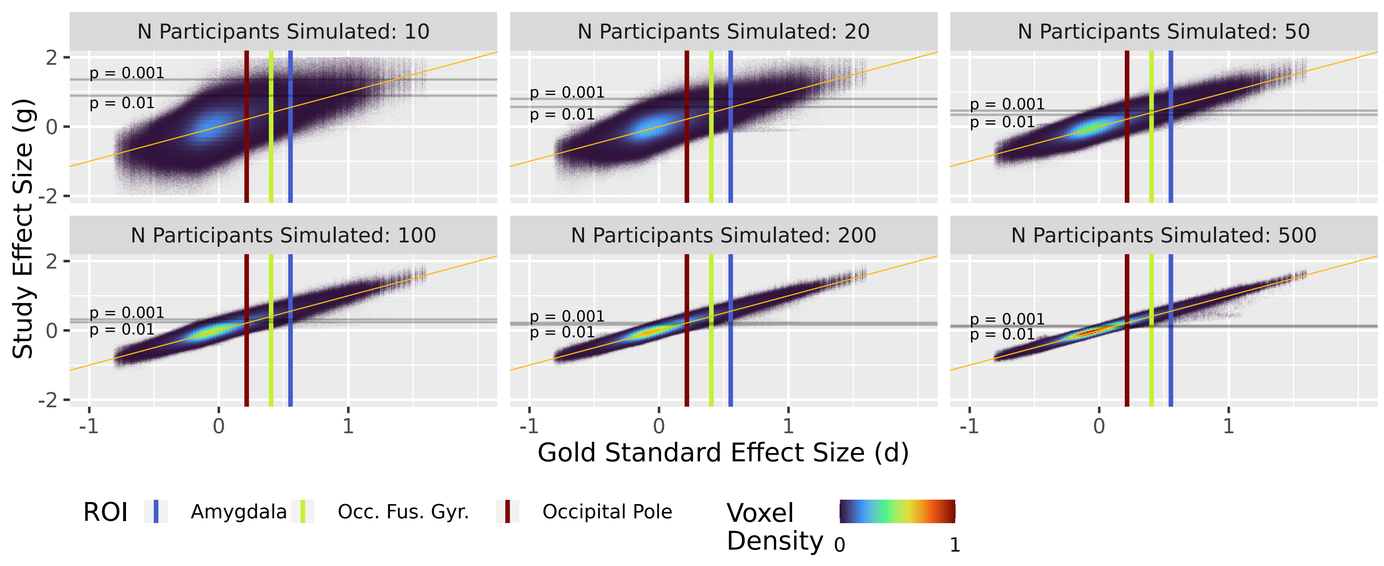


Peak Localization in Simulations



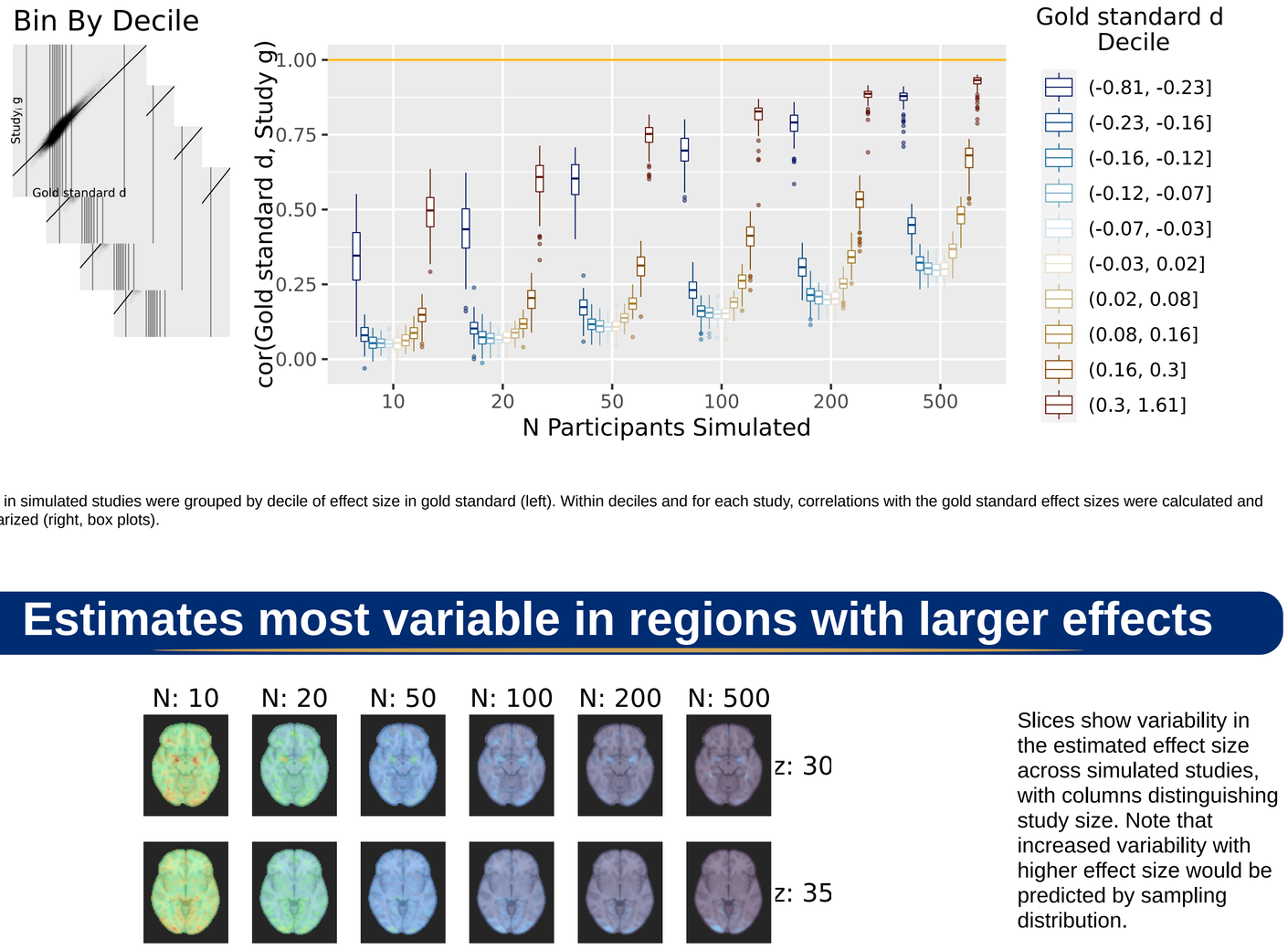
Study peaks were identified in unsmoothed, threshold-free cluster enhanced t-maps (FSL) after excluding voxels with $p > 0.05$ (permutation tests on the enhanced images, controlling for FWE). Panels corresponds to individual peaks in the gold standard (limited to 9 for display). Box plots summarize the distances from that peak to nearest peak in each simulated study. Voxels were 2 mm^3.

Reliability of Voxel-Wise Effect Sizes



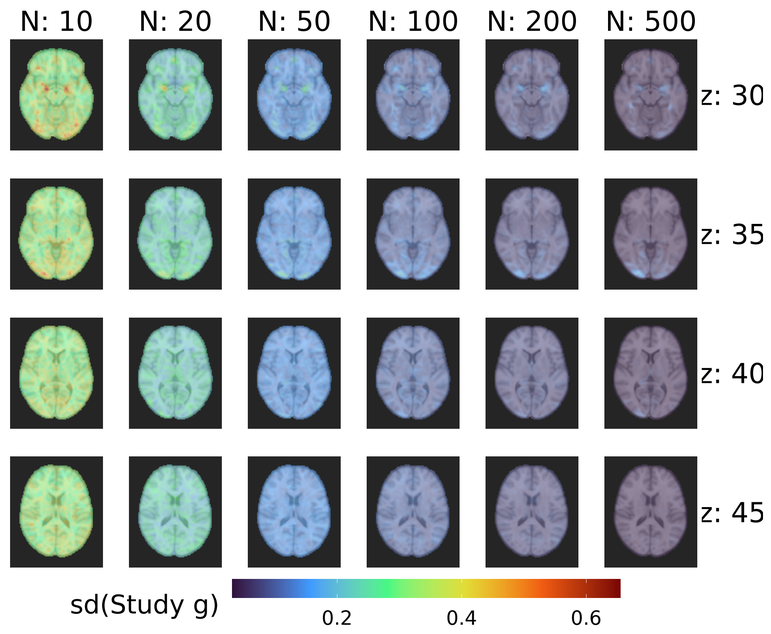
Points correspond to individual voxels in simulated studies (effect size in the gold standard by estimated effect in simulation). Color indicates degree of overplotting, normalized to one across panels. For reference, panels include a diagonal line indicating perfect estimation (gold), statistical significance thresholds (gray horizontal lines), and the average effect size of voxels within three (bilateral) regions of interest (colored vertical lines). The vertical axis is clipped at $\pm$ 2.

Correlations vary across deciles of effect size



Voxels in simulated studies were grouped by decile of effect size in gold standard (left). Within deciles and for each study, correlations with the gold standard effect sizes were calculated and summarized (right, box plots).

Estimates most variable in regions with larger effects



Slices show variability in the estimated effect size across simulated studies, with columns distinguishing study size. Note that increased variability with higher effect size would be predicted by sampling distribution.

References

1. Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
2. Alfaro-Almagro, F. *et al.* Image processing and quality control for the first 10,000 brain imaging datasets from UK biobank. *Neuroimage* **166**, 400–424 (2018).
3. Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F. & Weinberger, D. R. The amygdala response to emotional stimuli: A comparison of faces and scenes. *Neuroimage* **17**, 317–323 (2002).
4. Botvinik-Nezer, R. *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).