

# Revisiting doubt in neuroimaging research

A 2022 paper reported a caveat about brain–behavior relationships emerging from neuroimaging data, which then unintentionally cast an entire field and a neuroscientific method into doubt... again.

In 2017, *Nature Neuroscience* published an editorial<sup>1</sup> addressing the concerns of functional magnetic resonance imaging (fMRI) research after a paper suggested that the statistical significance of some fMRI results could be inflated<sup>2</sup>. The editorial reminded readers that neuroimaging findings in many subfields of neuroscience had been replicated across species, varying experimental paradigms, and methodological and analytical approaches. It discussed ways to promote reproducibility in MRI research: for example, transparent reporting of study design, data collection and data analyses, and improving data accessibility<sup>1</sup>. The authors of the provocative paper<sup>2</sup> similarly championed data and code accessibility as the way forward for fMRI researchers<sup>3</sup>.

Five years later, the neuroimaging community is back in the same unenviable position following the publication of a paper by Marek and colleagues<sup>4</sup>, which presented data suggesting that relationships between neuroimaging and behavioral phenotypes may not be robust or replicable unless the sample size is in the thousands. Numerous news outlets and online blogs have covered the paper, with headlines questioning the validity of neuroimaging findings based on much smaller groups of participants.

The use of openly shared neuroimaging datasets from thousands of participants — such as the Human Connectome Project (HCP)<sup>5</sup>, the Adolescent Brain Cognitive Development (ABCD) study<sup>6</sup> and the UK Biobank<sup>7</sup>, all used by Marek et al. — can certainly alleviate the concern about underpowered brain–behavior associations. The authors noted that this strategy resembles the path taken in genomics research, where issues regarding non-robust results from underpowered candidate-gene studies were resolved only when the field started to collect large datasets with tens of thousands of samples<sup>4</sup>.

However, comment pieces in defense of neuroimaging research are now emerging from the research community<sup>8,9</sup>, reminding readers that there are other ways to obtain robust and reproducible findings without having to recruit hundreds or thousands of participants. For example, investigators could gather ‘precision data’ by repeatedly measuring variables of interest from a much smaller sample of individuals<sup>8,9</sup>.

In this issue of *Nature Neuroscience*, Monica Rosenberg and Emily Finn put forth several additional suggestions, which include building, testing, and sharing models that predict (and not just correlate with) behavioral outcomes from patterns of brain features, as well as optimizing study designs to make these brain–behavior models more robust<sup>9</sup>. The recent comment pieces also touch upon the potential sociological ramifications of relying solely on sample sizes in the thousands to infer brain–behavior associations. For example, funding inequities may favor consortia-led studies over early-career researchers who are establishing independent labs<sup>8</sup>, and this could suppress scientific innovation and creativity<sup>9</sup>.

It is worth remembering that because of many factors — such as limits on funding, space allocation, availability of reagents, electrode placement, paradigm training, and ethical concerns — studies in humans and animals that use more direct, invasive techniques typically have sample sizes of ten or fewer. Notably, at the end of their paper, Marek and colleagues also supported study designs with smaller sample sizes<sup>4</sup>, but those details were lost in the public coverage of this work. So, what is it about neuroimaging research that causes alarm bells to ring when method-based caveats are published? Is it because brain activity is being examined in a non-invasive and indirect manner, which requires extensive processing and statistical analyses of myriad data points before a result can be interpreted? Or is it that this research is directly related to the human condition because most neuroimaging studies are conducted in humans rather than in animals? The extensive public coverage of a research paper that showed no empirical support for candidate genes for major depression<sup>10</sup> suggests that, regardless of the scientific field, cautionary papers about scientific findings from vast amounts of human data could be cause for concern. A lot of effort and resources have been devoted to collecting genetic and brain data from humans, all in the hope that we might better understand the biological bases of human health and disease. And research papers that demonstrate a need to refine our scientific processes in these efforts can be sobering to researchers and perhaps rather concerning if the news reaches the general public.

It is exactly at these times that the larger research community and the general public need to remember that science is an iterative process. The assumed answers to some of the most important research questions can and should be revisited in studies using newer techniques, and those answers may be confirmed, refined with further nuance, or even refuted. We all need to recognize the original intent of these cautionary papers: to highlight issues that researchers should take into account in order to interpret their data more carefully. Ironically, these papers are not always covered with the same level of nuance when discussed on social media, on online platforms, or in the news, which can instill doubt in these techniques and scientific fields.

So, how do we avoid throwing a scientific field and its techniques into doubt every time another caveat is reported? Perhaps we can take a step back from all of this to remind ourselves that the scientific process is conducted by humans; the answers we seek from this process and even the communication of study results will not always be correct or exact. Readers should remain open-minded, knowing that important nuances might be overlooked in social media, online posts, and news stories that discuss these caveats more broadly, and should avoid catastrophizing these issues. To assist this, researchers should always take great care to explain their findings to non-expert audiences, including journalists, without overstating claims. As researchers write cautionary papers, it is always important to clearly state any helpful nuances up front, as Marek et al. have done by emphasizing the continuing need for neuroimaging studies with smaller sample sizes at the end of their paper<sup>4</sup>. Journal editors and peer reviewers should look for these points beyond the main take-home messages of cautionary papers, prior to incorporating this information into their assessments of future neuroimaging work. Collectively, perhaps all of these things can help us to avoid another declaration of a reproducibility crisis in neuroimaging research. □

Published online: 5 July 2022  
<https://doi.org/10.1038/s41593-022-01125-2>

## References

1. Fostering reproducible fMRI research. *Nat. Neurosci.* **20**, 298–298 (2017).
2. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl Acad. Sci. USA* **113**, 7900–7905 (2016).
3. Eklund, A., Nichols, T. E. & Knutsson, H. Reply to Brown and Behrmann, Cox, et al., and Kessler et al.: Data and code sharing is the way forward for fMRI. *Proc. Natl Acad. Sci. USA* **114**, E3374–E3375 (2017).
4. Marek, S. et al. Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
5. Van Essen, D. C. et al. The WU-Minn Human Connectome Project: an overview. *NeuroImage* **80**, 62–79 (2013).
6. Casey, B. J. et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
7. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
8. Gratton, C., Nelson, S. M. & Gordon, E. M. Brain–behavior correlations: two paths toward reliability. *Neuron* **110**, 1446–1449 (2022).
9. Rosenberg, M. D. & Finn, E. S. How to establish robust brain–behavior relationships without thousands of individuals. *Nat. Neurosci.* <https://doi.org/10.1038/s41593-022-01110-9> (2022).
10. Border, R. et al. No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *Am. J. Psychiatry* **176**, 376–387 (2019).