Cluster failure:Why fMRI inferences for spatial extent have inflated false positive rates

Anders Eklund[1,2,3], Thomas E. Nichols[4], Hans Knutsson[1,2]

[1]Department of Biomedical Engineering, Linköping University, Sweden
[2]Center for Medical Image Science and Visualization (CMIV), Linköping University, Sweden
[3]Department of Computer and Information Science, Linköping University, Sweden
[4]Department of Statistics, University of Warwick, Coventry, UK

Functional magnetic resonance imaging (fMRI) is 25 years old, yet surprisingly its most common statistical methods have not been validated using real data. Here, we used resting-state fMRI data from 396 healthy controls to conduct 2 million task group analyses. Using this null data with different experimental designs, we estimate the incidence of significant results while controlling for multiple testing. In theory, we should find 5% false positives (for a significance threshold of 5%) but instead we found that the most common software packages for fMRI analysis (SPM, FSL, AFNI) can result in false positive rates of up to 70% for cluster inference. These results question the validity of some 40,000 fMRI studies and may have a large impact on the interpretation of neuroimaging results.

Methods

- Downloaded 396 resting state fMRI datasets from the 1000 functional connectomes project [1], Beijing 198 datasets, Cambridge 198 datasets
- Analyzed all 396 datasets in SPM, FSL and AFNI, scripts available at https://github.com/wanderine/ParametricMultisubjectfMRI
- Used 4 activity paradigms with this resting data, 2 block based (B1, B2) and 2 event related (E1, E2)
- Varied smoothing between 4 and 10 mm FWHM
- Ran 1,000 random group analyses (two-sample and one-sample t-tests) for each parameter combination
- Tested voxel inference as well as cluster inference, used 2 cluster defining thresholds (p = 0.01 and p = 0.001)
- Tested SPM, FSL FLAME1, FSL OLS, AFNI 3dttest, AFNI 3dMEMA and a non-parametric permutation test
- Familywise error (FWE) rate estimated as number of significant results (p = 0.05, corrected) divided by number of group analyses (1,000)

Results

Estimated FWE rates for cluster inference and two-sample t-tests, for the Beijing datasets, are given in Figure 1 (Cambridge results were similar). Corresponding results for voxel inference are given in Figure 2. Note how all permutation tests are within the 95% confidence interval; across all simulations only some one-sample t-tests generated invalid permutation tests, attributable to failures of the symmetry assumption.

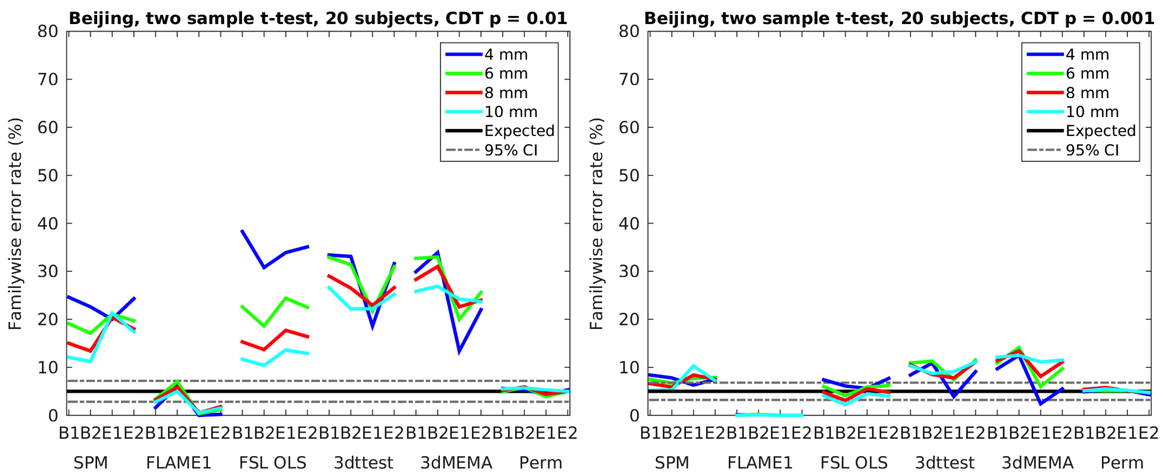

Figure 1

The parametric statistical methods work well, if conservatively, for voxel inference, but not for cluster inference. The main problems for cluster inference are that the data have a non-Gaussian spatial autocorrelation function (Figure 2), and that the spatial smoothness is not constant in the brain (Figure 3).

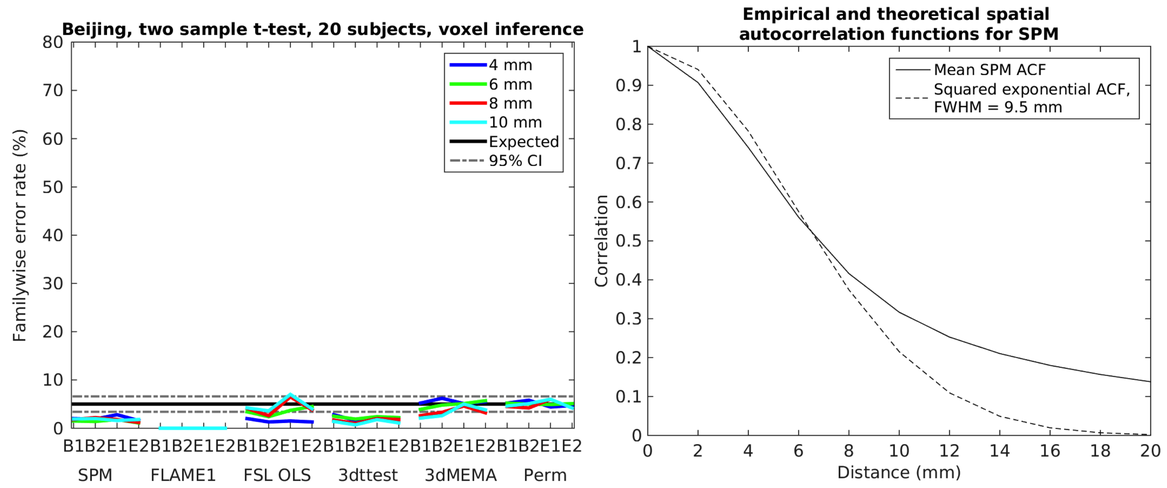

Figure 2

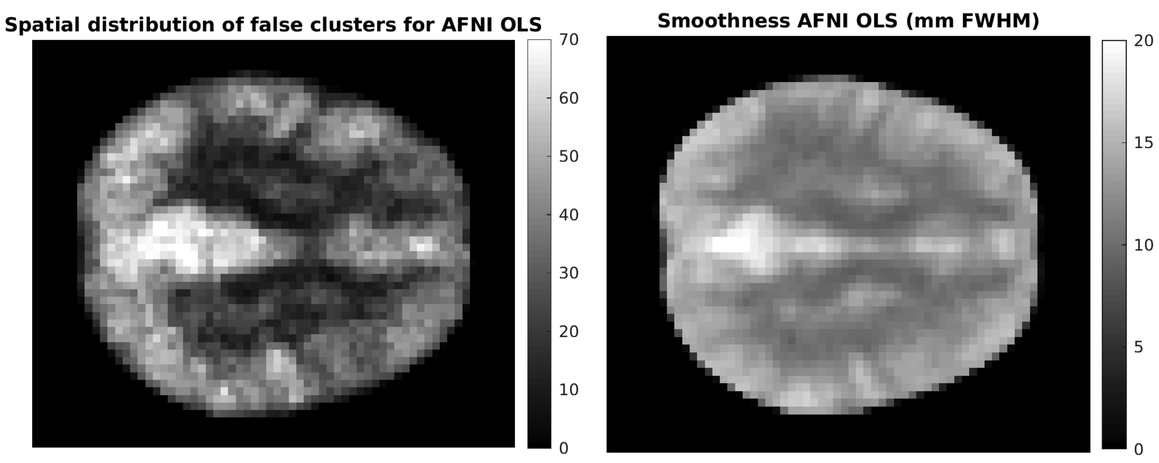

Figure 3

Figure 4 shows the ratio between non-parametric and parametric p-values for activity clusters obtained from four task datasets from OpenfMRI [2]. The non-parametric cluster p-value is always larger compared to the parametric p-value, especially for a cluster defining threshold of p = 0.01.

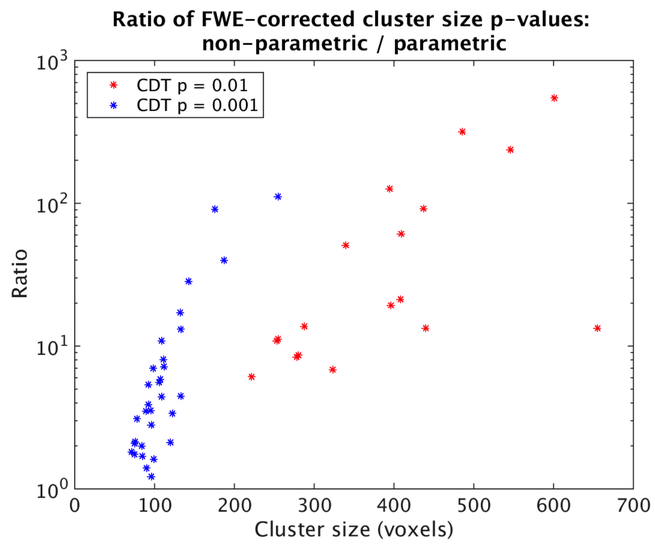

Figure 4

Discussion

Using mass empirical analyses with task-free fMRI data we have found that the parametric statistical methods used for group fMRI analysis with the packages SPM, FSL and AFNI can produce FWE-corrected cluster p-values that are erroneous, being spuriously low and inflating statistical significance. This calls into question the validity of countless published fMRI studies based on parametric cluster-wise inference. See the full paper [3] for all the details.

References

[1] Biswal, B. et al., "Toward discovery science of human brain function," *PNAS*, vol. 107, pp. 4734–4739, 2010.

[2] Poldrack, R. et al., "Toward open sharing of task-based fMRI data: the OpenfMRI project," *Frontiers in Neuroinformatics*, vol. 7, no. 12, 2013.

[3] A. Eklund, T. Nichols, and H. Knutsson, "Cluster failure: why fMRI inferences for spatial extent have inflated false positive rates," *PNAS*, 2016.

Organization for Human Brain Mapping (OHBM), Geneva, Switzerland, 2016