LETTER

Controversy in statistical analysis of functional magnetic resonance imaging data

Emery N. Brown[a,b,1] and Marlene Behrmann[c]

To test the validity of statistical methods for fMRI data analysis, Eklund et al. (1) used, for the first time, large-scale experimental data rather than simulated data. Using resting-state fMRI measurements to represent a null hypothesis of no task-induced activation, the authors compare familywise error rates for voxel-based and cluster-based inferences for both parametric and nonparametric methods. Eklund et al.'s study used three fMRI statistical analysis packages. They found that, for a target familywise error rate of 5%, the parametric methods gave invalid cluster-based inferences and conservative voxel-based inferences.

Eklund et al. (1) attribute the invalid cluster-based inferences to the incorrect assumption of squared exponential structure in the spatial autocorrelation function of the parametric models. The authors suggest nonparametric methods as a more appropriate way to achieve targeted error rates, and conclude that statistical methods for fMRI data analysis should be validated. In addition, Eklund et al. state that their findings "question the validity of some 40,000 fMRI studies and may have a large impact on the interpretation of neuroimaging results" (1). This sentence from the Significance section of the original paper was picked up by the press and yielded the alarming negative headline that fMRI analyses produce incorrect results because of a bug in a widely used data analysis package (2–4). Eklund et al. revised their extrapolation regarding the implication of their findings in a correction to their article (5) and report that their analysis might apply to 3,500 rather than 40,000 fMRI studies (6). However, before this revision was published, the original statements created considerable debate about data analysis and the accuracy of fMRI findings (2–4).

The overstatements of the original paper and the subsequent media attention cast doubt on fMRI as a technique for studying brain function, and possibly even caused damage to the field of cognitive neuroscience (2–4). In PNAS, Cox et al. (7) and Kessler et al. (8) offer clarifications about the original paper and its revision. Eklund et al. have added their rejoinder (9). Several scientific points have now been mostly resolved. The remaining question is: What else can be learned from this controversy?

fMRI is a highly valued methodology for understanding brain function and its relationship to behavior. During the last 25 y, significant scientific advances have been made using this technique. To ensure continued progress, fMRI experimentalists want to be assured that the instruments, experimental protocols, and data analysis paradigms have been vetted by experts and work correctly. At the same time, experimentalists must be well informed about the fMRI process, and have a solid understanding of how to apply and interpret commonly used statistical methods (10–12). The ease of analysis afforded by some of the software programs belies the complexity of the methods. This ease of use does not release experimentalists from their responsibility to validate findings using established statistical principles (12, 13). Judicious use of nonparametric methods can, as Eklund et al. (1) suggest, improve the current analysis paradigm in certain cases. However, application of nonparametric methods cannot be the universal solution, nor did Eklund et al. suggest that it could be.

The current discussion shows that the validity of fMRI data analysis paradigms has not been uniformly established and needs continued in-depth investigation. fMRI is a complex process that involves biophysics, neuroananatomy, neurophysiology, and statistics (experimental design, statistical modeling, and data analysis). fMRI data have a low signal-to-noise ratio (14, 15). As a consequence, all of the biophysics, neurophysiology, and neuroanatomy that underlie fMRI should be used to design experiments, formulate statistical models, and analyze the data to increase the signal-to-noise ratio and information extraction. Achieving more accurate fMRI data analyses is a challenging interdisciplinary task that requires concerted collaborations among physicists, statisticians, and neuroscientists who, together, can question the current approaches more deeply and construct more accurate analysis methods.

In an ideal fMRI statistical analysis, the relationships among the voxels would take account of the spatial and temporal properties of the experiment and the scanner

[a]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; [b]Department of Brain and Cognitive Sciences, Institute for Medical Engineering and Science and Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213

Author contributions: E.N.B. and M.B. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: enb@neurostat.mit.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1705513114

thermal noise (16). The experiment's spatial and temporal properties are dictated by the physiological changes (neural activity, blood flow, and blood oxygenation levels) induced by the particular behavioral task and background physiological activity and anatomy (white matter, gray matter, the ventricles, and blood vessels) of the relevant brain regions. The ideal fMRI acquisition scheme would be accompanied by a characterization of these spatial and temporal processes so that the subsequent data analysis can correctly take them into account (16). Improving fMRI statistical methods must combine research to decipher the meaning/origins of the blood oxygen level-dependent signal with characterizations of the spatio-temporal properties of task-related activity, background physiological activity, and scanner properties. Sharing data and methods would greatly expedite validation (9).

BRAIN 2025, the report of the NIH Brain Initiative, recommends fostering interdisciplinary collaborations among neuroscientists, physicists, engineers, statisticians, and mathematicians to properly collect, analyze, and interpret the data that result from the development of new neuroscience tools (https://www.braininitiative.nih.gov/2025/). The current exchange identifies fMRI as an existing tool that is perfect for pursuing such a collaboration. A possible goal could be to increase fMRI signal-to-noise ratios so that the technique can be used reliably to make inferences about an individual subject in a given paradigm.

Developing statistical methods based on detailed modeling of the fMRI process opens the door to using more direct, informative inference paradigms based on estimated effect sizes, confidence intervals, and Bayesian posterior assessments rather than more indirect approaches based on significance tests and *P* values. Linking statistical methodology development and fundamental fMRI research is crucial for developing more accurate analysis methods, attributing accurate scientific interpretations to results, and ensuring the reliability and reproducibility of fMRI studies. These points have been made before. However, their significance has perhaps not been considered to the extent required.

Acknowledgments

We thank Bruce Rosen, Michael Tarr, and Larry Wald for helpful comments.

1 Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA* 113:7900–7905. Erratum in *Proc Natl Acad Sci USA* 113:E4929.
2 Crew B (2016) A bug in fMRI software could invalidate 15 years of brain research. Available at www.sciencealert.com/a-bug-in-fmri-software-could-invalidate-decades-of-brain-research-scientists-discover. Accessed April 10, 2017.
3 Murphy K (2016) Do you believe in god or is that a software glitch? Available at https://www.nytimes.com/2016/08/28/opinion/sunday/do-you-believe-in-god-or-is-that-a-software-glitch.html?_r=0. Accessed April 10, 2017.
4 Biello D (2016) Much of what we know about the brain may be wrong: The problem with fMRI. Available at ideas.ted.com/much-of-what-we-know-about-the-brain-may-be-wrong-the-problem-with-fmri/. Accessed April 10, 2017.
5 Eklund A, Nichols TE, Knutsson H (2016) Correction for Eklund et al. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA* 113:E4929.
6 Nichols T (2016) Bibliometrics of cluster inference. Available at blogs.warwick.ac.uk/nichols/entry/bibliometrics_of_cluster/. Accessed April 10, 2017.
7 Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA (2017) FMRI clustering and false positive rates. *Proc Natl Acad Sci USA* 114:E3370–E3371.
8 Kessler D, Angstadt M, Sripada C (2017) Reevaluating "cluster failure" using nonparametric control of false discovery rate. *Proc Natl Acad Sci USA* 114:E3372–E3373.
9 Eklund A, Nichols TE, Knutsson H (2017) Reply to Cox et al. and Kessler et al.: Data and code sharing is the way forward for fMRI. *Proc Natl Acad Sci USA* 114:E3374–E3375.
10 Poldrack RA, et al. (2017) Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci* 18:115–126.
11 Munafo MR, et al. (2017) A manifesto for reproducible science. *Nature Human Behaviour* 1:0021.
12 Kass RE, et al. (2016) Ten simple rules for effective statistical practice. *PLOS Comput Biol* 12:e1004961.
13 Kass RE, Ventura V, Brown EN (2005) Statistical issues in the analysis of neuronal data. *J Neurophysiol* 94:8–25.
14 Parrish TB, Gitelman DR, LaBar KS, Mesulam MM (2000) Impact of signal-to-noise on functional MRI. *Magn Reson Med* 44:925–932.
15 Welvaert M, Rosseel Y (2013) On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. *PLoS One* 8:e77089.
16 Wald LL, Rosseel Y (2017) Impacting the effect of fMRI noise through hardware and acquisition choices-impliations for controlling false positive rates. *Neuroimage*, 10.1016/j.neuroimage.2016.12.057.