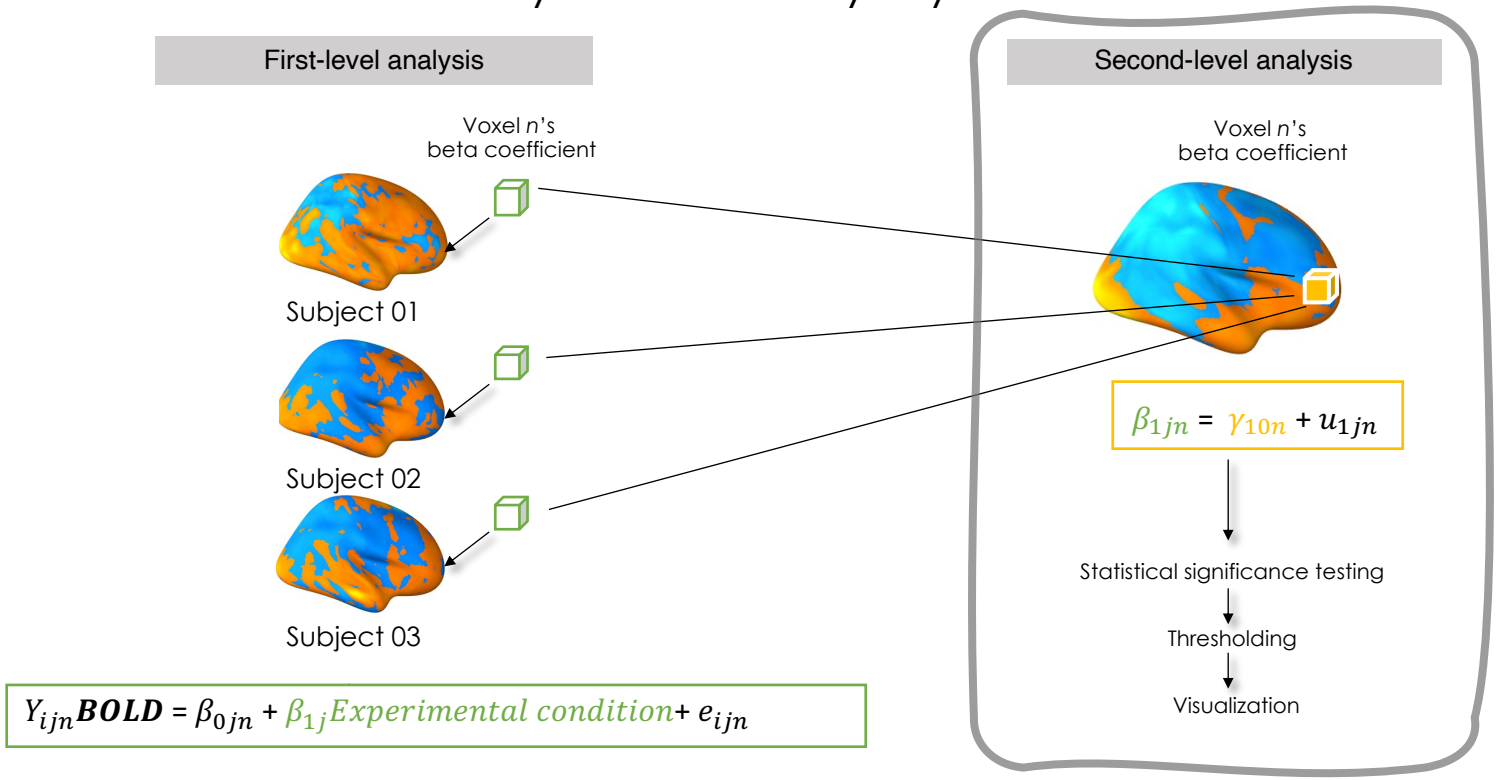
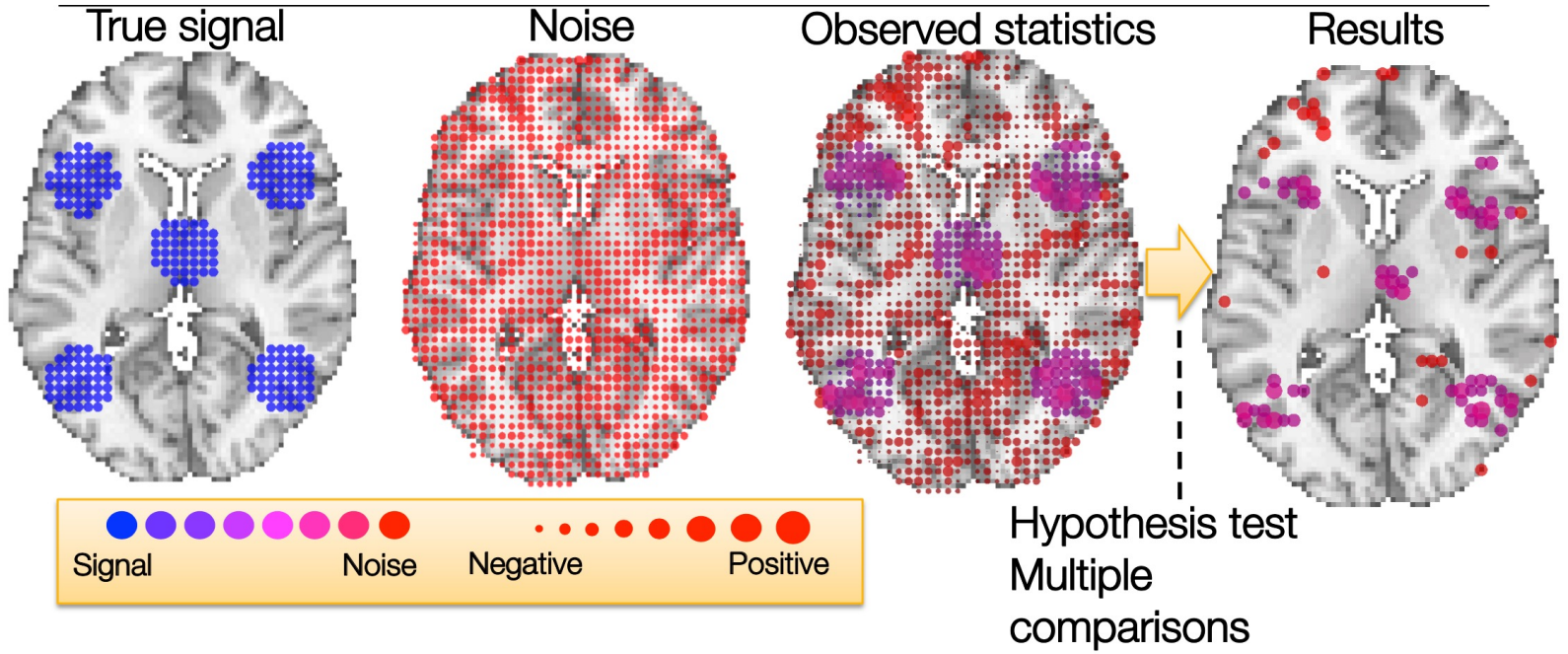
# Thresholding

- If we get the group-level statistical parametric map, how can we determine which regions or voxels respond to my task?
  - Which brain areas are commonly activated by my task?



# Thresholding

- If we get the group-level statistical parametric map, how can we determine which regions or voxels respond to my task?
  - Which brain areas are commonly activated by my task?




Slide credit: Martin Lindquist and Tor Wager

# Thresholding

- If we get the group-level statistical parametric map, how can we determine which regions or voxels respond to my task?
  - Which brain areas are commonly activated by my task?
- This brain mapping is good for
  - Making inferences on the presence of activity, to either a) test a theory, or b) characterize the pattern of brain responses to a task.
  - Limiting the false positive rate to a specified level.
  - Leverage **hypothesis testing** to provide evidence on a variety of theories:
    Is Area *r* involved in Task *x*?
- Therefore, thresholding is a procedure to look at which brain areas are significantly activated by task X

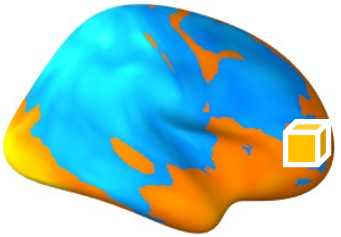
Slide credit: Martin Lindquist and Tor Wager

# Thresholding

- For a such things, we did first-level or second level analysis using GLM using
  - Ordinary least squares
  - Weighted least squares
  - Iterative generalized least squares
- We can get t-statistics, standard error, beta coefficient, p-value for each voxel
- Is it okay to use the standard p-value criterion such as under 0.05 ($p<0.05$)?

Second-level analysis

Voxel *n*'s
beta coefficient

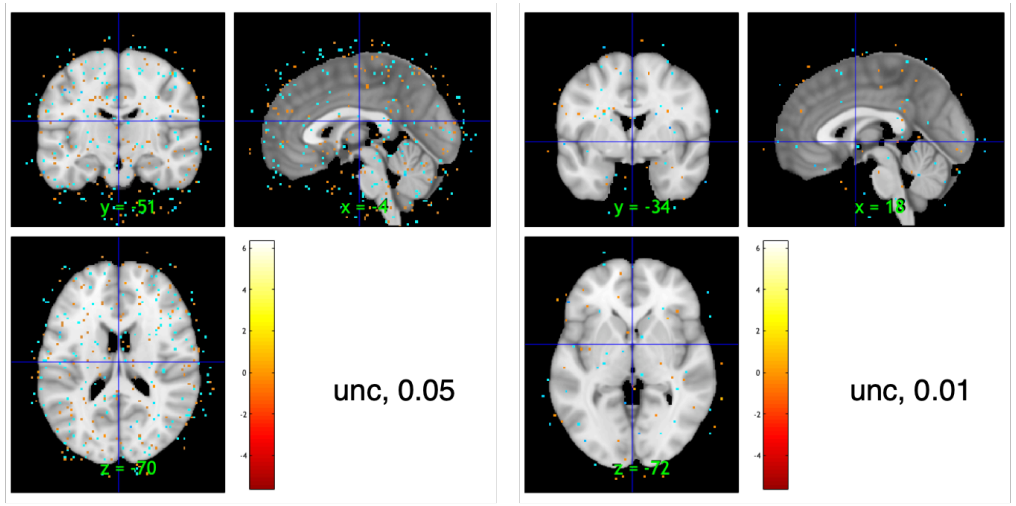


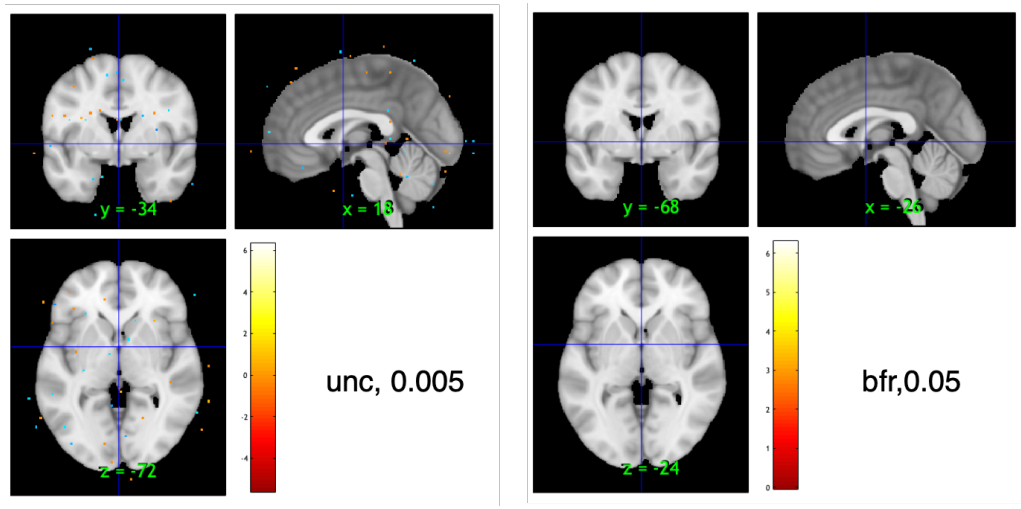
$$\beta_{1jn} = \gamma_{10n} + u_{1jn}$$

# Thresholding: examples

- I made fake fMRI datasets of two groups
  - Basically, the mean and standard deviation is the same
  - Therefore, if we compare these two group's brain activations, there should not exist significant voxels
- T-test was performed

- These should be happened. But, Why?

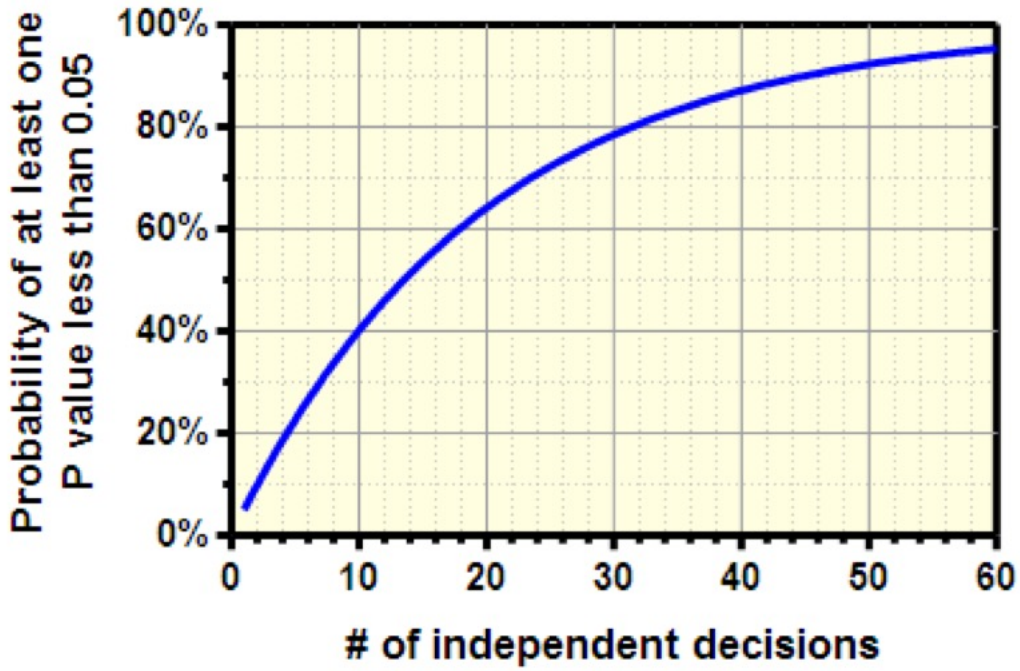


# Multiple comparison

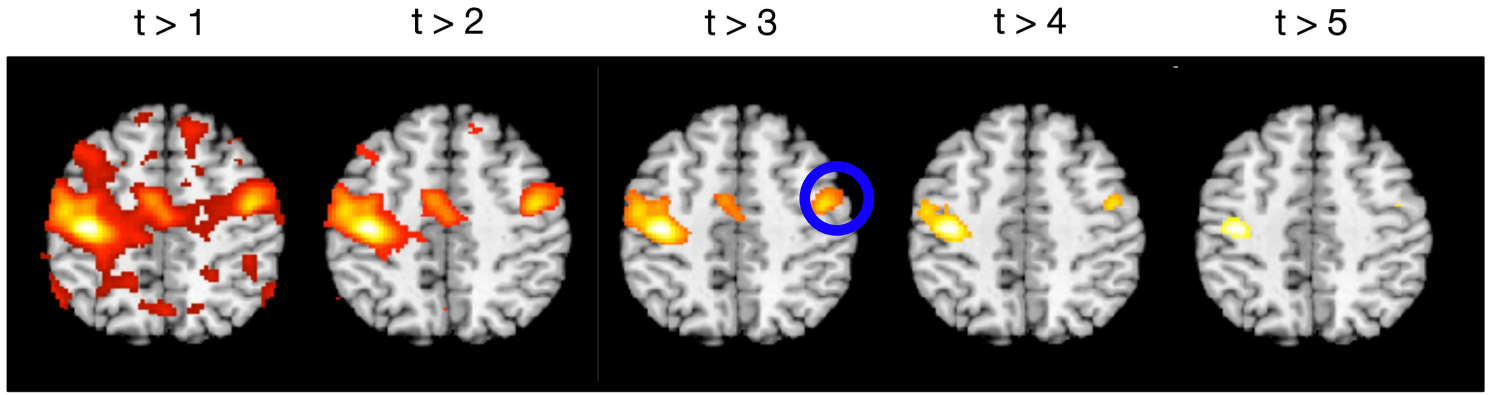
- What is the "**Multiple Comparisons problem**"
  - The increase in the number of **false-positive results** (i.e., Type I errors) with increasing number of statistical tests
  - It is of particular consequence for voxel-wise fMRI analyses, which may have many thousands of statistical tests

The more tests one performs, the greater the likelihood of getting at least one false positive



# Multiple comparison

- Measures of False positives
  - There exists several ways of quantifying the likelihood of obtaining false positives
    - In any data set with random noise, the number of false-positive results for n statistical tests is simply (n * alpha)
    - Which of 100,000 voxels are significant? – a=0.05 ⇒ 5,000 false positive voxels
  - Choosing a threshold is a balance between sensitivity (true positive rate) and specificity (true negative rate)




Slide credit: Martin Lindquist and Tor Wager

# Multiple comparison

- Calculating significance threshold
  - To overcome the problem of multiple compassions, fMRI researchers always reduce the desired alpha value, so that voxels are less likely to pass the significance threshold **by chance**
  - The two factors for selecting the alpha value

    1) the types of error they want to avoid

    2) the number of independent test in the data

# Family-Wise Error Rate

- **Family-Wise Error Rate** (FWER)
  - Probability of any false positive
  - The probability of making one or more type 1 errors (false-positive) in a family of tests, under the null hypothesis
  - It is of particular consequence for voxel-wise fMRI analyses, which may have many thousands of statistical tests
- FWER controlling methods: –
  - Bonferroni correction
  - Random Field Theory
  - Permutation Tests

Statistical Methods in fMRI, Martin Lindquist

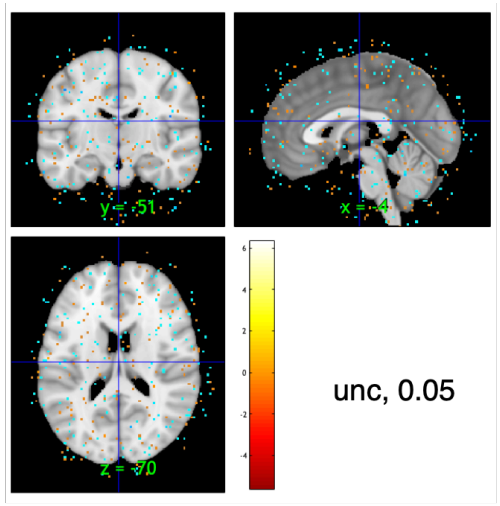


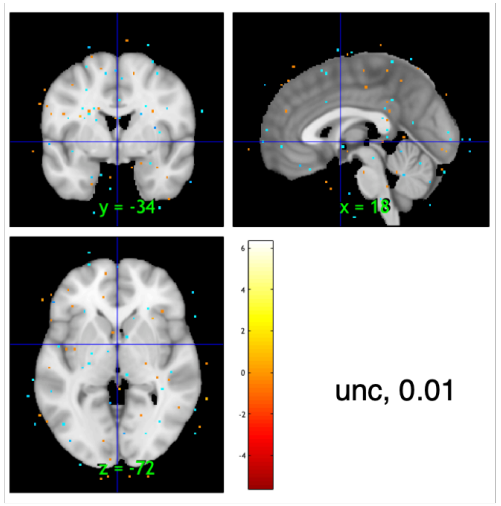
# Family-Wise Error Rate

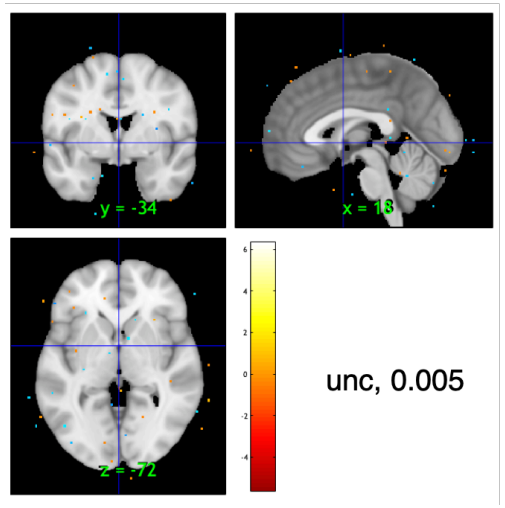
- A stringent method for controlling the FWER is **Bonferroni correction**

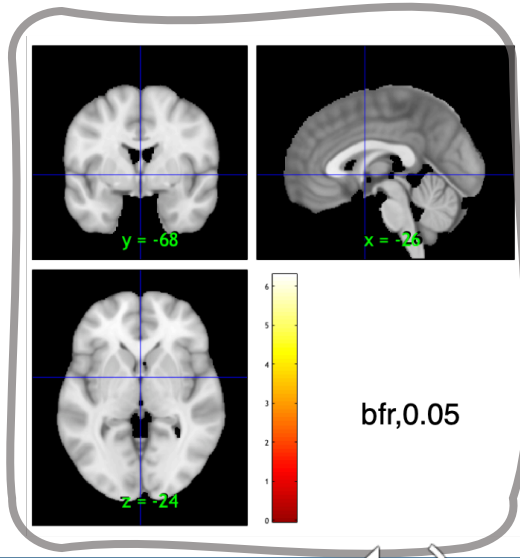
  : alpha (0.05) / number of voxels

  → if have 10,00 voxels, the corrected p-value is 0.000005 (0.05 ./ 10,000)
- Although this approach effectively decreases type I error, it also dramatically increases the probability of a Type 2 error, of failing to detect voxels with real activation



# False-discovery Rate

- **False Discovery Rate** (FDR)
  - A recent development in multiple comparison problems  (Benjamini and Hochberg, 1995)
  - While the FWER controls the probability of any false positives, the FDR controls the proportion of false positives among all rejected tests
- FDR controlling methods
  - Benjamini-Hochberg correction of the FDR
    - A procedure controlling the FDR ensures that on average the FDR is no bigger than a pre-specified rate q which lies between 0 and 1
    - An FDR-controlling technique guarantee controls of the FDR in the sense that FDR ≤ q

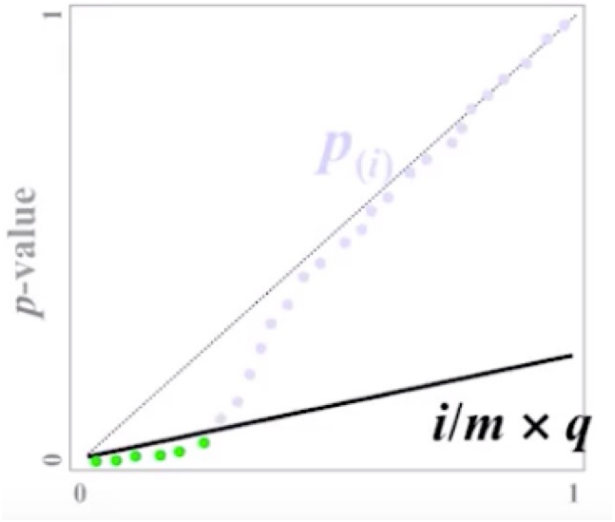
Slide credit: Martin Lindquist and Tor Wager



# False-discovery Rate

- Benjamini-Hochberg correction of the FDR
  - It is an alternative and popular approach for controlling a different quantity, the false discovery rate (FDR)

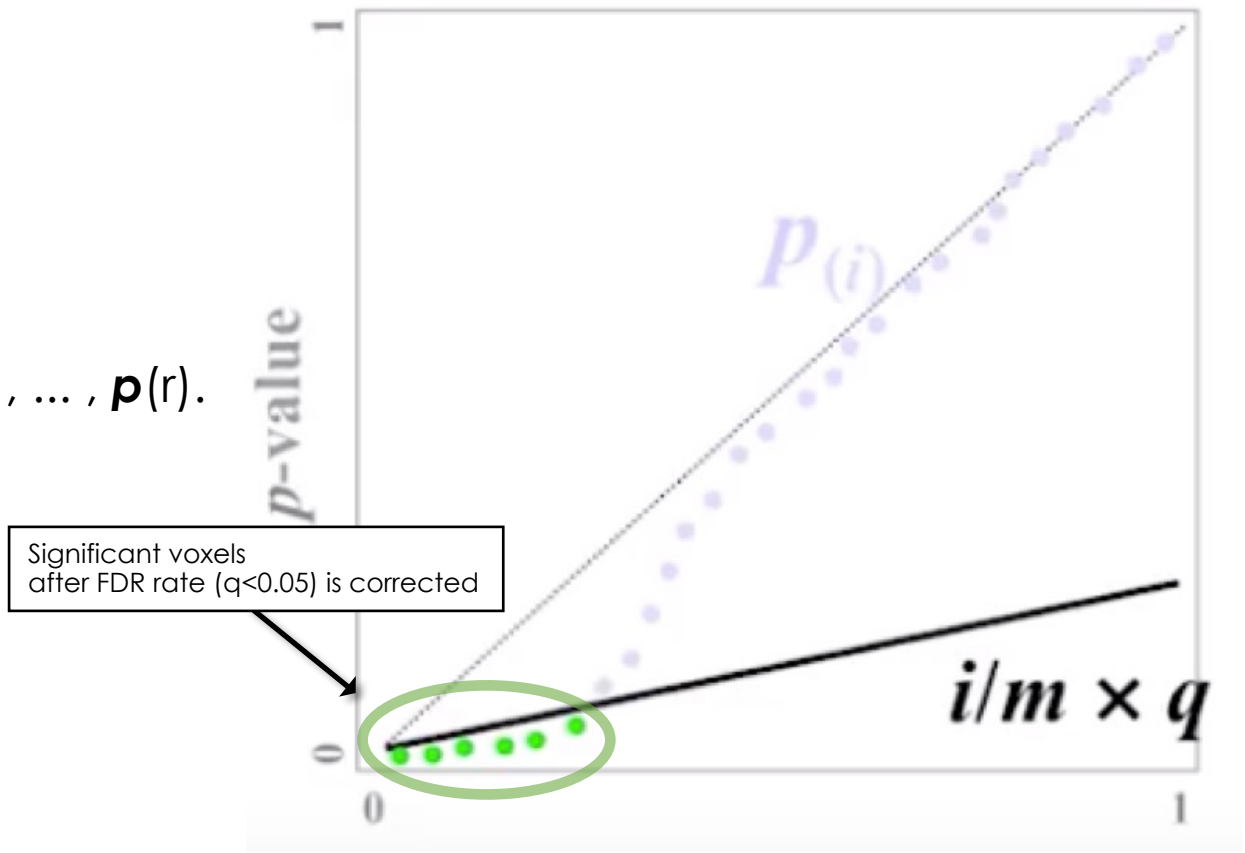
    : describes **the proportion of positive results** (i.e., discoveries) that are actually false positivity
  - Controlling the FDR rather than the FWER provides two primary advantages for fMRI research
    - Less-stringent correction
    - controls the proportion of false claims (activation clusters in a table), not tests (voxels)



# False-discovery Rate

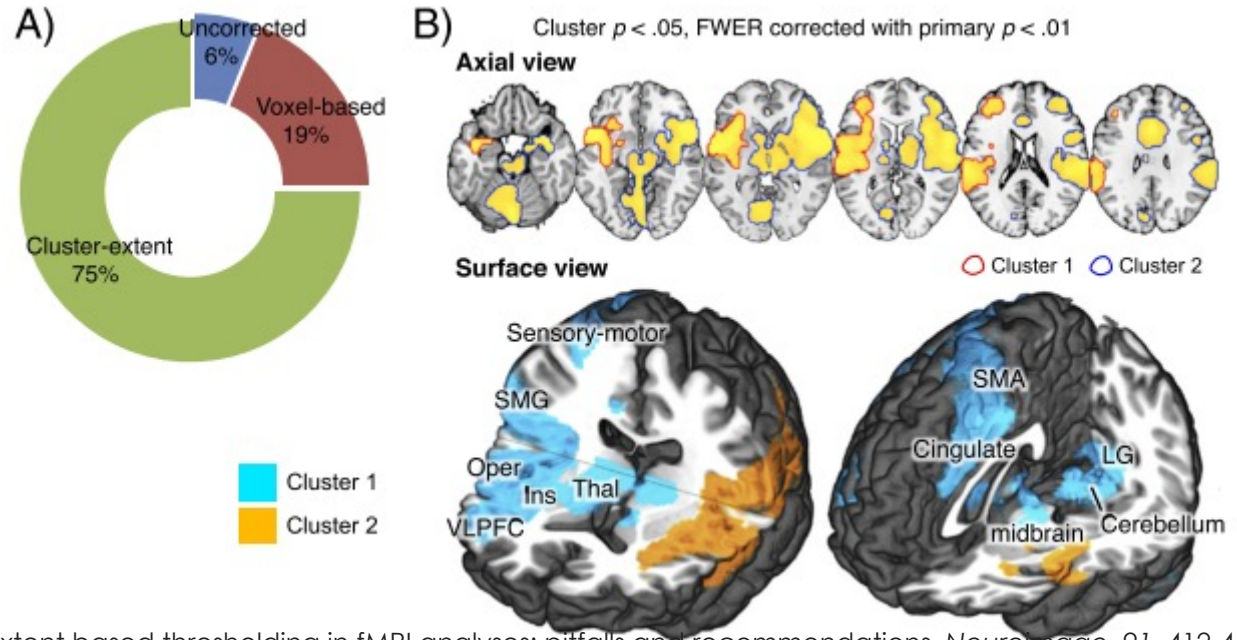
- Benjamini-Hochberg correction of the FDR
  1) Select desired limit q on FDR (e.g., 0.05)
  2) Rank p-values, $\boldsymbol{p}(1) \leq \boldsymbol{p}(2) \leq \ldots \leq \boldsymbol{p}(m)$
  3) Let $\boldsymbol{r}$ be largest $\boldsymbol{i}$ such that $\boldsymbol{p}(i) \leq \boldsymbol{i}/m \times q$
  4) Reject all hypotheses corresponding to $\boldsymbol{p}(1), \ldots, \boldsymbol{p}(r)$.

```
function [pID,pN] = FDR(p,aFDR)
% FORMAT [pID,pN] = FDR(p,aFDR)
%
% p     - vector of p-values
% aFDR  - alpha_FDR, aka q, the desired False Discovery Rate level
%
% pID - p-value threshold based on independence or positive dependence
% pN  - Nonparametric p-value threshold
%_______________________________________________________________________
% Author: T. Nichols
% Version: http://github.com/nicholst/matlab/tree/$Format:%h$
%          $Format:%ci$
```



# Cluster-level inference

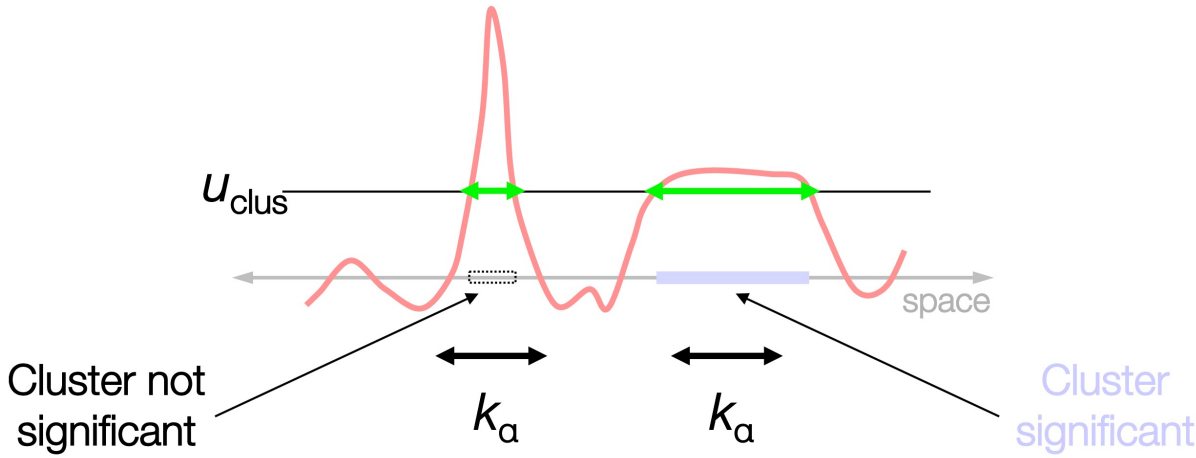
- Voxel-level multiple comparison correction did not consider spatial features of fMRI data
- **Cluster-based correction** is based on the assumption that in an FMRI dataset composed of several tens of thousands of voxels all abutting each other, there is likely to be some correlation in the observed signal between adjacent voxels (a group of contiguous voxels)




Woo, C. W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *NeuroImage*, *91*, 412-419.

# Cluster-level inference

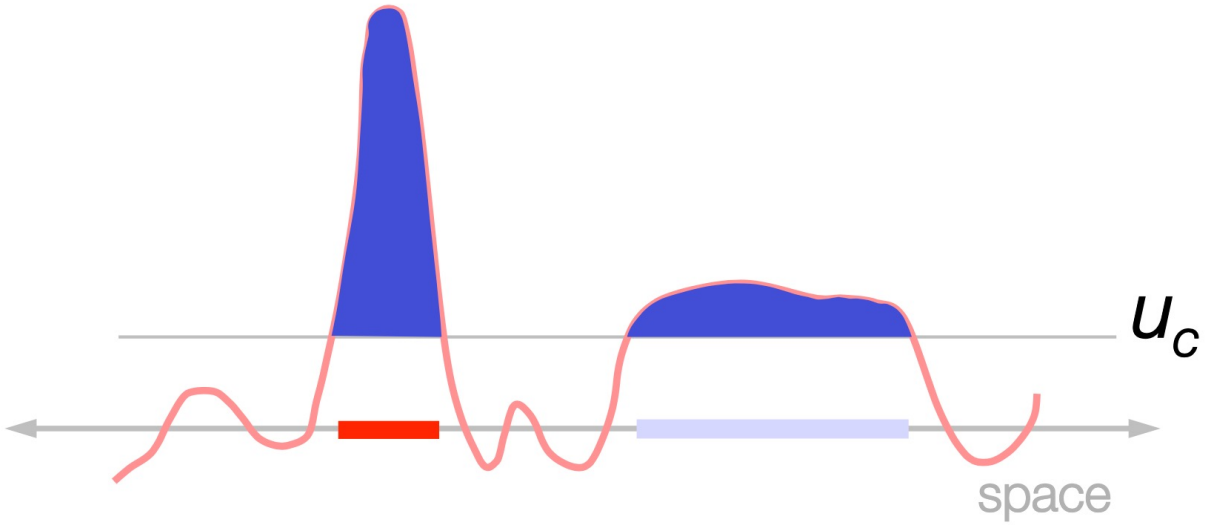
- **Two step-process**
  - Define clusters by arbitrary threshold $u_{clus}$
  - Retain clusters larger than a-level threshold $k_a$
- Typically, better sensitivity
- Worse spatial specificity
  - The null hypothesis of entire cluster is rejected
  - Only means that one or more voxels in cluster active




Slide credit: Martin Lindquist and Tor Wager

# Cluster-level inference

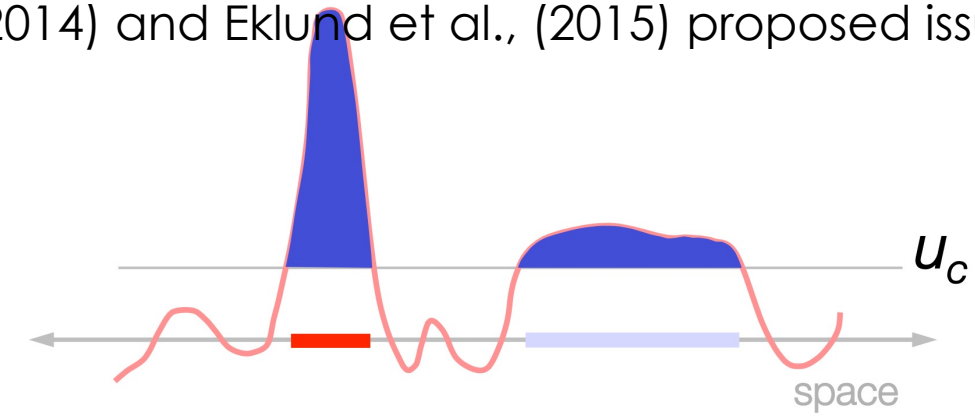
- Threshold-free cluster enhancement
  - Combining Cluster Size with Intensity Information
  - More powerfully combines **peak** & **height** (Hayasaka & Nichols, 2004)




Slide credit: Martin Lindquist and Tor Wager

# Cluster-level inference

- **Limitations**
  - If meaningful voxel size is smaller than cluster-size threshold we choose
  - If activation shape (or pattern) is non-spherical, as when running linear along the edge of a gyrus
  - Because of spatial correlation, significant voxels tend to cluster together, even if their significance results form noise process (how we can handle?)
  - Recently, Woo et al., (2014) and Eklund et al., (2015) proposed issues related this cluster-extent methods




Slide credit: Martin Lindquist and Tor Wager

# Summary

- FWER controlling methods:
  - Bonferroni correction
  - Random Field Theory
  - Permutation Tests
- FDR controlling methods
  - Benjamini-Hochberg correction of the FDR
- Cluster extent-based correction is for both FEWR and FDR



# Recommended Materials

- Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations (Woo et al., *Neuroimage*, 2014)
- **Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates** (Eklund et al., *PNAS*, 2016)
- Probabilistic TFCE: a generalized combination of cluster size and voxel intensity to increase statistical power (Spisak et al., *Neuroimage*, 2018)
- Andy's Brain Book (https://andysbrainbook.readthedocs.io/en/latest/index.html)
- Mumford brain stats (https://www.youtube.com/channel/UCZ7gF0zm35FwrFpDND6DWeA)
- Multiple Testing Toolbox in MATLAB (https://kr.mathworks.com/matlabcentral/fileexchange/70604-multiple-testing-toolbox)