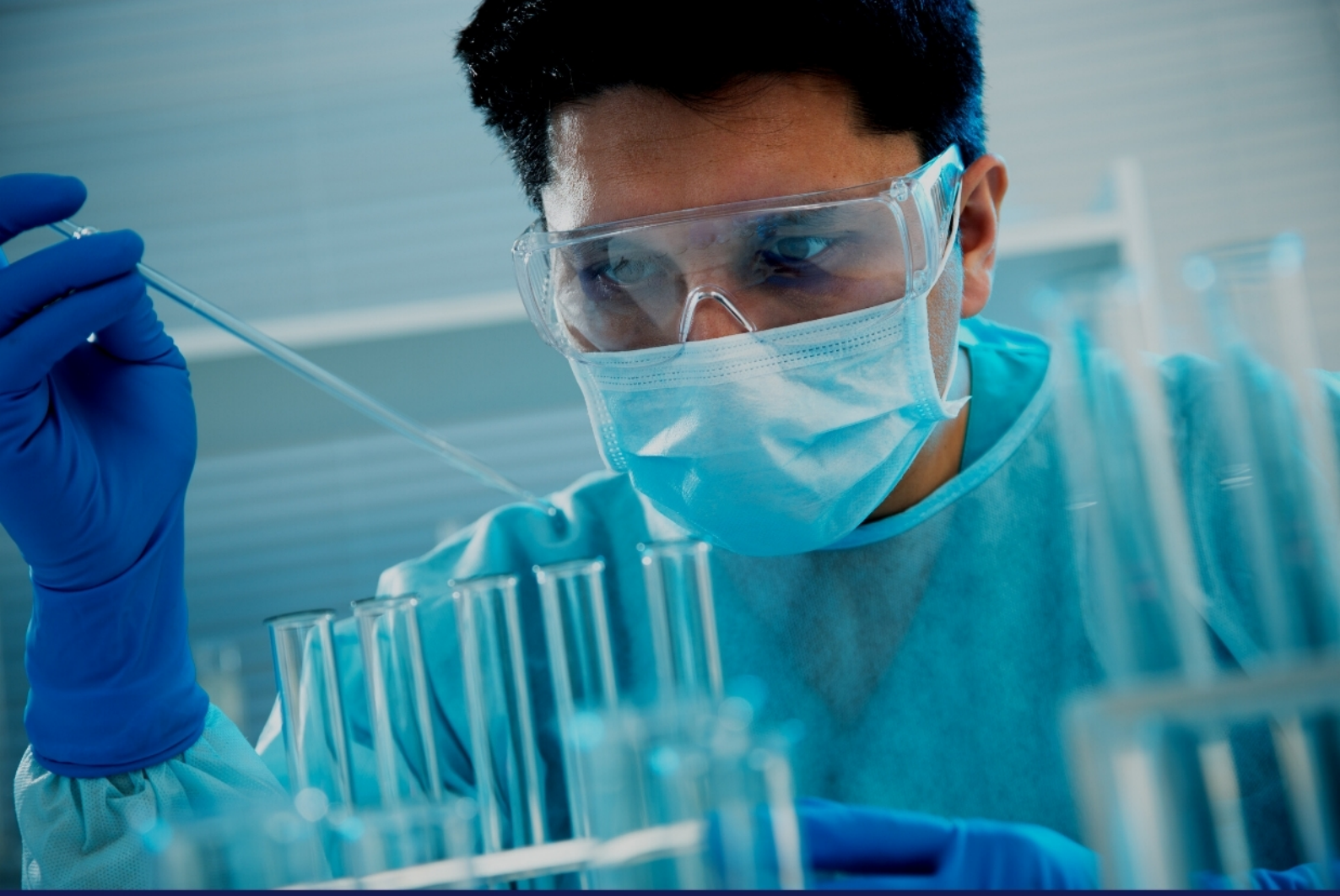

GENE EXPRESSION UNIVERSITY

October 06-29

THE ULTIMATE WEBINAR SERIES IN GENE EXPRESSION STUDIES

Register here

applied biosystems
by Thermo Fisher Scientific

WILEY

Received: 8 April 2020 | Revised: 16 July 2020 | Accepted: 12 August 2020

DOI: 10.1111/ejn.14954

EDITORIAL

EJN European Journal of Neuroscience FENS WILEY

The case for preregistering all region of interest (ROI) analyses in neuroimaging research

Abstract

In neuroimaging studies, small sample sizes and the resultant reduced statistical power to detect effects that are not large, combined with inadequate analytic choices, concur to produce inflated or false-positive findings. To mitigate these issues, researchers often restrict analyses to specific brain areas, using the region of interest (ROI) approach. Crucially, ROI analysis assumes the a priori justified definition of the target region. Nonetheless, reports often lack details about where in the timeline, ranging from study conception to the data analysis and interpretation of findings, were ROIs selected. Frequently, the rationale for ROI selection is vague or inadequately founded on the existing literature. These shortcomings have important implications for ROI-based studies, augmenting the risk that observed effects are inflated or even false positives. Tools like preregistration and registered reports could address this problem, ensuring the validity of ROI-based studies. The benefits could be enhanced by additional practices such as selection of ROIs using quantitative methods (i.e., meta-analysis) and the sharing of whole-brain unthresholded maps of effect size, as well as of binary ROIs, in publicly accessible repositories.

In the functional magnetic resonance imaging (fMRI) literature, false report probability (FRP), the probability that statistically significant findings are actually not true (Ioannidis, 2005), as well as the probability that reported effects are inflated (Ioannidis, 2008), are pervasive problems. These have been attributed to small sample sizes and resulting small statistical power to detect effects that are not large (Button et al., 2013), combined with selective reporting biases (Szucs & Ioannidis, 2017) and inadequate analytic choices (Eklund, Nichols, & Knutsson, 2016). fMRI studies often rely on whole-brain mass-univariate voxel-wise testing, paired with voxel- or cluster-wise corrections for multiple comparisons. Though voxel-wise corrections are more reliable in terms of controlling the false-positive rate, they have costs in reducing sensitivity (i.e., correctly identifying true effects). Conversely, many methods for cluster-wise correction still lead to inflated false-positive rates, that is, over the expected 5% (Eklund et al., 2016), though there is debate about the size of the inflation (Cox, Chen, Glen, Reynolds, & Taylor, 2017). The problem is further compounded when results from individual studies are used to estimate effects. Neuroimaging studies typically selectively report effects only for the subset of activated (i.e., statistically significant) voxels or clusters. Stringent multiple comparisons corrections leading to smaller p value thresholds combined with low statistical power imply that only voxels or clusters with high estimates will be statistically significant and consequently reported (Ioannidis, 2008; Reddan, Lindquist, & Wager, 2017). Most methods for neuroimaging meta-analyses are based on data in published reports and hence risk further overestimating the true effect. Moreover, when brain activity is analyzed using multivariate methods (Haynes, 2015), the resultant estimates, for example for prediction accuracy, may be problematic as well. Specifically, it has been demonstrated that both the estimate (Varoquaux et al., 2017) and the related standard error (Varoquaux, 2018) can fluctuate depending on the analytic choices, particularly with small samples.

To balance study sensitivity while also reducing the number of comparisons performed, researchers can restrict the search for significant results within specific brain areas, using the so-called regions of interest (ROI) approach. This method is ubiquitous in the fMRI literature (Poldrack, 2007). For instance, in an Activation Likelihood Meta-Analysis on specific phobia, we found that 28 out of 31 papers employed ROI approaches (Gentili, Messerotti Benvenuti, Lettieri, Costa, & Cecchetti, 2019). In principle, ROI analysis can be statistically appropriate and useful when the target brain region is defined a priori. Indeed, the approach requires that the ROI selection predates data exploration, and that its definition in terms of anatomical location is based on previous findings. Nonetheless, reports often lack details about when, in the timeline ranging from study conception to data analysis and interpretation of the

findings, have ROIs been selected. This ambiguity has crucial implications for ROI-based studies, as the ad hoc definition of ROIs, informed by the results of whole-brain voxel-wise analyses, inflates the FRP and the reported effect size (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009).

Preregistration of analyses, which requires researchers to commit to analytic steps without prior knowledge of the research outcomes, offers the opportunity to rigorously implement the concept of a priori (Nosek, Ebersole, DeHaven, & Mellor, 2018). Specifically, the definition of the ROI can be preregistered in a publicly accessible registry, such as the Open Science Framework (https://osf.io/), following guidelines recently proposed by the Committee on Best Practice in Data Analysis and Sharing of the Organization for Human Brain Mapping (Nichols et al., 2017; COBIDAS), summarized in a freely available document (https://osf.io/dvb2e/). The preregistration template unequivocally underscores the relevance of specifying the method used for ROI definition, the number of ROIs, and the modalities of analyzing brain activity from each. The selection of ROIs is thus time-stamped, confirming that decisions about its selection predated data acquisition or exploration. For studies using already collected, publicly available or shared datasets (e.g., Neuroimaging Analysis Replication and Prediction Study [NARPS]; Botvinik-Nezer et al., 2020), ROIs selection evidently cannot predate data acquisition. In such cases, preregistration of ROI selection fosters analytic "decision independence" (i.e., the decision that led to the selection of the ROI is not a function of unrepeateable features of the dataset analyzed; Srivastava, 2018). Specifically, it ensures ROIs are chosen before any processing step is performed, thereby limiting the possibility of questionable research practices. Some transparently reported data exploration may be warranted to ensure the feasibility of the study, for example, checking whether functional acquisitions characterized by partial brain coverage can be adequate to test study hypotheses. Other relevant analytic decisions (e.g., exclusion of subjects due to excessive movement) can be well-justified; however, it is crucial that they are not taken upon seeing results of interest.

P-hacking could take the form of choosing to report only a subset of ROI analyses out of a suite performed or, upon examining data, adding observations until the effect in an ROI becomes significant (Bruns & Ioannidis, 2016; Poldrack et al., 2017). If ROI definition is preregistered, researchers pursuing either of these approaches would at the very least need to be transparent, which might prompt critical peer-review and eventually even discourage *p*-hacking. Importantly, even though preregistration requires relevant choices to be taken before data exploration, deviations from protocol are possible when transparently reported and justified. Readers and reviewers could then weigh in regarding the judiciousness of the deviations and, depending on their magnitude, findings could be presented as exploratory or non-preregistered.

Furthermore, theoretical grounding of the a priori ROI selection may be problematic. For instance, in a meta-analysis on specific phobias, we showed that only 19 out of 31 studies explicitly justified the choice of ROI based on previous literature (Gentili et al., 2019). Frequently, the rationale supporting ROI selection is based on vague statements or on citing previous findings without motivating how these suggest an association between a well-defined brain area and a specific mental process. Across the neuroimaging literature, it is unclear how many published papers rely on an adequate definition of ROIs (Hong, Yoo, Han, Wager, & Woo, 2019). Hypothetically, the choice of any ROI may be justified based on the existing literature, leading to a "garden of forking paths" of potential data-dependent analyses (Gelman & Loken, 2014). Using previous non-preregistered reports to inform a new ROI-based study has to accommodate the possibility that some reports may be based on data-dependent ROI selection. In this case, the definition of the ROI in the new study, even when preregistered, could be influenced by estimates that are inflated or false positives. Thus, resultant findings could contradict or fail to replicate previous ones. Researchers might worry about difficulties in publishing such findings and consequently be wary of preregistering ROI definitions. Addressing this problem requires a commitment from journals to publish negative or contradictory findings, as well as replication failures.

Selection of ROIs based on a meta-analysis rather than a single study could also help overcome the issue of previous estimates being inflated or false positives. In this scenario, the ROI definition is determined by the spatial consistency of results coming from multiple studies and, while true effects presumably converge in the same anatomical location, false-positive results should be randomly distributed throughout the brain. Useful tools in this regard are represented by GingerALE (Müller et al., 2018) and Seed-based d Mapping (SDM; Albajes-Eizagirre, Solanes, Vieta, & Radua, 2019), which quantitatively aggregate neuroimaging findings, as well as by Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) and NeuroQuery (Dockès et al., 2020), which estimate the association between voxels and terms semantically related to the study hypothesis or to functional areas of interest (e.g., fusiform face area). The definition of ROIs could also be based on inherently vague anatomical terms (e.g., medial prefrontal cortex) or it may refer to a large patch of the cortex (e.g., precise location within the Superior Frontal Gyrus; Hong et al., 2019). Interestingly, Neurosynth and Neuroquery can be used to precisely define anatomical ROIs, which have no clear boundaries or are typically not included in anatomical atlases, as the temporo-parietal junction (Lettieri et al., 2019).

Will preregistration improve the quality of ROI selection? As long as the researchers' integrity and openness to being wrong are assumed, we believe it will. Conversely, preregistration could be thwarted or gamed (Yamada, 2018). In principle, a researcher could acquire and analyze the data, obtain voxel-wise results,

select the ROI according to the whole-brain statistical maps and subsequently "preregister" the analytic pipeline as if it were a priori. However, proceeding in this way involves substantial effort, heightened attention to avoid mismatches among specific descriptive details, such as recruitment and analysis times, as well as careful curation of any shared dataset to modify timestamps. More importantly, it also implies the researcher's willingness to deceive and act dishonestly. Therefore, preregistration per se does not guarantee protection against false-positive or inflated estimates due to ad hoc ROI selection and hypothesizing after results are known (i.e., harking; Poldrack et al., 2017), but it makes these less likely.

To further limit the possibility of producing inflated or false-positive results, researchers can opt for a registered report (RR; Chambers, 2013), a particular type of preregistered article, in which the methods are peer-reviewed even before any data is acquired (stage I submission). Indeed, if implemented correctly (Hardwicke & Ioannidis, 2018), RRs could ensure the validity of ROI studies and help prevent questionable research practices, as reviewers are heavily involved in the evaluation of the rationale, study hypotheses and, consequently, in the ROI selection. Moreover, RRs guarantee publication if the approved protocol is followed, thereby assuaging concerns about difficulties in publishing negative or contradictory findings. It is conceivable that, depending on the rationale of the stage I RR, reviewers might themselves suggest specific ROIs to be tested. Even in the extreme case of a highly motivated and dishonest researcher having already collected and analyzed the data, "conveniently" defined the ROI depending on whole-brain results, and then submitted the stage I RR, it is very unlikely that this methodological section, including ROI selection criteria, will be left unrevised. This is even more unlikely if the rationale for the ROI is weak, vague or not well grounded in the scientific literature. Hence, RRs are a promising tool to reduce FRP and inflation of true effects in the neuroimaging literature, especially when ROI-based analyses are involved. They could also potentially be more resilient to scientific misconduct compared to preregistration, provided stage I protocols are publicly archived as read-only, time-stamped versions, which has not always been the case so far (Hardwicke & Ioannidis, 2018).

In summary, to enhance reproducibility and ensure the validity of findings, we suggest that mass-univariate and multivariate ROI-based MRI studies adhere to the following recommendations:

1. Preregistration in a publicly accessible repository of the study design, methodology, and statistical analysis pipeline, including details about ROIs selection. Particularly, researchers should clearly and thoroughly report methods employed for ROI definition and the theoretical background underlying their choice. The authors should also share the ROI as a standard space binary mask in publicly available repositories;
2. ROI selection based on quantitative methods (i.e., meta-analysis) is preferred;
3. Performing and reporting an exploratory whole-brain analysis and providing voxel-wise unthresholded maps of the effect size, even when negative results are obtained. In this regard, data should be released following FAIR principles (Findable, Accessible, Interoperable and Reusable; Wilkinson et al., 2016). Specifically, in the neuroimaging field, the use of web-based repositories as NeuroVault (http://neurovault.org; Gorgolewski et al., 2015) ensures that unthresholded maps are not lost and remain accessible to the entire community. In addition, organizing data and meta-data following the brain imaging data structure (BIDS) standard (Gorgolewski et al., 2017) and the neuroimaging data model (NIDM) format (Keator et al., 2013) could help in increasing reusability of resources.

ACKNOWLEDGMENT

The present work was carried out within the scope of the research program Dipartimenti di Eccellenza (art.1, commi 314-337 legge 232/2016), which was supported by the MIUR to the Department of General Psychology, University of Padua.

LC is supported by Progetto di Attività Integrate (PAI) Project granted by IMT School for Advanced Studies Lucca.

CONFLICT OF INTEREST

The authors reported no conflict of interest.

AUTHOR CONTRIBUTIONS

CG conceptualized the comment and, together with IAC, wrote the first draft of the manuscript. LC, GH and GL substantially revised the manuscript. All authors approved this version for submission.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.14941

DATA AVAILABILITY STATEMENT

The manuscript does not include or report on data.

Claudio Gentili[1]
Luca Cecchetti[2]
Giacomo Handjaras[2]
Giada Lettieri[2]
Ioana A. Cristea[3]

[1]*Department of General Psychology, University of Padova, Padova, Italy*
[2]*Social and Affective Neuroscience Group, MoMiLab, IMT School for Advanced Studies Lucca, Lucca, Italy*
[3]*Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy*

Correspondence
Claudio Gentili, Department of General Psychology, University of Padova, Via Venezia 8, Padova, Italy.
Email: c.gentili@unipd.it

ORCID

Claudio Gentili https://orcid.org/0000-0002-2579-8755
Luca Cecchetti https://orcid.org/0000-0001-5184-6477
Giacomo Handjaras https://orcid.org/0000-0001-5677-8434
Giada Lettieri https://orcid.org/0000-0001-8027-9053
Ioana A. Cristea https://orcid.org/0000-0002-9854-7076

REFERENCES

Albajes-Eizagirre, A., Solanes, A., Vieta, E., & Radua, J. (2019). Voxel-based meta-analysis via permutation of subject images (PSI): Theory and implementation for SDM. *NeuroImage*, *186*, 174–184. https://doi.org/10.1016/j.neuroimage.2018.10.077

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M. … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*, 84–88. https://doi.org/10.1038/s41586-020-2314-9

Bruns, S. B., & Ioannidis, J. P. A. (2016). p-Curve and p-Hacking in observational research. *PLoS One*, *11*, e0149144. https://doi.org/10.1371/journal.pone.0149144

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. https://doi.org/10.1038/nrn3475

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*, 609–610.

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). fMRI clustering and false-positive rates. *Proceedings of the National Academy of Sciences*, *114*(17), E3370–E3371.

Dockès, J., Poldrack, R. A., Primet, R., Gözükan, H., Yarkoni, T., Suchanek, F., … Varoquaux, G. (2020). NeuroQuery, comprehensive meta-analysis of human brain mapping. *eLife*, *9*. https://doi.org/10.7554/eLife.53385

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 7900–7905. https://doi.org/10.1073/pnas.1602413113

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*, 460. https://doi.org/10.1511/2014.111.460

Gentili, C., Messerotti Benvenuti, S., Lettieri, G., Costa, C., & Cecchetti, L. (2019). ROI and phobias: The effect of ROI approach on an ALE meta-analysis of specific phobias. *Human Brain Mapping*, *40*, 1814–1828. https://doi.org/10.1002/hbm.24492

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., … Poldrack, R. A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Computational Biology*, *13*, e1005209. https://doi.org/10.1371/journal.pcbi.1005209

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., … Margulies, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinform*, *9*, 8. https://doi.org/10.3389/fninf.2015.00008

Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Mapping the universe of registered reports. *Nature Human Behaviour*, *2*, 793–796. https://doi.org/10.1038/s41562-018-0444-y

Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron*, *87*, 257–270. https://doi.org/10.1016/j.neuron.2015.05.025

Hong, Y. W., Yoo, Y., Han, J., Wager, T. D., & Woo, C. W. (2019). False-positive neuroimaging: Undisclosed flexibility in testing spatial hypotheses allows presenting anything as a replicated finding. *NeuroImage*, *195*, 384–395. https://doi.org/10.1016/j.neuroimage.2019.03.070

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, *2*, e124.

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648.

Keator, D. B., Helmer, K., Steffener, J., Turner, J. A., Van Erp, T. G., Gadde, S., … Nichols, B. N. (2013). Towards structured sharing of raw and derived neuroimaging data across existing resources. *NeuroImage*, *82*, 647–661. https://doi.org/10.1016/j.neuroimage.2013.05.094

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*, 535–540. https://doi.org/10.1038/nn.2303

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., … Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, *10*, 5568. https://doi.org/10.1038/s41467-019-13599-z

Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., … Eickhoff, S. B. (2018). Ten simple rules for neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *84*, 151–161.

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., … Yeo, B. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*, 299–303. https://doi.org/10.1038/nn.4500

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 2600–2606. https://doi.org/10.1073/pnas.1708274114

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Soc Cogn Affect Neurosci*, *2*, 67–70. https://doi.org/10.1093/scan/nsm006

Poldrack, R., Baker, C., Durnez, J., Gorgolewski, K., Matthews, P., Munafò, M., … Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*. https://doi.org/10.1038/nrn.2016.167

Reddan, M. C., Lindquist, M. A., & Wager, T. D. (2017). Effect size estimation in neuroimaging. *JAMA Psychiatry*, *74*, 207–208.

Srivastava, S. (2018). Sound inference in complicated research: A multi-strategy approach. *PsyArXiv*.

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, *15*, e2000797. https://doi.org/10.1371/journal.pbio.2000797

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, *180*, 68–77.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145*, 166–179.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Front Psychol*, *9*, 1831. https://doi.org/10.3389/fpsyg.2018.01831

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670. https://doi.org/10.1038/nmeth.1635