Neuroimaging Neuroinformatics: Sample Size and Other Evolutionary Topics

David N. Kennedy¹

Published online: 27 April 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

The field of neuroinformatics has come a long way since this journal was founded in 2003. Many neuroinformatics-based tools, resources and initiatives have been developed, and continue to evolve in order to provide critical infrastructure support to expand the study of neuroscience. Over these first fifteen years of this journal, Drs. De Schutter, Ascoli and I have striven to bring a broad perspective to this diverse field. However, to keep the spirit of the journal fresh and up to date, the time has come to initiate some gentle evolution of the leadership of the journal. In order to better serve the neuroimaging subdomain, we are making some adjustments to the leadership structure of the journal. Specifically, as of July 1, 2018, we are pleased to announce that John Van Horn will be joining me in a new role of Co-Associate Editors for neuroimaging neuroinformatics. This change will help to strengthen our commitment to the neuroimaging area. While I will be stepping down from the Co-Editor-in-Chief role, I am confident that this new editorial configuration will continue to support the overall neuroinformatics discipline with greater efficiency.

During my tenure as a Co-Editor-in-Chief, I have been exceedingly pleased with the progress in the field, and the role that the journal has played in fostering this progress. *Neuroinformatics* has successfully been on the forefront of implementing publication practices that promote more reproducible and open scholarly communication, including introduction of the Information Sharing Statement,¹ Software Original Articles,² Data Original Articles,³ etc.

¹ Kennedy D. N. (2017). The information sharing statement grows some teeth. *Neuroinformatics*, 15(2):113–114.

² De Schutter E., Ascoli G. A., Kennedy D. N. (2009). Review of papers describing neuroinformatics software. *Neuroinformatics*, 7(4):211–2.

³ Kennedy, D. N., Ascoli, G. A., & Schutter, E. D. (2011). Next steps in data publishing. *Neuroinformatics*, 9(4), 317–320.

✉ David N. Kennedy
David.Kennedy@umassmed.edu

¹ University of Massachusetts Medical School, Worcester, MA, USA

So as a parting perspective of a soon to be ex-Co-Editor-in-Chief, let me expound upon a number of topics from these years:

1) The ‘Curse of Innovation’

Sometimes ‘innovation’ seems to be more valued than ‘utility’. This problem is seen both in manuscripts submission guidelines (for any journal) as well as in many grant funding guidelines. Particularly in cases where ground truth is not known, a common summary of the main point of many manuscripts and grants is often: “I can’t show that [my approach] is better, but I can argue that it’s different”. We, as a field, need to make sure that our innovations are also improvements, and to make sure that we couch the improvements relative to the work that has preceded them.

2) Best-case Scenario

The sub-discipline of neuroimaging, as a field, ought to offer one of the best ecosystems in which the advancements in technology can finally be brought to establish an end-to-end information management system (from data acquisition to published claim) that can be put into routine use. Massive amounts of neuroimaging data are being acquired; there are good data formats and standards; there are ample data and results hosting platforms freely available⁴; there is a large, but tractable, number of outcome measures (and tools to generate these) that are routinely used. Despite lots of investment in the neuroimaging informatics infrastructure, the majority of neuroimaging publications are not associated with a specific data archive, a precise analysis workflow specification, and complete results. There remains a ‘last mile’ problem, whereby the barriers to accomplishing these important reproducibility tasks remain too labor intensive for

⁴ Eickhoff, S., Nichols, T. E., Van Horn, J. D. & Turner, J. (2016). A. Sharing the wealth: Neuroimaging data repositories. *Neuroimage*, 124, 1065–1068.

routine use. One would think that the neuroimaging sub-field has the greatest potential to overcome these limitations, and, due to the size of the field, the most to gain from doing so.

3) Everything (at least in psychiatric neuroimaging) is Underpowered

Can we just admit that everything in psychiatric neuroimaging is underpowered?⁵ If we admit this, there are then a number of logical consequences that we know, but tend to forget. In the underpowered regime, our studies are more subject to false positive and false negative findings. As we tend to publish positive findings, these (sometimes false positive findings) aggregate in the literature and there is little opportunity to update/refine/dampen out these observations so they tend to persist.⁶ There are two main lines of solution to this type of problem: promote the publication of negative findings (in order to provide a counter observation to the false positives), and promote the pooling of data in order to create subsequent (less underpowered) observations in order to observe how the original findings evolve as a function of amount of data (i.e. number of subjects).

With respect to data pooling, a subsequent tenet is, therefore, that the funders of these original underpowered studies should be obligated to support the capability of pooling of these observations with future observations; not to do so dooms the initial result to be forever underpowered. But of course, it is routinely announced that data management costs are permitted by most funders. So the corollary of this tenet is that it is the research investigators themselves who should be obligated to support the future pooling of their observations; since not to do so dooms the initial result to be forever underpowered, and would constitute an irresponsible use of funds and a breach of the implied contract with the research study participant to maximize the value of their contribution. While some data pooling advocates are awaiting the time when the ‘Data Management Plan’ (or Resource Sharing Plan) becomes officially score-able (and actionable) in U.S. National Institutes of Health (NIH) grants, it can be argued that, to the savvy reviewer, it already is. Within the context of the NIH review, there is a provision for an explicit evaluation of ‘rigor and reliability’ in the context of the

‘Approach’ section review. I challenge more reviewers to ask themselves: “How can it be rigorous or reliable to not provide for the preservation and pooling of newly acquired data?” The NIH already provides for the evaluation of the ‘significance’ of a study. I challenge more reviewers to ask themselves: “How can a research result be ‘significant’ if provision is not made for the preservation and integration of that result?”

4) The 10% Argument

It has been argued (and I’ve not seen it refuted), that the ‘cost’ of data preservation may be expected to be roughly 10% of the cost of the primary data acquisition.⁷ For a given funding amount, to do (and preserve) 90% of an underpowered study that will generate a small number of potentially false-positive findings that can be integrated with other studies must be more valuable than doing 100% of an underpowered study that will generate a small number of potentially false-positive findings that will then be lost to future use. In general, funding agencies to do not want to become committed to devoting vast parts of their limited resources to the perpetual funding of infrastructure. The 10% preservation cost, when borne by the clinical and basic science proposals that incur the data acquisition costs, preserves the appearance of funding the critical science of our times, while also guiding resources to the needed infrastructure to preserve and amplify the investment.

5) Take Control of the Future!

The funding agencies only have so much money; the community cannot ask (or wait) for a new source of money to be thrown at solving the data preservation and research reproducibility issues. The community, encompassing each of the many roles that an individual plays, has to solve it with the resources they already have. Grant reviewers have to hold the grant applicants responsible for planning for the sustainability of their research activity; the grant agencies have to monitor and hold accountable their funding recipients to the sustainability of their research activities; the reviewers of grants and publications have to ask if this finding is being reported in a reproducible fashion; the journals, editors and publishers have to promote the best practices of reproducibility; we all have to be part of the solution to guarantee that all available resources are being used to foster a field of neuroscience that builds upon itself, and results in stronger, better results, faster.

⁵ Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.

⁶ Jennings R. G., Van Horn J. D. (2012). Publication bias in neuroimaging research: implications for meta-analyses. *Neuroinformatics*, 10(1):67–80.

⁷ Kennedy, D. N. (2014). Data persistence insurance. *Neuroinformatics*, 12(3):361–3.