

# News & views

## Human neuroscience

# Design tips for studies of brain–behaviour links

Roselyne J. Chauvin & Nico U. F. Dosenbach

Brain-wide association studies, which aim to link features of the brain to a person’s characteristics, have a replicability problem. Besides increasing sample size, what else can be done to make these studies more reproducible?

Population brain-wide association studies (BWAS) are used to relate inter-individual differences in structural and functional neuroimaging measurements to cognitive and psychiatric traits. But these studies have been plagued by poor replicability, unless they are based on sample sizes in the thousands<sup>1</sup>. Writing in *Nature*, Kang *et al.*<sup>2</sup> describe further design principles, such as sampling schemes that boost inter-individual variance, to improve replicability.

In 2023 alone, the US National Institutes of Health (NIH) invested around US\$1.9 billion in neuroimaging research (see [go.nature.com/48yzt6u](https://go.nature.com/48yzt6u)). There are two ongoing mega studies: ABCD (Adolescent Brain Cognitive Development)<sup>3</sup>, which is collecting information such as brain scans from 12,000 US children longitudinally (over time); and the UK Biobank<sup>4</sup>, which is amassing neuroimaging, genetic and medical data on 100,000 UK citizens.

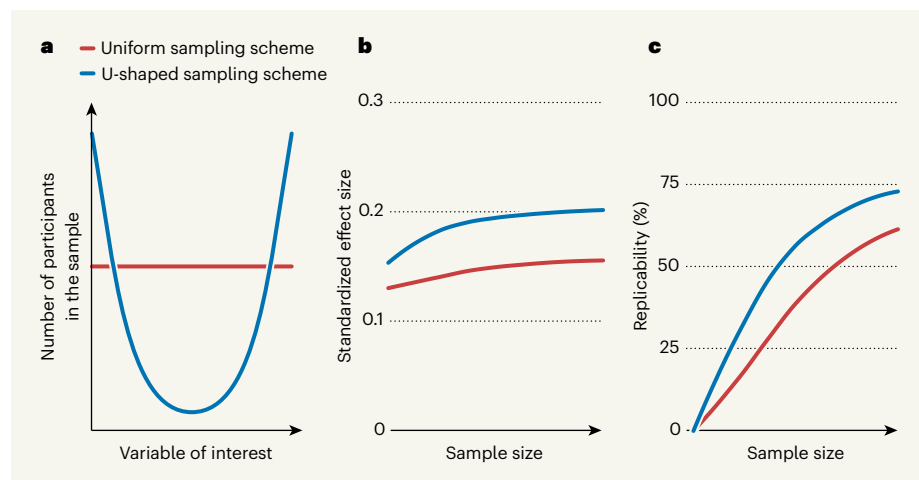
Unlike these large projects, most earlier BWAS have had fewer than 100 participants, with a median sample size of fewer than 25 (ref. 1). A minor panic ensued<sup>5,6</sup>, therefore, when several reports from mega-study data revealed that actual BWAS effect sizes – the strength of the relationship between a structural or functional measurement of the brain and a given trait – were smaller than previously thought, requiring sample sizes to be in the thousands to become replicable<sup>1,7–10</sup>. As happened in the early days of genome-wide association studies (GWAS), which are used to identify links between genetic variants and observable traits, the combination of small samples, sampling variability and publication bias had led to the reporting of inflated effect sizes.

There is consensus that, all else being equal,

larger samples are better for BWAS than are smaller ones. However, high neuroimaging costs make it impossible to generate a sample size of more than 1,000 with the funding provided by a standard investigator-initiated research grant. Non-BWAS neuroimaging studies, such as classical functional magnetic resonance imaging (fMRI), which tracks brain activations tied to the demands of a specific task, can be reliable when averaged across tens of people, or even in an individual, given sufficient repetition. Precision-mapping studies in which the same individual is sampled repeatedly or longitudinally (or

both), while eschewing group-averaging and cross-sectional analyses (in which data are examined from a representative population at a specific point in time), have revealed principles of brain organization and plasticity with only a single participant<sup>11–13</sup>. Does this mean that BWAS must remain restricted to consortium mega-studies, or meta-analytic samples such as the Lifespan Brain Chart Consortium (LBCC, consisting of more than 100 studies)<sup>14</sup>, whereas individual researchers must write grants for task-focussed fMRI and precision mapping?

To search for design features of BWAS that could improve replicability at a given sample size, Kang *et al.* combined data sets from projects such as the UK Biobank, ABCD and LBCC, amassing more than 100,000 brain scans. They report that sampling schemes that increase the variance between participants in a variable of interest (such as age) can improve BWAS replicability. The true relationship between variables (such as height versus age) is set by biology and could, hypothetically, be measured exactly by sampling all humans, but that is of course unrealistic. However, over-sampling the tails of a distribution of interest relative to its centre (in the case of ageing, for example, by enrolling only people aged 18–24 and 79–85), will on average return a larger standardized measurement of effect size for a given number of participants, compared with



**Figure 1 | Considerations for designing brain-wide association studies.** Population brain-wide association studies (BWAS) aim to identify associations between traits and structural or functional features of the brain. By performing meta-analyses of BWAS, Kang *et al.*<sup>2</sup> sought to find aspects of BWAS design, other than sample size, that maximize the standardized measure of effect size (the strength of the associations), and therefore the study’s replicability. **a–c**, For example, Kang *et al.* found that sampling schemes in which the participant population is ‘oversampled’ at the extreme ends of the distribution of the variable of interest (**a**) can increase between-participant variability, and therefore the effect size (**b**) and chance of replicability (**c**) at a given sample size, compared with a uniform sampling scheme. (Adapted from Extended Data Fig. 1 and Fig. 2 from ref. 2)

sampling equally across the distribution. Such an approach should directly translate into a greater chance of replication (Fig. 1).

Longitudinal sampling also tends to improve average replicability, for a given sample size, but with several qualifications. Most importantly, in longitudinal and repeated sampling designs, the within-participant effects and between-participant effects must be estimated separately. Failure to do so could decrease replicability if within-participant variance – for example in cognitive metrics that are susceptible to changes in neurological state (such as sleep) – is not linked to variance in neuroimaging metrics (such as structural measures that are not affected by sleep). If within-participant and between-participant variance are properly accounted for, spacing repeated measurements more widely in time further enhances the benefits of longitudinal sampling. Kang *et al.* show that most benefits of a longitudinal design, for the specific case of relating ageing to structural brain metrics, were achieved with just two time points.

The staggering number of thoughtful analyses that Kang *et al.* conducted on a trove of neuroimaging data brings into focus broader principles for better BWAS sampling strategies. First, researchers must maximize the variance that they are interested in. Second, they must minimize ‘nuisance’ variance, which is not biologically or clinically meaningful – head movement during scanning, for example. Third, they must know where variance comes from. Is it between-participant variance or within-participant variance? Within-participant variance can be due to day-to-day changes in neurological state (sleepy versus well-rested, for example) or consistent changes across longer timescales (with age, for example). Variance might also arise from errors in measurement. Relative to functional metrics, structural ones have lower measurement error and are less susceptible to state effects (such as sleep).

On the non-brain side of the BWAS equation, variables also differ in developmental or ageing effects, measurement error and susceptibility to state changes. Hence, the optimal BWAS design is not one-strategy-fits-all and still depends on the specifics. It is also important to be aware that tail oversampling and longitudinal designs do not protect against the risk of reporting inflated brain-behaviour associations. The dangers of using statistically underpowered samples for BWAS remain, but for a given sample size, such variance-optimizing strategies can further increase the likelihood of replication.

Following the studies that argued for larger samples, this impressive work by Kang and colleagues is a key methodological next step towards making BWAS more replicable. Other milestones on the path towards consensus best practices for BWAS are still missing. Kang *et al.* bring powerful evidence in favour of more-complex sampling schemes (tail oversampling and longitudinal sampling), but those might also raise the cost per participant. For example, oversampling the tails requires pre-existing knowledge about the underlying distribution, which in some cases requires measuring non-brain variables in a larger sample, from which to later sub-sample. Therefore, the fiscal trade-offs need to be estimated for different sampling schemes and other study parameters to define BWAS designs with the best chance of making replicable discoveries per unit of funding.

Another key consideration when using targeted BWAS sampling schemes is accidentally inflating or inducing confounding correlations among variables that are not under investigation. For example, when studying the associations between body weight and the brain, oversampling the extreme tails could introduce eating disorders as a major confounder. Undersampling the centre of a distribution could backfire if some of the brain associations are non-linear – if the relationship is U-shaped, for example. Therefore, researchers need to identify all variables with strong brain associations that could act as confounders if they are not tracked as part of the sampling scheme.

Neuroimaging consortium mega-studies and mega meta-analysis samples have already proved extremely valuable, and if GWAS is any indication, they will continue to grow. Mega-studies that attempt to achieve a representative sampling of the underlying population for many variables simultaneously have many advantages. They can be used by myriad researchers to investigate a great variety of questions that can be answered using BWAS, including ones that had not been thought of when the study was conceived. Hence, there is a trade-off between engineering BWAS samples for specific questions and general-use cases that are more amenable to exploratory science.

Longitudinal designs are recommended for BWAS of development and ageing, disease progression and studying episodic disorders (such as depression). Tail oversampling seems best suited for BWAS of clinical cohorts, especially when key non-brain metrics are already available. Although some might wonder whether more complex sampling schemes are akin to

*P*-hacking (manipulations to falsely achieve statistical significance) or other questionable research practices that inflate reported effect sizes, the findings described by Kang and colleagues, if properly applied, can do the opposite. They enable researchers to understand the contributions of different aspects of the experimental design to the true signal.

Kang *et al.* are opening the door for specialty BWAS samples (longitudinal and clinical cohorts), but the extra costs of working with patients and repeated scans will probably push budgets beyond the standard funding for investigator-initiated grants. Therefore, teamwork seems to be a requirement for generating adequate BWAS samples. For researchers not yet on a BWAS team, there is still precision mapping in small samples, which has also proved extremely fruitful<sup>11,15–18</sup>.

As for small-sample BWAS, researchers should not fall prey to the human cognitive bias of viewing small samples as more representative than they are. Instead, they must heed psychologists Amos Tversky and Daniel Kahneman, who admonished scientists<sup>19</sup> back in 1971 to accept that sometimes “there is simply no point in running the study unless ... sample size is multiplied”.

**Roselyne J. Chauvin and Nico U. F.**

**Dosenbach** are in the Department of Neurology, Washington University School of Medicine, St. Louis, Missouri 63110, USA.

**N. U. F. D.** is also in the Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, Missouri, USA.

e-mails: chauvin@wustl.edu;

ndosenbach@wustl.edu

1. Marek, S. *et al.* *Nature* **603**, 654–660 (2022).
2. Kang, K. *et al.* *Nature* <https://doi.org/10.1038/s41586-024-08260-9> (2024).
3. Casey, B. J. *et al.* *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
4. Bycroft, C. *et al.* *Nature* **562**, 203–209 (2018).
5. Spisak, T., Bingel, U. & Wager, T. D. *Nature* **615**, E4–E7 (2023).
6. *Nature Neurosci.* **25**, 833–834 (2022).
7. Rosenblatt, M. *et al.* *Nature Hum. Behav.* **8**, 2018–2033 (2024).
8. Tervo-Clemmens, B. *et al.* *Nature* **615**, E8–E12 (2023).
9. Chen, J. *et al.* *Nature Commun.* **13**, 2217 (2022).
10. Liu, S., Abdellaoui, A., Verweij, K. J. H. & van Wingen, G. A. *Nature Hum. Behav.* **7**, 1344–1356 (2023).
11. Pritschet, L. *et al.* *Nature Neurosci.* **27**, 2253–2260 (2024).
12. Laumann, T. O. *et al.* *Lancet Neurol.* **20**, 255–256 (2021).
13. Laumann, T. O. *et al.* *Neuron* **87**, 657–670 (2015).
14. Bethlehem, R. A. I. *et al.* *Nature* **604**, 525–533 (2022).
15. Siegel, J. S. *et al.* *Nature* **632**, 131–138 (2024).
16. Lynch, C. J. *et al.* *Nature* **633**, 624–633 (2024).
17. Gordon, E. M. *et al.* *Nature* **617**, 351–359 (2023).
18. Gordon, E. M. *et al.* *Neuron* **95**, 791–807 (2017).
19. Tversky, A. & Kahneman, D. *Psychol. Bull.* **76**, 105–110 (1971).

The authors declare no competing interests.