

MGH/HST Athinoula A. Martinos
Center for Biomedical Imaging



MASSACHUSETTS
GENERAL HOSPITAL

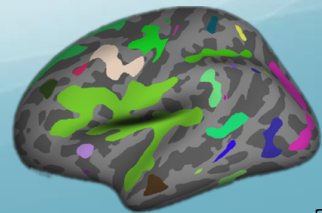


Harvard-MIT
Health Sciences & Technology

Correction for multiple comparisons in FreeSurfer

Overview

- The problem of multiple comparisons
- Clusters and Cluster-forming threshold (CFT)
- Parametric Correction
- Inflated False Positive Rates (Eklund, et al, 2016)
- Non-parametric Correction (Permutation)
- Using the software

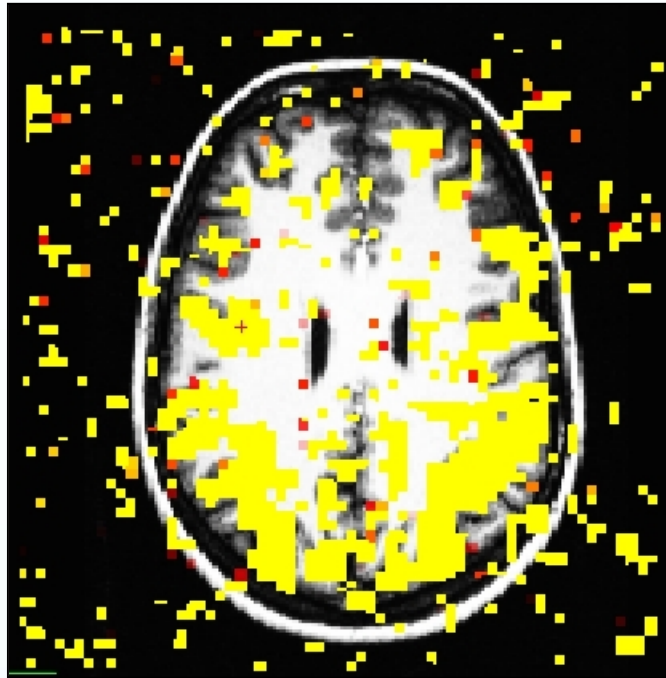


What does a p-value mean?

fMRI

$p < 0.10$

Sig > 1

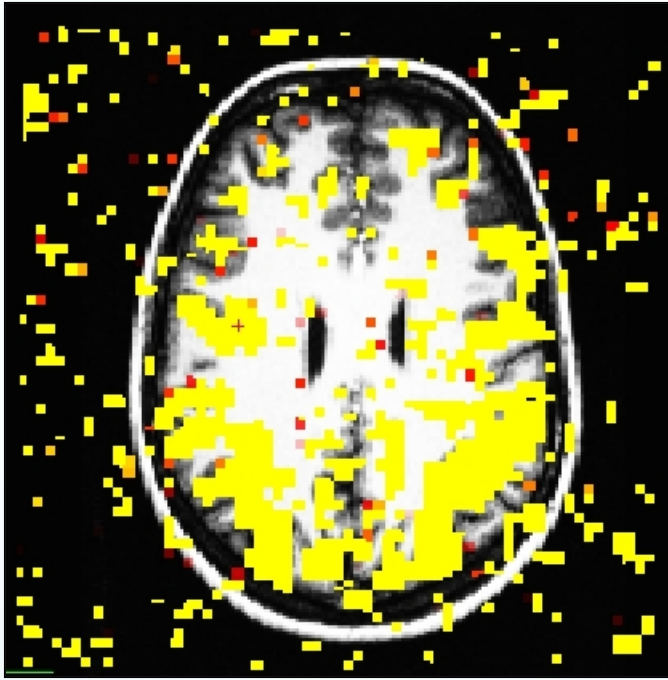


Task

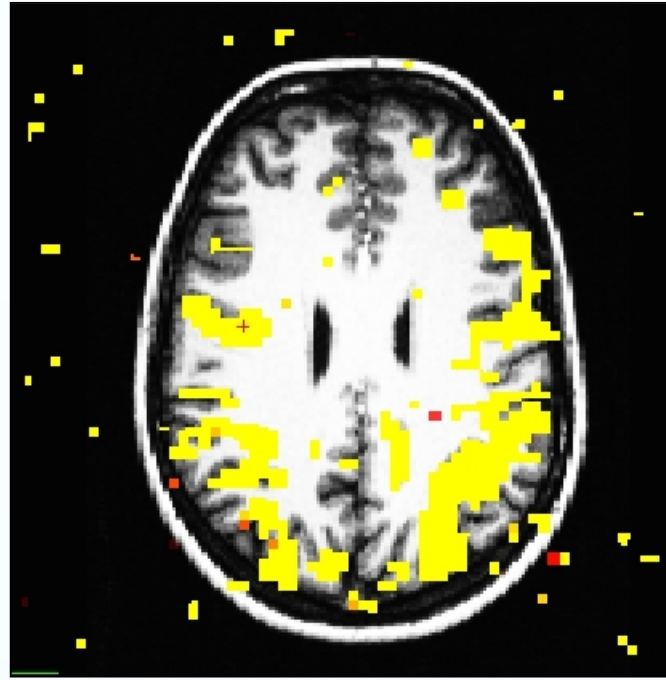
- flashing checker board
- auditory tone
- button press

p-value: probability of a positive under the Null Hypothesis (false positive, eg, background)

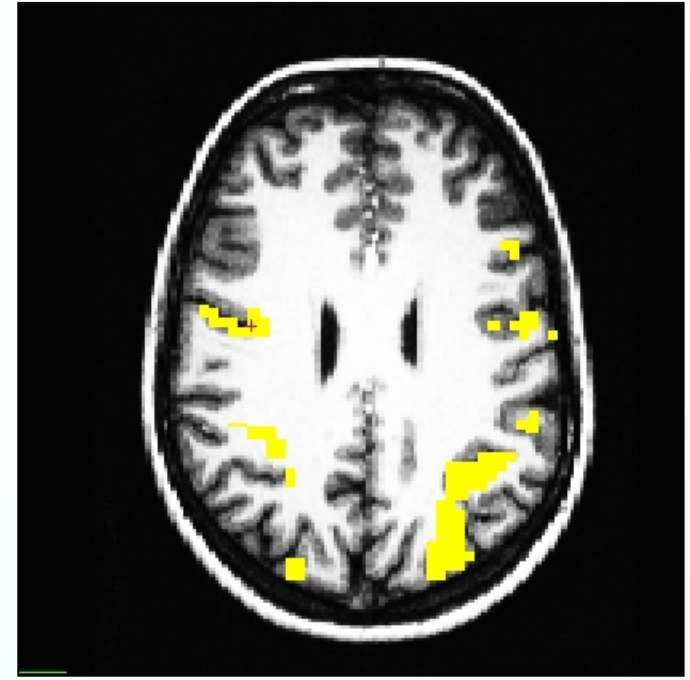
Problem of Multiple Comparisons



$p < 0.10$



$p < 0.01$

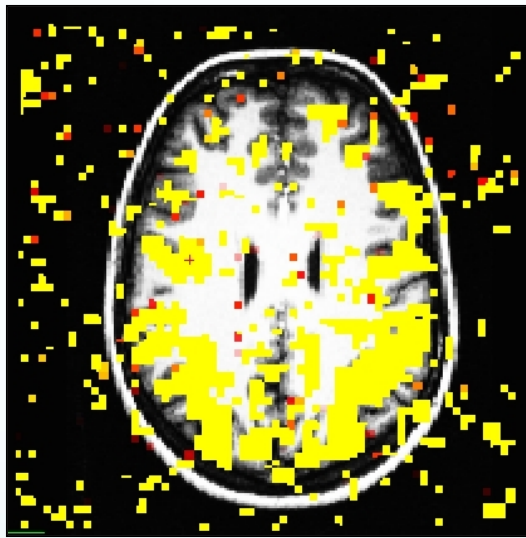


$p < 10^{-7}$

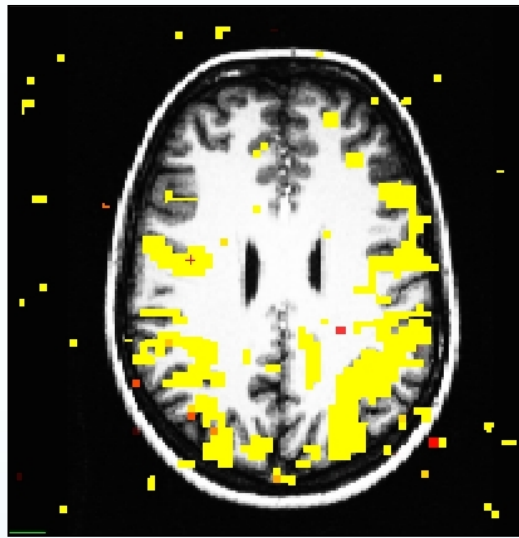
Problem: how can you control false positives adequately while not destroying all your activation?

Clustering

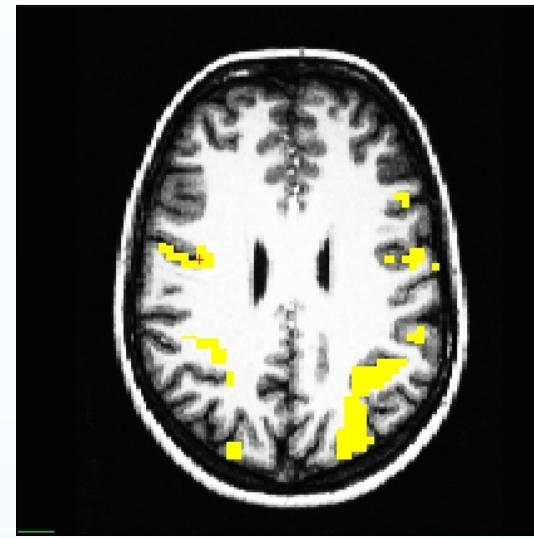
$p < 0.10$



$p < 0.01$



$p < 10^{-7}$

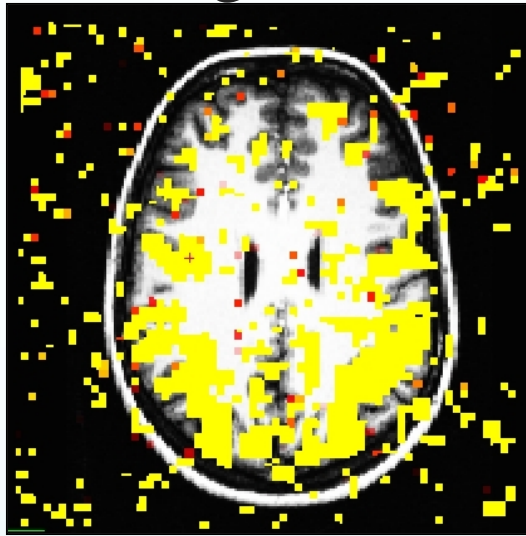


- True signal tends to be clustered
- False Positives tend to be randomly distributed in space
- Cluster – set of spatially contiguous voxels that are above a given threshold.

Voxel-wise Significance in FS

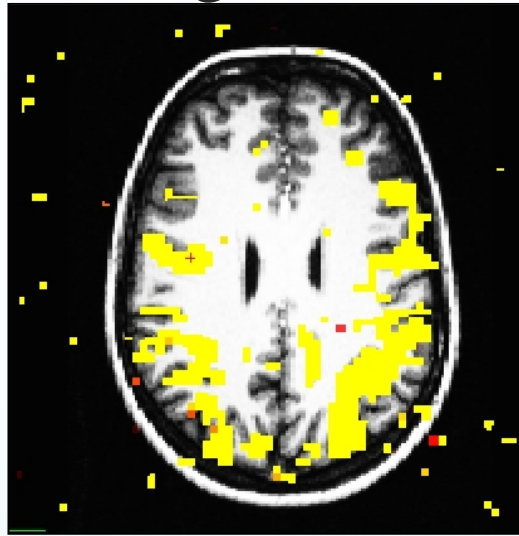
$$p < 0.10$$

$$\text{Sig} > 2$$



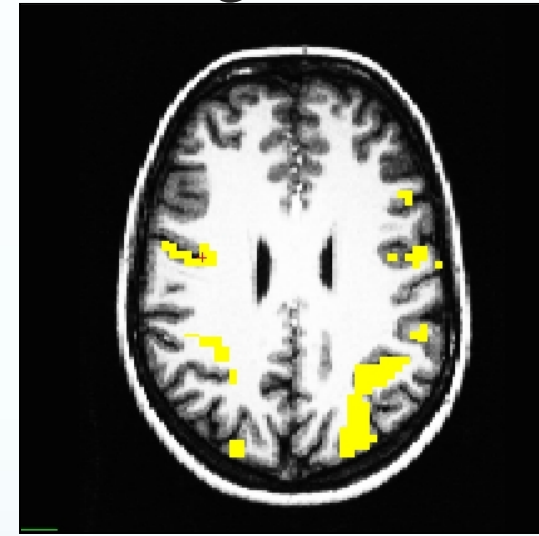
$$p < 0.01$$

$$\text{Sig} > 3$$



$$p < 10^{-7}$$

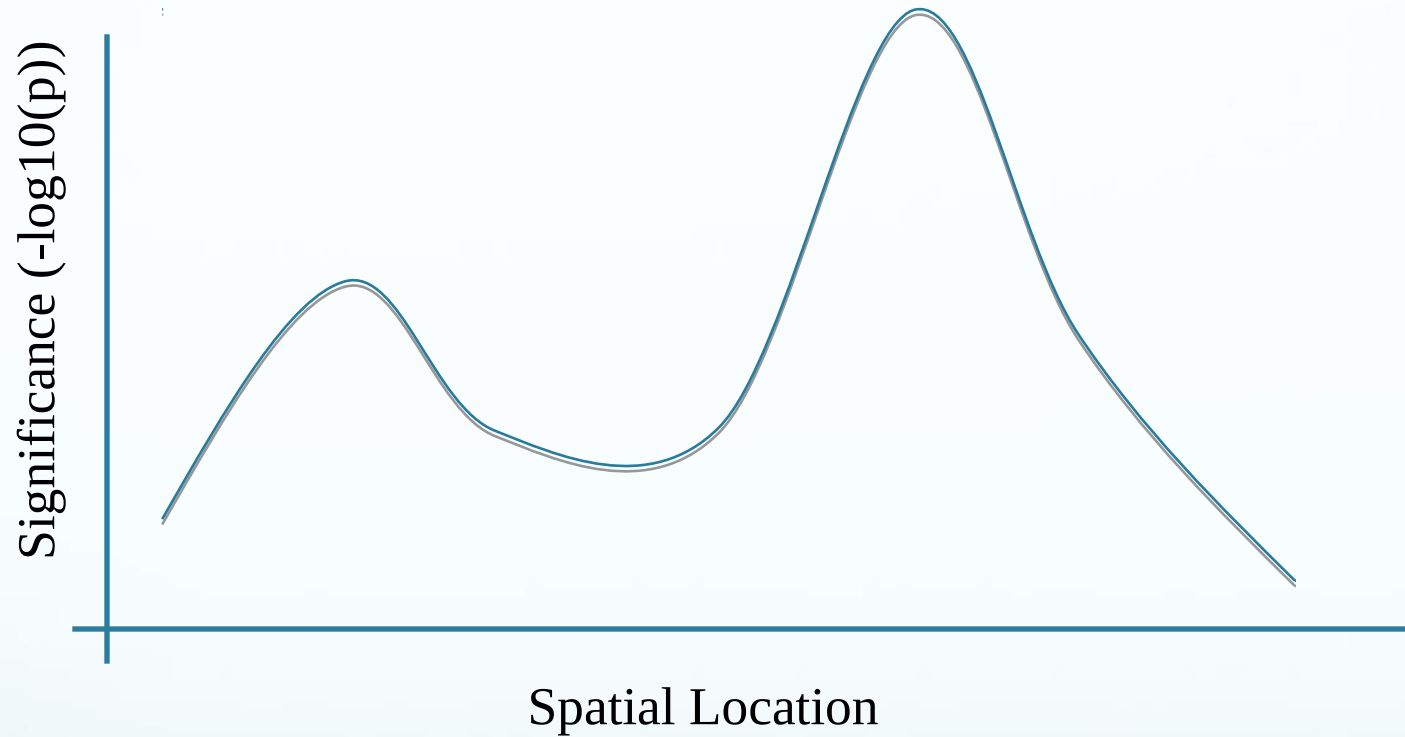
$$\text{Sig} > 7$$



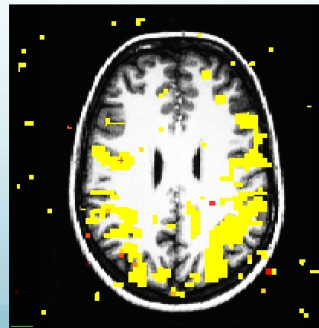
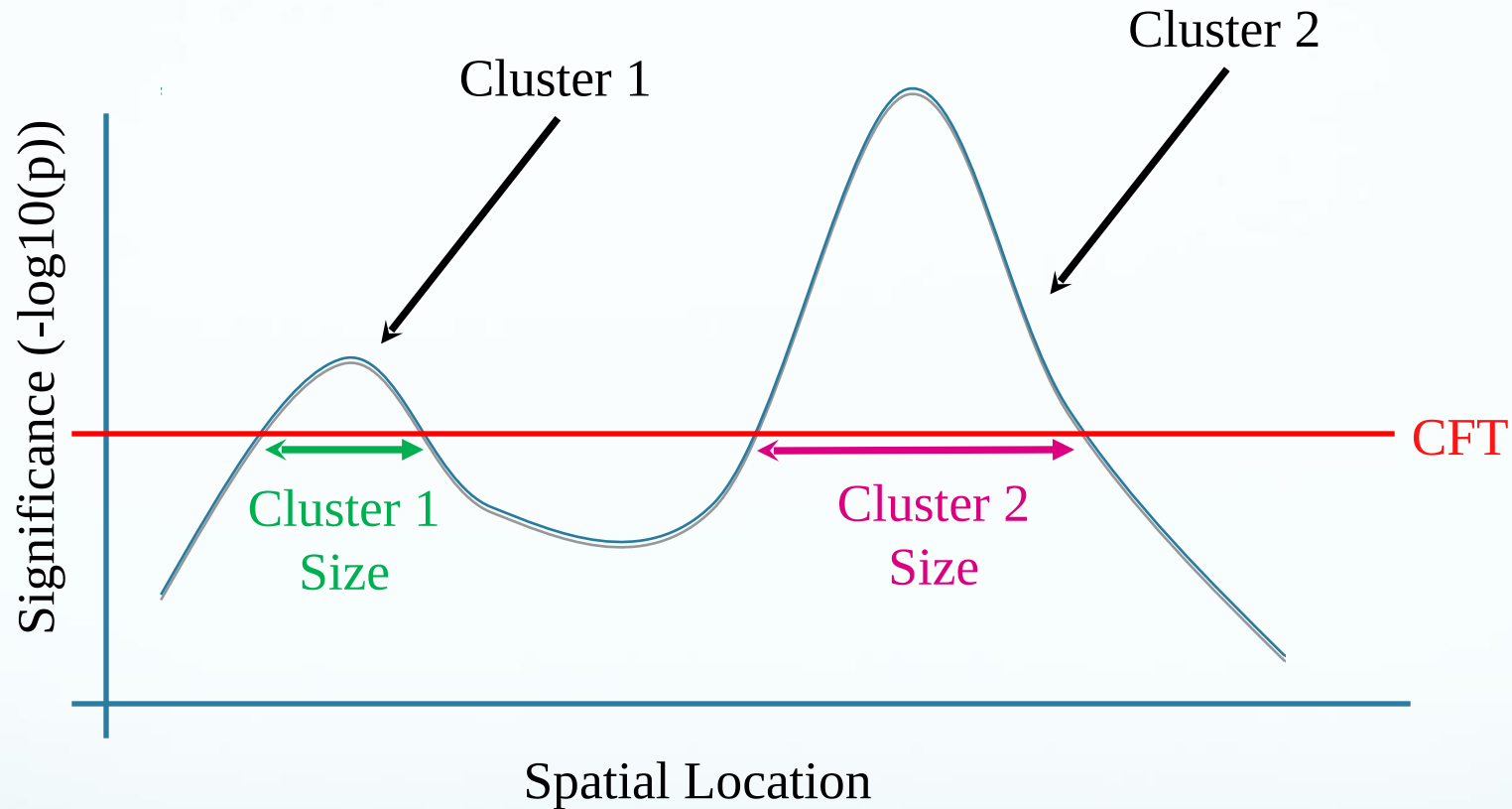
$\text{Sig} = -\log_{10}(\text{pvalue})$

Bigger is better

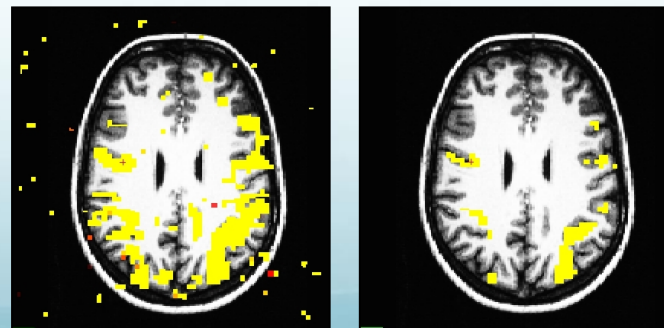
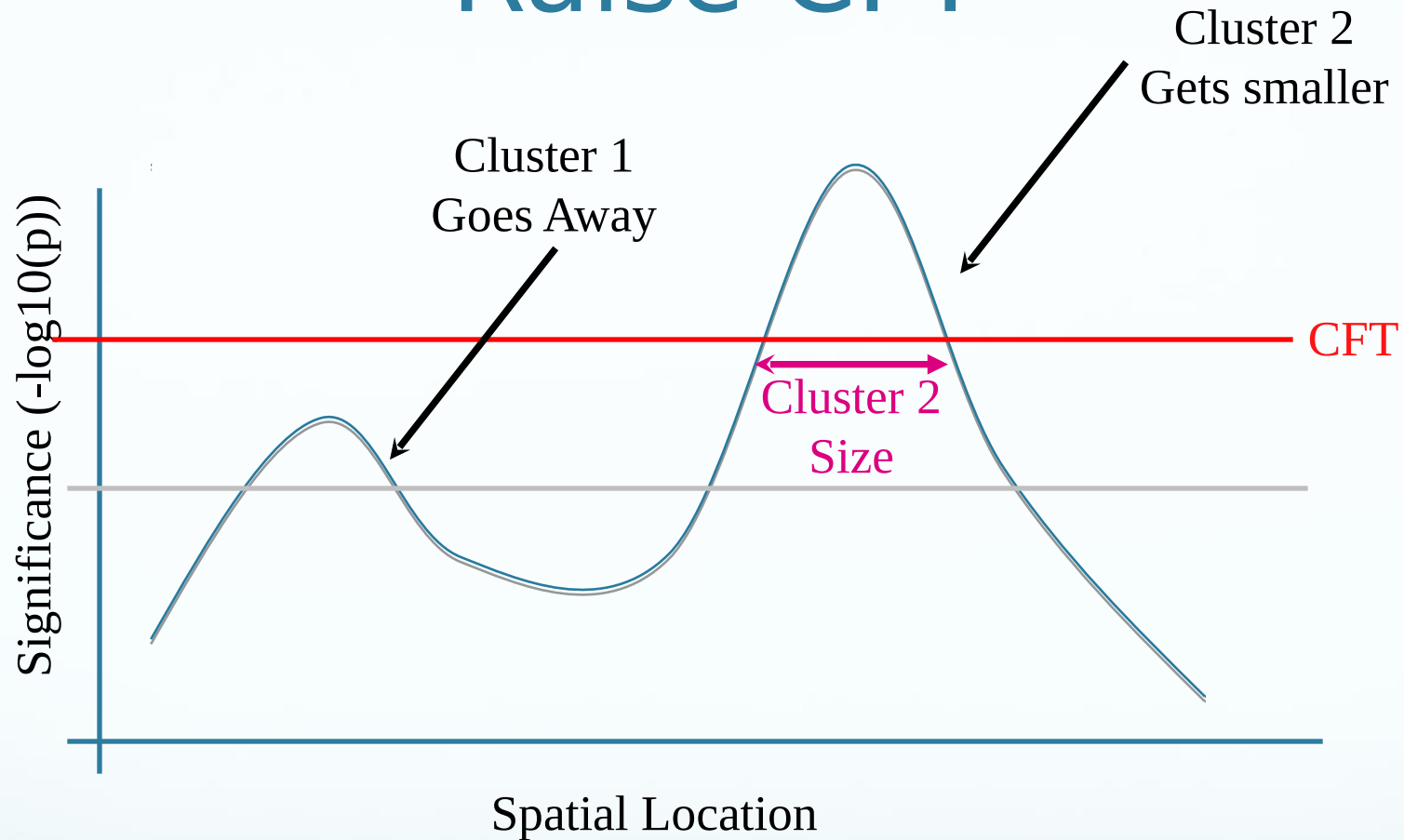
Cluster-forming Threshold (CFT)



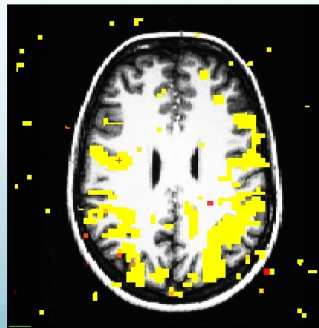
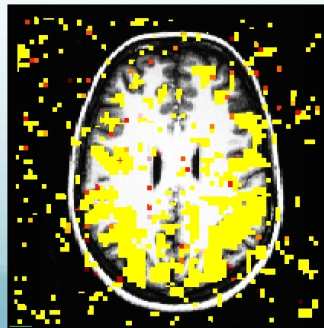
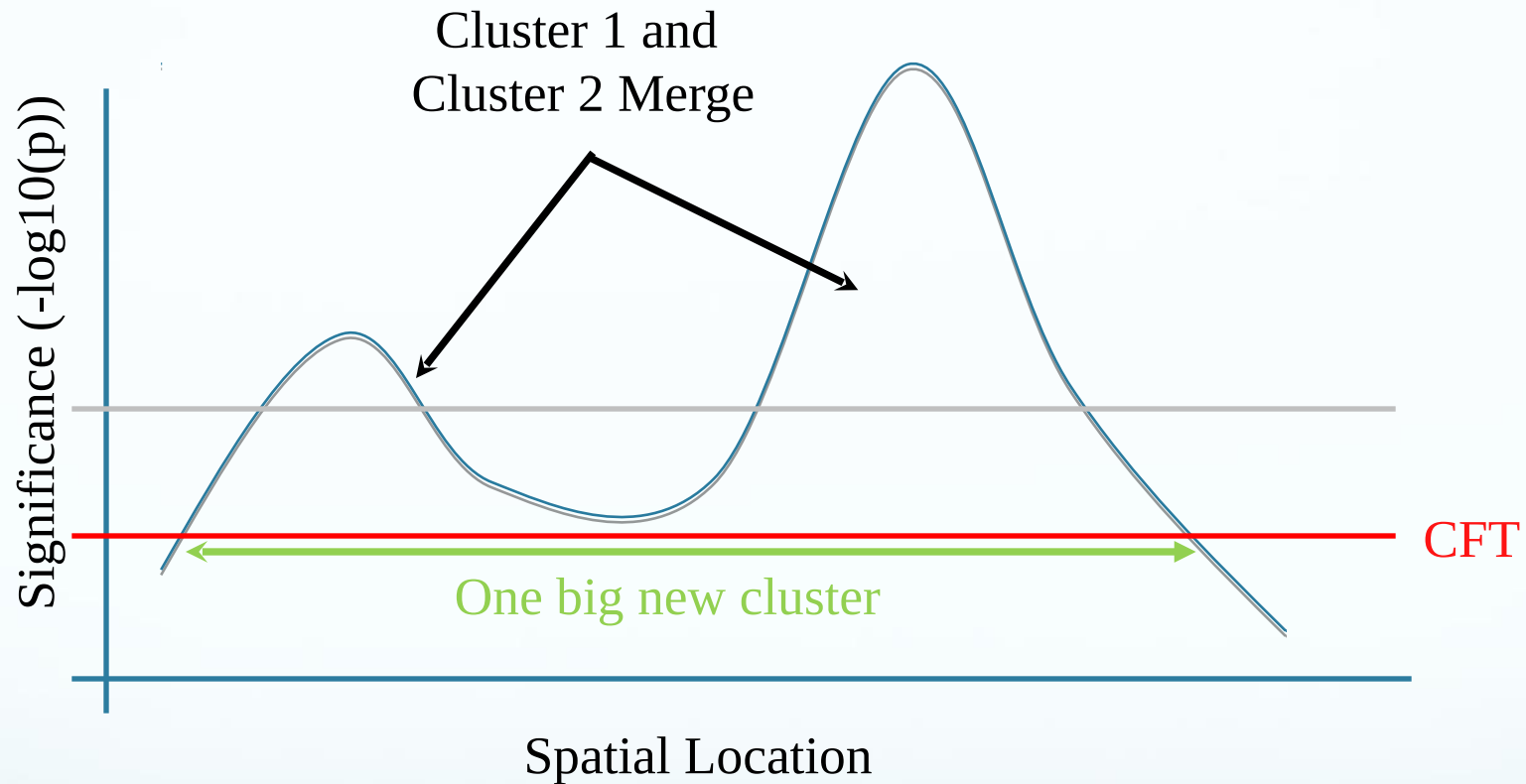
Cluster-forming Threshold (CFT)



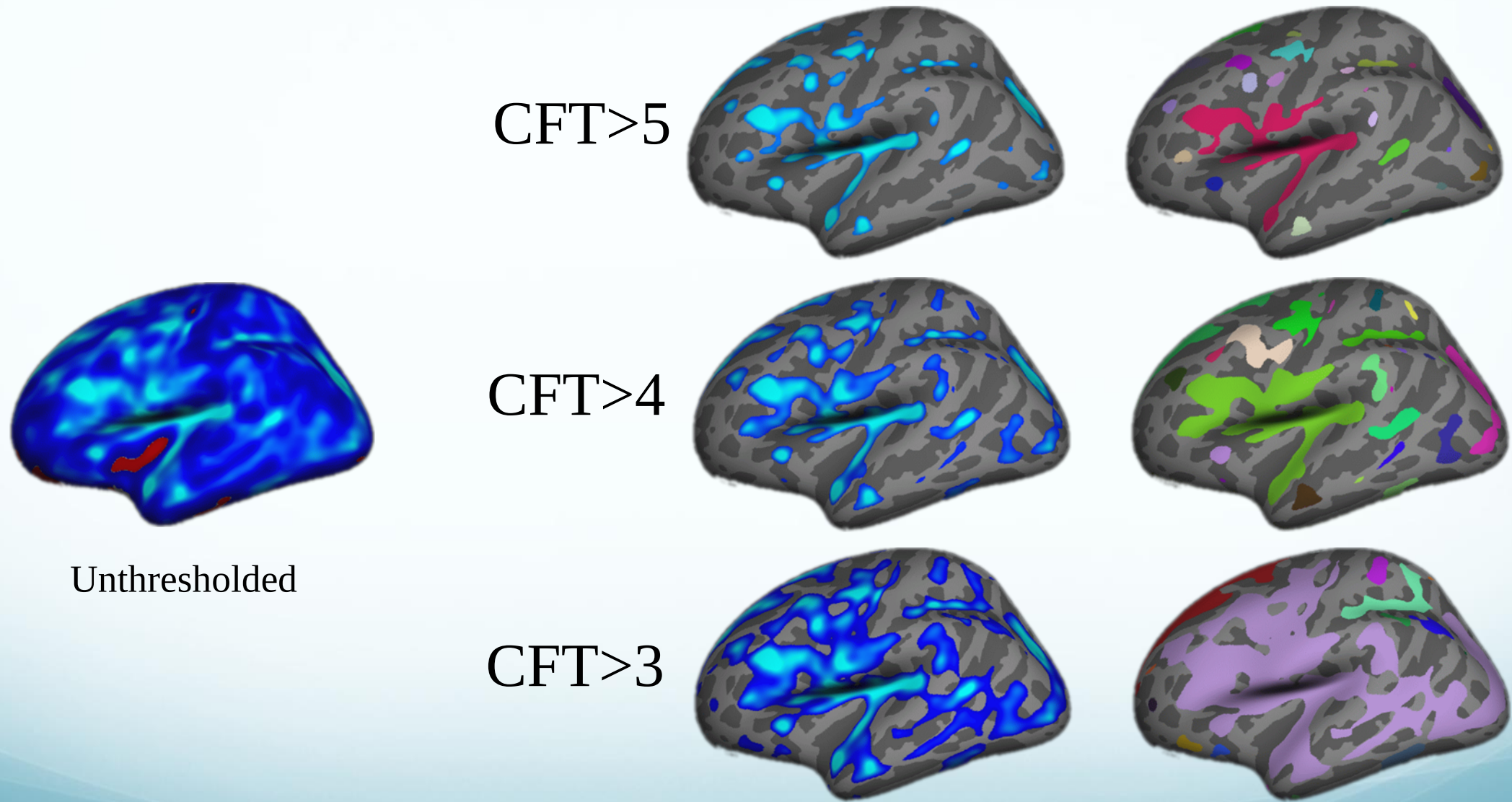
Raise CFT



Lower CFT



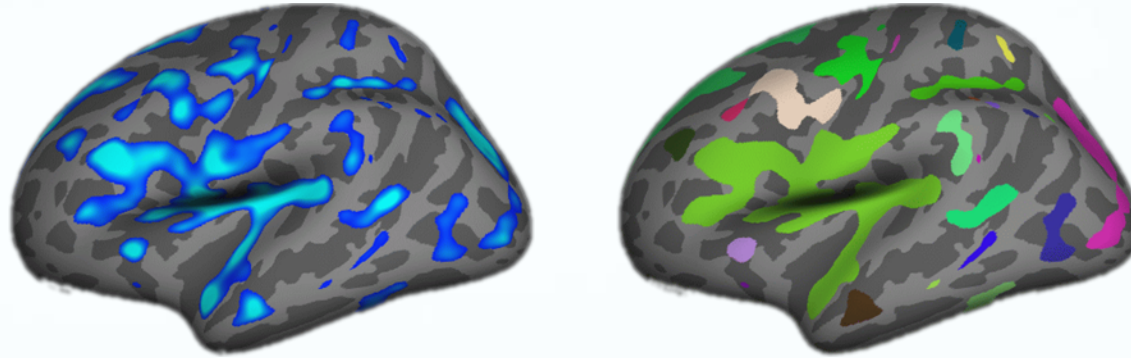
Cluster-forming Threshold (CFT)



As threshold lowers, clusters may expand or merge and new clusters can form.
There is no way to say what the threshold is best.

Cluster Table, Uncorrected

CFT>4



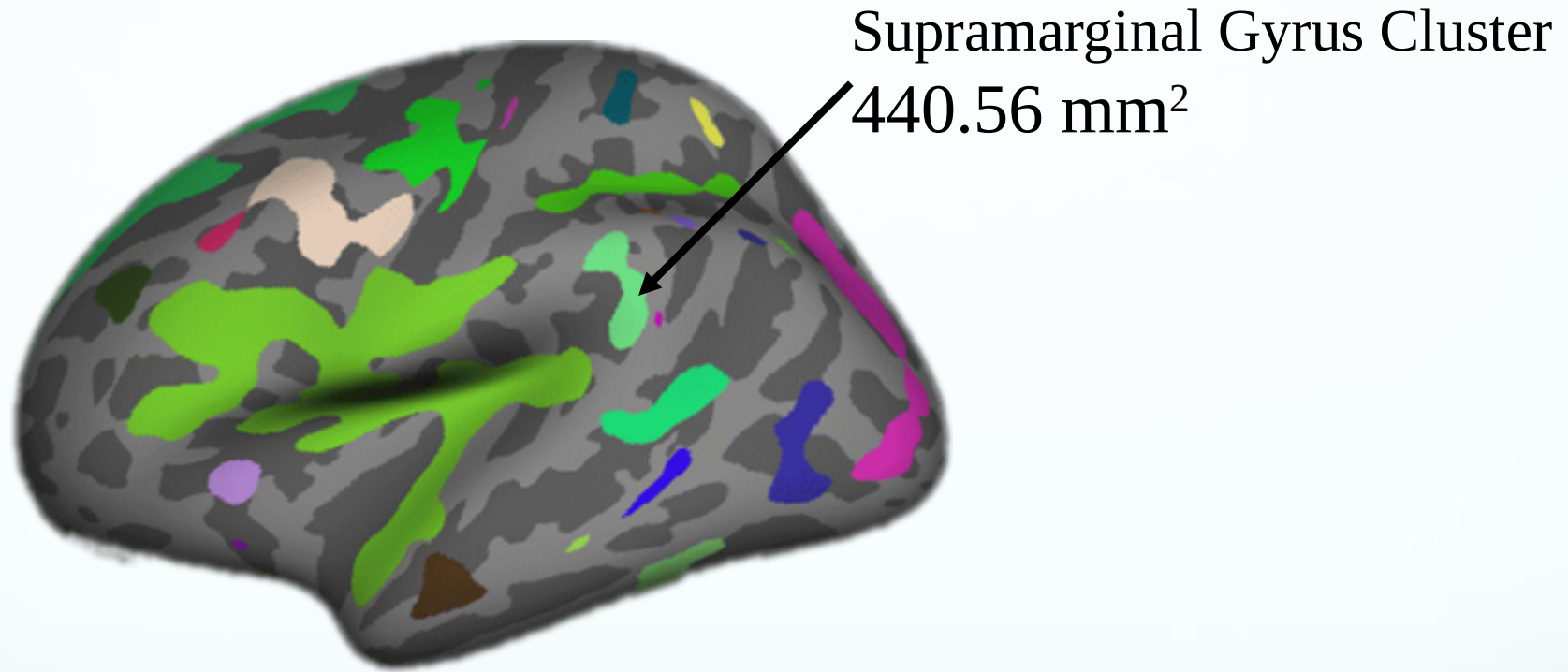
38 clusters

ClusterNo	Area(mm ²)	X	Y	Z	Structure
Cluster 1	3738.82	-11.1	34.5	27.2	superiorfrontal
Cluster 2	5194.19	-32.4	-23.3	15.7	insula
Cluster 3	1271.30	-25.9	-75.0	19.0	superiorparietal
Cluster 4	775.38	-44.4	-9.7	51.3	precentral
Cluster 5	440.56	-33.0	-36.8	37.5	supramarginal

...

How likely is it to get a cluster of a certain size under the null hypothesis?

Clusterwise Correction

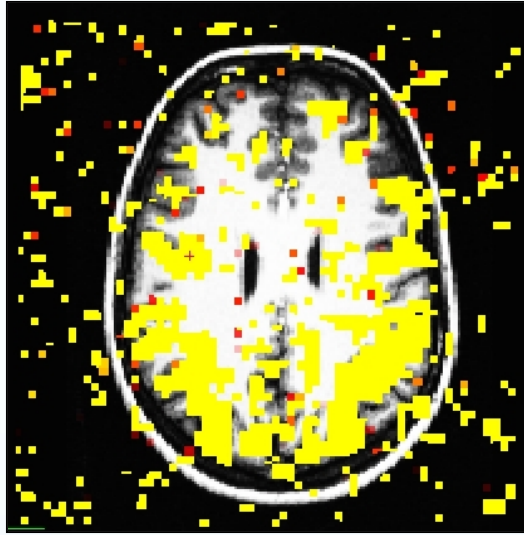


How likely is it to get a cluster 440.56mm² or bigger by chance?

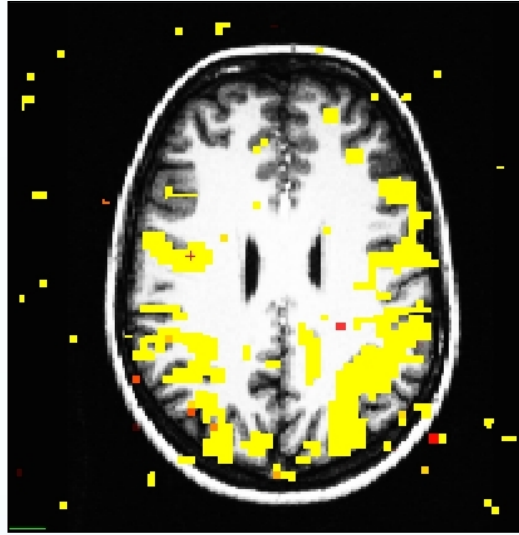
How likely is it to get a cluster of a certain size under the null hypothesis?

Cluster size under the Null?

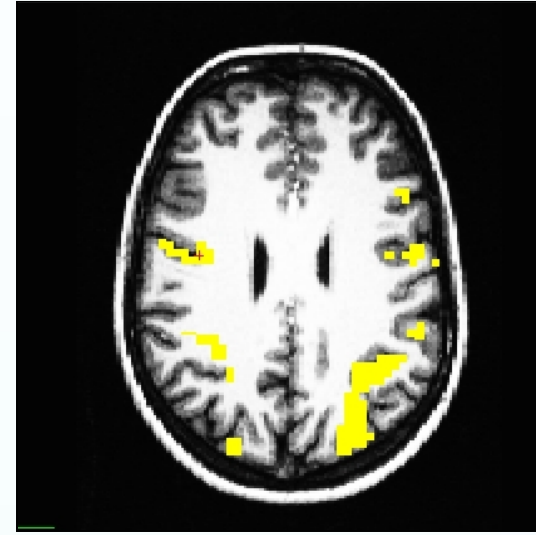
CFT > 2



CFT > 3



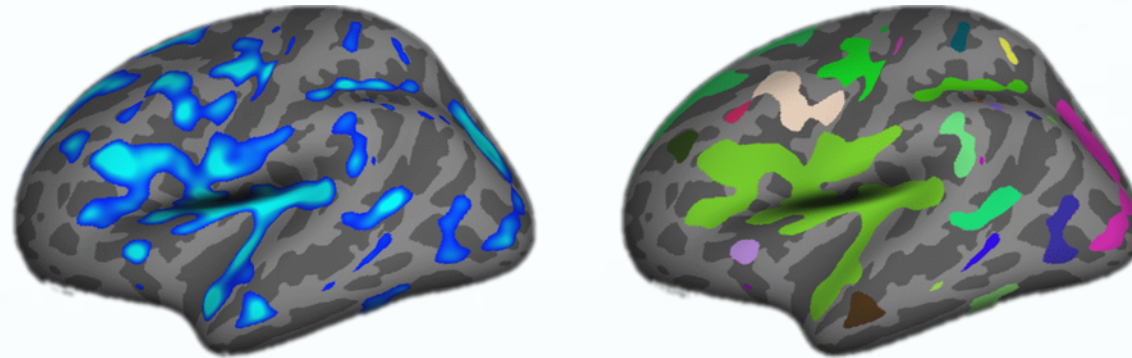
CFT > 7



Going back to fMRI, the background is the Null
How often do you see a cluster of a given size or bigger in the background? That is the p-value of the cluster!
Generally not so easy, especially on the surface.

Cluster Table, Corrected

CFT>4



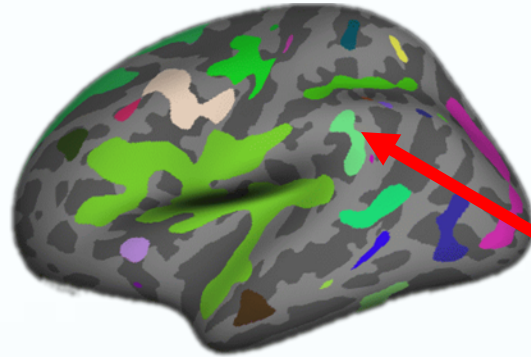
22 clusters out of 38 have cluster p-value < 0.05

ClusterNo	Area(mm ²)	X	Y	Z	Structure	Cluster P
Cluster 1	3738.82	-11.1	34.5	27.2	superiorfrontal	.0001
Cluster 2	5194.19	-32.4	-23.3	15.7	insula	.0003
Cluster 3	1271.30	-25.9	-75.0	19.0	superiorparietal	.0050
Cluster 4	775.38	-44.4	-9.7	51.3	precentral	.0100
Cluster 5	440.56	-33.0	-36.8	37.5	supramarginal	.0400

...

This table would go into your results section.

What does a cluster p-value mean?



ClusterNo	Area(mm ²)	X	Y	Z	Structure	Cluster P
Cluster 5	440.56	-33.0	-36.8	37.5	supramarginal	.04

What does a cluster p-value of .04 mean? If there is no effect (Null) and you were to repeat this experiment 100 times, you would expect to declare a false positive 4 times (4%). Generally, the cluster p-value must be less than .05 for publication.

How to compute a cluster p-value?

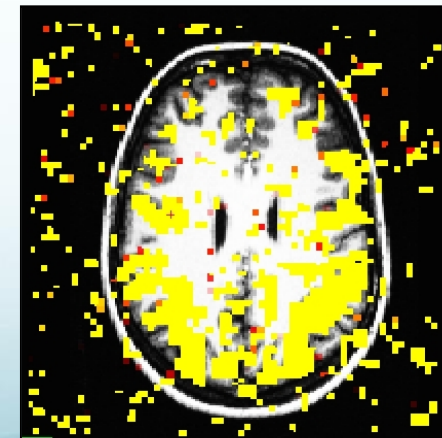
How likely is it to get a cluster of a certain size or bigger by chance, ie, under the Null (nothing is really happening)?

- Parametric:

- Gaussian Random Field Theory (RFT)
- Monte Carlo (MC) Simulation (Gaussian smoothness)

- Non-Parametric:

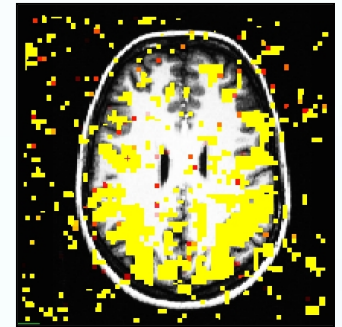
- Permutation



Monte Carlo Simulations (parametric)

- Traditional method in FreeSurfer
 1. Synthesize white Gaussian noise (z-map) on surface (fsaverage)
 2. Smooth by a given FWHM (Gaussian smoothing)
 3. Threshold at a given CFT
 4. Record size of largest cluster
 5. Return to #1, repeat 10,000 times

Process is the Null Hypothesis

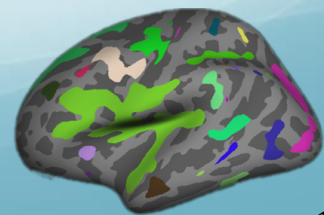


Pre-computed and distributed tables with FS

- FWHM=1-30mm
- CFT Sig thresholds (1.3, 2, 2.3, 3.0, 3.3, 4.0)
- Sign (positive, negative, unsigned (abs))
- Application: measure FWHM in data (fwhm.dat)

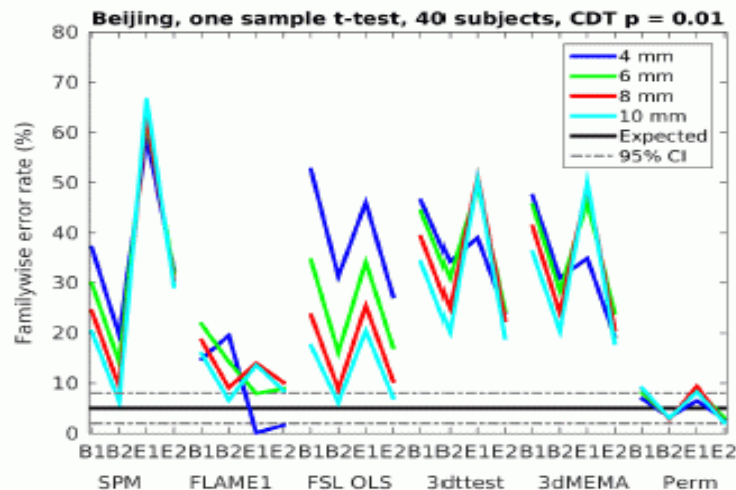
MC Simulations: Application

- Analyze your data (mri_glmfit)
 - Measures FWHM in data (fwhm.dat, eg 10.4mm)
- Run mri_glmfit-sim
 - Set CFT threshold (eg, 2) and sign (eg, abs)
 - Gets list of cluster sizes (eg, 440.56mm²)
 - Goes to MC table (FWHM, threshold, sign, size)
 - Gets prob of cluster of that size under MC assumptions
 - Eliminates clusters with $p > .05$
 - Any clusters left? Declare a positive
 - $p < .05$ means False Positive Rate=5%
- Quick – a few seconds
- But ...



fMRI: Eklund, et al, 2016 PNAS

- Resting state fMRI analyzed as task data (Null)
- Full fMRI analysis pipeline
- Group analysis (N=20), Clusters $p < .05$ (RFT)
- Expected False Positive Rate of 5%
- Actual rates were 10-60% (smoothing, CFT/CDT)
- Non-gaussian Smoothness

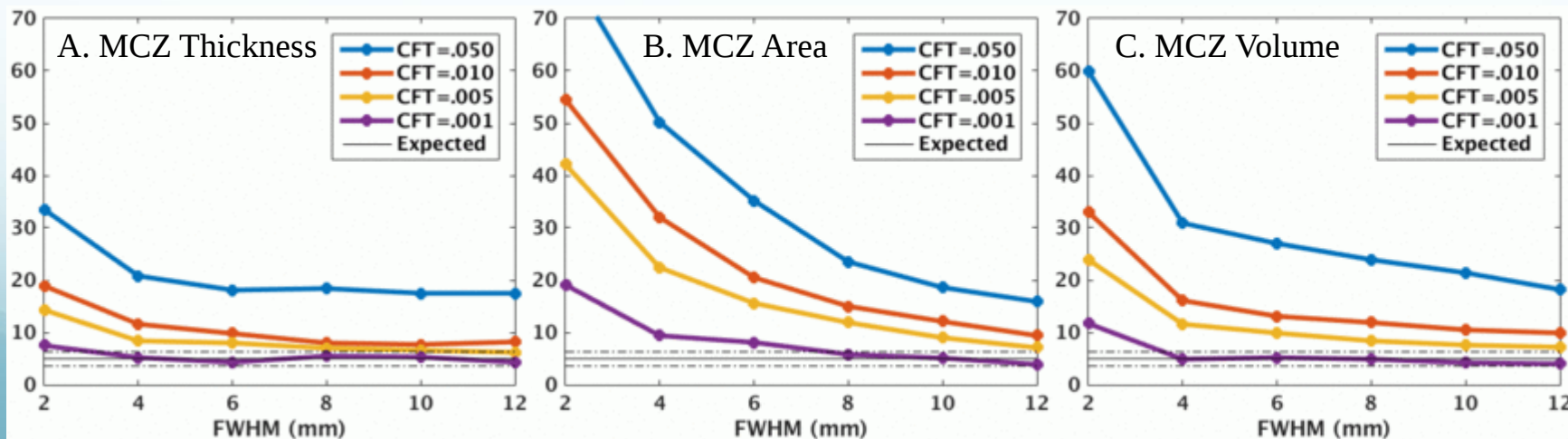


Cluster Failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Eklund, Nichols, Knutsson, 2016. PNAS

CDT=Cluster Defining Threshold = CFT

Greve and Fischl, 2017, NI

- Analyzed 500 young subjects in FS
- Thickness, vertex-wise surface area, and vertex-wise volume
- Randomly chose 20, randomly assigned to two groups
- Test for a group difference (should not be there, Null)
- Use MCSim to declare clusters to be positive ($p < .05$)
- Elevated false positive rates (Area, Volume worse)
- Non-gaussian Smoothness



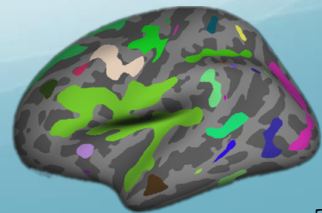
Permutation Simulations

If there is no effect of group, then group membership can be randomly changed.

1. Permute (real) data (ie, randomly change labels) – Null
2. Analyze using GLM
3. Threshold at a given CFT
4. Record size of largest cluster
5. Return to #1, repeat 1000-10,000 times (table)

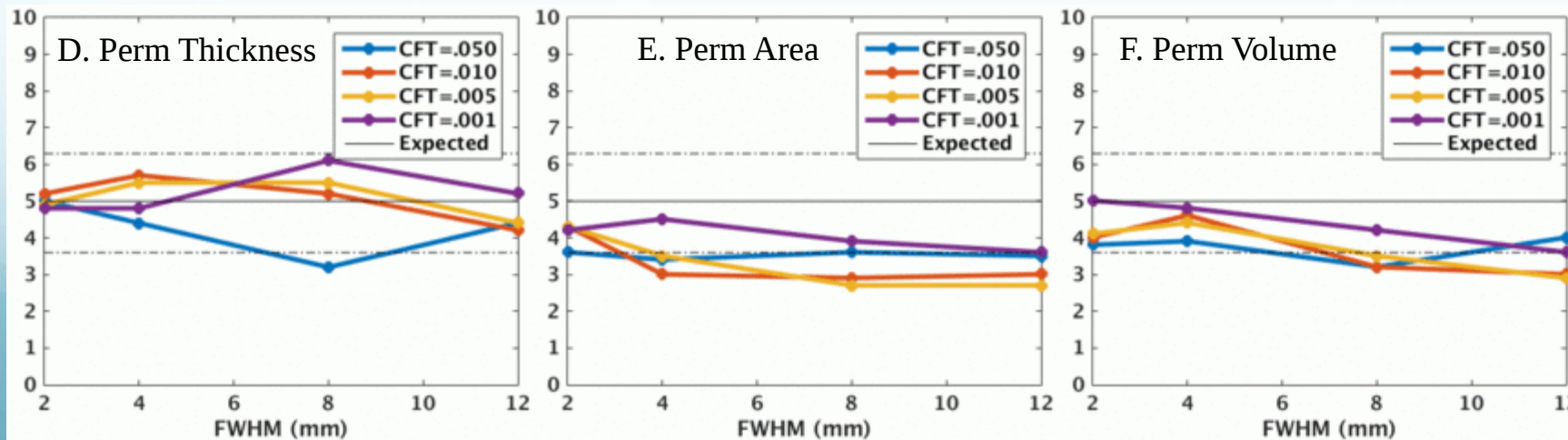
- Apply

1. Analyze real data using true order in GLM
2. Threshold at the given CFT
3. Extract clusters, get cluster p-value from table
4. Eliminate clusters with $p > .05$



Greve and Fischl, 2017, NI

- Permutation works!
- Takes about 15-30min to run simulation (less if parallel)
- Permutation is complex
 - Exchangeability
 - Shuffling vs Sign Flipping
 - Presence of a covariate (eg, age)



How to do this in FreeSurfer

Surface-based Correction for Multiple Comparisons

- 2D Cluster-based Correction at $p < .05$

```
mri_glmfit-sim  
--glmdir lh.gender_age.glmdir  
--perm 1000 2 pos  
--2spaces  
--cwp .05  
--bg N
```

Patch to allow for non-orthogonal designs using the ter Braak method

- ftp://surfer.nmr.mgh.harvard.edu/pub/dist/freesurfer/6.0.0-patch/mri_glmfit-sim

Surface-based Correction for Multiple Comparisons

- 2D Cluster-based Correction at $p < .05$

mri_glmfit-sim

```
--glmdir lh.gender_age.glmdir  
--perm 1000 2 pos  
--2spaces  
--cwp .05  
--bg N
```

Original mri_glmfit command:

```
mri_glmfit  
--y lh.thickness.sm10.mgh  
--fsgd gender_age.txt  
--C age.mtx -C gender.mtx  
--surf fsaverage lh  
--cortex  
--eres-save
```

```
--glmdir  
lh.gender_age.glmdir
```

```
lh.gender_age.glmdir/  
beta.mgh - parameter estimates  
rvar.mgh - residual error variance  
age/  
sig.mgh - -log10(p), uncorrected  
gamma.mgh, F.mgh  
gender/  
sig.mgh - -log10(p), uncorrected  
gamma.mgh, F.mgh
```

Surface-based Correction for Multiple Comparisons

- 2D Cluster-based Correction at $p < .05$

```
mri_glmfit-sim  
--glmdir lh.gender_age.glmdir  
--perm 1000 2 pos  
--2spaces  
--cwp .05  
--bg N
```

- Permutation simulation 1000 iterations
- positive contrast
- Cluster Forming Threshold = 2

Surface-based Correction for Multiple Comparisons

- 2D Cluster-based Correction at $p < .05$

```
mri_glmfit-sim  
--glmdir lh.gender_age.glmdir  
--perm 1000 2 pos  
--2spaces  
--cwp .05  
--bg N
```

Doing analysis with left hemi but right hemi will be done separately. Need to correct for full search space.

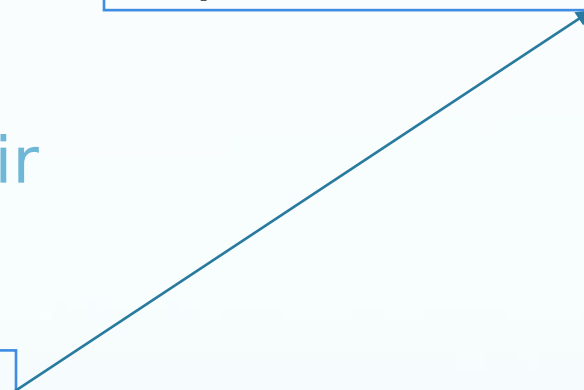


Surface-based Correction for Multiple Comparisons

- 2D Cluster-based Correction at $p < .05$

```
mri_glmfit-sim  
--glmdir lh.gender_age.glmdir  
--perm 1000 2 pos  
--2spaces  
--cwp .05  
--bg N
```

Cluster-wise threshold $p < .05$
Only report on clusters with $p < .05$



Surface-based Correction for Multiple Comparisons

- 2D Cluster-based Correction at $p < .05$

```
mri_glmfit-sim  
--glmdir lh.gender_age.glmdir  
--perm 1000 2 pos  
--2spaces  
--cwp .05  
--bg N
```

Speed up by running multiple (N) simulations in parallel. You must have a computer with multiple processors.

For single thread (N=1), the process can take about 20min

Correction for Multiple Comparisons Output

```
mri_glmfit-sim  
--glmdir  
  lh.gender_age.glmdir  
--perm 1000 2 pos  
--cwp .05  
--2spaces
```

lh.gender_age.glmdir

age

gender

sig.mgh – pre-existing uncorrected p-values

perm.th20.pos.sig.cluster.mgh – map of significance of clusters

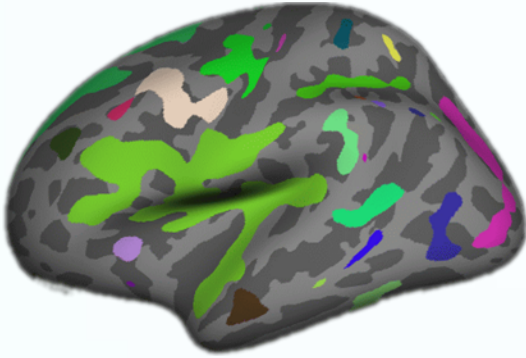
perm.th20.pos.sig.ocn.annot – annotation of significant clusters

perm.th20.pos.sig.cluster.summary – text file of cluster table (clusters, sizes, MNI305 XYZ, and their significance)

- Only shows clusters $p < .05$, change -cwp to a larger value to get more (ie, less sig) clusters



Corrected Outputs



cache.th20.pos.sig.[ocn.annot](#) – annotation of significant clusters

cache.th20.pos.sig.[cluster.summary](#) -- text file of cluster table

ClusterNo	Area(mm ²)	X	Y	Z	Structure	Cluster P
Cluster 1	3738.82	-11.1	34.5	27.2	superiorfrontal	.0001
Cluster 2	5194.19	-32.4	-23.3	15.7	insula	.0003
Cluster 3	1271.30	-25.9	-75.0	19.0	superiorparietal	.0050
Cluster 4	775.38	-44.4	-9.7	51.3	precentral	.0100
Cluster 5	440.56	-33.0	-36.8	37.5	supramarginal	.0400

...

Permutation Analysis of Linear Models (PALM)

- Toolbox developed by Anderson Winkler
- fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM
- Allows for very elaborate models
- Approximations when using
- Ability to speed up simulations
- Matlab or octave
- FreeSurfer interface: fspalm

```
fspalm -glmdir glmdir ...
```

<http://surfer.nmr.mgh.harvard.edu/fswiki/FsPalm>

Winkler AM, et al. Non-Parametric Combination and related permutation tests for neuroimaging. Hum Brain Mapp. 2016 Apr;37(4):1486-511.

Winkler AM, et al, . Multi-level block permutation. Neuroimage. 2015;123:253-68.

Winkler AM, et al, . Faster permutation inference in brain imaging. Neuroimage. 2016 Jun 7;141:502-516

False Discover Correction Possible

- False Discovery Rate (FDR)
 - Built into QDEC
 - Genovese, et al, NI 2002
- `mri_fdr --help`

Tutorial

1. Command-line Stream

- Create an FSGD File for a thickness study
- Age and Gender
- Run
 - `mris_preproc`
 - `mri_surf2surf`
 - `mri_glmfit`
 - `mri_glmfit-sim`

2. QDEC – same data set

QDEC – An Interactive Statistical Engine GUI

Query – Select subjects based on Match Criteria

Design – Specify discrete and continuous factors

Estimate – Fit Model

Contrast – Automatically Generate Contrast Matrices

Interactive – Makes easy things easy (that used to be hard)

Some limitations

- No Query
- Two Discrete Factors (Two Levels)
- Two Continuous Factors
- Surface only
- No correction for multiple comparisons

-qcache

For QDEC to work interactively, you need to run:

```
recon-all -s <sid> -qcache
```

(or as additional flag in your regular processing)

This will map and smooth **thickness** maps to **fsaverage**,
use *-target <id>* to specify your own target
and *-measure <surfmeas>* to specify curv, area, sulc etc.

QDEC – Spreadsheet

qdec.table.dat – spreadsheet with subject information – can be huge!

fsid	gender	age	diagnosis	Left-Cerebral-White-Matter-Vol
011121_vc8048	Female	70	Demented	202291
021121_62313-2	Female	71	Demented	210188
010607_vc7017	Female	73	Nondemented	170653
021121_vc10557	Male	75	Demented	142029
020718_62545	Male	76	Demented	186087
020322_vc8817	Male	77	Nondemented	149810

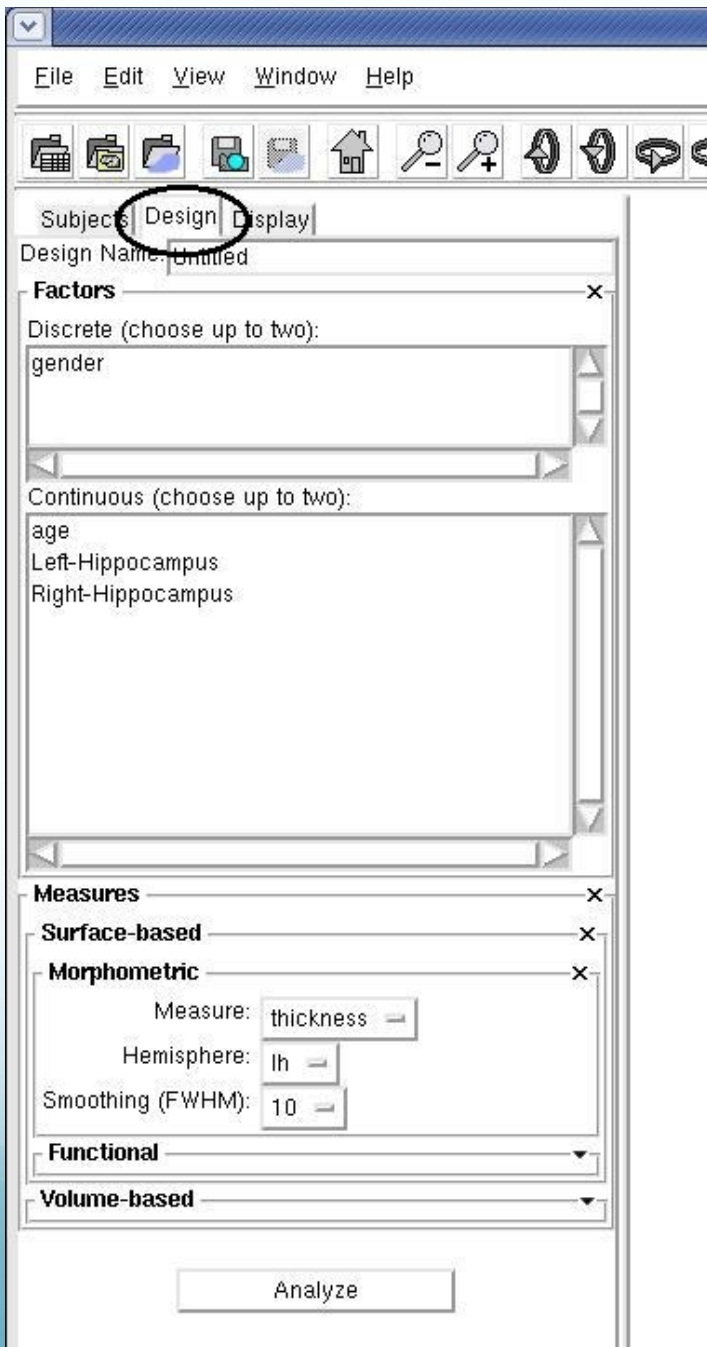
gender.levels

Female
Male

diagnosis.levels

Demented
Nondemented

Discrete Factors need a
factorname.level file



QDEC GUI

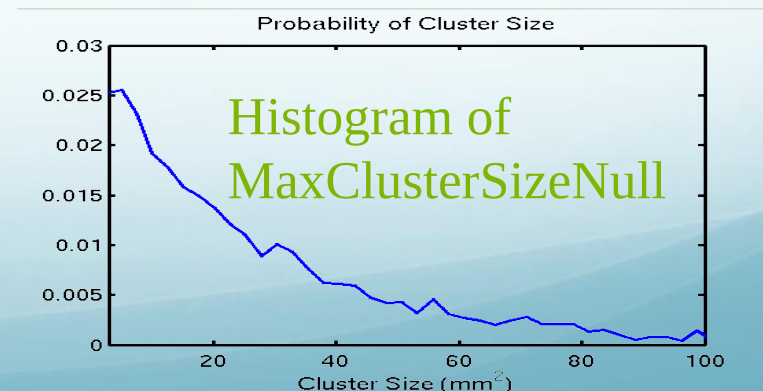
- Load QDEC Table File
 - List of Subjects
 - List of Factors (Discrete and Cont)
- Choose Factors
- Choose Input (cached):
 - Hemisphere
 - Measure (eg, thickness)
 - Smoothing Level
- “Analyze”
 - Builds Design Matrix
 - Builds Contrast Matrices
 - Constructs Human-Readable Questions
 - Analyzes
 - Displays Results

End of Presentation

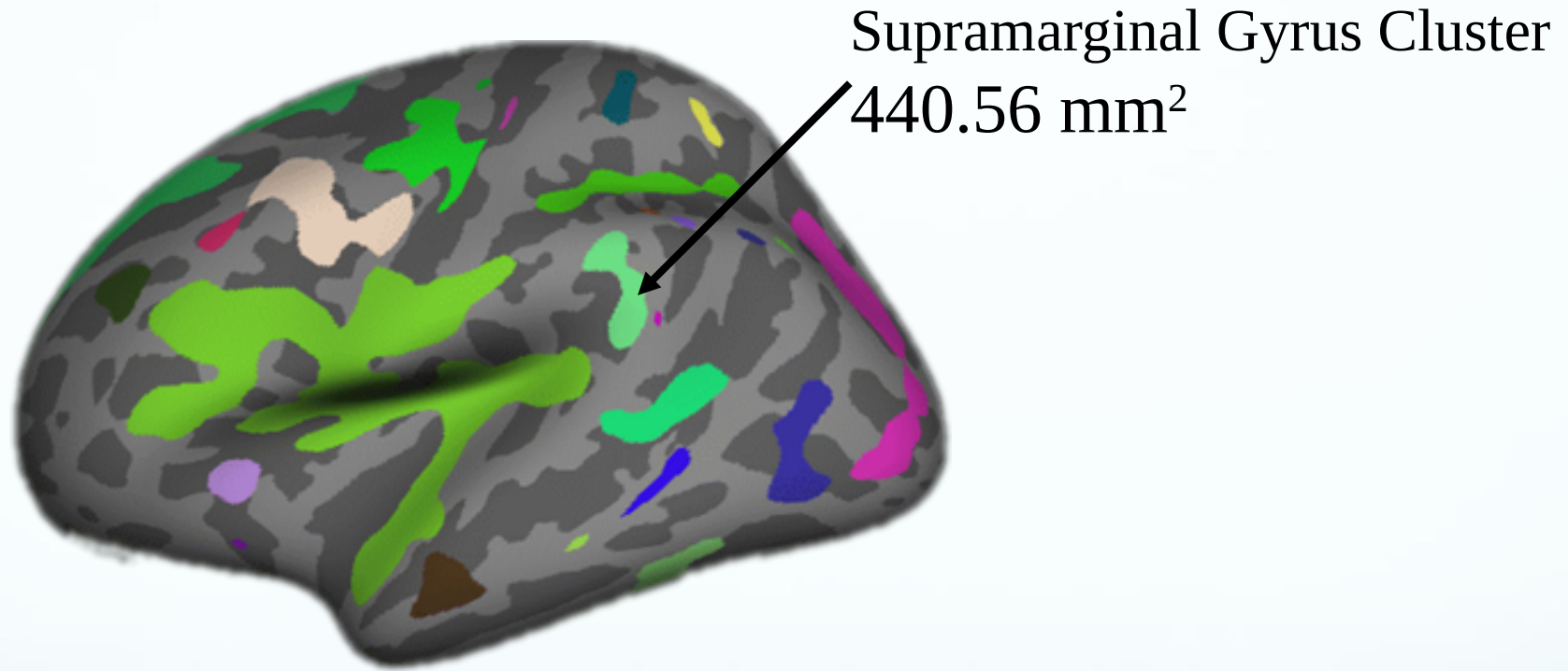
Cluster-based Correction for Multiple Comparisons

1. Simulate data under Null Hypothesis:
 - Synthesize Gaussian noise and then smooth (Monte Carlo), like random field theory
 - Permute rows of design matrix (Permutation, orthog.)
2. Analyze, threshold, cluster, get **MaxClusterSizeNull**
3. Repeat 10,000 times – gives a list of 10000 **MaxClusterSizeNulls** under the null
4. Analyze real data, get **ClusterSize** (eg, 440.56 mm²)
5. Count number of times **MaxClusterSizeNull** > **ClusterSize**
 $P(\text{cluster}) = \#(\text{MaxClusterSizeNull} > \text{ClusterSize}) / 10000$

mri_glmfit-sim



Clusterwise Correction



Probability of getting a cluster 440.56mm² or bigger by chance is $p=.04$
This is the clusterwise p-value.