Comparison of the Signal-to-Noise Sensitivity of Different False Discovery Rate Methods

M. Murugavel[1], J. M. Sullivan[1], K. G. Helmer[2]

[1]Dept. of Mechanical Engineering, Worcester Polytechnic Institute, Worcester, MA, United States, [2]Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA, United States

Introduction:

Controlling the False Discovery Rate[1](FDR) has become a popular method of correcting for multiple comparisons in fMRI data analysis. These methods differ in their assumptions on the underlying correlation structure between fMRI voxels. These methods control the proportion of false positives relative to the number of activations to within a user specified value, q($0<q<1$); such that $E(FDR) \leq q$. In this study, we compare five algorithms[1-5] using a p-value map generated from experimental data. Given the large range in the number of voxels deemed to reject the null hypothesis, we have undertaken a study of the response of various methods to the data signal-to-noise ratio using a simulated data set that includes the effects of machine and physiological noise.

Method:

The experimentally-derived data set consisted of a 64×64, 28 slice, 120 TR, multi-epoch, data set originating from a standard block-design hand-flex task. The original image intensity time-course data was correlated on a pixel-by-pixel basis to a paradigm waveform. The resulting correlation coefficients were converted to p-values. This p-value array was passed to each of the five FDR algorithms using a standard value of 0.05 for the level of control parameter, q. The simulated fMRI dataset was created from a single volume of a 64×64, 28 slice image volume taken from the imaging study that produced the experimental p-value array. This imaging volume was replicated 120 times and a 764 voxel ROI in slices 16-20 was chosen for the region to which simulated signal would be added. Different simulated datasets were created for different levels of signal intensity for the ROI at 0.5, 0.8, 0.9, 2, and 3% of baseline signal levels. Physiological noise at about 3 % of base level described by sinusoidal function with 72 beats/minute (average human heart rate) and 14 cycles/minute (average human breathing rate) were introduced for the whole brain region. A 1% overall random noise over the base intensity was also added to the entire volume. The noise levels were chosen such that correlation coefficients from the simulated data were on the order of those for experimental data (<0.7).

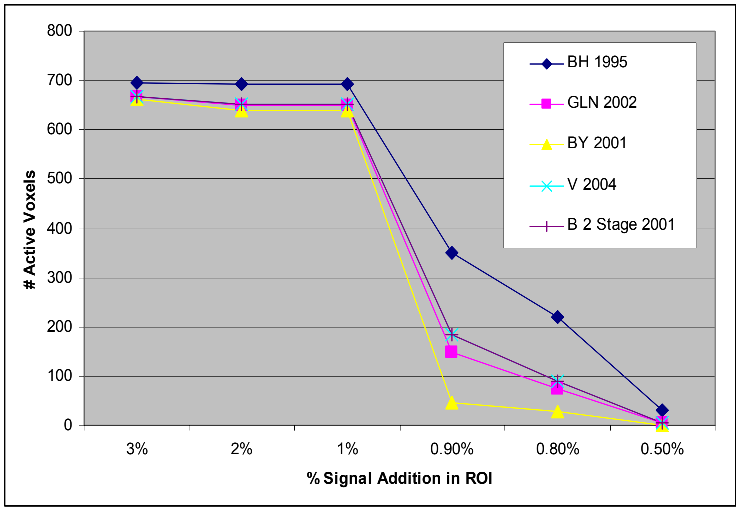
Method (applied on human hand flex data)	p-value threshold	Number of activated Pixels
BH 1995 [1]	1.6702E-02	5908
GLN 2002 [2]	1.1142E-03	4082
BY 2001 [3]	1.0483E-04	2946
V 2004 [4]	1.9138E-03	4391
B 2 stage 2001 [5]	1.8900E-03	4385

Results:

The p-value threshold values and the number of activated voxels for each algorithm are listed in the table for the experimental data set. For the simulated dataset study, the number of positive voxel activations is plotted in the figure below as a function of the level of signal activation applied to the ROI within the brain, for the five different algorithms. The FDR algorithm was applied to the entire brain volume in both cases.

Conclusions :

The different FDR methods attempt to control the number of false positives relative to all detections and some have significantly higher statistical power than another. We observe marked differences in p-value thresholds and in the number of activated voxel, between different algorithms in the case of the hand-flex data. In the simulated data set we know *a priori*, the number of active voxels. Note that the Benjamini and Hochberg[1] method which assumes that all voxels are uncorrelated shows the greatest number of active voxels at each signal level. However, methods such as that by Benjamini and Yekutieli[3] which assume nothing about the correlation structure of the data give the least number of activations at each signal level. The number of activations for all methods converges as the signal intensity is increased. These results show that significantly different results can be obtained with different FDR methods depending upon the signal-to-noise ratio of the underlying data. Therefore, different FDR methods may be optimal for different tasks that have different strengths of response (e.g., hand flex versus working memory).



References: 1. Benjamini, Y, and Y. Hochberg. 1995. JRSSB 57:289-300. **2.** Genovese, C.R., Lazar, N.A. and Nichols, T.E. (2002) *NeuroImage* 15, 870--878. **3.** Benjamini, Y. and D. Yekutieli. 2001. Annals of Statistics 29:1165-1188. **4.** Ventura, V., C.J. Paciorek, and J.S. Risbey. 2004. Journal of Climate 17:4343-4356. **5.** Benjamini, Y., A. Krieger, and D. Yekutieli. 2001. Department of Statistics and Operations Research Internal Report, Tel Aviv University.