Cohort composition and sample size impact the replicability of multivariate brain-behavioural associations

Michelle Wang[1], Brent McPherson[1], Jean-Baptiste Poline[1]

[1] McConnell Brain Imaging Centre, The Neuro (Montreal Neurological Institute-Hospital), Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada

Motivation

Recent work in neuroimaging suggests that **thousands of samples** are needed to get replicable **brain-behaviour associations** with multivariate methods like **Canonical Correlation Analysis** (CCA) (Marek *et al*. 2022).

These results were for a general population sample. **No one has investigated the replicability of brain-behaviour CCA in more specific disease cohorts**.

The use of **cross-validation (CV)** could also make results more replicable (Spisak *et al*. 2023), but there has been limited work on cross-validated CCA so far.

Methods

Data

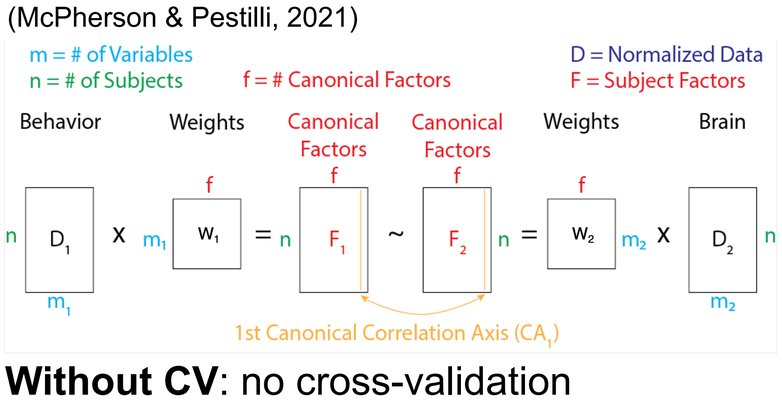
40514 participants from the UK Biobank

Behavioural: cognitive assessment scores

Brain: diffusion magnetic resonance imaging (dMRI) measures

Canonical Correlation Analysis

(McPherson & Pestilli, 2021)



Without CV: no cross-validation

With CV: ensemble model (50 instances)

Data sampling

30 sample sizes (50–20257, log-spaced)

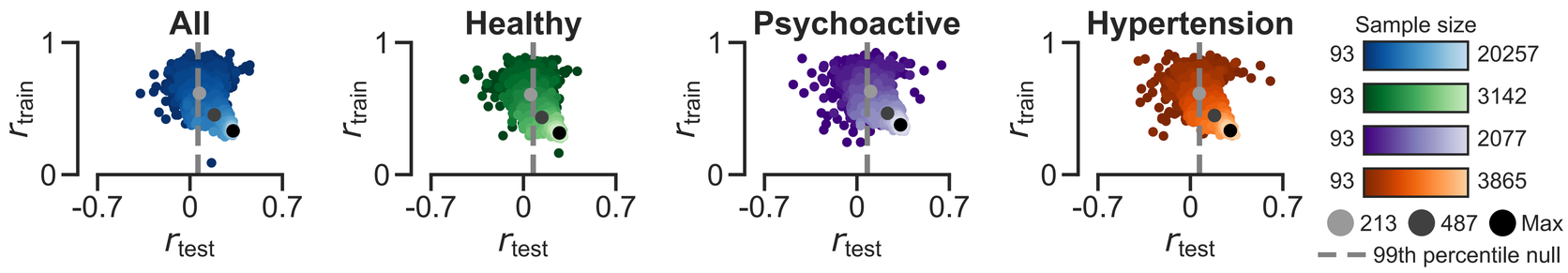
All participants + 3 targeted subgroups

- **Healthy** (*N*=6676)
- **Psychoactive** substance use (*N*=4725)
- **Hypertension** (*N*=7768)

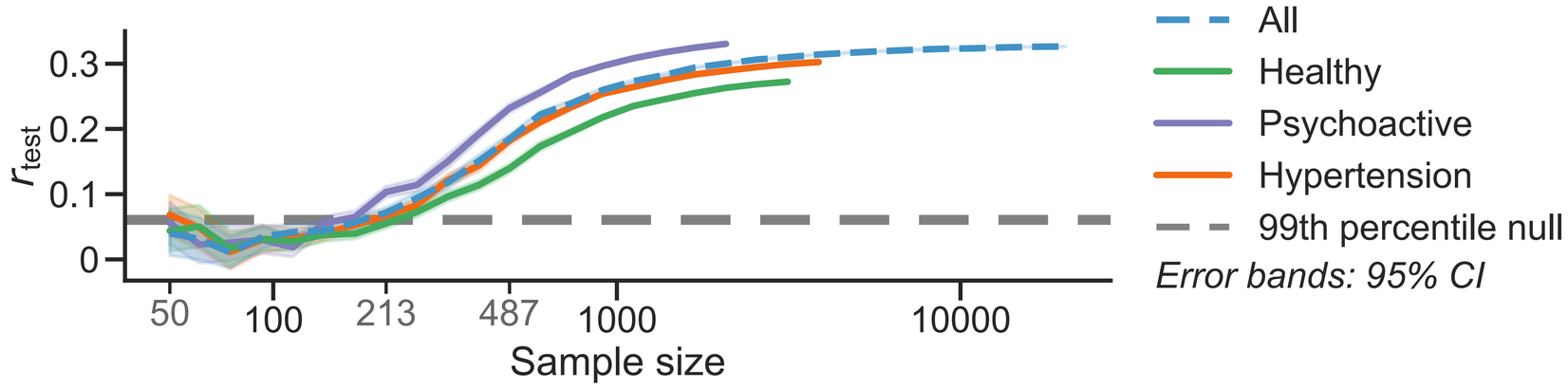
200 random samples per condition

Test set: largest sample size for condition

Replicable brain-behavioural CCA on the UK Biobank requires hundreds – not thousands – of samples

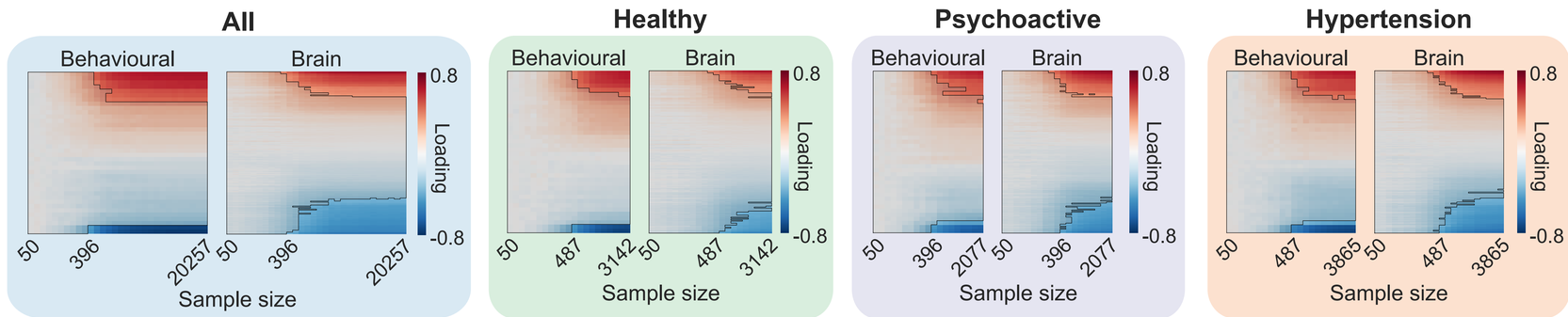


Correlation strengths depend on cohort composition

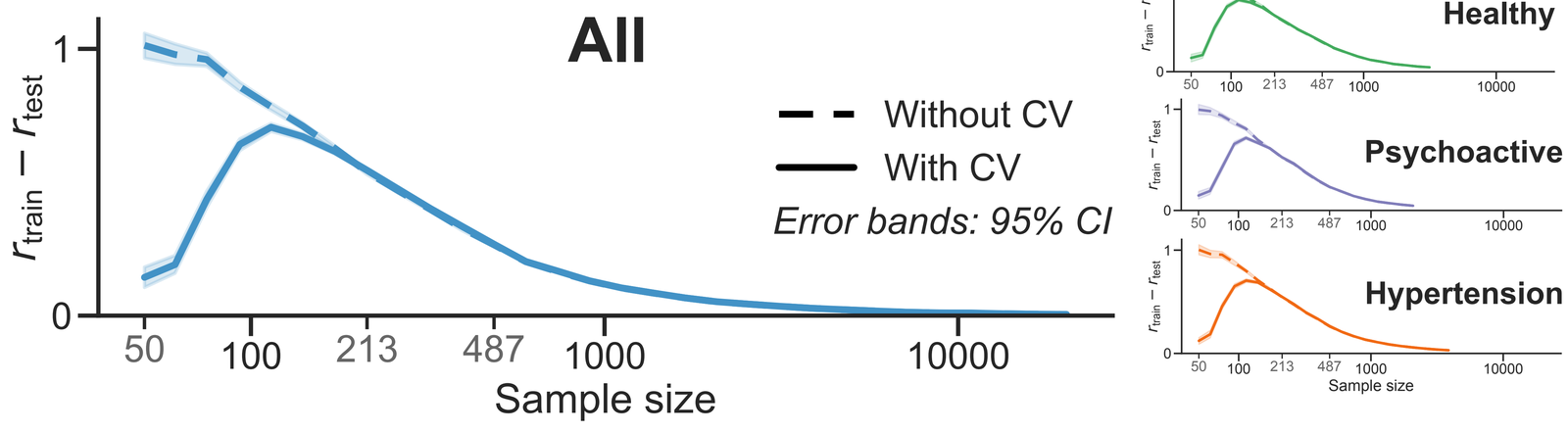


Variable loading order is preserved across sample sizes

Loadings obtained by **correlating** canonical factors with original datasets



Cross-validation reduces effect size inflation, but only at low sample sizes



References

Marek, S. *et al*. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654–660.

McPherson, B. C., & Pestilli, F. (2021). A single mode of population covariation associates brain networks structure and behavior and predicts individual subjects' age. *Communications Biology*, *4*(1), 1–16.

Spisak, T., Bingel, U., & Wager, T. D. (2023). Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, *615*(7951), Article 7951.

Acknowledgements

Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture
Montreal Neurological Institute-Hospital
McConnell Brain Imaging Centre
National Institutes of Health
THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH
QUEBEC PARKINSON NETWORK
HBHL NEUROHUB
Chan Zuckerberg Initiative
Fondation Brain Canada Foundation