# False Positives in Imaging Genetics Using Nonstationary Cluster-Size Inference

Matt Silver[1], Giovanni Montana[1], Thomas Nichols[2]

1. Imperial College Statistics, London, United Kingdom.  2. University of Warwick Statistics, Coventry, United Kingdom.

## Introduction

Imaging genetic studies present a particular statistical challenge due to the high dimensionality of each domain's data. Specifically some issues have been raised about the adequate control of false positives in imaging genetics [1]. A previous study assessed the effect of a set of 'null' SNPs on brain structure and function [2]. Using voxel-wise tests, they found that the empirical family-wise error rate (FWE) was at or below 5%, indicating that false positive risk was controlled.

Statistical tests based on the size of significant clusters of contiguous voxels are also widely used, due to their typically increased statistical power over voxel-wise tests [3]. With VBM (voxel-based morphometry) data, an adjustment for non-stationarity is made to account for variations in smoothness across the brain. While non-stationary RFT tests have been validated with simulations in small datasets [4], they have not yet been evaluated with the large datasets typical to imaging genetics.

## Methods

Real image and genotype data on 181 subjects with mild cognitive impairment was obtained from the Alzheimer's Disease Neuroimaging Initiative. 700 null SNPs were selected from chromosome 3, a chromosome with no well-known AD risk genes, specifically not APOE, PSEN1, PSEN2 or SORL1. The 700 were selected to satisfy minimum allele frequency greater than 5%, and negative Hardy-Weinberg Equilibrium test at $P<0.05/700$.

We used standard 'optimized' VBM preprocessing (SPM5), with 6 mm and 12 mm Gaussian kernel smoothing, and cluster-forming thresholds ($u_c$) corresponding to α 0.01, 0.05 and 0.001 uncorrected. Non-stationary cluster size inference was used with a nominal rejection rate of 0.05 familywise corrected.

A total of 700 SPM analyses were fitted, and the empirical familywise error rate determined by the proportion of the 700 that had one or more (necessarily false) positives. To rule out any possible true association, we also permuted the SNP data 10 times, running an additional 7000 analyses.

## Results

Figure 1 shows the image of local FWHM smoothness estimated from the 181 subjects. The lack of stationarity is evident, with FWHM smoothness varying by over a factor of 2 from the smooth to rough regions of cortex. Table 1 shows the empirical false positive rates; as expected, application of a *stationary* cluster-size test resulted in false positive rates far in excess of the nominal 5%, especially with 6mm smoothing. Unexpected, however, was that *nonstationary* cluster-size tests only controlled false positives for the case of 12mm

| | | 6mm smoothing | | 12mm smoothing | | Meyer-Lindenberg et al. (2008)[2] | |
| | $α_c$ | Observed(%) | Permuted(%)[1] | Observed(%) | Permuted(%)[1] | Observed(%) | Permuted(%)[3] |
|---|---|---|---|---|---|---|---|
| *t*-tests | | | | | | | |
| *cluster size* | | | | | | | |
| stationary | 0.001 | 10.7 | 9.2±1.1 | 3.4 | 3.6±0.7 | *stationary test not supposed to work & doesn't* | |
| | 0.01 | 23.3 | 24.8±2.2 | 9.4 | 9.4±1.5 | | |
| | 0.05 | 47.4 | 46.7±1.8 | 20.7 | 18.6±1.5 | | |
| non-stationary | 0.001 | 10.0 | 8.1±1.0 | 3.9 | 3.8±0.8 | *nonstationary supposed to work but doesn't!* | |
| | 0.01 | 19.9 | 21.2±1.6 | 9.1 | 9.3±1.5 | | |
| | 0.05 | 45.0 | 43.2±2.2 | 20.1 | 18.4±1.6 | | |
| *voxel-wise* | - | 3.4 | 2.7±0.6 | 3.0 | 2.7±0.6 | 1.9 | 1.1±0.4 |

(header spanning: "Rejection Rates")

**Table 1**. Empirical cluster size test FWE false positive rates, based on 700 null SNPs before ("Observed") and after permutation ("Permuted"). Nominal 5% error control is only attained at the highest smoothness level and highest cluster-forming threshold. Voxel-wise FWE false positive rates are also shown, which exhibit good control.
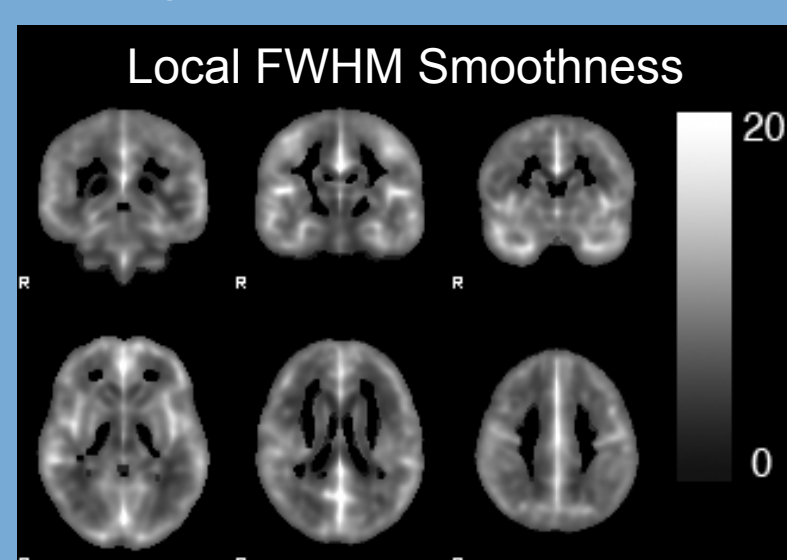


**Figure 1. Local FWHM** on 12mm-smoothned VBM data. Substantial variation in FWHM (3.8mm min, 27.7mm max) demonstrates the importance of correcting for nonstationarity.
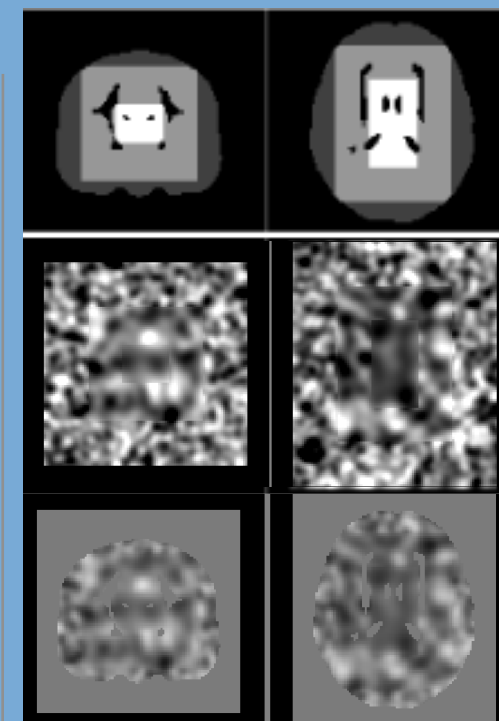


**Figure 2**. Simulation of nonstationary Gaussian images. The top row shows the pattern of nonstationarity simulated; for '6mm' smoothness setting outer-to-inner regions have 4, 6, & 9 mm FWHM, respectively, and for the '12mm' smoothness setting the regions have 8, 12 and 18 mm FWHM smoothness. The middle row shows the three different levels of smoothness assembled 'cookie cutter' fashion; the discontinuity at the boundaries of smoothness are evident. The bottom row shows the final simulated nonstationary data, after a small (1.5mm) smoothing to eliminate these discontinuities, and application of the gray matter analysis mask.

**Table 2**. Monte Carlo cluster size test FWE false positive rates. Using the very same null SNPs as with the real data, Gaussian nonstationary data was simulated and fit. Nominal 5% error control is only attained at both low and high smoothness, and for cluster-forming thresholds 0.001 and 0.01.

| | | 6mm smoothing | | 12mm smoothing | |
| | $α_c$ | Stationary | Non-stationary[1] | Stationary | Non-stationary[2] |
|---|---|---|---|---|---|
| *t*-tests | | | *Again, stationary test doesn't work, as expected* | | |
| *cluster size* | | | | | |
| stationary | 0.001 | 0.5 | 5.5 | 2.3 | 8.2 |
| | 0.01 | 1.0 | 11.7 | 1.6 | 13.0 |
| | 0.05 | 13.0 | 35.9 | 6.9 | 21.9 |
| non-stationary | 0.001 | 0.6 | 1.1 | 2.7 | 2.6 |
| | 0.01 | 0.9 | 0.3 | 1.4 | 1.9 |
| | 0.05 | 9.3 | 10.7 | 6.4 | 7.4 |
| *voxel-wise* | | *With simulated data, now nonstationary does work!* | | | |
| FWE | - | 3.7 | 3.6 | 5.1 | 4.3 |
| FDR | - | 4.7 | 4.0 | 2.9 | 2.7 |

(header spanning: "Rejection Rates")

12mm smoothing with a cluster-forming threshold of 0.001. Otherwise tests were very anticonservative (i.e. invalid), with rejection rates ranging from 9.1- 45%.

As this contradicted previous work showing the non-stationary cluster-size test to be valid at $u_c$ corresponding to 0.01, we ran simulations with both stationary and non-stationary noise (Figure 2 shows the pattern of nonstationarity induced). Table 2 shows the results using simulated data, showing that false positive rates are controlled for both 0.001 and 0.01 cluster-forming threshold when the non-stationary cluster size-test is applied to nonstationary data. Further simulations (not shown) with more complicated patterns of nonstationarity (based on observed FWHM image), and with non-Gaussian simulated noise images also showed good performance.

## Conclusions

Our results suggest that any study—not just genetics studies—using VBM and RFT cluster-size tests should use great caution with any cluster forming threshold less than 0.001, as there is likely inflated false positive risk. Nonparametric permutation, of course, is valid for any threshold and should be consulted as a reference when RFT results are in question.

## References

1. Meyer-Lindenberg, A. (2006), 'Intermediate phenotypes and genetic mechanisms of psychiatric disorders', Nature Reviews Neuroscience, vol. 7, pp. 818-827.
2. Meyer-Lindenberg, A. (2008), 'False positives in imaging genetics', NeuroImage, vol. 40, no. 2, pp. 655-661.
3. Friston, K. (1996), 'Detecting activations in pet and fMRI: Levels of inference and power', NeuroImage, vol. 4, pp. 223-235.
4. Hayasaka, S. (2004), 'Nonstationary cluster-size inference with random field and per- mutation methods', NeuroImage, vol. 22, no. 2, pp. 676-687.