# Statistical inferences in fMRI

Cyril Pernet, PhD
BRIC/SINAPSE – University of Edinburgh

# Overview

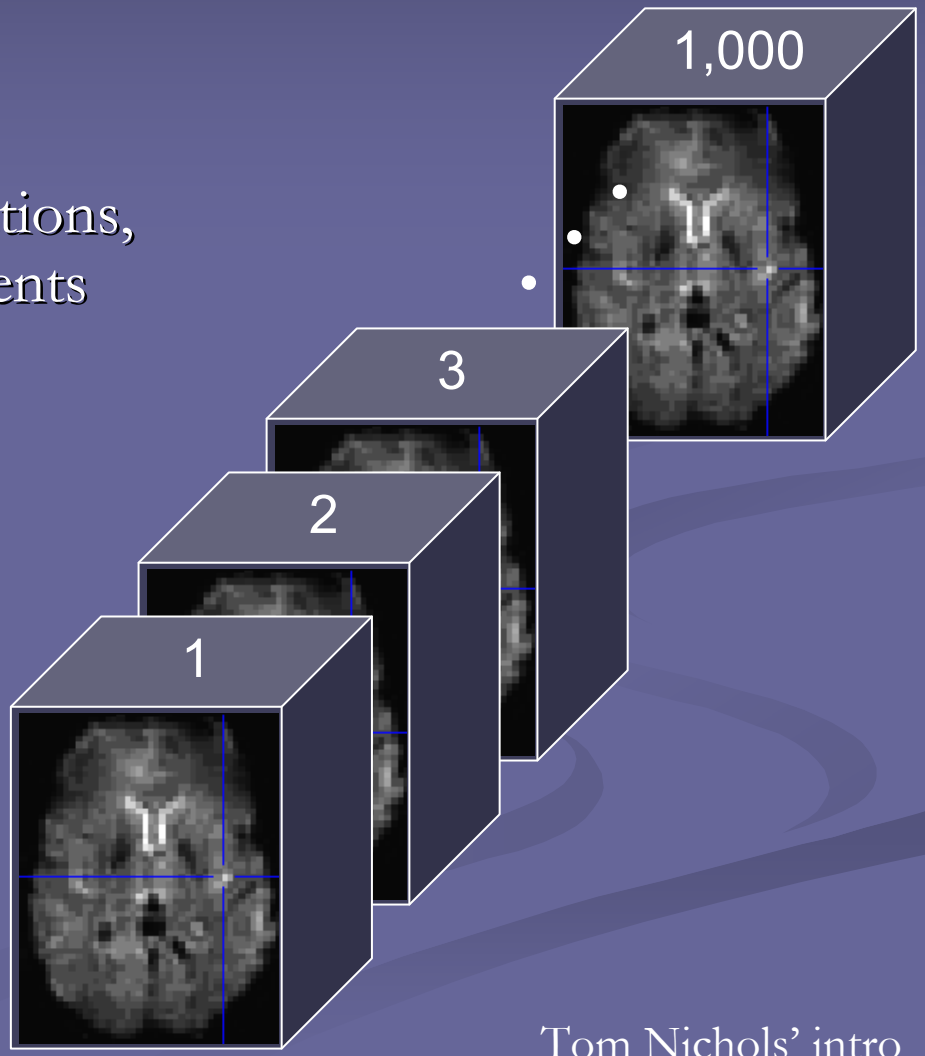
- Multiple comparisons correction procedures
- Levels of inferences (set, cluster, voxel)
- Circularity issues

# Multiple comparison correction

Avoiding false positives

# What Problem?

- 4-Dimensional Data
  - 1,000 multivariate observations, each with > 100,000 elements
  - 100,000 time series, each with 1,000 observations
- Massively Univariate Approach
  - 100,000 hypothesis tests
- Massive MCP!




Tom Nichols' intro

# What Problem?

- Typical brain ~ 130000 voxels
- @ p = .05, it is expected = 6500 false positives!
- @ a more conservative value like p = .001 we still expect 130 false positives.

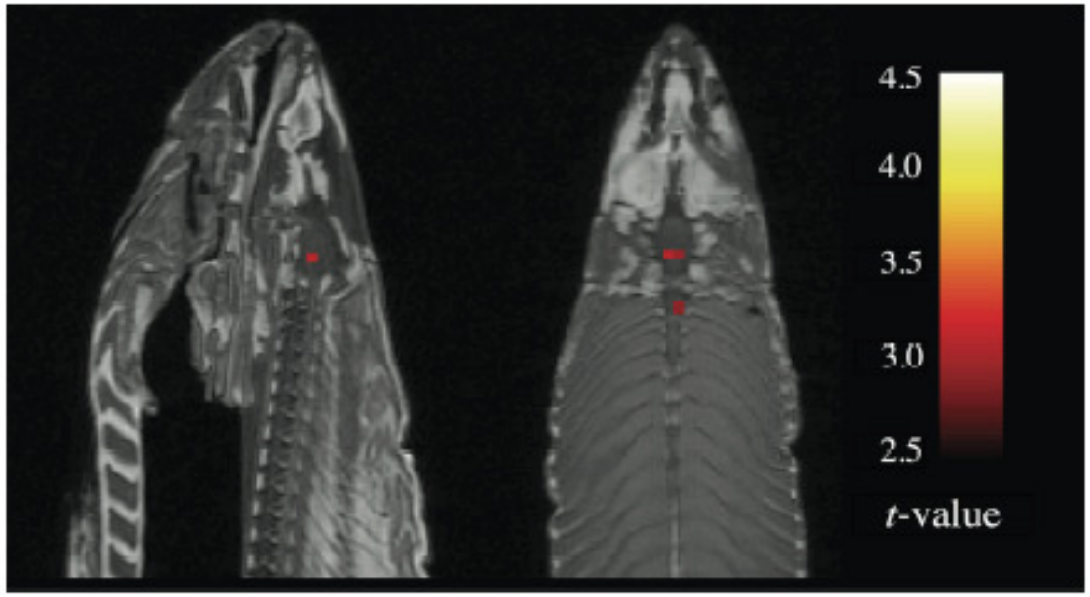
- Using extend threshold k without correction is not enough as it, by chance, can cluster as well.

# What Problem?

- Bennet et al., 2009

- Task: take a decision about emotions on pictures
- Design: blocks of 12 sec activation/rest
- Analysis: standard data processing with SPM
- Subject: a dead salmon!

# What Problem?




A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

- The cluster was 81mm$^3$ ! – after multiple comparison corrections all false activations were removed.

# Solutions for MCP

- Height Threshold
- Familywise Error Rate (FWER)
  - Chance of *any* false positives; Controlled by Bonferroni & Random Field Methods
- False Discovery Rate (FDR)
  - Proportion of false positives *among* rejected tests
- Bayes Statistics

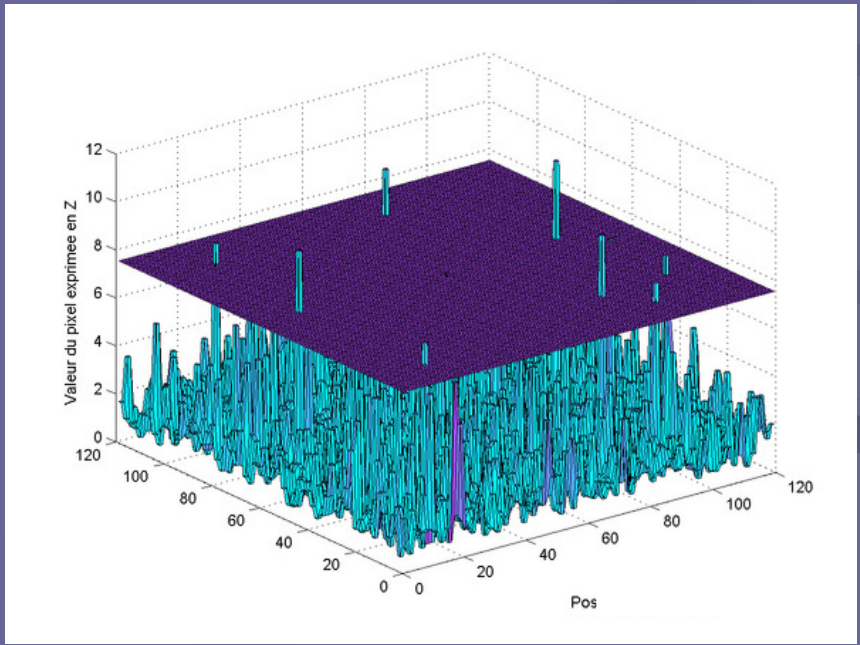
# From single univariate to massive univariate

| Univariate stat | Functional neuroimaging |
|---|---|
| 1 observed data | Many voxels |
| 1 statistical value | Family of statistical values |
| Type 1 error rate (chance to be wrong rejecting H0) | Family-wise error rate |
| Null hypothesis | Family-wise null hypothesis |

# Height Threshold

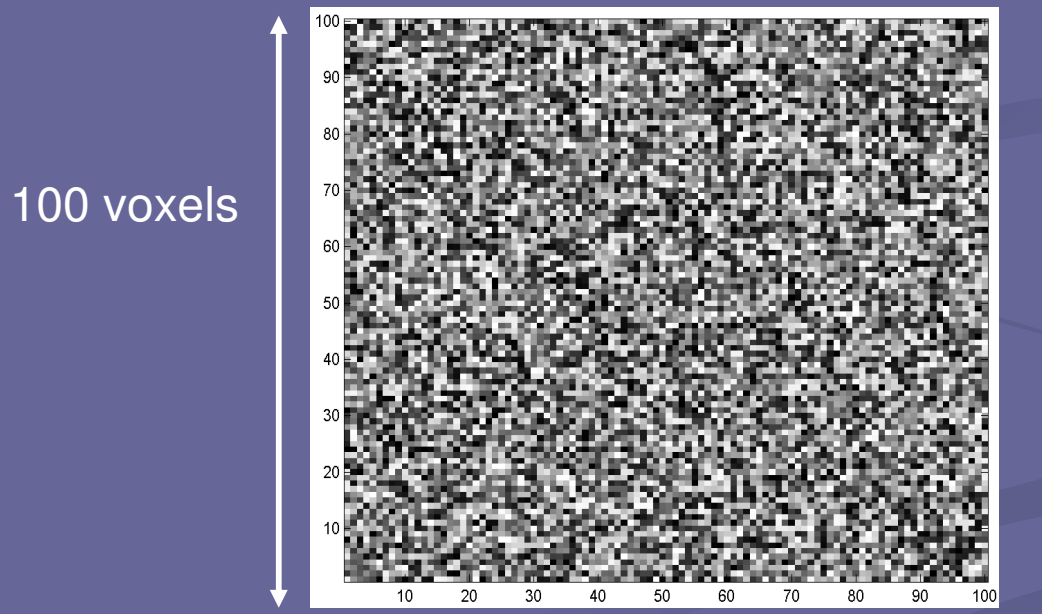
- Choose locations where a test statistic Z (T, F, ...) is large to threshold the image of Z at a height z
- The problem is how to choose this threshold z to exclude false positives with a high probability (e.g. 0.95)?

To control for family wise error on must take into account the nb of tests



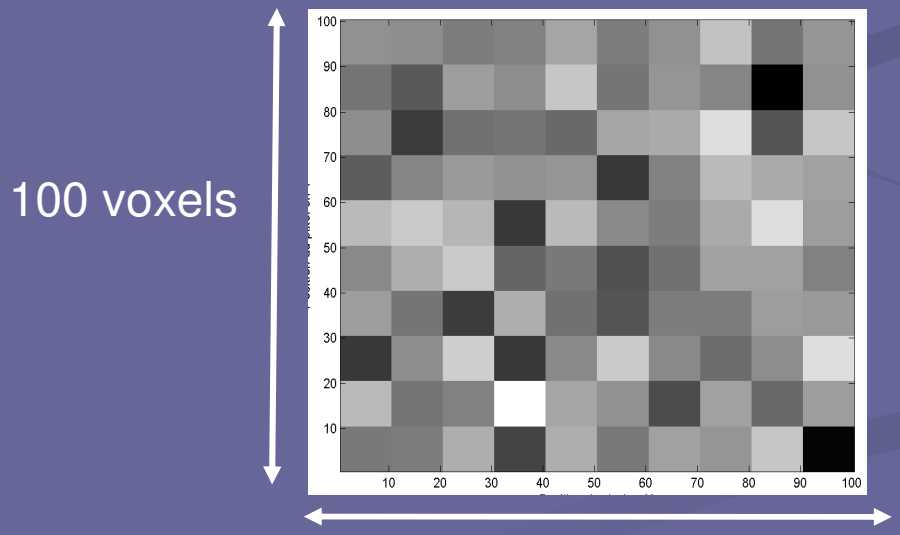
# Bonferroni

- 10000 Z-scores ; alpha = 5%
- alpha corrected = .000005 ; z-score = 4.42



# Bonferroni

- 10000 Z-scores ; alpha = 5%
- 2D homogeneous smoothing – 100 independent observations
- alpha corrected = .0005 ; z-score = 3.29

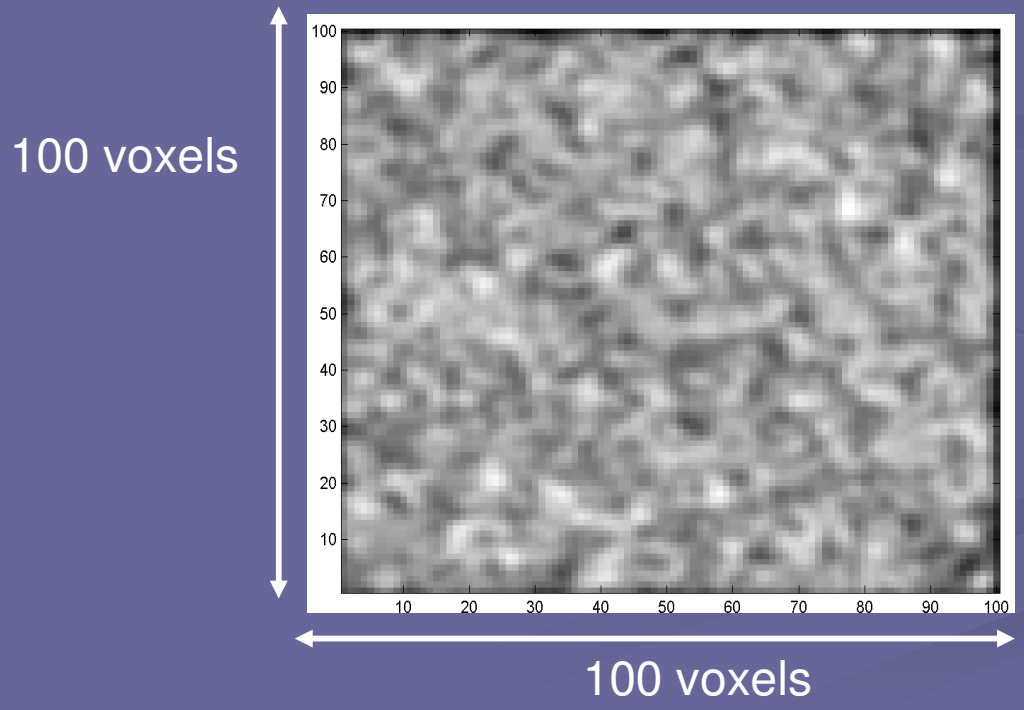


# Solutions for MCP

- An important feature of neuroimaging data is that we have a family of stat values that has topological features (Bonferroni for instance consider tests as independent)
- Why considering data as a smooth lattice? (Chumbley et al., 2009 NeuroImage 44)
  - fMRI/PET are projection methods of data points onto the whole space – MEEG forms continuous functions in time and are smooth by the scalp (space)
  - Neural activity propagate locally through intrinsic/lateral connections and is distributed via extrinsic connections / Hemodynamic correlates are initiated by diffusing signals (e.g. NO)

# Random Field Theory

- 10000 Z-scores ; alpha = 5%
- Gaussian kernel smoothing –
- How many independent observations ?



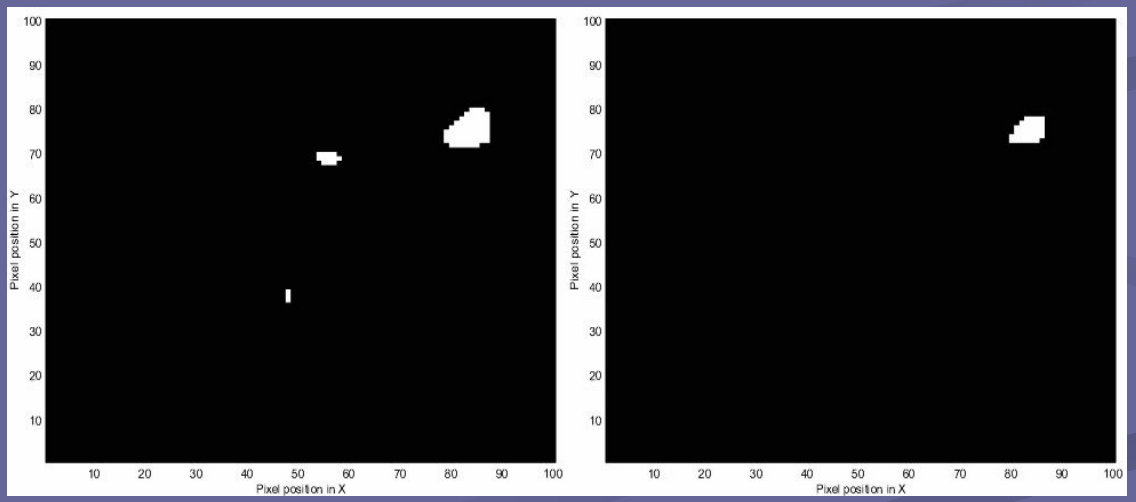
# Random Field Theory

- RFT relies on theoretical results for smooth statistical maps (hence the need for smoothing), allowing to find a threshold in a set of data where it's not easy to find the number of independent variables. Uses the expected Euler characteristic (EC density)

- 1 Estimation of the smoothness = number of resel (resolution element) = f(nb voxels, FWHM)
- 2 expected Euler characteristic = number of clusters above the threshold
- 3 Calculation of the threshold

# Random Field Theory

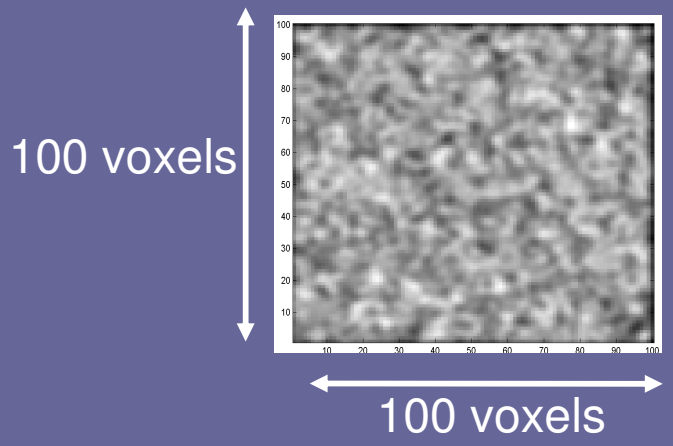
- The Euler characteristic can be seen as the number of blobs in an image after thresholding (p value that you select in SPM)
- At high threshold, EC = 0 or 1 per resel: E[EC] ≈ $p^{FWE}$




$E[EC] = R \cdot (4 \log_e 2) \cdot (2\pi)^{-2/3} \cdot Z_t \cdot e^{-1/2 Z^2_t}$ for a 2D image, more complicated in 3D

# Random Field Theory

- For 100 resels, the equation gives E[EC] = 0.049 for a threshold Z of 3.8, i.e. the probability of getting one or more blobs where Z is greater than 3.8 is 0.049




| $\alpha$ | number of resels in the image | Bonferroni | | RFT |
|---|---|---|---|---|
| | | threshold | score Z | score Z |
| 0.05 | 100 | $\frac{0.05}{100}$ | 3.3 | |
| | | | | 3.8 |

- If the resel size is much larger than the voxel size then E[EC] only depends on the nb of resels otherwise it also depends on the volume, surface and diameter of the search area (i.e. shape and volume matter)

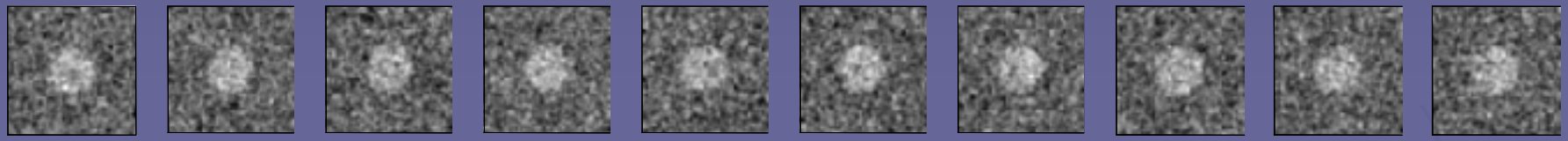
# False discovery Rate

- Whereas family wise approach corrects for any false positive, the FDR approach aim at correcting among positive results only.

1. Run an analysis with alpha = x%
2. Sort the resulting positive data
3. Threshold to remove the false positives

# False discovery Rate

Signal+Noise



FEW correction

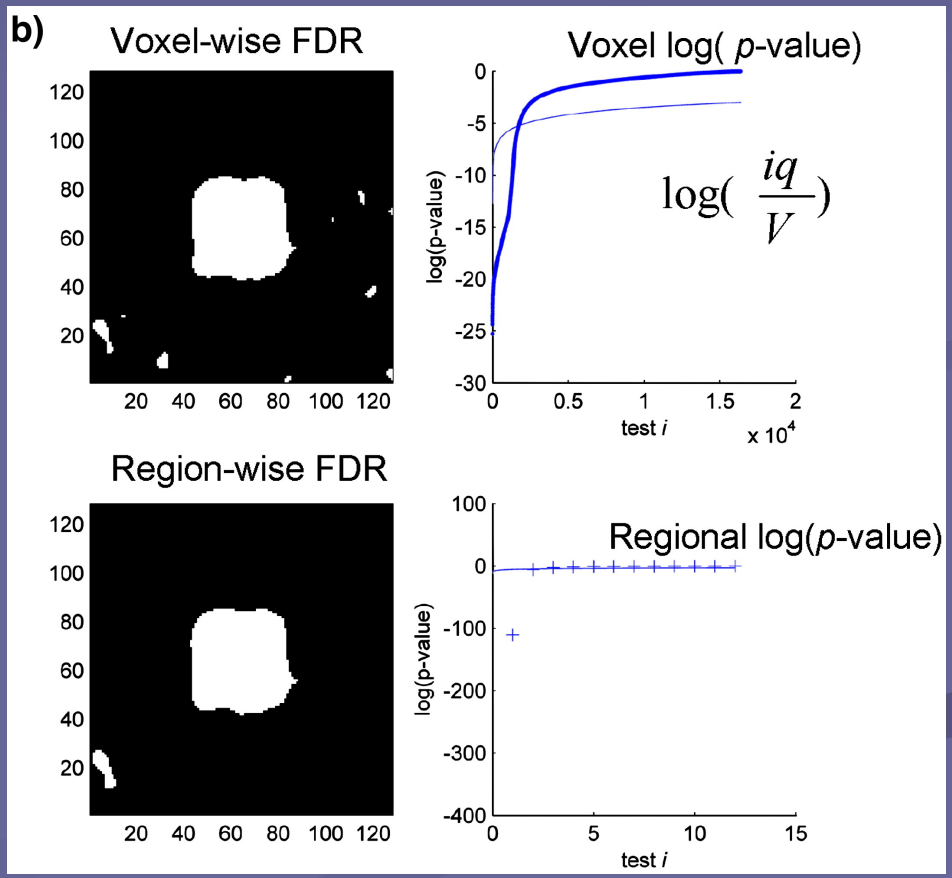
FDR correction

# False discovery Rate

takes the spatial structure into account

Under H0 the nb of voxels per cluster is known → uncorrected p value for clusters → apply FDR on the clusters (volume-wise correction)

Assumes that the volume of each cluster is independent of the number of clusters



# Levels of inference

Voxel, cluster and set

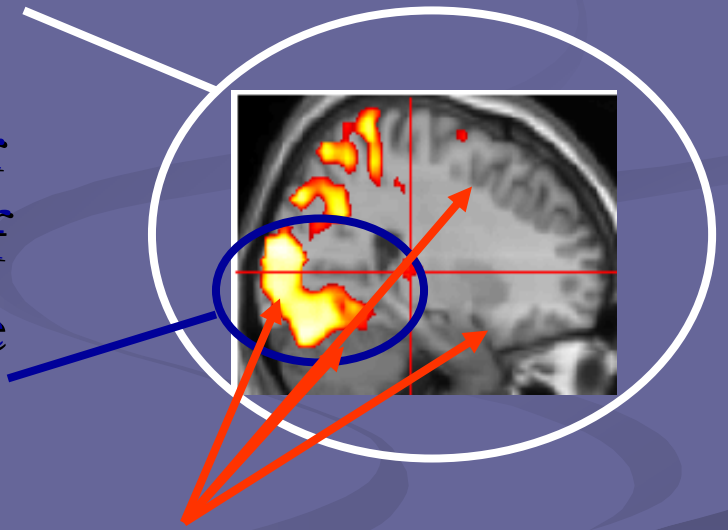
# Levels of inference

- 3 levels of inference can be considered:
  - Voxel level (prob associated at each voxel)
  - Cluster level (prob associated to a set of voxels)
  - Set level (prob associated to a set of clusters)

- The 3 levels are nested and based on a single probability of obtaining c or more clusters (set level) with k or more voxels (cluster level) above a threshold u (voxel level): $P_w(u,k,c)$

# Levels of inference

- Set level: we can reject H0 for an omnibus test, i.e. there are some significant clusters of activation in the brain.
- Cluster level: we can reject H0 for an area of a size k, i.e. a cluster of 'activated' voxels is likely to be true for a given spatial extend.



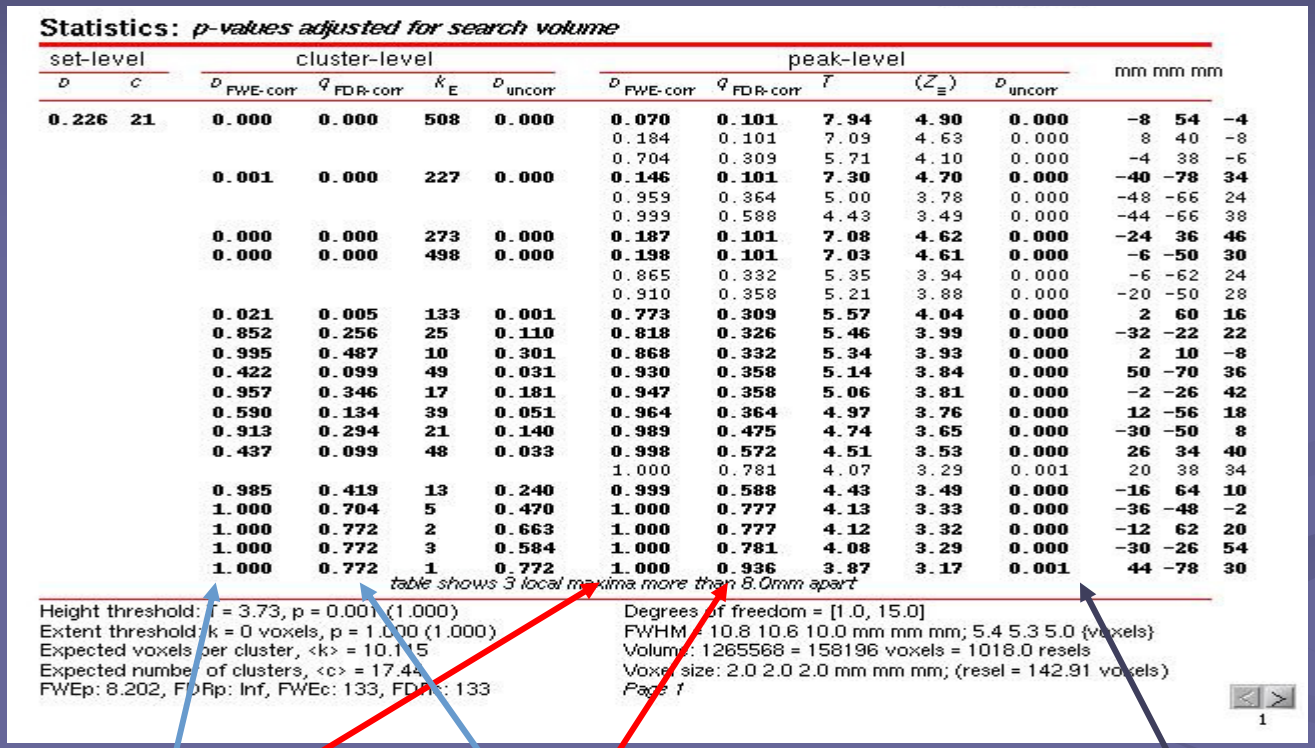
- Voxel level: we can reject H0 at each voxel, i.e. a voxel is 'activated' if exceeding a given threshold

# Levels of inference

- Each level of inference is valid, but the inferences are different – e.g. a set might be enough to check that subjects activated regions selected a priori for a connectivity analysis – clusters might be good enough if hypotheses are about the use of different brain areas between groups

- Both voxel and cluster levels need to address the multiple comparison problem. If the activated region is predicted in advance, the use of corrected p values is unnecessary and inappropriately conservative – a correction for the number of predicted regions (Bonferroni) is enough

# Levels of inference



**Statistics:** *p-values adjusted for search volume*

| set-level | | cluster-level | | | | peak-level | | | | | mm mm mm | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | $c$ | $p_{FWE\text{-}corr}$ | $q_{FDR\text{-}corr}$ | $k_E$ | $p_{uncorr}$ | $p_{FWE\text{-}corr}$ | $q_{FDR\text{-}corr}$ | $T$ | $(Z_\equiv)$ | $p_{uncorr}$ | | | |
| **0.226** | **21** | **0.000** | **0.000** | **508** | **0.000** | **0.070** | **0.101** | **7.94** | **4.90** | **0.000** | **-8** | **54** | **-4** |
| | | | | | | 0.184 | 0.101 | 7.09 | 4.63 | 0.000 | 8 | 40 | -8 |
| | | | | | | 0.704 | 0.309 | 5.71 | 4.10 | 0.000 | -4 | 38 | -6 |
| | | **0.001** | **0.000** | **227** | **0.000** | **0.146** | **0.101** | **7.30** | **4.70** | **0.000** | **-40** | **-78** | **34** |
| | | | | | | 0.959 | 0.364 | 5.00 | 3.78 | 0.000 | -48 | -66 | 24 |
| | | | | | | 0.999 | 0.588 | 4.43 | 3.49 | 0.000 | -44 | -66 | 38 |
| | | **0.000** | **0.000** | **273** | **0.000** | **0.187** | **0.101** | **7.08** | **4.62** | **0.000** | **-24** | **36** | **46** |
| | | **0.000** | **0.000** | **498** | **0.000** | **0.198** | **0.101** | **7.03** | **4.61** | **0.000** | **-6** | **-50** | **30** |
| | | | | | | 0.865 | 0.332 | 5.35 | 3.94 | 0.000 | -6 | -62 | 24 |
| | | | | | | 0.910 | 0.358 | 5.21 | 3.88 | 0.000 | -20 | -50 | 28 |
| | | **0.021** | **0.005** | **133** | **0.001** | **0.773** | **0.309** | **5.57** | **4.04** | **0.000** | **2** | **60** | **16** |
| | | **0.852** | **0.256** | **25** | **0.110** | **0.818** | **0.326** | **5.46** | **3.99** | **0.000** | **-32** | **-22** | **22** |
| | | **0.995** | **0.487** | **10** | **0.301** | **0.868** | **0.332** | **5.34** | **3.93** | **0.000** | **2** | **10** | **-8** |
| | | **0.422** | **0.099** | **49** | **0.031** | **0.930** | **0.358** | **5.14** | **3.84** | **0.000** | **50** | **-70** | **36** |
| | | **0.957** | **0.346** | **17** | **0.181** | **0.947** | **0.358** | **5.06** | **3.81** | **0.000** | **-2** | **-26** | **42** |
| | | **0.590** | **0.134** | **39** | **0.051** | **0.964** | **0.364** | **4.97** | **3.76** | **0.000** | **12** | **-56** | **18** |
| | | **0.913** | **0.294** | **21** | **0.140** | **0.989** | **0.475** | **4.74** | **3.65** | **0.000** | **-30** | **-50** | **8** |
| | | **0.437** | **0.099** | **48** | **0.033** | **0.998** | **0.572** | **4.51** | **3.53** | **0.000** | **26** | **34** | **40** |
| | | | | | | 1.000 | 0.781 | 4.07 | 3.29 | 0.001 | 20 | 38 | 34 |
| | | **0.985** | **0.419** | **13** | **0.240** | **0.999** | **0.588** | **4.43** | **3.49** | **0.000** | **-16** | **64** | **10** |
| | | **1.000** | **0.704** | **5** | **0.470** | **1.000** | **0.777** | **4.13** | **3.33** | **0.000** | **-36** | **-48** | **-2** |
| | | **1.000** | **0.772** | **2** | **0.663** | **1.000** | **0.777** | **4.12** | **3.32** | **0.000** | **-12** | **62** | **20** |
| | | **1.000** | **0.772** | **3** | **0.584** | **1.000** | **0.781** | **4.08** | **3.29** | **0.000** | **-30** | **-26** | **54** |
| | | **1.000** | **0.772** | **1** | **0.772** | **1.000** | **0.936** | **3.87** | **3.17** | **0.001** | **44** | **-78** | **30** |

*table shows 3 local maxima more than 8.0mm apart*

Height threshold: T = 3.73, p = 0.001 (1.000)
Extent threshold: k = 0 voxels, p = 1.000 (1.000)
Expected voxels per cluster, <k> = 10.115
Expected number of clusters, <c> = 17.44
FWEp: 8.202, FDRp: Inf, FWEc: 133, FDRc: 133

Degrees of freedom = [1.0, 15.0]
FWHM = 10.8 10.6 10.0 mm mm mm; 5.4 5.3 5.0 {voxels}
Volume: 1265568 = 158196 voxels = 1018.0 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 142.91 voxels)
*Page 1*

1

RFT (Gaussian Random Fields)
-> Prob of cluster
-> Prob of voxel

Using p=.001 this creates an excursion set
Prob clusters of that size
Prob peak that height
→ after FDR correction

Uncorrected (bad)

# Circularity issues in fMRI

# Definition

- Refers to the problem of selecting data for analysis
- How data (areas usually) are selected, analysed and sorted is key to avoid circularity

- Put forward by Vul et al. 2009, *Perspectives on Psychological Science.* 4
- Better explained in Kriegeskorte et al., 2009 *Nat. Neuroscience* 12

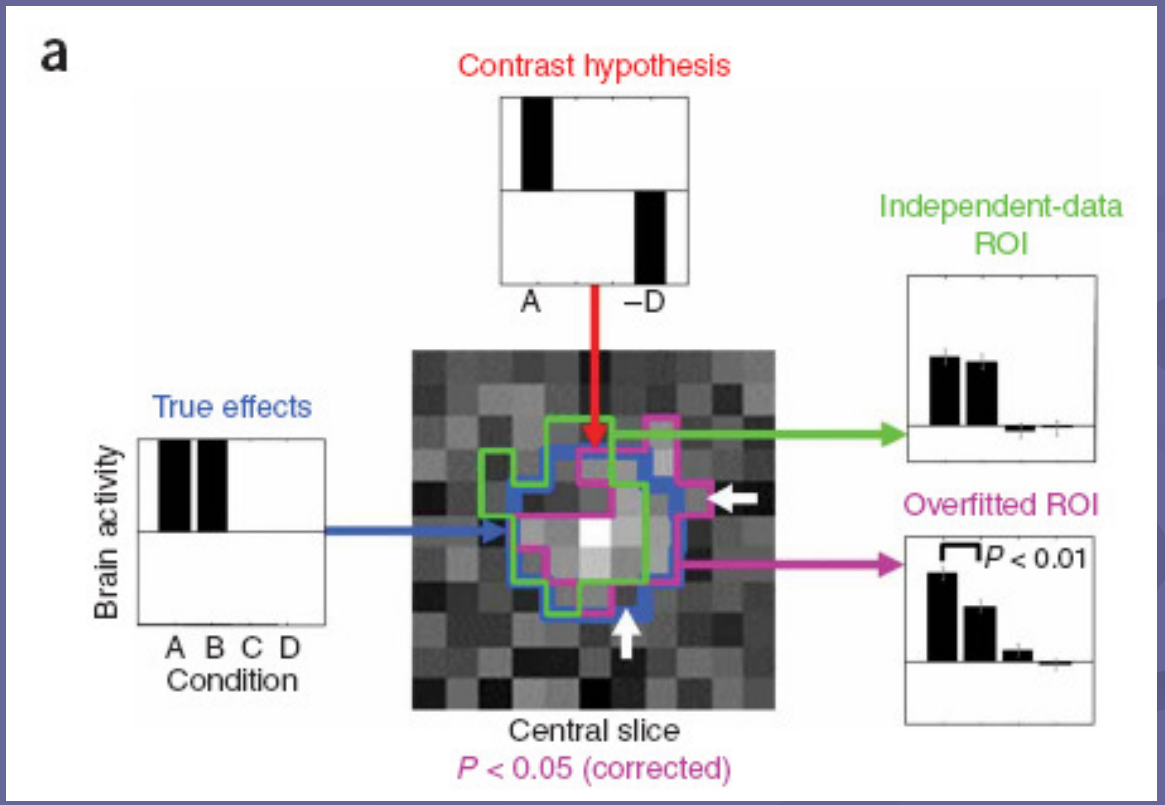
# Circularity

- Double dipping pblm: "data are first analyzed to select a subset and then the subset is reanalyzed to obtain the results. In this context, assumptions and hypotheses determine the selection criterion and selection can, in turn, distort the results."

- Take a gp of subjects and measures RTs, then take 2 subgroups from the same subjects and re-do some analysis?? → increases the diff.
- Take fMRI data and get activated areas, extract ROI and re-do some analyses??

# Circularity

- Selection and tests must be independent – non independence create spurious effects



# Circularity

- Independence of the selection and tests

1. Anatomic ROI, analysis of fMRI
2. SPM, minimal requirement is orthogonality of the contrasts (e.g. find regions using A+B>0 C=[1 1] and test A vs B C=[1 -1]) but if $N_A$ and $N_B$ are different there is still a bias when testing A-B (across subjects independence is ensured by $C_{selection}{}^{T}(X^{T}X)^{-1}C_{test}$)
3. Select using a subset of data, test with another one

# Enough for today ☺


Thanks for your attention