

Editorials

Balancing sample size and scan duration in functional MRI studies: A reflection on Ooi et al.

Janine D. Bijsterbosch, Ph.D.¹  ^a

¹ Department of Radiology, Washington University in St Louis School of Medicine, St. Louis, MO, USA

Keywords: functional MRI, scan duration, prediction, commentary

<https://doi.org/10.52294/001c.157712>

Aperture Neuro

Vol. 6, 2026

In the era of ‘big data’, there is a push for neuroimaging datasets with more subjects and/or longer scans. At the same time, funding pressures interact with the public perception of science to demand increased transparency and return on investment for research spending. In a recent paper, Ooi et al. developed a timely cost-benefit analysis to address this tension between cost-effectiveness and big data. The main takeaways from the Ooi et al. paper are that sample size and scan duration are interchangeable and for many studies the most cost-effective approach is to ensure that functional magnetic resonance imaging (fMRI) scans are at least 30 minutes long. This commentary discusses when and why increased scan duration is helpful, and how the balance between sample size and scan duration interacts with broader goals of external validation, representation, and individual differences in brain-behavior modeling.

In Ooi et al., the authors present a compelling and comprehensive cost effectiveness analysis to optimize the balance between scan duration and sample size for brain-behavior prediction studies.¹ This paper follows prior work from Marek et al., which suggested that a sample size of at least 2,000 subjects is required for robust brain-wide associations, given realistic effect sizes of $r=0.1$.² Ooi et al. find that sample size and scan duration are broadly interchangeable and therefore sample size requirements can be partly mitigated with longer functional magnetic resonance imaging (fMRI) scans. The goal of this perspective is to summarize the core takeaways from Ooi et al. and discuss some nuances and extensions of this important study.

The core recommendation for most neuroimaging studies resulting from the analyses in Ooi et al. is a duration of at least 30 minutes per subject for fMRI scans.¹ This recommendation is longer than several prior studies that focused on reliability³⁻⁵ and in-line with more recent estimates.⁶ Importantly, Ooi et al. move beyond reliability by quantifying the percentage of multivariate prediction accuracy achievable as a function of scan duration out of the ‘maximum’ prediction accuracy based on the full available scan time (focusing on phenotypes with maximum accuracy $r>0.1$). Here, [Table 1](#) provides a summary of the general recommendation for balancing per-subject scan duration and sample size and scenarios that warrant prioritizing longer per-subject scans (left) or larger sample sizes (right).

WHY ARE LONGER SCANS HELPFUL?

A key advantage of longer fMRI scans is the resulting increase in temporal degrees of freedom, which is known to improve the reliability of resting state measures of interest including partial correlation⁹ and subject-specific resting state network maps derived from independent component analysis (ICA) and dual regression.¹⁰ More generally, any analysis pipeline that includes a within-subject multiple linear regression with a relatively large number of regressors (such as the number of brain regions in partial correlation analysis or the number of ICA maps in dual regression) benefits from increased temporal degrees of freedom as achieved by longer scans and/or faster repetition times. As such, longer scan times improve the statistical power and thereby enhance the accuracy of within-subject estimates of functional connectivity.¹⁰

DO LONGER SCANS IMPACT EXTERNAL VALIDATION?

Ooi et al. focus on the size of the training sample in their cost-benefit analysis of sample size versus scan duration. In addition to well-powered training data and traditional cross-validation techniques,¹¹ there is increasing awareness of the need to replicate prediction results in out-of-

^a Corresponding Author:
Janine Bijsterbosch, PhD
janine.bijsterbosch@wustl.edu

Table 1. Overview of recommendations.

Typical study recommendation: 30-minute fMRI scans (ideally with N=900+)	
When to prioritize longer scans:	When to prioritize larger sample size:
Recruiting from rare (sub-) population	Prospective studies (e.g., ABCD, UK Biobank)
Mitigating low quality data points (e.g., in children) ⁷	High patient heterogeneity
Non-stationarity or dynamics studies	Interest in subtype/clustering analyses
Precision studies	External validation usage ⁸
Interest in subcortical brain regions	Normative studies

sample and out-of-distribution data.¹² Importantly, recent work emphasized the importance of the sample size of datasets used for such external validation.⁸ Rosenblatt et al. replicated the finding from Ooi et al. that longer scans improved prediction accuracy for training and internal validation data, but longer scan duration did not improve performance for external validation data. As such, large cohorts play an important role in the external validation of predictive models.

CAN LONGER SCANS HELP DIVERSITY AND REPRESENTATION?

There is increasing awareness of the need for representation of diverse and intersectional sociodemographic identities in neuroimaging studies.^{13,14} An appealing suggestion from Ooi et al. is that underrepresented subpopulations (e.g., racial minorities) could be scanned for longer than overrepresented groups within the same cohort to achieve diverse representation.¹ However, preliminary work has reported higher inter-individual variability in neuroimaging and behavior among Black participants compared to White participants,¹⁵ which may not be captured in a longer-scan study design without adding more participants. Therefore, future work is needed to understand the trade-offs between mixed scan-durations within-study, recruitment strategies and costs, sampling schemes, and other approaches to improve the diverse representation in neuroimaging datasets.

HOW DO LONGER SCANS AND/OR BIGGER SAMPLES INFORM INDIVIDUAL DIFFERENCE RESEARCH?

The study by Ooi et al. focuses on cost effectiveness of fMRI studies. As noted by the authors, there are study design choices that intrinsically require a (much) larger sample size such as prospective epidemiological studies where only a small proportion of participants is expected to endorse any particular symptom or behavior of interest.¹⁶ On the other hand, some study design choices require a (much) longer scan duration such as precision studies.¹⁷ Intriguingly, these two extremes of study design choices share a common goal of mapping individual differences in symptoms or behaviors onto individualized neural patterns.¹⁸

A key challenge for all individual difference research is that the degree of variance in brain-behavior associations

may differ between groups of individuals, leading to idiosyncrasies that are challenging to capture. For example, normative modeling approaches have shown a lack of overlap in normative deviations across patients,^{19–21} suggesting particularly high variance. Similarly, the so-called ‘Anna Karenina effect’ suggests the presence of substantially greater variance among patients than among healthy controls.^{22,23} Although scanning fewer people for longer may maximize cost effectiveness, scanning more participants enables a comprehensive mapping of the full breadth of idiosyncrasies across individuals. Alternatively, precision studies offer an opportunity for in-depth mapping of intricate idiosyncrasies in a limited set of individuals.

One popular approach for parsing high degrees of heterogeneity within a patient group is to assume the presence of biological subtypes (or ‘biotypes’).²⁴ A broad range of data-driven clustering approaches have been developed to identify biotypes from neuroimaging and/or clinical data.²⁵ Although all clustering methods will return a solution that minimizes the cost function, it is more challenging to determine whether resulting biotypes are clinically and/or etiologically meaningful.^{26,27} To assess biotype validity, post-hoc comparisons between biotypes are often performed to assess differences in treatment effects or other clinically or etiologically relevant variables that were not used for biotype determination. Notably, depending on the inclusion/exclusion criteria, cluster dimensionality, and validation measure missingness, these post-hoc analyses can suffer from relatively low power even when starting from a large sample size.²⁸ As such, cohorts with larger sample sizes are needed to robustly study heterogeneity and subtype validity.

WHERE DOES BEHAVIOR FIT IN?

Even with 30-minute scans, Ooi et al. show that N=900 is required to achieve 80% of the maximum prediction accuracy and N=2,500 is required to achieve 90% of the maximum prediction accuracy. As such, there is a need to explore additional avenues for optimizing brain-behavior associations. Ooi et al. show that phenotypic reliability affects the overall prediction accuracy (i.e., effect size) but was not related to the tradeoff between sample size and scan duration.¹ This is expected because the overall accuracy is capped by the joint reliability of brain and phenotypic measures,^{29,30} yet scan time is independent of the out-of-scanner phenotypic measures and therefore only

impacts the reliability of brain measures. As an extension of the work by Ooi et al., it is likely that brain-behavior association effect sizes can be improved by predicting the averaged phenotypic measure across several repeats instead of a one-off phenotypic assessment.³¹ Analogous to longer scans, repeated phenotypic measures are expected to enhance the accuracy of within-subject estimates of phenotypic traits. Combined with long scans, this has the potential to raise the joint reliability and thereby increase brain-behavior association effect sizes. Future work is needed to explore equivalent strategies to longer scans in the phenotypic space and its contributions to improving maximum prediction accuracy.

CONCLUSION

In conclusion, the paper by Ooi et al. provides an important framework and online calculator to maximize the cost efficiency of fMRI studies by optimizing the balance between scan duration and sample size. The recommendation of 30-minute long fMRI scans is suitable for many studies. At the same time, there are valid reasons for prioritizing

longer scans (e.g., precision studies) or for prioritizing more participants (e.g., prospective studies, external validation, and subtyping studies). In an era of declining public trust in higher education and increasing pressures on research funding, objective efforts to maximize returns on investment such as the study by Ooi et al. are critically important to guide study design and funding decisions, improve transparency, and inform advocacy.

.....

FUNDING SOURCES

Janine Bijsterbosch was supported by the NIH (NIMH R01 MH128286 & NIMH R01 MH132962).

CONFLICTS OF INTEREST

Janine Bijsterbosch reports no conflicts of interest.

Submitted: November 05, 2025 CDT. Accepted: November 18, 2025 CDT. Published: May 13, 2026 CDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Ooi LQR, Orban C, Zhang S, et al. Longer scans boost prediction and cut costs in brain-wide association studies. *Nature*. 2025;644(8077):731-740. doi:[10.1038/s41586-025-09250-1](https://doi.org/10.1038/s41586-025-09250-1)
2. Marek S, Tervo-Clemmens B, Calabro FJ, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022;603(7902):654-660. doi:[10.1038/s41586-022-04492-9](https://doi.org/10.1038/s41586-022-04492-9)
3. Birn RM, Molloy EK, Patriat R, et al. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage*. 2013;83:550-558. doi:[10.1016/j.neuroimage.2013.05.099](https://doi.org/10.1016/j.neuroimage.2013.05.099)
4. Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J Neurophysiol*. 2010;103(1):297-321. doi:[10.1152/jn.00783.2009](https://doi.org/10.1152/jn.00783.2009)
5. Tomasi DG, Shokri-Kojori E, Volkow ND. Temporal evolution of brain functional connectivity metrics: Could 7 min of rest be enough? *Cereb Cortex*. 2017;27(8):4153-4165. doi:[10.1093/cercor/bhw227](https://doi.org/10.1093/cercor/bhw227)
6. Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. Influences on the test-retest reliability of functional connectivity MRI and its relationship with behavioral utility. *Cereb Cortex*. 2017;27(11):5415-5429. doi:[10.1093/cercor/bhx230](https://doi.org/10.1093/cercor/bhx230)
7. Rai S, Godfrey KJ, Graff K, et al. How much is “enough”? Considerations for functional connectivity reliability in pediatric naturalistic fMRI. *Imaging Neurosci (Camb)*. 2025;3(IMAG.a.117). doi:[10.1162/IMAG.a.117](https://doi.org/10.1162/IMAG.a.117)
8. Rosenblatt M, Tejavibulya L, Sun H, et al. Power and reproducibility in the external validation of brain-phenotype predictions. *Nat Hum Behav*. 2024;8(10):2018-2033. doi:[10.1038/s41562-024-01931-7](https://doi.org/10.1038/s41562-024-01931-7)
9. Mejia AF, Nebel MB, Barber AD, et al. Improved estimation of subject-level functional connectivity using full and partial correlation with empirical Bayes shrinkage. *Neuroimage*. 2018;172:478-491. doi:[10.1016/j.neuroimage.2018.01.029](https://doi.org/10.1016/j.neuroimage.2018.01.029)
10. Feinberg DA, Moeller S, Smith SM, et al. Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One*. 2010;5(12):e15710. doi:[10.1371/journal.pone.0015710](https://doi.org/10.1371/journal.pone.0015710)
11. Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*. 2020;77(5):534-540. doi:[10.1001/jamapsychiatry.2019.3671](https://doi.org/10.1001/jamapsychiatry.2019.3671)
12. Yeung AWK, More S, Wu J, Eickhoff SB. Reporting details of neuroimaging studies on individual traits prediction: A literature survey. *Neuroimage*. 2022;256(119275):119275. doi:[10.1016/j.neuroimage.2022.119275](https://doi.org/10.1016/j.neuroimage.2022.119275)
13. Ricard JA, Parker TC, Dhamala E, Kwasia J, Allsop A, Holmes AJ. Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data. *Nat Neurosci*. 2023;26(1):4-11. doi:[10.1038/s41593-022-01218-y](https://doi.org/10.1038/s41593-022-01218-y)
14. Dhamala E, Ricard JA, Uddin LQ, et al. Considering the interconnected nature of social identities in neuroimaging research. *Nat Neurosci*. 2025;28(2):222-233. doi:[10.1038/s41593-024-01832-y](https://doi.org/10.1038/s41593-024-01832-y)
15. Fields C, Rosenblatt M, Aina J, et al. Racialized heteroscedasticity as a source of algorithmic bias in psychiatric prediction models. *Biol Psychiatry*. 2025;97(9):S45. doi:[10.1016/j.biopsych.2025.02.121](https://doi.org/10.1016/j.biopsych.2025.02.121)
16. Allen NE, Lacey B, Lawlor DA, et al. Prospective study design and data analysis in UK Biobank. *Sci Transl Med*. 2024;16(729):eadf4428. doi:[10.1126/scitranslmed.adf4428](https://doi.org/10.1126/scitranslmed.adf4428)
17. Gratton C, Kraus BT, Greene DJ, et al. Defining Individual-Specific Functional Neuroanatomy for Precision Psychiatry. *Biol Psychiatry*. 2020;88(1):28-39. doi:[10.1016/j.biopsych.2019.10.026](https://doi.org/10.1016/j.biopsych.2019.10.026)
18. Gratton C, Nelson SM, Gordon EM. Brain-behavior correlations: Two paths toward reliability. *Neuron*. 2022;110(9):1446-1449. doi:[10.1016/j.neuron.2022.04.018](https://doi.org/10.1016/j.neuron.2022.04.018)
19. Wolfers T, Beckmann CF, Hoogman M, Buitelaar JK, Franke B, Marquand AF. Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol Med*. 2020;50(2):314-323. doi:[10.1017/S0033291719000084](https://doi.org/10.1017/S0033291719000084)
20. Wolfers T, Doan NT, Kaufmann T, et al. Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry*. 2018;75(11):1146-1155. doi:[10.1001/jamapsychiatry.2018.2467](https://doi.org/10.1001/jamapsychiatry.2018.2467)

21. Verdi S, Kia SM, Yong KXX, et al. Revealing individual neuroanatomical heterogeneity in Alzheimer disease using neuroanatomical normative modeling. *Neurology*. 2023;100(24):e2442-e2453. doi:[10.1212/WNL.0000000000207298](https://doi.org/10.1212/WNL.0000000000207298)
22. Finn ES, Glerean E, Khojandi AY, et al. Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage*. 2020;215:116828. doi:[10.1016/j.neuroimage.2020.116828](https://doi.org/10.1016/j.neuroimage.2020.116828)
23. Murphy M, Öngür D. Schizophrenia, variability, and the Anna Karenina principle. *bioRxiv*. Published online October 1, 2025:2025.09.30.679570. doi:[10.1101/2025.09.30.679570](https://doi.org/10.1101/2025.09.30.679570)
24. Williams LM, Whitfield Gabrieli S. Neuroimaging for precision medicine in psychiatry. *Neuropsychopharmacology*. 2024;50(1):246-257. doi:[10.1038/s41386-024-01917-z](https://doi.org/10.1038/s41386-024-01917-z)
25. Chen P, Zhang S, Zhao K, Kang X, Rittman T, Liu Y. Robustly uncovering the heterogeneity of neurodegenerative disease by using data-driven subtyping in neuroimaging: A review. *Brain Res*. 2024;1823(148675):148675. doi:[10.1016/j.brainres.2023.148675](https://doi.org/10.1016/j.brainres.2023.148675)
26. Wen Z, Hammoud MZ, Siegel CE, et al. Neuroimaging-based variability in subtyping biomarkers for psychiatric heterogeneity. *Mol Psychiatry*. 2025;30(5):1966-1975. doi:[10.1038/s41380-024-02807-y](https://doi.org/10.1038/s41380-024-02807-y)
27. Hannon K, Jarukasemkit S, Balogh L, et al. Comparing data-driven subtypes of depression informed by clinical and neuroimaging data: A Registered Report. *Biol Psychiatry Glob Open Sci*. 2025;100473:100473. doi:[10.1016/j.bpsgos.2025.100473](https://doi.org/10.1016/j.bpsgos.2025.100473)
28. Hannon K, Easley T, Zhang W, et al. Parsing clinical and neurobiological sources of heterogeneity in depression. *Biol Psychiatry*. Published online May 8, 2025. doi:[10.1016/j.biopsych.2025.04.025](https://doi.org/10.1016/j.biopsych.2025.04.025)
29. Gell M, Eickhoff SB, Omidvarnia A, et al. How measurement noise limits the accuracy of brain-behaviour predictions. *Nat Commun*. 2024;15(1):10678. doi:[10.1038/s41467-024-54022-6](https://doi.org/10.1038/s41467-024-54022-6)
30. Nikolaidis A, Chen AA, He X, et al. Suboptimal phenotypic reliability impedes reproducible human neuroscience. *bioRxiv*. Published online July 23, 2022:2022.07.22.501193. doi:[10.1101/2022.07.22.501193](https://doi.org/10.1101/2022.07.22.501193)
31. Easley T, Chen R, Hannon K, Dutt R, Bijsterbosch J. Population modeling with machine learning can enhance measures of mental health - Open-data replication. *Neuroimage: Reports*. 2023;3(2):100163. doi:[10.1016/j.ynirp.2023.100163](https://doi.org/10.1016/j.ynirp.2023.100163)