FALSE DISCOVERY RATE AND FUNCTIONAL MAGNETIC RESONANCE IMAGING: A QUANTITATIVE ANALYSIS

ANDERSON MARCELO WINKLER* HUMBERTO REMIGIO GAMBA* HELLEN MATHEI DELLA JUSTINA* ARNOLFO CARVALHO NETO**
*** Imaging and Electronic Instrumentation Laboratory (LABIEM), Federal Technological University of Parana (UTFPR)**
**** Discipline of Medical Radiology, Department of Internal Medicine, Federal University of Parana (UFPR)**
www.labiem.cpgei.cefetpr.br - andersonwinkler@gmail.com

Introduction

The most common strategy to analyse fMRI data (though not necessarily optimum) is to:

- Apply some transformation to the images, such as affine transformations, warpings, phase shifts and smoothing in order to reduce potential confounding effect of head motion, noise, acquisition timing or else, to conform the subjects' brain to a standard brain;
- Perform a voxelwise statistical analysis of the data, fitting the obtained response to a predefined model of cerebral activation. The estimated parameters are tested against the null hypothesis of non-activation, resulting in a statistic image (known as statistical parametric map);
- Set a threshold. Voxels which statistical score is above this threshold are labeled as "active"; the remaining are labeled as "inactive."

A fundamental question, however, remains controversial: how to choose the most adequate threshold? Very high thresholds may obscure legitimately active areas, labeling them as "inactive". Too low thresholds, on the other hand, result in beautiful, but meaningless images, as many inactive areas may become falsely labeled as "active."

There are some methods available for choosing the threshold when dealing with this multiple comparisons problem, and the ideal choice is somewhat disputed. One recent approach is to set up the threshold using the *false discovery rate*.

The false discovery rate (FDR)

FDR is the proportion of incorrect rejections of the null hypotesis (false positives) among those tests where the null hypothesis is true. This proportion is the maximum amount of false positives that the researcher is willing to accept. The procedure to control the FDR at level q consists in calculate the uncorrected p-value for each voxel and order them so that the ordered p-values are $p_1 \leq p_2 \leq \ldots \leq p_V$. Then, find the largest R so that $p_R \leq (R.q)/(V)$. The voxels $v_1, \ldots, v_R$ are declared active. Figure 1 (left) shows a graphical perspective; the graph is zoomed (right) to show the threshold in greater detail.

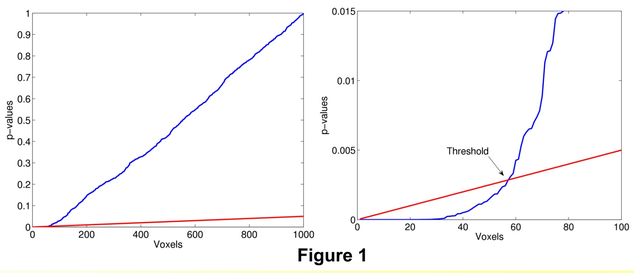

Figure 1

Epidemiological viewpoint

Solving for the multiple comparisons problem is rather similar to the epidemiological problem of finding the ideal cut-off value for diagnostic tests. In both cases, the results of the individual tests can be summarised in a contingency table:

	Truly active	*Truly inactive*
Declared active	Truly positive (A)	Falsely positive (B)
Declared inactive	Falsely negative (C)	Truly negative (D)

From the table, the following terms can be defined:

- Sensitivity (S): A/A+C
- Specificity (E): D/B+D
- Positive predictive value (PPV): A/A+B
- Negative predictive value (NPV): D/C+D
- False discovery rate (FDR): B/A+B = 1-PPV

Objective

The aim of this study is to evaluate the performance of FDR in simulated non-smoothed 2D random fields, in terms of sensitivity (S), specificity (E), positive and negative predictive values (PPV and NPV), as well FDR itself. This analysis does not account for spatial distribution of active areas, which would eventually become a confounding in the analysis.

Method

Gaussian 2D random-fields [$N(0;1)$], sized 64x64 voxels were simulated. To each random field, a patch of "true activation" was added, shifting the voxel statistic within the range 0.5 to 8.0. The size of the patch varied from 1 to 4096. Figure 2 shows examples of (A) noise, (B) noise plus small and weak, (C) large and strong and (D) small and strong patches of "activation."

Each simulation was repeated 100 times to minimise the effect of outliers. The statistic scores were converted to p-values, and the FDR procedure was applied. The calculated cut-off was used to threshold the statistic image.

The values of S, E, PPV, NPV, as well the false discovery rate itself were computed for every patch size and magnitude, and averaged between the 100 simulations. The FDR was controlled at $q = 0.05$ in all cases, as typical for fMRI studies.

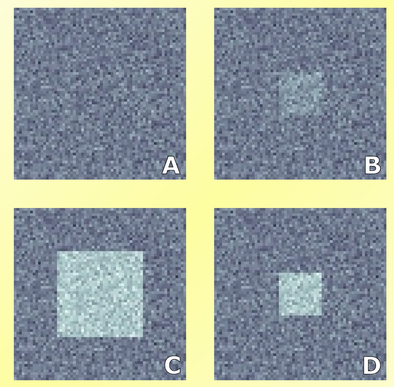

Figure 2

Results

The results were summarised in the Figures 3-7.

Many vertices are missing in some of these meshes. This is consequence of eventual divisions by zero, most common for weak or very small patches of "cerebral activity."

Note the erratic behaviour of the FDR for small and weak signals, although, even in these unfavourable settings, the threshold is quite well-behaved, not surpassing the bound q.

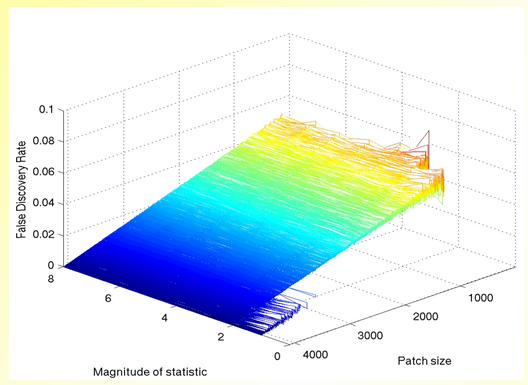

Figure 3: False discovery rate

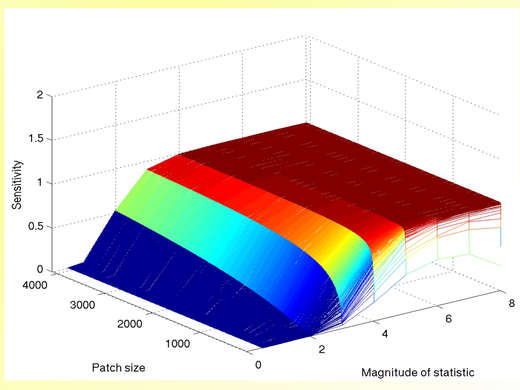

Figure 4: Sensitivity

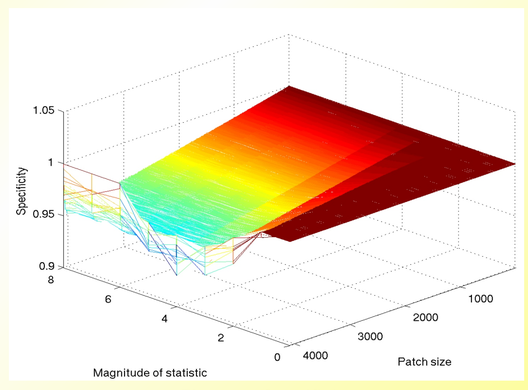

Figure 5: Specificity

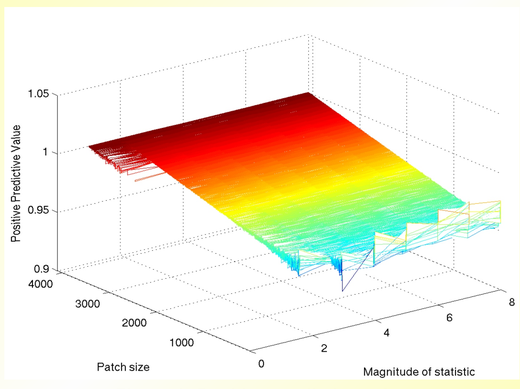

Figure 6: Positive predictive value

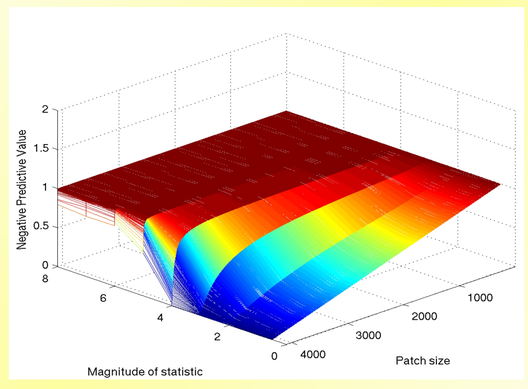

Figure 7: Negative predictive value

Discussion and Conclusions

The main improvement introduced with the use of FDR procedure is that, on average, the proportion of false discoveries is lower than the set bound level q. The consequence is that the PPV, a measure of the trustworthiness of the result obtained at each voxel, is controlled at the level 1-q, getting at the essence of what the researcher usually wants to control.

The results, however, show that the NPV can assume very low levels when FDR is applied to images with extensive areas of activation, or not strongly active areas, when many active voxels may wrongly be declared as inactive.

Another shortcoming of FDR is its low sensitivity to identify activity in images with weak and sparse signals. This is, however, also inherent to other modalities of thresholding, such as Bonferroni or based on random field theory.

References

[1] Benjamini Y. and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, vol 57, pp 289-300, 1995;

[2] Genovese C. R., Lazar N. A. and Nichols, T. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, vol 15, pp 870-878, 2002.

UTFPR
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ