doi:10.1093/brain/awaa095 BRAIN 2020: 143; 1–6 | e41

BRAIN
A JOURNAL OF NEUROLOGY

LETTER TO THE EDITOR

Reply: The influence of sample size and arbitrary statistical thresholds in lesion-network mapping

Alexander L. Cohen[1,2,3] and Michael D. Fox[2,4,5]

1 Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
2 Berenson-Allen Center for Non-Invasive Brain Stimulation and Division of Cognitive Neurology, Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
3 Computational Radiology Laboratory, Department of Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
4 Athinoula A. Martinos Centre for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA
5 Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Correspondence to: Alexander Li Cohen, MD, PhD
Department of Neurology
Boston Children's Hospital
300 Longwood Avenue
Boston, MA 02115, USA
E-mail: alexander.cohen2@childrens.harvard.edu

Sir,

We thank Sperber and Dadashi (2020) for their interest in our recent paper (Cohen *et al.*, 2019) and the techniques of lesion network mapping (Boes *et al.*, 2015) and coordinate network mapping (Darby *et al.*, 2019) more generally. In their letter, they highlight two methodological concerns: the size of the normative connectome and the threshold for defining a region as 'functionally connected' (Sperber and Dadashi, 2020). At present, both parameters represent 'researcher degrees-of-freedom' for which there is no accepted standard or single correct answer (Simmons *et al.*, 2011; Wicherts *et al.*, 2016). As such, these parameters have varied across our published studies, raising concern that they were chosen in an arbitrary manner. However, when organized chronologically, trends in this variability are apparent (Table 1). Over time, we have moved to a larger connectome, increased statistical power, and increased statistical rigor. Further, many of our papers have used multiple thresholds or even different normative connectomes to ensure our conclusions were robust to these parameter choices.

However, this variability raises good questions about which parameters investigators should choose if they seek to implement lesion network mapping in their own laboratories. As such, we thank Sperber and Dadashi for the opportunity to discuss these issues in greater detail.

Variability in normative connectome size

Our earliest lesion network mapping studies used resting state functional connectivity data from 98 healthy young controls to identify connectivity with each lesion location (Boes *et al.*, 2015; Fischer *et al.*, 2016; Laganiere *et al.*, 2016; Darby *et al.*, 2017; Fasano *et al.*, 2017). We chose this connectome dataset for convenience, as it was high-quality, locally available, and part of a large ongoing data collection initiative (Holmes *et al.*, 2015).

To combine connectivity measures for these 98 subjects, we used a random effects analysis to generate a 'T map' that reflected the statistical probability of connectivity with the lesion location, accounting for variance across subjects. An alternative we considered was to average connectivity maps across the 98 subjects, generating an average '*r*-map' of Pearson's correlation coefficients. This *r*-map would be consistent across connectomes with differing number of subjects, but it ignores the variability across subjects and does not allow one to easily assess the statistical significance of identified connections represented in the T map.

As this connectome dataset grew in size, we upgraded our network mapping pipeline to take advantage of an expanded 1000-subject connectome (Darby *et al.*, 2018*a*, *b*; Joutsa

Advance access publication May 4, 2020

Table 1 Chronological list of our lesion and coordinate network mapping studies

Study	Symptom of interest	Size of connectome	Sensitivity analysis			Specificity analysis
			Threshold *P*-level	Primary *t*-level threshold	Additional *t*-level thresholds reported	Were non-threshold dependent analyses performed?
Boes *et al.* (2015)	Multiple deficits	98	0.00005 (uncorrected)	4.25	Yes, multiple	No
Fischer *et al.* (2016)	Coma	98	0.00005 (uncorrected)	4.25	Yes, 9, 3.75, 2	Yes
Laganiere *et al.* (2016)	Hemichorea-hemiballismus	98	0.00005 (uncorrected)	4.25	Yes, multiple	Yes
Darby *et al.* (2017)	Capgras delusions	98	0.00005 (uncorrected)	4.25	Yes, multiple	Yes
Fasano *et al.* (2017)	Freezing gait	98	0.05 (uncorrected)	2	Yes, 4.25, 3	Yes
Darby *et al.* (2018*a*)	Criminal behaviour	98 and 1000	10^{-17} (uncorrected)	12	No	Yes
Darby *et al.* (2018*b*)	Free will	1000	10^{-6} (uncorrected)	5	No	Yes
Joutsa *et al.* (2018*a*)	Parkinson syndrome	1000	10^{-11} (uncorrected)	7	No	Yes
Joutsa *et al.* (2018*b*)	Tremor relief	1000	10^{-6} (FWE-corrected)	7	No	No
Joutsa *et al.* 2019	Holmes tremor	1000	0.05 (FWE-corrected)	5	No	Yes
Darby *et al.* (2019)	Neurogenerative disease	1000	0.05 (FWE-corrected)	5	Yes, 10, 7	Yes
Corp *et al.* (2019)	Cervical dystonia	1000	10^{-6} (FWE-corrected)	7	No	Yes
Weil *et al.* (2019)	Parkinson's dementia	1000	10^{-6} (FWE-corrected)	7	No	Yes
Ferguson *et al.* (2019)	Amnesia	1000	10^{-6} (FWE-corrected)	7	No	Yes
Burke *et al.* (2020)	Migraine	1000	10^{-6} (FWE-corrected)	7	No	Yes
Kim *et al.* (2019)	Hallucinations	1000	10^{-6} (FWE-corrected)	7	Yes, 12, 5	Yes
Padmanabhan *et al.* (2019)	Depression	1000	Leave-one-dataset-out cross validation of unthresholded maps			Yes
Cohen *et al.* (2019)	Prosopagnosia	1000	10^{-11} (FWE-corrected)	9	Yes, multiple	Yes
Snider *et al.* (2020)	Consciousness	1000	Logistic regression of unthresholded maps			Yes

Studies are presented in order of when they were conducted, not necessarily when they were published. All studies that required a *t*-value threshold examined multiple thresholds, and many included these analyses in the final manuscript. All studies except one brief communication included analyses that were not dependent on a *t*-value threshold.

et al., 2018*a*, *b*, 2019; Cohen *et al.*, 2019; Corp *et al.*, 2019; Ferguson *et al.*, 2019; Kim *et al.*, 2019; Padmanabhan *et al.*, 2019; Weil *et al.*, 2019; Burke *et al.*, 2020; Snider *et al.*, 2020). Our first lesion network mapping paper using this expanded connectome compared results to those obtained using our original 98 subject cohort (Darby *et al.*, 2018*a*). Lesion networks generated using both normative connectomes were very similar, with nearly identical spatial distributions of positive and negative connectivity (Darby *et al.*, 2018*a*). Average r maps using the 98 and 1000 subject connectomes were also nearly identical, consistent with other work showing that group connectivity estimates stabilize above 150–200 subjects (Cremers *et al.*, 2017; Dansereau *et al.*, 2017; Cui and Gong, 2018; Turner *et al.*, 2018; Varoquaux, 2018). However, *t*-values were much larger using our larger connectome, reflecting the increase in statistical power. When seeking to highlight only the most significant connections, a higher *t*-value threshold was required (Darby *et al.*, 2018*a*).

Taking advantage of this increased statistical power, we found that a range of statistically significant *t*-value thresholds could be used, as has been done for years with functional MRI activation results, to focus on only the strongest findings or a broader network of slightly less significant findings (Darby *et al.*, 2018*a*, *b*; Joutsa *et al.*, 2018*a*, 2019; Cohen *et al.*, 2019; Corp *et al.*, 2019; Ferguson *et al.*, 2019; Kim *et al.*, 2019; Padmanabhan *et al.*, 2019; Weil *et al.*, 2019; Burke *et al.*, 2020; Snider *et al.*, 2020). We also began using improved statistical methods, including family-wise error rate (FWE) multiple comparison correction, non-threshold dependent specificity analyses, and non-parametric permutation testing for assessing statistical significance (Winkler *et al.*, 2014; Eklund *et al.*, 2016).

Now, efforts are underway to generate human connectomes from 100 000 subjects (Miller *et al.*, 2016). Using this connectome for lesion network mapping would again result in increased *t*-values, to the point that almost every voxel would likely be considered 'statistically significant'. As more and larger connectomes become available, it will be important to re-examine approaches for combining connectivity maps across large cohorts of subjects; and the notion of a 'statistically significant' connection may become irrelevant. Instead, metrics such as the 'top 25%' of connections may become more useful, an approach that has been used in graph theory analyses (Bullmore and Sporns, 2009; Power *et al.*, 2011).

Variability in the t-value threshold

As noted above, we have used different *t*-value thresholds across different papers using the 1000 subject connectome to define a region as 'connected'. This threshold variability is not unique to lesion network mapping, but is an issue for many types of functional connectivity analyses (Reijneveld *et al.*, 2007; Stam and Reijneveld, 2007; Buckner *et al.*, 2009; Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Wang *et al.*, 2010; Power *et al.*, 2011; Fornito *et al.*, 2015).

Given the difficulty inherent in choosing a threshold, it is worth asking why we use a threshold at all. When we first introduced lesion network mapping (Boes *et al.*, 2015), our primary goal was to show that the method added value compared to traditional lesion mapping. Using a threshold for lesion network maps (connected or not), facilitated comparison to traditional lesion mapping (lesioned or not). By binarizing lesion network maps, we were able to use the same statistical tools used in traditional lesion mapping and directly compare the two results (Boes *et al.*, 2015).

A second motivation for using a threshold is that it can help simplify result interpretation, especially when presenting a new technique or concept. Concluding that '95% of lesion locations causing symptom X are connected to region Y' is easy to understand. This result also means that connectivity with region Y (at the same threshold) defines a brain network that encompasses 95% of lesion locations causing symptom X. While this number is threshold-dependent and could change to 98% with a lower *t*-threshold or 90% with a higher *t*-threshold, choosing a threshold can help illustrate the lesion network mapping concept.

Finally, there are statistical advantages to binarizing a lesion network map. Because of the spatial autocorrelation inherent to seed-based connectivity analyses, a lesion network will have extremely high 'connectivity' at the lesion location itself. This results in high variance at each lesion location when a cohort of different lesion locations causing the same symptom is combined. As such, using a one-sample *t*-test to combine these maps, as suggested by Sperber and Dadashi (2020), is biased away from the lesion sites themselves, which are likely important locations within the affected network (Rorden and Karnath, 2004). By thresholding and binarizing lesion network maps prior to combining them, commonalities across a group of lesions can be identified without penalizing the lesion locations themselves.

Given that there is potential value in using a *t*-value threshold, yet multiple different thresholds could reasonably be used and values will change with the size of the connectome, how does one choose a threshold? Below we provide some recommendations:

Methodological recommendations for lesion network mapping

First, use normative connectomes large enough to provide stable group-level connectivity estimates, ideally larger than 150 subjects. If possible, we recommend trying to match the size of the connectome used in prior work, as this allows for more direct comparisons across studies.

Second, ensure any conclusions based on threshold-dependent analyses hold true across a range of different thresholds. We have used this approach in almost all recent papers, including Cohen *et al.* (2019), which is demonstrated visually in Fig. 1. Consistent with this principle, results that do not hold true across different thresholds should be made

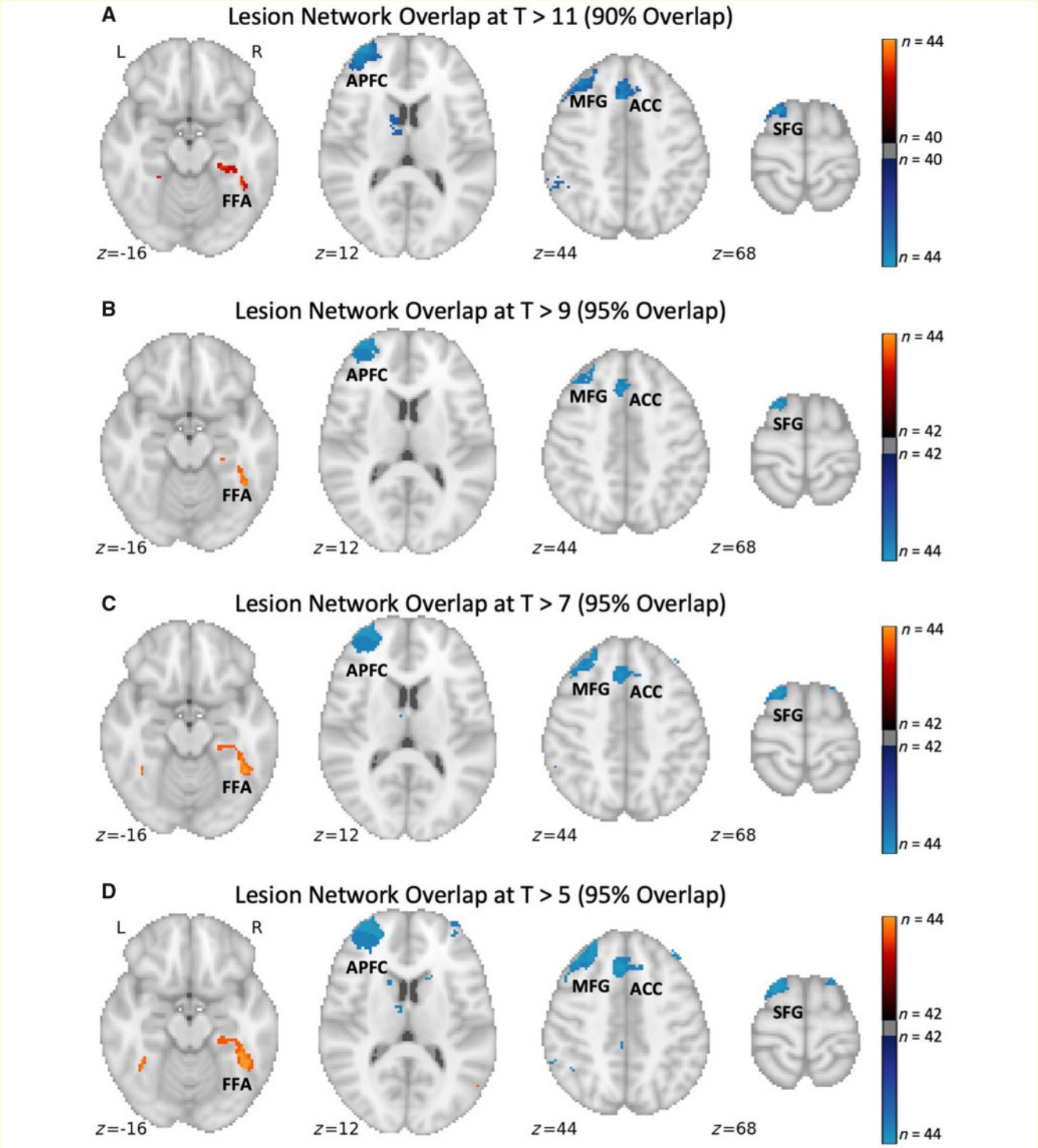

Figure 1 Lesion network mapping of prosopagnosia is consistent across a range of thresholds. Regions of peak lesion network overlap in the fusiform face area (FFA), anterior prefrontal cortex (APFC), middle frontal gyrus (MFG), anterior cingulate cortex (ACC), and superior frontal gyrus (SFG) are consistent across a range of *t*-value thresholds. At higher thresholds the percentage overlap decreases, but the topography of the lesion network and central findings remain unchanged. L = left; R = right.

explicit and de-emphasized relative to results that are more robust, an approach we have also used in prior papers (Fischer *et al.*, 2016; Laganiere *et al.*, 2016; Fasano *et al.*, 2017).

Third, for simplicity, we often use a single threshold for presenting our 'primary' lesion network overlap results, and we include results at other thresholds as secondary analyses. When deciding which threshold to use for the 'primary'

analysis, we choose the highest threshold that still demonstrates the relevant finding, as this will improve specificity. For example, in Cohen *et al.* (2019) we could have chosen a lower *t*-threshold that was still 'statistically significant' and obtained the same result (Fig. 1); however, the resulting prosopagnosia network would have been slightly less specific to prosopagnosia lesions. This principle is responsible for much of the variation in *t*-value threshold across our recent studies (Table 1).

Fourth, complement any threshold-dependent analysis with a threshold-independent analysis. In the vast majority of our papers we use unthresholded maps for assessing the specificity of lesion network mapping results (Table 1). For example, the conclusion that '95% of lesions causing symptom X are connected to region Y' may be threshold dependent, but the conclusion that 'lesions causing symptom X are significantly more connected to region Y then lesions not causing symptom X' is not dependent on this parameter.

Fifth, whenever possible, analyse data from multiple independent cohorts to test whether lesion network mapping results, which may have used an arbitrary threshold, are reproducible across cohorts, independent of thresholds, e.g. Fig. 8 in Cohen *et al.* (2019).

Finally, if one is uncomfortable with choosing a threshold for lesion network analyses, one can avoid using a threshold at all, a choice we have made in some recent lesion network mapping reports (Padmanabhan *et al.*, 2019; Snider *et al.*, 2020). One study used logistic regression to identify connections that covaried with the severity of loss of consciousness (Snider *et al.*, 2020) while another focused on connections that varied significantly with the presence or severity of depression (Padmanabhan *et al.*, 2019). This latter study is also notable for confirming reproducibility across five independent datasets (Padmanabhan *et al.*, 2019).

Future considerations

We anticipate that as the field of lesion network mapping and coordinate-based network mapping grows, there will be parallel growth in methodological studies seeking to identify the best approaches, similar to what has occurred for the field of functional connectivity in general (Ciric *et al.*, 2017; Murphy and Fox, 2017; Parkes *et al.*, 2018).

In summary, we thank Sperber and Dadashi for their interest in lesion network mapping and for the opportunity to discuss these methodological issues. There is no simple, or single, solution to the question of how 'best' to perform lesion network mapping. Presently, we feel that thresholding lesion connectivity maps provides some utility; however, we look forward to methodological advances that may circumvent this researcher degree of freedom.

Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this work.

Funding

A.L.C. was supported by NIH T32MH112510 and the Child Neurology Foundation. M.D.F. was supported by the Sidney R. Baer, Jr. Foundation, NIH K23NS083741, the Dystonia Foundation, and the Nancy Lurie Marks Foundation.

Competing interests

The authors report no competing interests.

References

Boes AD, Prasad S, Liu H, Liu Q, Pascual-Leone A, Caviness VS, et al. Network localization of neurological symptoms from focal brain lesions. Brain 2015; 138: 3061–75.

Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J Neurosci 2009; 29: 1860–73.

Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 2009; 10: 186–98.

Burke MJ, Joutsa J, Cohen AL, Soussand L, Cooke D, Burstein R, et al. Mapping migraine to a common brain network. Brain 2020; 143: 541–53.

Ciric R, Wolf DH, Power JD, Roalf DR, Baum GL, Ruparel K, et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 2017; 154: 174–87.

Cohen AL, Soussand L, Corrow SL, Martinaud O, Barton JJS, Fox MD. Looking beyond the face area: lesion network mapping of prosopagnosia. Brain 2019; 142: 3975–90.

Corp DT, Joutsa J, Darby RR, Delnooz CCS, van de Warrenburg BPC, Cooke D, et al. Network localization of cervical dystonia based on causal brain lesions. Brain 2019; 142: 1660–74.

Cremers HR, Wager TD, Yarkoni T. The relation between statistical power and inference in fMRI. PLoS One 2017; 12: e0184923.

Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. Neuroimage 2018; 178: 622–37.

Dansereau C, Benhajali Y, Risterucci C, Pich EM, Orban P, Arnold D, et al. Statistical power and prediction accuracy in multisite resting-state fMRI connectivity. Neuroimage 2017;

Darby RR, Horn A, Cushman F, Fox MD. Lesion network localization of criminal behavior. Proc Natl Acad Sci USA 2018a; 115: 601–6.

Darby RR, Joutsa J, Burke MJ, Fox MD. Lesion network localization of free will. Proc Natl Acad Sci USA 2018b; 115: 10792–7.

Darby RR, Joutsa J, Fox MD. Network localization of heterogeneous neuroimaging findings. Brain 2019; 142: 70–9.

Darby RR, Laganiere S, Pascual-Leone A, Prasad S, Fox MD. Finding the imposter: brain connectivity of lesions causing delusional misidentifications. Brain 2017; 140: 497–507.

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; 113: 7900–5.

Fasano A, Laganiere SE, Lam S, Fox MD. Lesions causing freezing of gait localize to a cerebellar functional network. Ann Neurol 2017; 81: 129–41.

Ferguson MA, Lim C, Cooke D, Darby RR, Wu O, Rost NS, et al. A human memory circuit derived from brain lesions causing amnesia. Nat Commun 2019; 10(1): 3497.

Fischer DB, Boes AD, Demertzi A, Evrard HC, Laureys S, Edlow BL, et al. A human brain network derived from coma-causing brainstem lesions. Neurology 2016; 87: 2427–34.

Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. Nat Rev Neurosci 2015; 16: 159–72.

Holmes AJ, Hollinshead MO, O'Keefe TM, Petrov VI, Fariello GR, Wald LL, et al. Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. Sci Data 2015; 2: 1–16.

Joutsa J, Horn A, Hsu J, Fox MD. Localizing parkinsonism based on focal brain lesions. Brain 2018a; 141: 2445–56.

Joutsa J, Shih LC, Fox MD. Mapping holmes tremor circuit using the human brain connectome. Ann Neurol 2019; 86: 812–20.

Joutsa J, Shih LC, Horn A, Reich MM, Wu O, Rost NS, et al. Identifying therapeutic targets from spontaneous beneficial brain lesions. Ann Neurol 2018b; 84: 153–7.

Kim NY, Hsu J, Talmasov D, Joutsa J, Soussand L, Wu O, et al. Lesions causing hallucinations localize to one common brain network. Mol Psychiatry 2019.

Laganiere S, Boes AD, Fox MD. Network localization of hemichorea-hemiballismus. Neurology 2016; 86: 2187–95.

Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 2016; 19: 1523–36.

Murphy K, Fox MD. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. Neuroimage 2017; 154: 169–73.

Padmanabhan JL, Cooke D, Joutsa J, Siddiqi SH, Ferguson M, Darby RR, et al. A human depression circuit derived from focal brain lesions. Biol Psychiatry 2019; 86: 749–58.

Parkes L, Fulcher B, Yücel M, Fornito A. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. Neuroimage 2018; 171: 415–36.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, et al. Functional Network Organization of the Human Brain. Neuron 2011; 72: 665–78.

Reijneveld JC, Ponten SC, Berendse HW, Stam CJ. The application of graph theoretical analysis to complex networks in the brain. Clin Neurophysiol 2007; 118: 2317–31.

Rorden C, Karnath H-O. Using human brain lesions to infer function: a relic from a past era in the fMRI age? Nat Rev Neurosci 2004; 5: 813–9.

Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 2010; 52: 1059–69.

Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 2011; 22: 1359–66.

Snider SB, Hsu J, Darby RR, Cooke D, Fischer D, Cohen AL, et al. Cortical lesions causing loss of consciousness are anticorrelated with the dorsal brainstem. Hum Brain Mapp 2020; 41: 1520–31.

Sperber C, Dadashi A. The influence of sample size and arbitrary statistical thresholds in lesion-network mapping. Brain 2020; 143: e40.

Stam CJ, Reijneveld JC. Graph theoretical analysis of complex networks in the brain. Nonlinear Biomed Phys 2007; 1: 1–3.

Turner BO, Paul EJ, Miller MB, Barbey AK. Small sample sizes reduce the replicability of task-based fMRI studies. Commun Biol 2018; 1: 62.

Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage 2018; 180: 68–77.

Wang J, Zuo X, He Y. Graph-based network analysis of resting-state functional MRI. Front Syst Neurosci 2010; 4: 16.

Weil RS, Hsu JK, Darby RR, Soussand L, Fox MD. Neuroimaging in Parkinson's disease dementia: connecting the dots. Brain Commun 2019; 1: 1–17.

Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen M. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid P-hacking. Front Psychol 2016; 7: 1832.

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage 2014; 92: 381–97.