

UMass Chan Medical School

eScholarship@UMassChan

Open Access Publications by UMMS Authors

2019-11-13

Introduction to the special issue on reproducibility in neuroimaging

Russell A. Poldrack
Stanford University

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: <https://escholarship.umassmed.edu/oapubs>

 Part of the [Biostatistics Commons](#), [Investigative Techniques Commons](#), [Laboratory and Basic Science Research Commons](#), and the [Neuroscience and Neurobiology Commons](#)

Repository Citation

Poldrack RA, Whitaker K, Kennedy DN. (2019). Introduction to the special issue on reproducibility in neuroimaging. Open Access Publications by UMMS Authors. <https://doi.org/10.1016/j.neuroimage.2019.116357>. Retrieved from <https://escholarship.umassmed.edu/oapubs/4070>

Creative Commons License



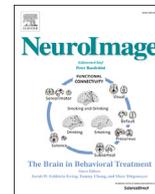
This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Open Access Publications by UMMS Authors by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Editorial

Introduction to the special issue on reproducibility in neuroimaging

The last decade has seen increasing attention to the problem of scientific reproducibility, across a broad range of scientific fields (Camerer et al., 2016; Morrison, 2014; Open Science Collaboration, 2015). Within the field of neuroimaging, there has been a particular focus on issues of analytic variability (Bowring et al., 2019; Carp, 2012) statistical power (Button et al., 2013; Poldrack et al., 2017), and test-retest reliability (Bennett and Miller, 2013), all of which have raised alarms regarding the potential for irreproducible results. In addition, failed replications (Boekel et al., 2015; Dinga et al., 2019) and meta-analytic null results (Müller et al., 2017) have raised particular concern about studies of group and individual differences. This special issue was developed in light of these emerging concerns, with the goal of highlighting and encouraging work that aims to both quantify and improve the reproducibility of neuroimaging research. Here we provide a brief overview of the papers within this special issue.

The terms “reproducibility”, “replicability”, and “reliability” are used in variable ways by many groups (Nichols et al., 2017). Here we use “reproducibility” as a blanket term encompassing all aspects of the ability to reproduce a result, from same data/same analysis to different data/different analysis. We use the term “replication” to refer specifically to the ability for a finding to be reproduced - qualitatively found again - in a separate dataset. By “reliability” we specifically mean the degree to which a measurement is stable across multiple repeated measurements.

1. Quantifying reliability

A number of the papers submitted for this special issue focused on quantifying the test-retest reliability of neuroimaging measurements across repeated measures. Several of these focused on functional connectivity using fMRI. Nobel et al. (this issue) performed a systematic review that combined results across 25 published studies of test-retest reliability of edge-level functional connectivity, which showed poor reliability on average and highlighted a number of features that promoted greater reliability. Bhadwar et al. (this issue) took the opposite approach, assessing connectivity in a single individual across multiple sites and scanner vendors. This effort showed reasonable reliability that was moderated by both site and vendor effects, highlighting possible limitations on between-site generalizability. Baria et al. (this issue) assessed the reliability of resting state fMRI “fingerprinting” (Finn et al., 2015), finding that an optimal reconstruction method led to higher fingerprinting performance both within and across sites. Tu et al. (this issue) also assessed the ability to fingerprint individuals across multiple sessions, replicating previous successful fingerprinting results. They further showed that connectivity could be used to predict pain thresholds both within and across sessions.

<https://doi.org/10.1016/j.neuroimage.2019.116357>

Available online xxx

1053-8119/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Other papers assessed the reliability of structural MRI measures. Buonincontri et al. (this issue) assessed the test-retest reliability as well as the between-site reproducibility of a method for reconstructing multiple parametric maps from a single excitation (known as “MR fingerprinting”, not be confused with “connectome fingerprinting”). They report high levels of test-retest reliability as well as acceptable levels of between-site reproducibility. Drenthen et al. (this issue) assessed the retest reliability of multi-slice GRASE for use in quantitative myelin water imaging. They validated the multi-slice measures in relation to the standard whole-brain GRASE method, and further showed that this method achieved acceptable reliability even using parallel acceleration to substantially reduce scan time. Lancione et al. (this issue) assessed the specific effect of echo time (TE) on reproducibility of quantitative susceptibility mapping across 3T and 7T scanners. Their results show that it is possible to generate reproducible maps across scanners by calibrating the echo time across scanners. Lerma-Usabaig et al. (this issue) assessed the utility of fractional anisotropy in homologous pairs of tracts as a diagnostic metric across nine datasets, showing that this metric was substantially more precise compared to standard measures of diffusion anisotropy.

2. Replication

The last decade has seen increasing interest in replication studies, but it remains difficult to publish such studies. For this reason, we solicited direct replication studies as part of our call for papers, ultimately including two such studies. Geller et al. (this issue) attempted to replicate a previous finding that had shown that binarized measures of tract connection were strongly associated with post-stroke language function compared to a continuous measure of lesion load. They failed to find an over difference between the binary and continuous measures. Kampa et al. (this issue) assessed replicability between two samples within a single fMRI study of stress resilience, finding that replicability at the whole-brain level was good whereas replicability at the level of individual regions of interest was lower.

Other studies examined the nature of replication and the factors that affect it. Hong et al. (this issue) surveyed studies that had claimed to replicate previous findings, in order to assess what was meant by “replication”. They found that most studies did not provide quantitative evidence for anatomical replication, and that many of the reported peaks were distant from the original peak, suggesting that claimed replication does not always imply true replication. Xia et al. (this issue) examined the reproducibility of case-control differences in major depressive disorder using resting fMRI. They found reliable differences between groups in several measures, but these measures were only reproducible when relatively large samples were included in the analysis. Jollans et al. (this

issue) assessed the ability of several different machine learning regression methods to predict continuous outcomes from both real and simulated fMRI data. They found that methods varied in their predictive accuracy across the same data, and that different methods performed better or worse depending on the size of the effect and the size of the sample. Like Xia, they found that prediction in the case of small effect sizes was only accurate when sample sizes were large. Thus, one generalizable outcome from these papers is greater attention to larger sample sizes (cf. Button et al., 2013; Poldrack et al., 2017).

3. Confounds

A final set of papers focused on effects of confounds in neuroimaging analysis. Quax et al. focused on a specific confound in decoding of stimulus location from MEG data due to the effects of eye movements. Their results suggest that controlling for the confounding effects of eye movements is essential in any MEG decoding study. Hyatt et al. reviewed the use of covariates in structural neuroimaging studies, finding a broad range of different strategies used across the literature. In an analysis of data from the Human Connectome Project, they found that some covariates had strong effects on outcomes whereas the effects of others were minimal. They highlight the need for pre-registration to ensure that flexibility in confound modeling does not inflate error rates.

4. Conclusion

Reproducibility is a multi-faceted and ongoing concern for neuroimaging researchers. The papers in this special issue highlight the different ways that the field is tackling these issues head on. What particularly ties these articles together is a drive to efficiently and effectively utilize neuroimaging research to better understand the human mind and brain in health and disease, and across the lifespan. We hope that they inspire additional work aimed at assessing and improving reproducibility.

References

- Bennett, C.M., Miller, M.B., 2013. fMRI reliability: influences of task and experimental design. *Cognit. Affect Behav. Neurosci.* 13, 690–702.
- Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., Forstmann, B.U., 2015. A purely confirmatory replication study of structural brain-behavior correlations. *Cortex* 66, 115–133.

- Bowring, A., Maumet, C., Nichols, T.E., 2019. Exploring the impact of analysis software on task fMRI results. *Hum. Brain Mapp.* 40, 3362–3384.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafó, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376.
- Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351, 1433–1436.
- Carp, J., 2012. On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Front. Neurosci.* 6, 149.
- Dinga, R., Schmaal, L., Penninx, B.W.J.H., van Tol, M.J., Veltman, D.J., van Velzen, L., Mennes, M., van der Wee, N.J.A., Marquand, A.F., 2019. Evaluating the evidence for biotypes of depression: methodological replication and extension of. *Neuroimage Clin.* 22, 101796.
- Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671.
- Morrison, S.J., 2014. Reproducibility project: cancer biology: time to do something about reproducibility. *Elife* 3, e03981.
- Müller, V.I., Cieslik, E.C., Serbanescu, I., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2017. Altered brain activity in unipolar depression revisited: meta-analyses of neuroimaging studies. *JAMA Psychiatry* 74, 47–55.
- Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T.T., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20, 299–303.
- Open Science Collaboration, 2015. Psychology. Estimating the reproducibility of psychological science. *Science* 349, aac4716.
- Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafó, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* 18, 115–126.

Russell A. Poldrack*

Department of Psychology, Stanford University, Stanford, CA, 94305, USA

Kirstie Whitaker

The Alan Turing Institute, London, NW1 2DB, UK

Department of Psychiatry, University of Cambridge, Cambridge, CB2 0SZ, UK

David Kennedy

Eunice Kennedy Shriver Center, Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, 01655, USA

* Corresponding author.

E-mail address: russpold@stanford.edu (R.A. Poldrack).