Journal for Reproducibility in Neuroscience

Five best practices for fMRI research: towards a biologically grounded understanding of mental phenomena

Colleen Mills-Finnerty, PhD

Palo Alto Veterans Administration & Stanford Dept. of Psychiatry & Behavioral Science.
cmfinn@stanford.edu

The replication crisis in science has not spared functional magnetic resonance imaging (fMRI) research. A range of issues including insufficient control of false positives (1, 2), code bugs (3), concern regarding generalizability and replicability of findings (4-7), inadequate characterization of physiological confounds (8, 9), over-mining of repository datasets (10), and the small sample sizes/low power of many early studies have led to hearty debate in both the field and the press about the usefulness and viability of fMRI (11, 12). Others still see enormous potential for fMRI in diagnosing conditions that do not otherwise lend themselves to non-invasive biological measurement, from chronic pain to neurological and psychiatric illness (13). How do we reconcile the limitations of fMRI with the hype over its potential? Despite many papers hailed by the press as the nail in the coffin for fMRI, from the dead salmon incident of 2009 (14) to cluster failure more recently (2), funders, researchers, and the general public do not seem to have reduced their appetite for pictures of brain maps, or gadgets with the word "neuro" in the name. Multiple blogs exist for the sole purpose of criticizing such enterprise (see Table 3).

The replicability crisis should certainly give 'neuroimagers' pause, and reason to soul-search. It is more important than ever to clarify when fMRI is and when it is not useful. The method remains the best noninvasive imaging tool for many research questions, however imperfect and imprecise it may be. However, to address past limitations, I argue neuroimaging researchers planning future studies need to consider the following five factors: power/effect size, design optimization, replicability, physiological confounds, and data sharing. I believe we can rapidly improve the quality of fMRI research if researchers incorporate the following five guidelines and if reviewers incorporate these criteria into their evaluations of neuroimaging manuscripts. Note that this is intended as a starting point, not a comprehensive proposal. Perhaps these practices will help us make faster progress towards a biologically grounded understanding of mental phenomena.

1. Conduct a power analysis

One of the most important factors in any study is establishing the power needed to detect an effect, if one truly exists. This is especially crucial for fMRI studies as it speaks to feasibility - spending $500/hour to scan a sample too small to draw inferences from is costly both monetarily and scientifically. Recently tools have become available to conduct power analysis for simple fMRI study designs. The 'fmripower' tool is an application that provides a power curve based on a region of interest (ROI) in the brain and potential sample size (see Table 1). Researchers can use results from previous studies (e.g. Zstat maps of a statistical contrast) to project the necessary sample size for a future, similar study. It is highly recommended to either obtain existing statistical maps or collect a pilot sample of data to use to calculate power. While many published studies have modest sample sizes, increasingly the field has recognized that this has held back progress by contributing to high rates of false positive results and being biased only to detect larger than average effect sizes (15). If a properly powered sample is not possible, clearly labelling work as pilot studies can help with transparency.

Table 1. Tools to Improve fMRI Research

Best Practice Consideration	Tools/Solutions
Power Analysis	fMRIPower Tool
Data Simulation	fMRISim from Brainiak
Research Design Efficiency	Efficiency Tutorial for FSL FEAT User Guide
Custom Hemodynamic Response Functions	Hemodynamic Tools
Preregistration	As Predicted Open Science Framework PLoS One
Data and Code Sharing Repositories	European Open Science Cloud Figshare Kaggle Mendeley

DOI: https://doi.org/10.31885/jrn.2.2021.1517

	Open Aire OpenNeuro Open Science Framework Zenodo
Imaging Data Standardization	BIDS
Automated meta-analysis	Neurosynth

2. Pilot, pilot, pilot.

Since "timing is everything" in fMRI, it is critical to thoroughly pilot studies. Experimental design is a complicated issue that could easily be its own article. We recommend all 'neuroimagers' familiarize themselves with the basics of MRI physics, the hemodynamic response function (HRF), and the assumptions underlying linear modeling methods - and consult closely and regularly with experts in these domains. Every study will be uniquely influenced by factors ranging from habituation (neural and behavioral suppression after repeated presentations of stimuli or types of stimuli), head motion, and participant fatigue. These factors, in turn, can vary by scanning protocol and research population - children with attention deficit hyperactivity disorder (ADHD) performing a working memory task, for example, will likely experience an hour long scan very differently than adult expert meditators, for example.

Tools like fMRIsim and the design efficiency tool (Table 1) in FMRIB Software Library (FSL) can be used to optimize designs during piloting. FMRIsim can be used to generate the ideal schedule of trial order and duration to maximize signal detection ability, while respecting logistic confines such as length limits of the scan. The FSL design efficiency tool can be used to evaluate whether the trials in a task are spaced appropriately as well as if it is statistically distinguishable from one another, or if there is too much temporal overlap and the study design is unlikely to produce interpretable results.

Whenever possible, piloting should also be done in the intended research population to account for the above issues. Even small procedural changes, such as behavioral performance measured inside versus outside the scanner, can lead to behavioral and potentially neural differences (16). Ideally, imaging centers would permit a limited number of hours (5-10) for piloting at no cost, as it is in the interest of both the researcher and the imaging center for the data to be as high quality as possible. When this is not possible, we recommend researchers either seek seed grants, or include piloting costs in grant budgets. Absent these options, pre-planning to pause analysis after a set number of scans to check for quality control and confounds (e.g. significant habituation or fatigue effects) can reduce the risk of discovering issues too late.

An additional issue is that the assumptions of linear modeling may not match the realities of fMRI data, as cognitive and psychophysical responses can take nonlinear forms. The HRF can also vary across brain regions, which may necessitate use of custom region or network specific HRFs (Table 1). While these factors are outside the scope of this commentary, researchers are urged to always consider whether nonlinear analysis methods and/or customized HRFs are appropriate for their question.

3. Predefine the hypothesis testing and data analysis plan in advance (when appropriate).

Given the thousands of potential ways to analyze any given data set, one way to reduce researcher degrees of freedom is to predefine analysis in advance. This can be formalized through preregistration of hypotheses, publishing the study protocol, and/or submitting a registered report where methods are peer reviewed before data analysis, and findings are published in a second stage. Many software programs allow a template analysis workflow to be generated, and this can be included in a code repository to promote transparency. Manuscripts reporting fMRI data should specify how quality control assessment was conducted, rate of data exclusions due to poor scan quality, and what measures were taken to mitigate the impact of data quality for individuals as well as the group statistics (e.g. outlier deweighting). For those new to neuroimaging, mapping out the analysis plan with an experienced fMRI researcher can often help prevent major data quality, study design, or interpretability issues. Carefully documenting procedures such that anyone reading a manuscript or accessing a study repository can replicate the study is another overlooked practice that can increase replicability and transparency.

Journals and funders can reinforce the importance of these practices by making data availability statements required for projects, and either requiring or strongly encouraging data be available in a public repository rather than "upon reasonable request" as is often the default (17). Proposals that predefine analysis plans and data and code sharing practices should be given higher appraisals for rigor, reproducibility, and data quality.

Some researchers feel these practices are overly restrictive, and they are not yet widely adopted (with total published registered reports numbering in the low hundreds), although the number of journals offering registered reports has grown to over 200 since 2013 (18). Journals vary in how deviations from predefined protocols are handled, but generally,

there is some flexibility as long as the authors are transparent about the reasoning. Note, however, that the purpose of predefining workflows is to avert the issue of researcher degrees of freedom and therefore too many changes post-hoc can nullify this advantage.

There is also an argument that preregistration can be overly rigid and stifle creativity; experiments that failed to confirm hypotheses have often produced unexpected but important findings (Gary Glover, personal communication, August 21, 2020). For work that is exploratory, or where it is difficult to define testable hypotheses, data driven approaches may be more appropriate. For example, splitting data into a "testing set" and a "hold out set" allows the researcher to demonstrate that patterns found in the test set are robust enough to apply to an unseen sample. Such analyses can even be preregistered as "exploratory reports." For some novel investigations, it may be infeasible or compromise the science to predefine data collection or analysis plans. In such a case, we return to the points in Best Practice #2: Pilot, pilot, pilot.

Additionally, meta-analysis techniques can be used to inform exploratory analyses, such as by identifying networks of brain regions to reduce degrees of freedom in analysis. Automated tools for this exist, such as Neurosynth (19, see Table 1). For topics not represented in the Neurosynth repository of over 14,000 studies (as of March 2021), a systematic review or meta-analysis can help inform decisions about data-driven analysis techniques.

4. Openly share data and code.

Practices such as data and code sharing promote transparency and replicability. Providing data and analysis code in a public repository allows reviewers of manuscripts to directly assess data quality, appropriateness of the analysis plan, and fidelity of the results. Open repositories also help identify errors in data acquisition or processing that may influence results (20). Increasingly, funders and journals are requiring data availability statements or mandating data sharing. It is important to incorporate these requirements into study design, such as by specifying in the consent form that anonymized data will be included in a repository. Additionally, ensure that data is truly anonymized by removing protected health information (PHI) such as dates of scans from all files.

Open data and code also allows researchers to test the reliability, generalizability, and replicability of findings, which can lead to important dialogues about the burden of evidence necessary, for example, when making claims about translation of findings from the scanner to the clinic (see references 4-7). Table 2 includes a reading list of papers laying out methodological considerations, with several relevant especially to researchers using neuroimaging to study individual differences, neurological or psychiatric disorders, and those who hope to translate findings from the scanner into clinical practice.

Table 2. Selected articles discussing methodological best practices relevant to neuroimaging.

Article Focus	Citation
Defines criteria for data reusability (Findability, Accessibility, Interoperability, and Reusability)	Wilkinson MD. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Nature Publishing Group*. https://doi.org/10.1038/sdata.2016.18
Framework for improving neuroimaging analysis workflows	Gorgolewski KJ, Alfaro-Almagro F, Auer T, Bellec P, Capotă M, Chakravarty MM., *et al.* (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Computational Biology*, *13*(3), e1005209. https://doi.org/10.1371/journal.pcbi.1005209
Limitations of replicability of task fMRI studies, implication for biomarker discovery	Elliott ML, Knodt AR, Ireland D, Morris M L, Poulton R, Ramrakha S, *et al.* (2020). What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychological Science*, *31*(7), 792–806. https://doi.org/10.1177/0956797620916786
Considerations and challenges for individual differences research using neuroimaging	Dubois J, Adolphs R. (2016). Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences*, Vol. 20, pp. 425–443. https://doi.org/10.1016/j.tics.2016.03.014

Recommendations for advancing causal inference based on fMRI connectivity data	Reid AT, Headley DB, Mill RD, Sanchez-Romero R, Uddin LQ, Marinazzo D, *et al* (2019). Advancing functional connectivity research from association to causation. *Nature Neuroscience*. https://doi.org/10.1038/s41593-019-0510-4
Psychiatric neuroimaging best practices	Saggar M, Uddin LQ. (2019). Pushing the boundaries of psychiatric neuroimaging to ground diagnosis in biology. *ENeuro*, *6*(6). https://doi.org/10.1523/ENEURO.0384-19.2019
Considerations for machine learning analysis of neuroimaging	Haynes JD. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270. https://doi.org/10.1016/j.neuron.2015.05.025
Statistical considerations when re-analyzing open data	Thompson WH, Wright J, Bissett PG, Poldrack, RA. (2020). Dataset decay and the problem of sequential analyses on open datasets. *ELife*, *9*, 1–17. https://doi.org/10.7554/eLife.53498

5. Account for relevant physiological factors.

Since the blood oxygen level dependent (BOLD) signal is not a direct index of neural activity it is subject to confounding by other biological processes. Head motion, eye blinks, cardiac rhythm, task-locked respiration patterns, and age-related structural and vascular differences are just a few factors that can contribute to variability in BOLD signal (8, 21, 22). For example, head motion that is higher in a clinical group can lead to inflated parameter estimates due to motion artifact, which then create the spurious appearance of case-control differences at the group level. Inadequately accounting for these factors can produce misleading results.

Transparently including parameter choices and rationale in a preregistration, registered report, or methods section can increase replicability of studies by making it clear what choices were made and why. At a minimum, for each of the major sources of physiological noise - motion, respiration, and heart rate - researchers should include a rationale for including or not including each one as a confound (e.g. due to feasibility issues).

Unfortunately dealing with physiological confounds is not always as easy as regressing them out, such as in cases where such effects are correlated with behavior. In these cases, the signal modulations are entangled with the processes of interest and therefore regressing them out may reduce power and increase type II error. This means that even if a true effect exists, it is concluded that there is no effect (i.e., false negative). It may therefore be prudent to compare results with and without potential confounds removed to ensure that the baby is not thrown out with the bath water.

While these guidelines are not a balm that will solve all the larger issues of fMRI (what *is* the BOLD signal, anyway?) they have the potential to drastically improve the quality and replicability of science. Therefore, despite the career costs that may be incurred (23), it is recommended that all current and prospective 'neuroimagers' consider the above guidelines to ensure the future viability of the science.

Table 3. A sampling of sources of criticism of fMRI.

Source Type	Web Address
Journal editorial	https://www.nature.com/articles/nn.4521.pdf
Journal editorial	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5410776/
Blog	https://www.discovermagazine.com/author/neuroskeptic
Blog	https://neurocritic.blogspot.com/
Blog	http://neurobonkers.com/
Blog	https://blogs.scientificamerican.com/guest-blog/controversial-science-of-brain-imaging/
Blog	https://medium.com/swlh/the-limitations-and-reliability-of-fmri-60275559e203#:~:text=fMRI%20research%20also%20receives%20criticism,behavior%20(e.g.%2C%20neuroscience).
Satire	https://twitter.com/CousinAmygdala?s=20

News article	https://www.vox.com/2016/9/8/12189784/fmri-studies-explained
News article	https://www.yalescientific.org/2014/04/debunking-science-fmri-a-not-so-reliable-mind-reader/
News Article	https://today.duke.edu/2020/06/studies-brain-activity-aren%E2%80%99t-useful-scientists-thought

Acknowledgements: I would like to thank Gary Glover, PhD for his input on fMRI best practices, and Halee Staggs, BS for assistance editing the commentary.

References

1. Woo CW, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRIanalyses: pitfalls and recommendations. Neuroimage. 2014 May;91:412-19. Available from: https://doi.org/10.1016/j.neuroimage.2013.12.058
2. Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci USA. 2016 Jul;113(28):7900-05. Available from: https://doi.org/10.1073/pnas.1602413113
3. Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA. fMRI clustering in AFNI: false-positive rates redux. Brain Connect. 2017 Apr;7(3):152-71. Available from: https://doi.org/10.1089/brain.2016.0475
4. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med. 2016 Dec;23(1):28-38. Available from: https://doi.org/10.1038/nm.4246
5. Dinga R, Schmaal L, Penninx BWJH, van Tol MJ, Veltman DJ, van Velzen L, et al. Evaluating the evidence for biotypes of depression: methodological replication and extension of Drysdale et al. Neuroimage Clin. 2019;22:101796. Available from: https://doi.org/10.1016/j.nicl.2019.101796
6. Dinga R, Schmaal L, Marquand AF. A closer look at depression biotypes: correspondence relating to Grosenick et al. (2019). Biol Psychiatry: Cog Neurosci Neuroimaging. 2020 May;5:554-5. Available from: https://doi.org/10.1016/j.bpsc.2019.09.011
7. Grosenick L, Liston C. Reply to: a closer look at depression biotypes: correspondence relating to Grosenick et al. (2019). Biol Psychiatry: Cog Neurosci Neuroimaging. 2020;5(5):556. Available from: https://doi.org/10.1016/j.bpsc.2019.11.002
8. Tsvetanov KA, Henson RNA, Rowe JB. Separating vascular and neuronal effects of age on fMRI BOLD signals. Philos Trans R Soc Lond B Biol Sci. 2020 Nov;376:20190613. Available from: https://doi.org/10.1098/rstb.2019.0631
9. Özbay PS, Chang C, Picchioni D, Mandelkow H, Moehlman TM, Chappel-Farley MG, et al. Contribution of systemic vascular effects to fMRI activity in white matter. Neuroimage. 2018 Aug;176:541-9. Available from: https://doi.org/10.1016/j.neuroimage.2018.04.045
10. Thompson WH, Wright J, Bissett PG, Poldrack RA. Dataset decay and the problem of sequential analyses on open datasets. ELife. 2020 May;9:e53498. Available from: https://doi.org/10.7554/eLife.53498
11. Stockton N. 2017 June. Don't be so quick to flush 15 years of brain scan studies. Iowa USA: Wired; 2016 Jul 8. Available from: https://www.wired.com/2016/07/dont-quick-flush-15-years-brain-scan-studies/
12. Morris E. Why we need guidelines for brain scan data. Iowa USA: Wired; 2019 Sept 17. Available from: https://www.wired.com/story/why-we-need-brain-scan-data-guidelines/
13. D'Esposito M. 2019. Are individual differences in human brain organization measured with functional MRI meaningful? Proc Natl Acad Sci USA. 2019 Nov;116(45):22432-4. Available from: https://doi.org/10.1073/pnas.1915982116
14. Bennett CM, Miller MB, Wolford GL. Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: an argument for multiple comparisons correction. Neuroimage. 2009 Jul;47(1):S125.
15. Cremers HR, Wager TD, Yarkoni T. 2017. The relation between statistical power and inference in fMRI. PLoS One. 2017 Nov;12(11):e0184923. Available from: https://doi.org/10.1371/journal.pone.0184923
16. Mills-Finnerty C, Hanson C, Khadr M, Hanson S.J. Computations and connectivity underlying aversive counterfactuals. Brain Connect. 2020 Nov;10(9):467-78. Available from: https://doi.org/10.1089/brain.2020.0766
17. Miyakawa T. No raw data, no science: another possible source of the reproducibility crisis. Mol Brain. 2020 Feb;13:24. Available from: https://doi.org/10.1186/s13041-020-0552-2
18. Chambers C. What's next for registered reports? Nature. 2019 Sept;573:187-9. Available from: https://doi.org/10.1038/d41586-019-02674-6
19. Yarkoni T, Poldrack R, Nichols T, Van Essen D, Wager T. NeuroSynth: a new platform for large-scale automated synthesis of human functional neuroimaging data. Front Neuroinform Conference Abstract: 4th INCF Congress of Neuroinformatics. 2011 Sept. Available from: https://doi.org/10.3389/conf.fninf.2011.08.00058
20. Bowring A, Maumet C, Nichols TE. Erratum: exploring the impact of analysis software on task fMRI results. Hum Brain Mapp. 2020 Dec;42(5):1564-78. Available from: https://doi.org/10.1002/hbm.25302
21. Murphy K, Birn RM, Bandettini PA. 2013. Resting-state fMRI confounds and cleanup. Neuroimage. 2013 Oct;80:349-59. Available from: https://doi.org/10.1016/j.neuroimage.2013.04.001
22. Park HD, Barnoud C, Trang H, Kannape OA, Schaller K, Blanke O. 2020. Breathing is coupled with voluntary action and the cortical readiness potential. Nat Commun. 2020 Feb;11:289. Available from: https://doi.org/10.1038/s41467-019-13967-9
23. Mills-Finnerty C. Barriers to reproducibility: misalignment of career incentives and open science best practices. J Rep Neurosci. 2020 Aug;10. Available from: https://doi.org/10.31885/jrn.1.2020.304