# Equitable Thresholding And Clustering (ETAC)

## Robert W Cox

SSCC / NIMH / NIH / DHHS / USA / EARTH

Ahu Akivi

https://afni.nimh.nih.gov

2018

# Crude Outline

- Start slow: Summary of 2-level group analysis and cluster-ization

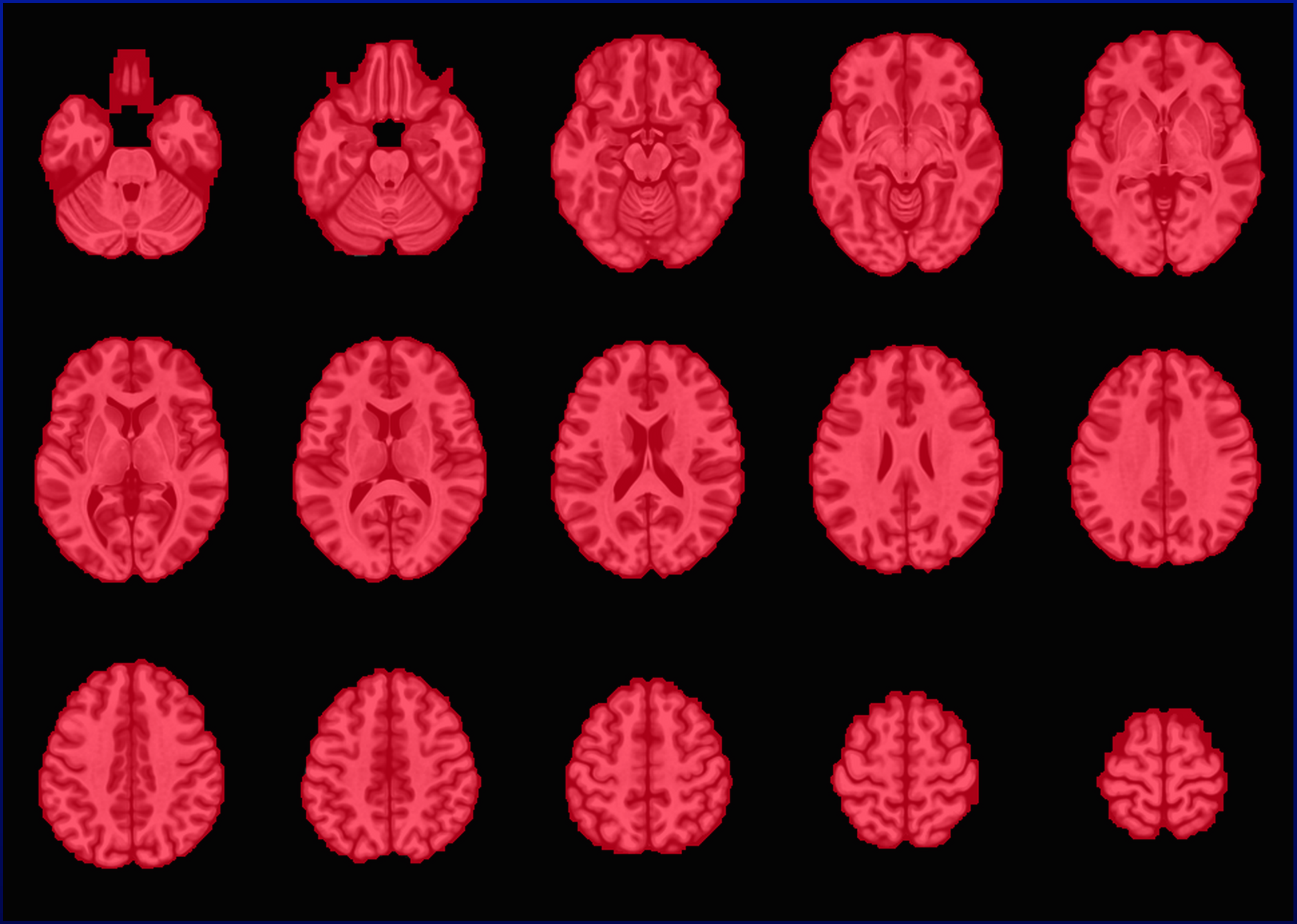
😱 The Great Cluster Panic of 2016
  - AFNI responses

😀 Equitable Thresholding and Clustering
  - What I mean by that phrase
  - Some results
  - Work yet to come

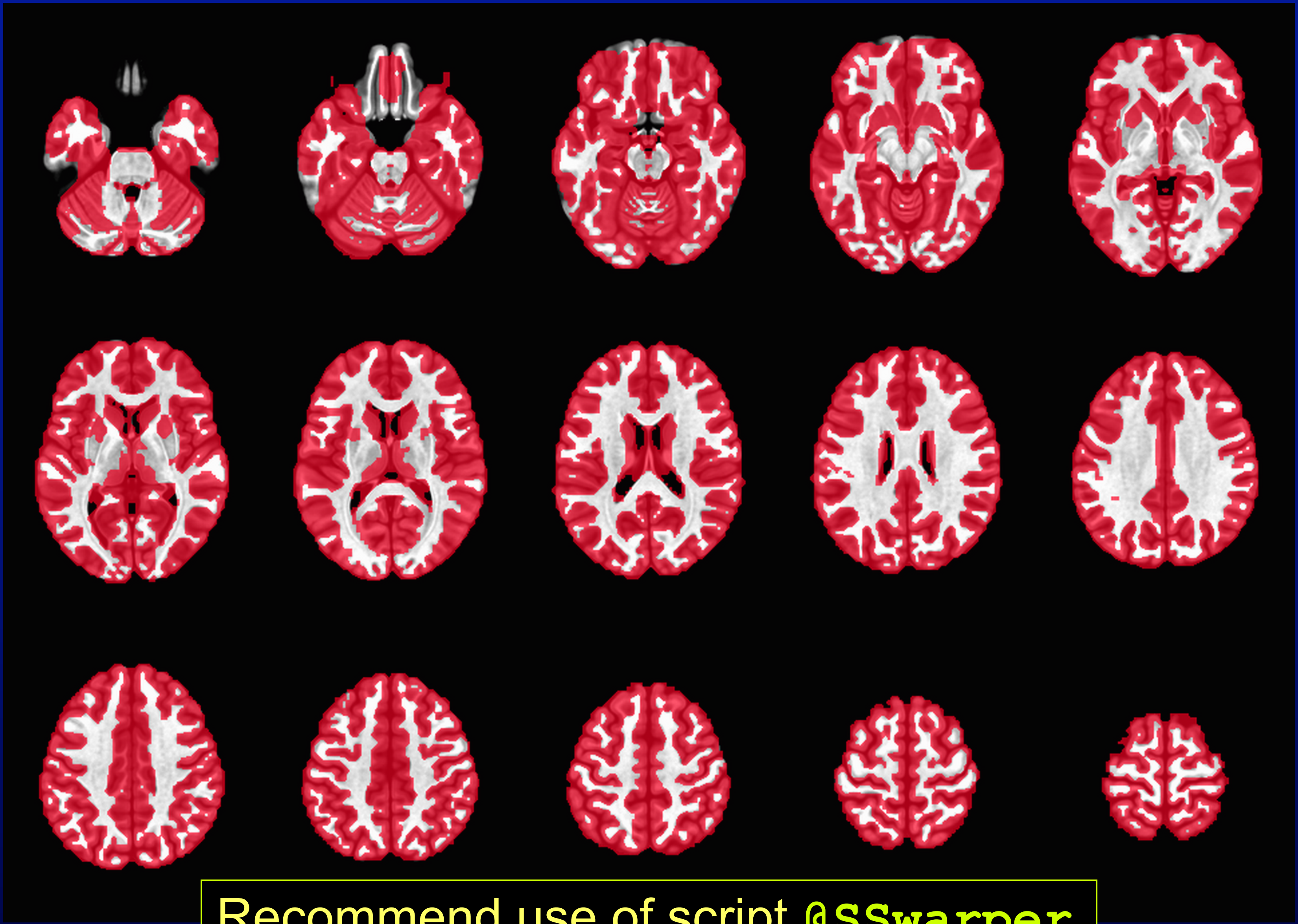
# Voxel-Wise Group Analysis

- Do first level time series analysis on each subject's data separately
  - Transformed to common template (*e.g.*, MNI)
    - Best with nonlinear transformation (**`3dQwarp`**)
      - Can restrict analysis to dilated gray matter mask
- Second level group analysis on voxel $\beta$ values = % signal change (*not* ROIs)
  - Can be as simple as *t*-tests (**`3dttest++`**)
  - Or a complicated model such as Linear Mixed Effects (**`3dLME`**), *etc.*

# Aside: Whole Brain Mask



# Dilated GM Mask (≈60% of WB)



Recommend use of script `@SSwarper`

# Group Spatial Inference - 1

- ***Goal***: control *global* **F**alse **P**ositive **R**ate (FPR) – to 5% level (*e.g.*)
  - **FPR** = **FWE** = **F**amily-**W**ise **E**rror
    - = rate of errors across the family of voxel tests
    - "error" = when *anything* is found in noise-only data *vs* the **null hypothesis** (*i.e.*, no "activity")
- Different approach: to control the **F**alse **D**iscovery **R**ate (**FDR**, voxel-wise)
  - = fraction of "discoveries" that are "errors"
  - *Not* what I'm going to talk about here
    - Difficult to allow for inter-voxel correlation in noise

# Group Spatial Inference - 2

- Voxel-wise thresholding on group *t*-statistic is usually super conservative (to get global FPR≈5%)
  - Can estimate *false non-discovery rate* (FNDR of voxels) using adaptation of voxel-wise FDR algorithm
    - Not highly accurate, nor widely used in FMRI
    - An algorithm for this estimate is hidden in **AFNI**
  - Typically 60-90% (or more)
    - Depends on number of subjects (*i.e.*, statistical power) – figure above is for ≈20 subjects

# Group Spatial Inference - 3

- *A Solution*: form clusters of neighboring voxels, each above a lower (less strict) voxel-wise *t*-statistic (or *z*-statistic)
  - With a larger voxel-wise *p*-value (=smaller *t*)
- *Then*: threshold on cluster-size as well
  - Or some other cluster-FOM (Figure of Merit)
    - *e.g.*, Sum over cluster of voxel-wise $z^2$
    - Reject small/weak isolated clusters
  - Given voxel-wise *p*, adjust cluster-FOM threshold to get desired global FPR ⇛⇛⋯

# Group Spatial Inference - 4

- Dual threshold method (voxel then cluster) can be weak (low power to detect)
- *A Solution*: use spatial blurring ≈ average nearby voxel $\beta$ ("Coef") values together, in each subject, *before* group statistics
  - To reduce noise and reinforce commonality
  - To reduce effective number of independent statistical tests (but lose spatial resolution)
  - To select the *minimum* spatial scale of what we are hunting for

# Old ClustSim - 1

- Spatial correlation of “noise” in FMRI data means no exact formula for cluster-FOM threshold, for a given *p* threshold
- So: Assume Gaussian-shape for spatial auto-correlation function (ACF) *of noise*
  - Fit Gaussian width parameter (Forman 1995)
  - Use approximate formula (**SPM**) *or* Monte-Carlo simulation (**AFNI**) to get cluster-size threshold
    - SPM method possible due to Gaussian ACF

# Old ClustSim - 2

1) Generate random noise-only dataset with Gaussian ACF (with chosen FWHM)
2) Threshold at various per-voxel *p*-values
3) Find largest cluster *in brain mask*
4) Repeat steps 1-3 10,000+ times
5) For each per-voxel *p*-value, cluster-size threshold is largest cluster size which occurs only in 5% (*e.g.*) of cases

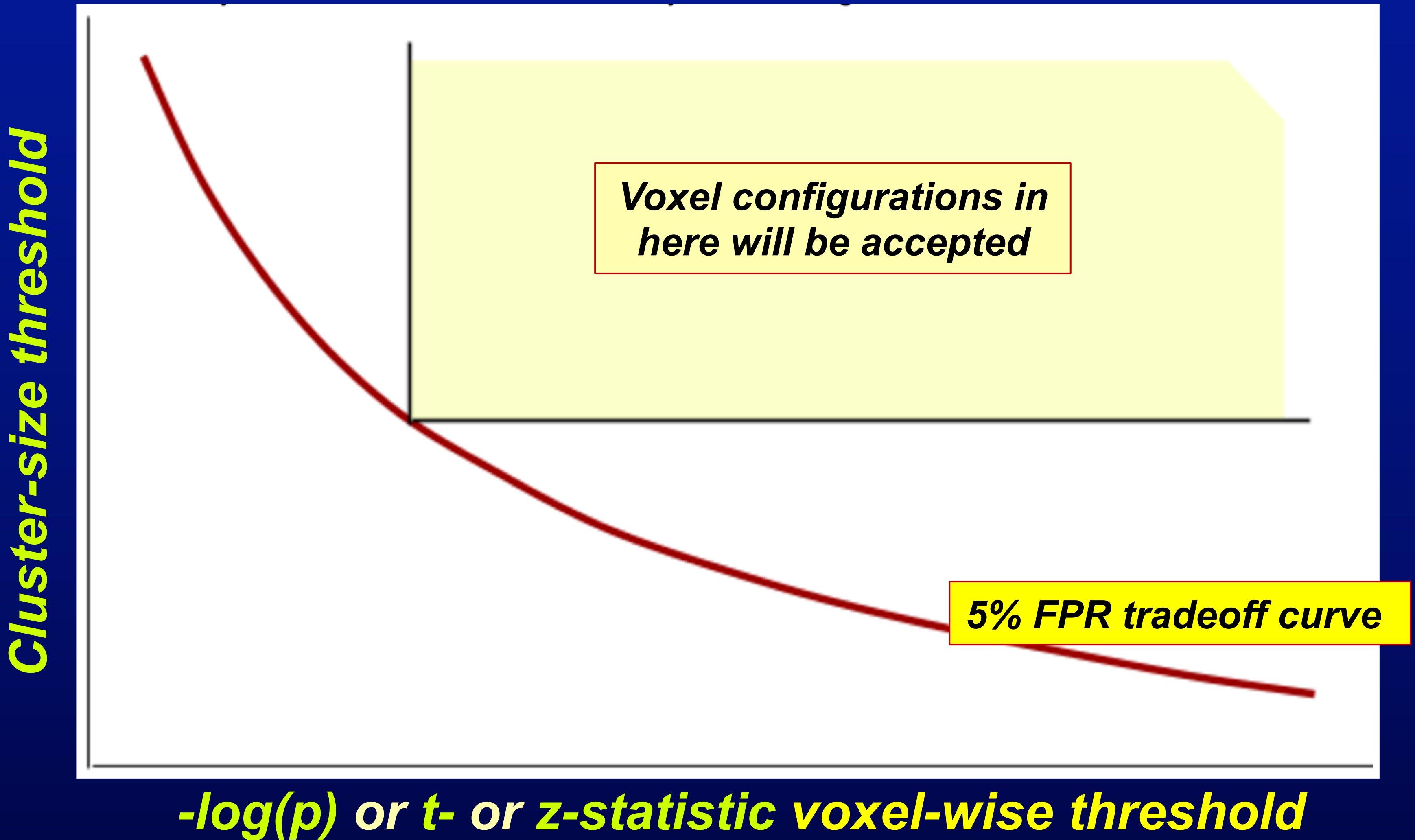
# ClustSim - 4

- **3dClustSim** outputs tables like this:

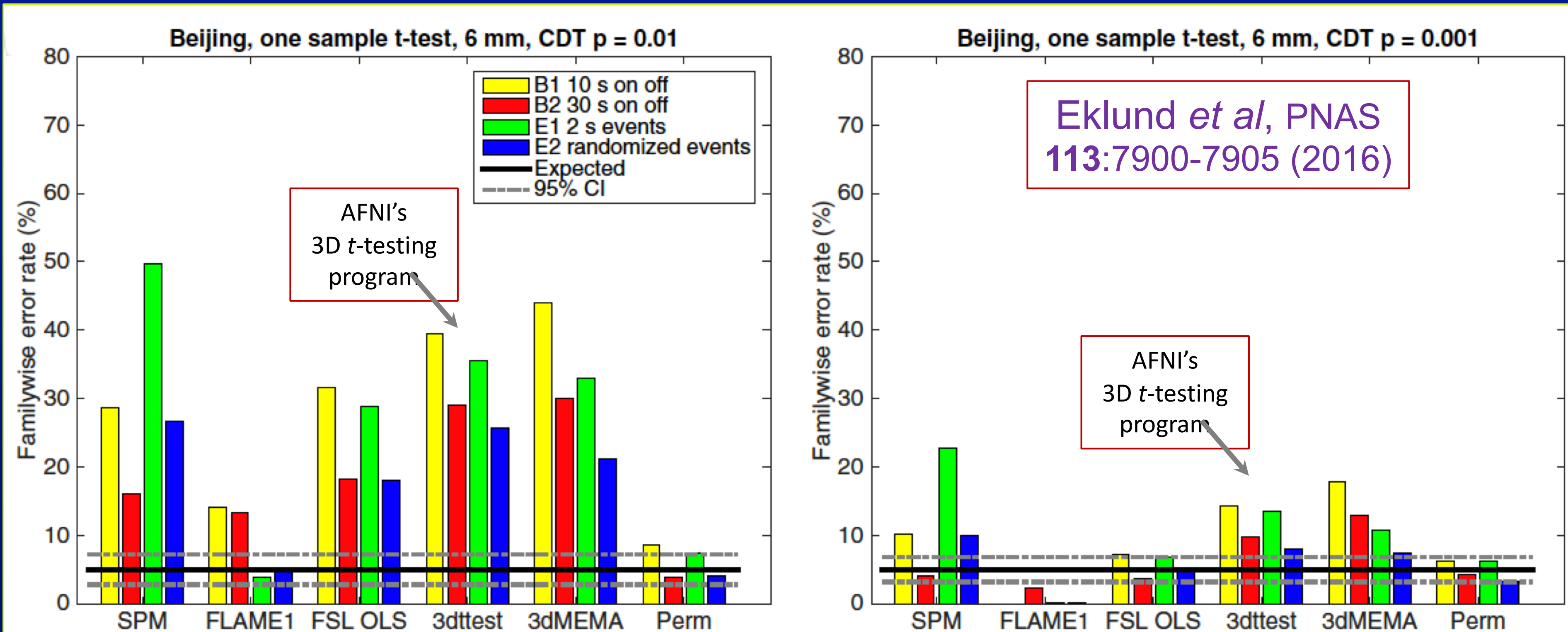
```
# CLUSTER SIZE THRESHOLD(pthr,alpha)
# -NN 2   | alpha=Prob(Cluster > given size)
#  pthr   |.10000 .05000 .02000 .01000
# ------  | ------ ------ ------ ------
0.010000     50.3   57.2   66.3   73.6
0.005000     34.4   39.5   46.3   51.6
⇛0.002000    22.1   25.7   30.4   34.1
0.001000     16.0   19.0   22.8   26.0
0.000500     12.0   14.5   17.4   20.1
0.000200      8.1   10.0   12.6   14.6
0.000100      6.1    7.7    9.9   11.6
```

# ClustSim - 5

- High *t* threshold ⇛ small cluster threshold



# The Great Cluster Panic - 2016




- FPR ≫ 5%: *notably* for voxel-wise *p*=0.01
- A lot of doom-crying about this in 2016:

  "*Could Invalidate 15 Years of Brain Research*"

# FPR: Testing *Some* Method

- **Eklund *et al***: use rsFMRI (FCON-1000) as null data
  - Analyze each of 198 x 2 subject collections (Beijing and Cambridge) with fake task timings
    - 2 x Block design, 2 x Event-related design
    - 4 x spatial blur levels (4, 6, 8, 10 mm)

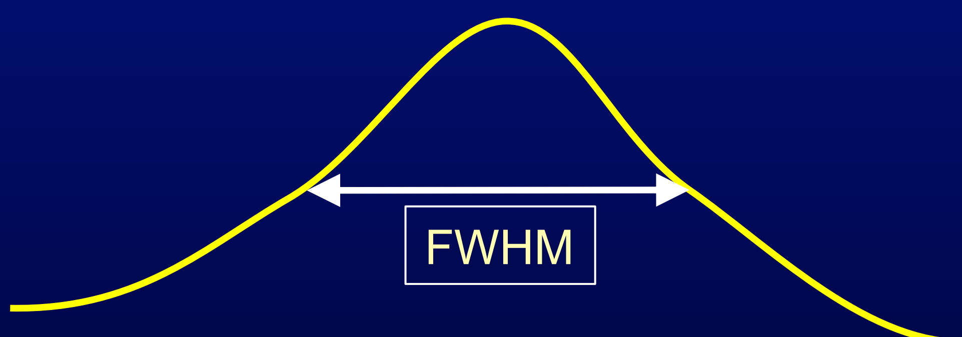
    } 16 basic cases
  - Carry out 1- and 2-sample t-tests between subsets of these collections – 1000 random subsets (per case, per collection, per diverse variations)
  - Count clusters surviving the given software, get FPR estimate
- Scripts and tabular results available on GitHub

# 1 Fix + 3 Solutions in AFNI

0) Fix `3dClustSim` *bug* found by Eklund
1) Extend ACF model in `3dClustSim` to be more complicated than a Gaussian shape (the mixed model)
2) Eliminate ACF modeling by extending `3dClustSim` to directly use *residuals* from `3dttest++` via randomization
3) Generalize cluster-thresholding model in several more directions: **ETAC**

# 0) Bugs and Flaws

- **AFNI**'s cluster-size threshold calculating program (`3dClustSim`) had a bug
  - A big deal in the PNAS paper (and popular press)
  - Not actually that important (*cf* 5 slides ahead)
  - Forman method for FWHM estimate = another flaw (**FHWM** = **F**ull **W**idth at **H**alf **M**aximum)
    - Using statistics of nearest-neighbor differences of noise to estimate FWHM of noise correlation

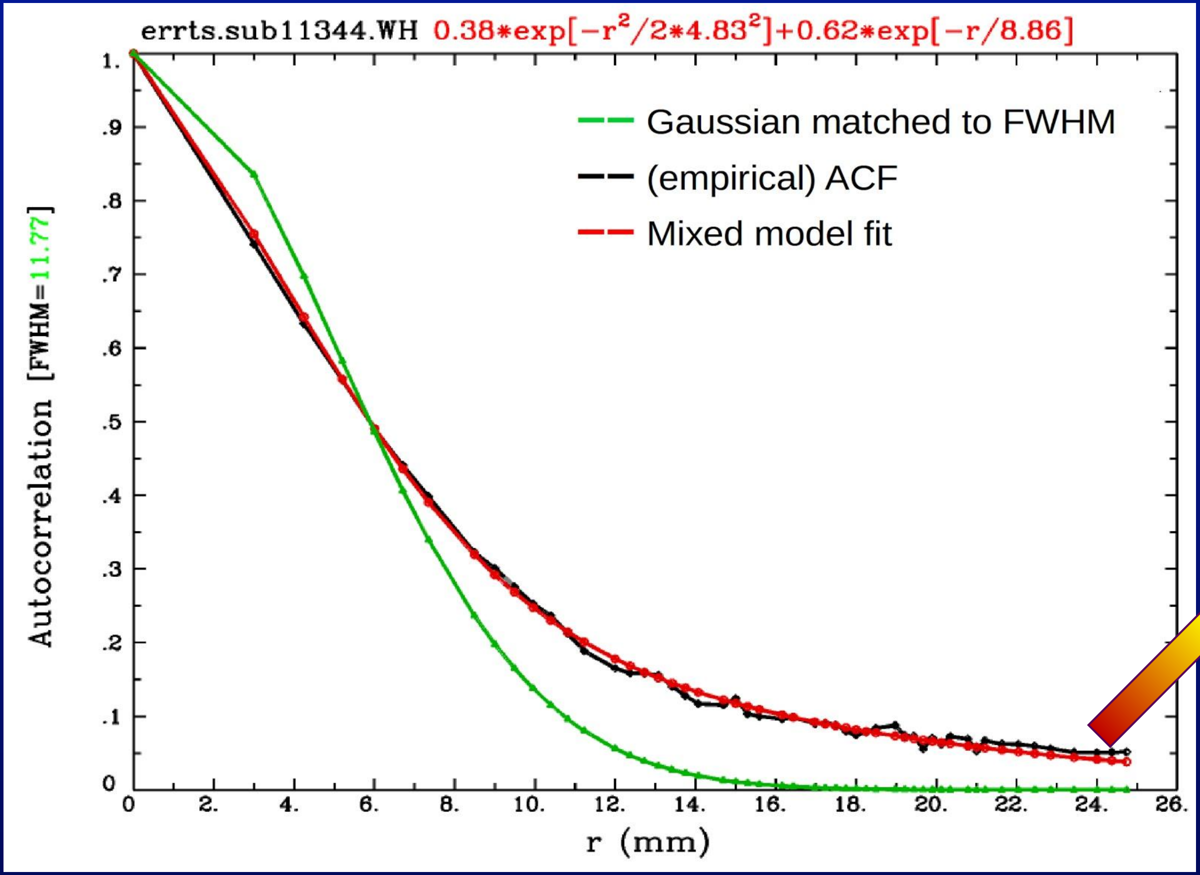


# 0) Bugs and Flaws

- However, there was/is a *much* bigger **flaw**
  - Shared with FSL and SPM for unnumbered years
  - Assumption of **Gaussian shape** for spatial autocorrelation function (ACF) of the noise
    - ACF(*r*) describes how noise in one voxel is correlated with noise in another voxel (distance *r* away)
- We are interested in clusters caused by true differences in signal
- But we also have to study clusters caused by noise (signal fluctuations)
  - Estimate probability of results being “bad luck”

# 1) NonGaussianity in ACF

- ACF from single subject datasets has long tails – nonGaussian shape + 1st difference fail




Modify 3dClustSim to use mixed ACF model (Gaussian plus mono-exponential) with 3 parameters (*a*,*b*,*c*) instead of 1 (FWHM)

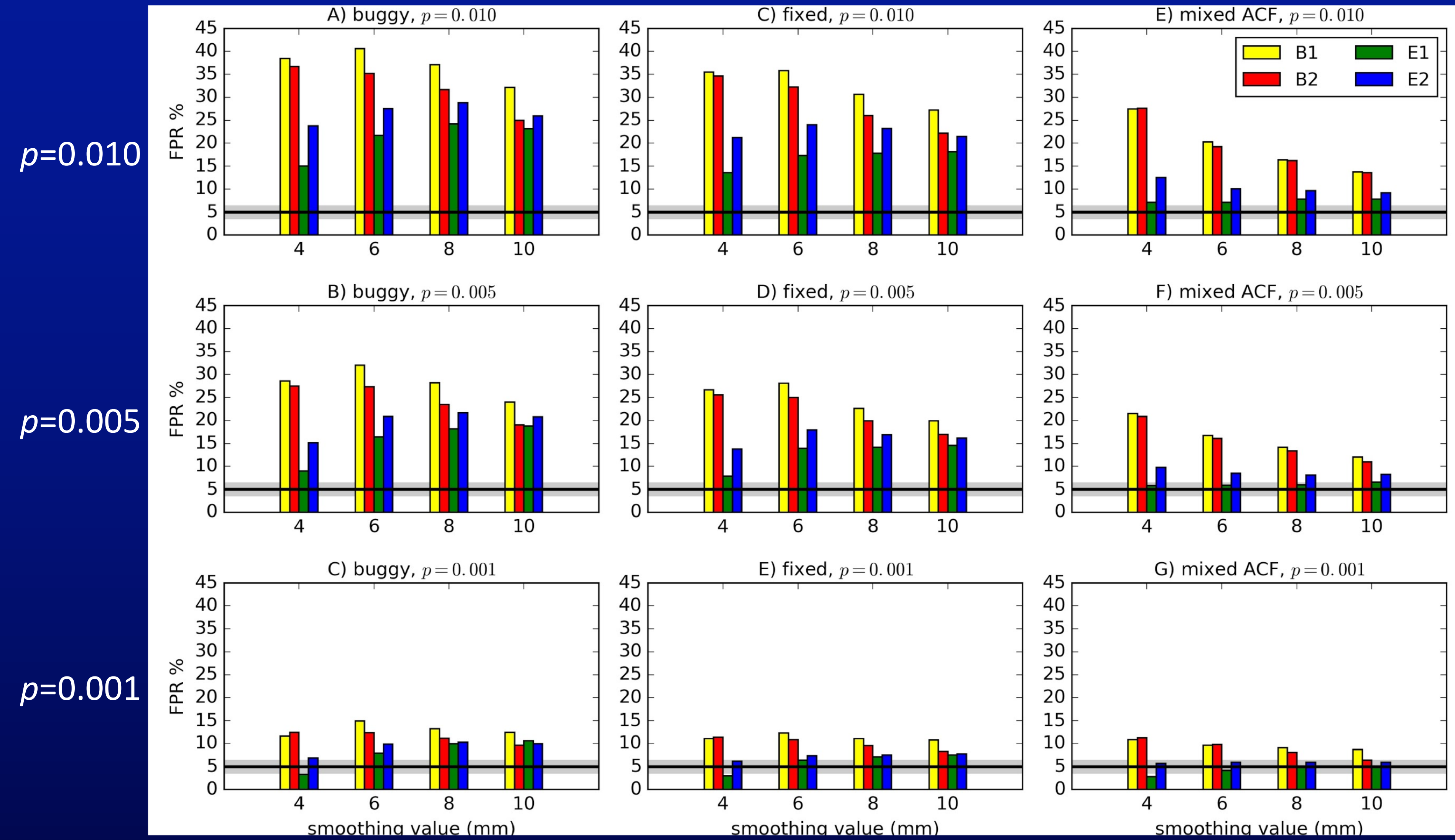
$$\text{ACF}(r) = a\,\exp[-r^2/(2b^2)] + (1-a)\exp[-r/c]$$

# 1) Updated ClustSim

- Program **3dFWHMx** now estimates the mixed model (*a*,*b*,*c*) ACF parameters
  - *No longer shows Forman estimates*
- Program **3dClustSim** takes ACF parameters *and*
  - Simulates random noise-only 3D dataset with mixed model ACF
    - A little slower than Gaussian ACF approach
  - Otherwise, the same method as before:
    - Builds tables of cluster sizes found

# 0 & 1) AFNI Results Redux

Pre-bug fix | Post-bug fix | Mixed-model ACF

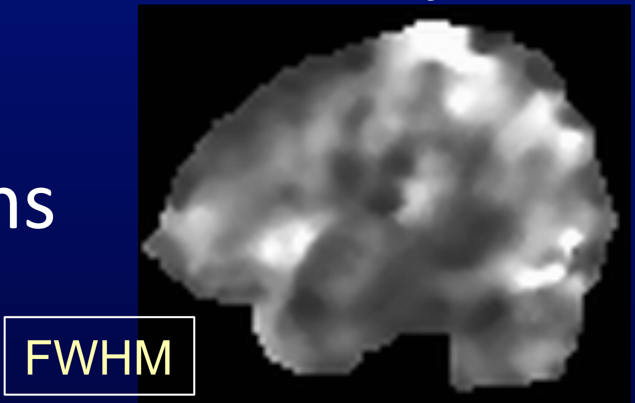


# 1) How to: ACF method

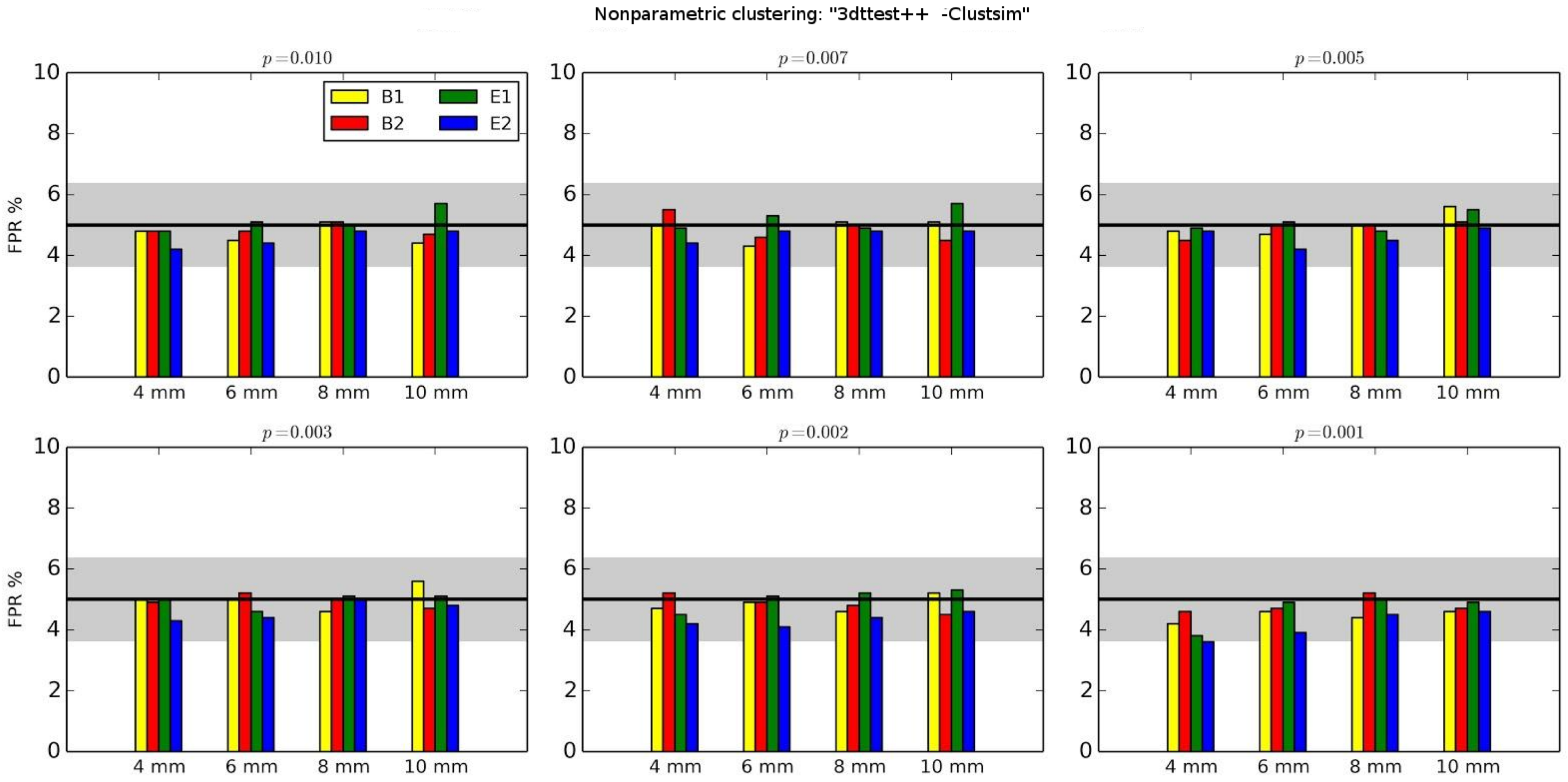
- Run **`3dFWHMx`** with '-acf' option to get (*a*,*b*,*c*) for each subject, from residuals dataset errts*+tlrc.HEAD
  - This calculation is done now in **`afni_proc.py`**
  - Average each of the 3 ACF parameters across subjects (not automatic)
- Use **`3dClustSim`** with '-acf' option (giving it the 3 averaged parameters) to get cluster size threshold tables for group analysis
  - This method is OK, if per-voxel $p \leq 0.002$

# ¿Why Is Model-Based FPR Still High?

- Using ACF mixed model improved results
  - So the wider ACF and longer tails are a part of the original problem – but not all of it
- Too short tails in the group *t*-statistics, caused by outlier subjects in the data
  - Also explained a part of it – *but not very much*
- Spatial ACF is not stationary (same everywhere)
  - Over-wide in some places
  - Drives up FPR in those regions



# 2) A Different Solution:
# *Nonparametric* Clustering in **AFNI**




- *t*-test residuals are permuted/randomized (10000 times)
- 10000 re-*t*-tests computed from residuals fed to `3dClustSim`

# 2) How to: Nonparametric Clustering

- **Only for *t*-tests at this time**
  - Re-running many **`3dLME`** cases (*e.g.*) is too slow
- **`3dttest++`** with the **`–Clustsim`** option
- Gives excellent FPR control ☺
- Has stringently large cluster-size thresholds ☹
  - Seems to be needed to deal with the extra-wide spatial ACF in some regions (notably, midline)
  - Cluster-size threshold is nonlinear in smoothness
  - Leads to the idea of making the cluster-size threshold depend on spatial location ⇛⇛ ⋯

# (Semi-) Arbitrary Choices

- I've mentioned two parameters that must be chosen by the researcher in the "usual" methods:
  - Voxel-wise *p*-value for first-level thresholding
    - Typical values range from 0.001 to 0.01
  - Amount of spatial blurring to add to data
    - Typical values range from 4 to 10 mm
- But there are no "best" values ☹
  - **ETAC** can rescue you! (from these choices) ☺

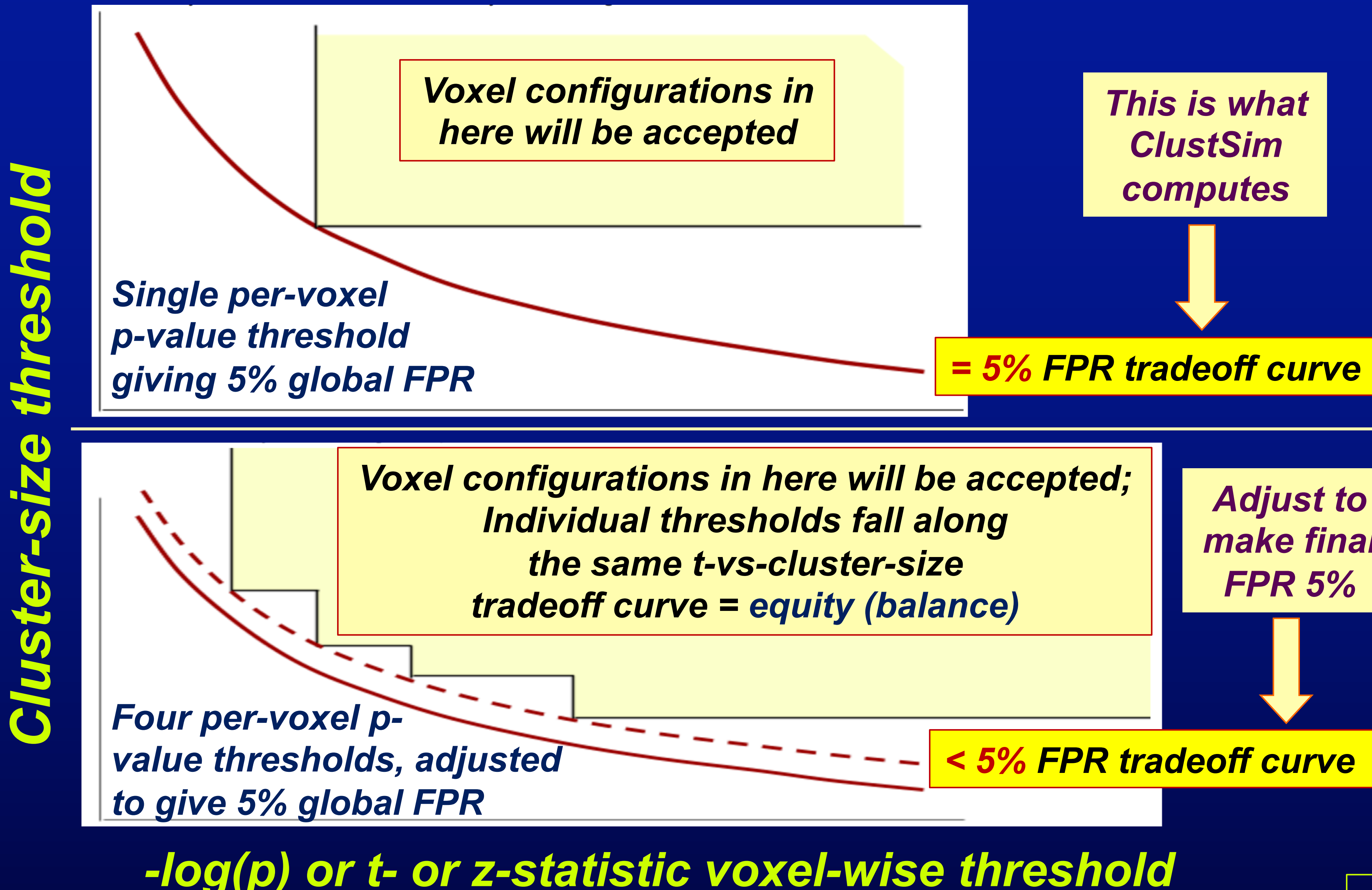
# 3) ETAC ☺ ☺ ☺

- **E**quitable **T**hresholding **A**nd **C**lustering
- Uses multiple sub-methods at same time
  - **Equity** = **balancing** FPRs of sub-methods

1) Voxel-wise thresholding at multiple *p*-values, then cluster-FOM thresholding
2) Multiple cases of spatial blurring
3) Different cluster-FOM thresholds in different brain regions (*vs* global thresh)

- No model for ACF: uses randomization

# Equity: Multi-Thresholding



# Equity: Across Methods

- Balancing can apply to any multi-choice method for selecting voxel clusters
  - Each sub-method has a cluster-FOM threshold adjustable to get desired FPR
  - Balance = choose each sub-method's cluster-FOM threshold to have the same global FPR $\alpha_0 < \alpha_{Goal}$ (e.g., 5%)
- **ETAC** method (set union): accept a voxel if it survives at least one sub-method
  - Adjust $\alpha_0$ up or down to get final FPR = $\alpha_{Goal}$

# Equity: Across Blur Cases

- Blurring at (*e.g.*) 4, 6, 8, 10 mm
- Potential to detect both small intense clusters and larger weak clusters
  - Blur = 10 mm might "wash out" small cluster
  - Blur = 4 mm might not reduce noise enough to find larger weak cluster
- Combined with multi-thresholding (different *p*-values), reduces number of arbitrary choices to make in thresholding

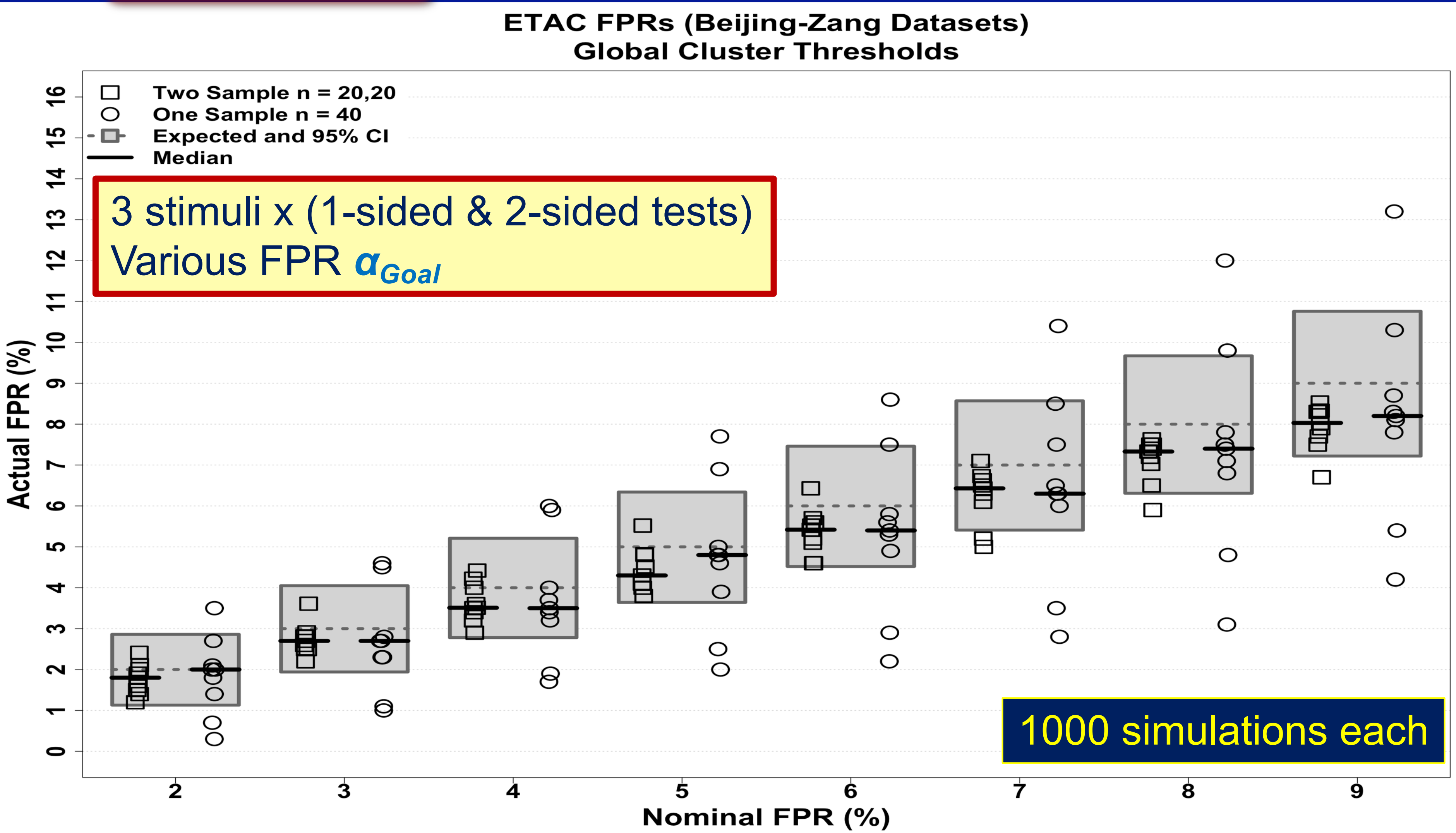
# Equity: Across Space

- Smoothness (ACF) of noise varies across the brain
  - Using same cluster threshold everywhere will make FPR non-uniform
  - *Could* try to differentially smooth to make ACF more uniform (not implemented in **AFNI**)
- **ETAC** method: Use different cluster-FOM thresholds at different locations
  - For each sub-method, produce a 3D map of the cluster-FOM threshold to use

# Global and Local ETAC

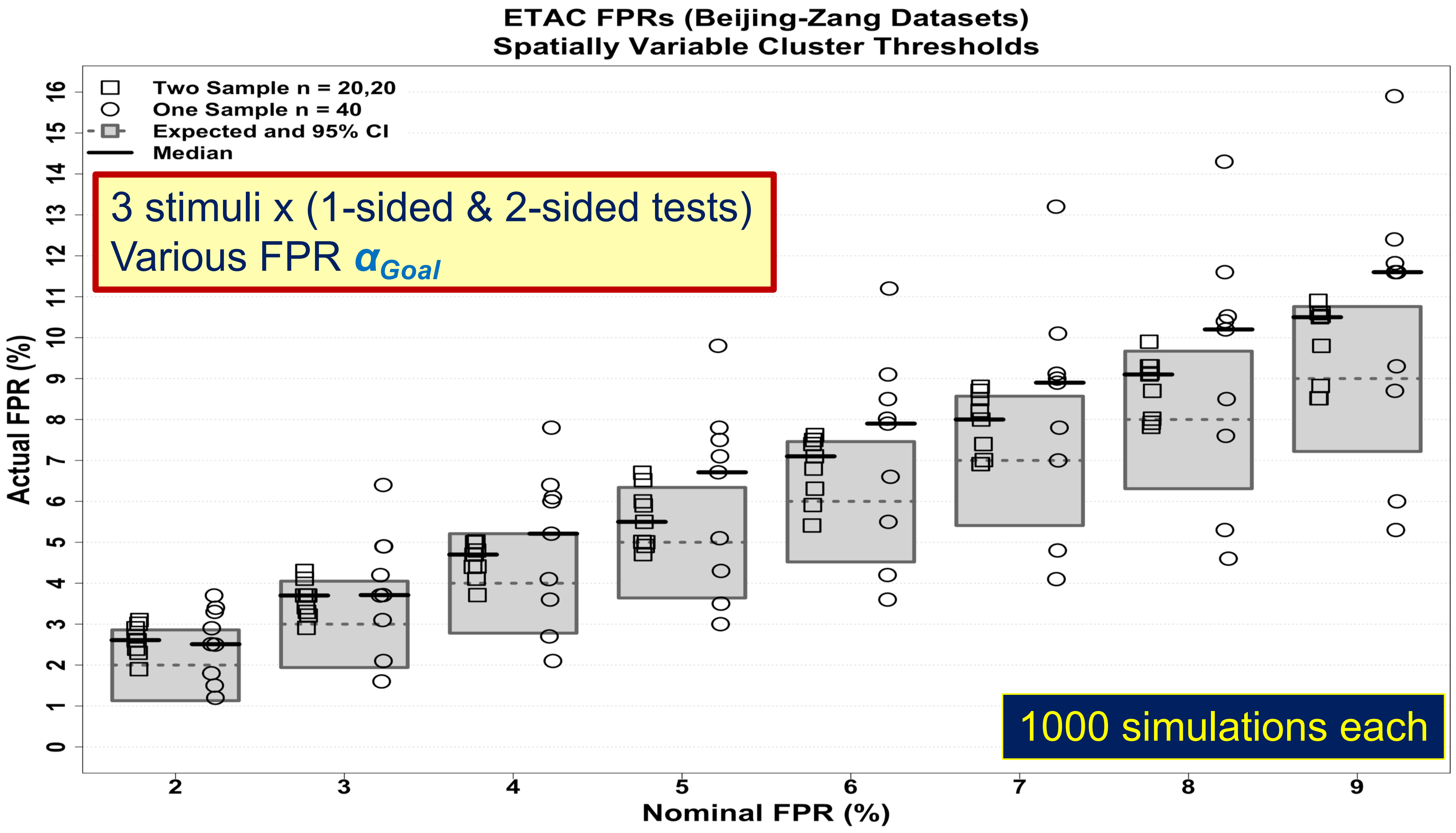
- **Global** = apply ETAC across multiple *p*-thresholds and multiple blurs
  - Get a table of cluster-size (or FOM) thresholds to use, one cluster threshold per *p*/blur combination (sub-method)
- **Local** = also apply ETAC across voxels
  - Get a 3D dataset of cluster thresholds for each sub-method
- Applied via program `3dMultiThresh`

# Global ETAC: FPR Control



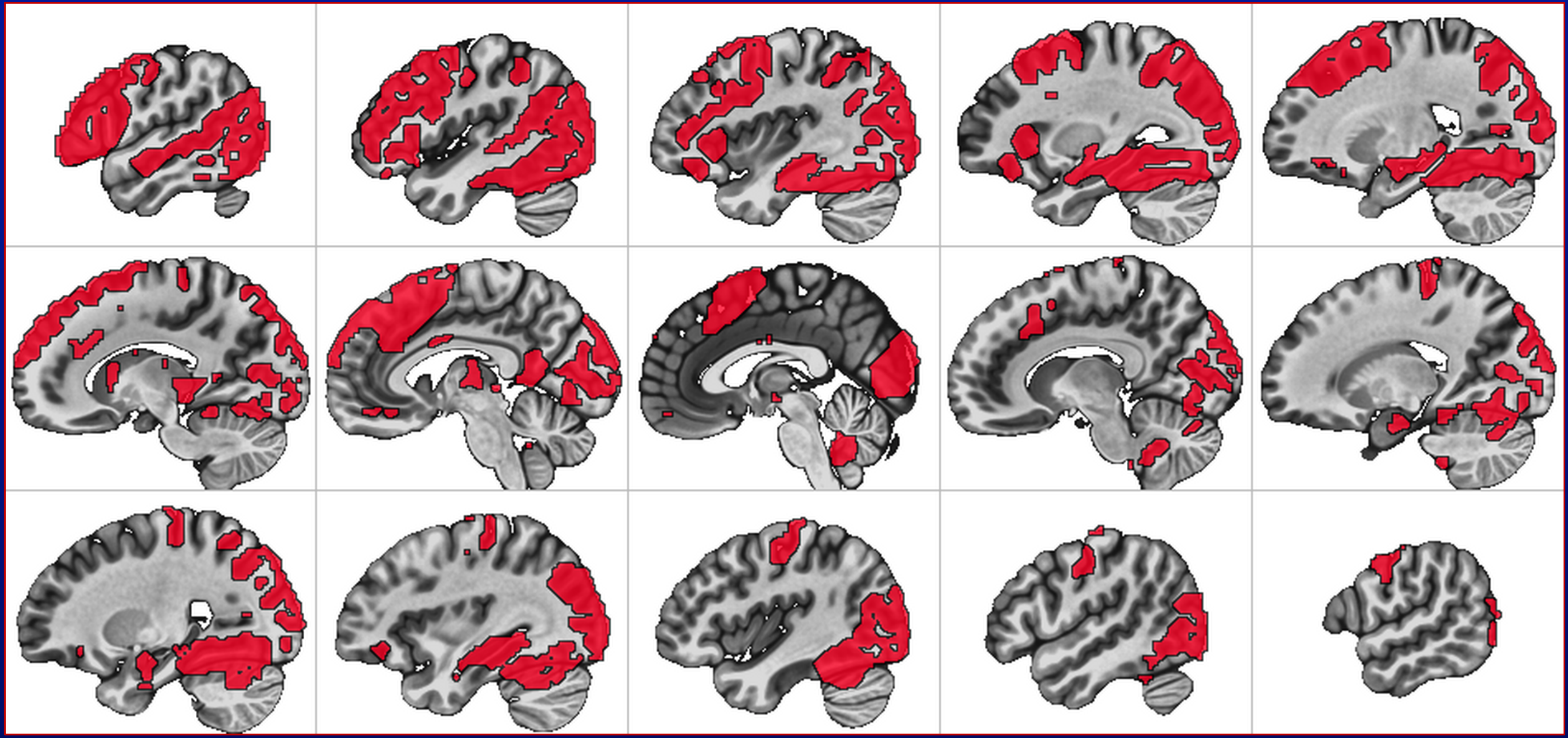

$p = 0.01, 0.005, 0.002, 0.001$ blur $= 4, 7, 10$

# Local ETAC: FPR Control




$p = 0.01, 0.005, 0.002, 0.001$ blur$= 4, 7, 10$

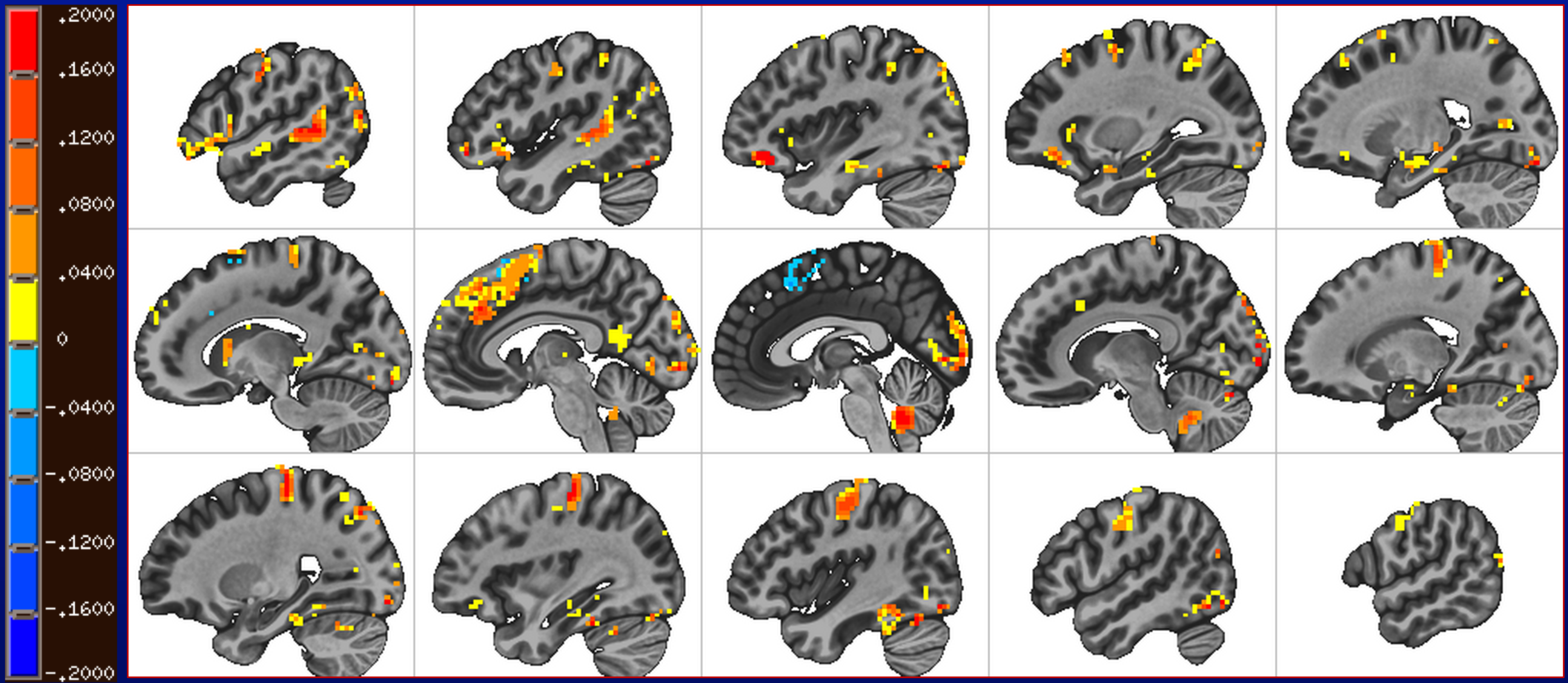
# ETAC activation mask (2% FPR, all 81 subjects)



UCLA Phenomics study (*pamenc vs control task*)
20 (out of 81) subjects per test
⇛ data from OpenFMRI web site

# Task Detection Power: ETAC *minus* Single Threshold

500 simulations




UCLA Phenomics study (*pamenc vs control task*)
20 (out of 81) subjects per test
⇛ data from OpenFMRI web site

# Using ETAC

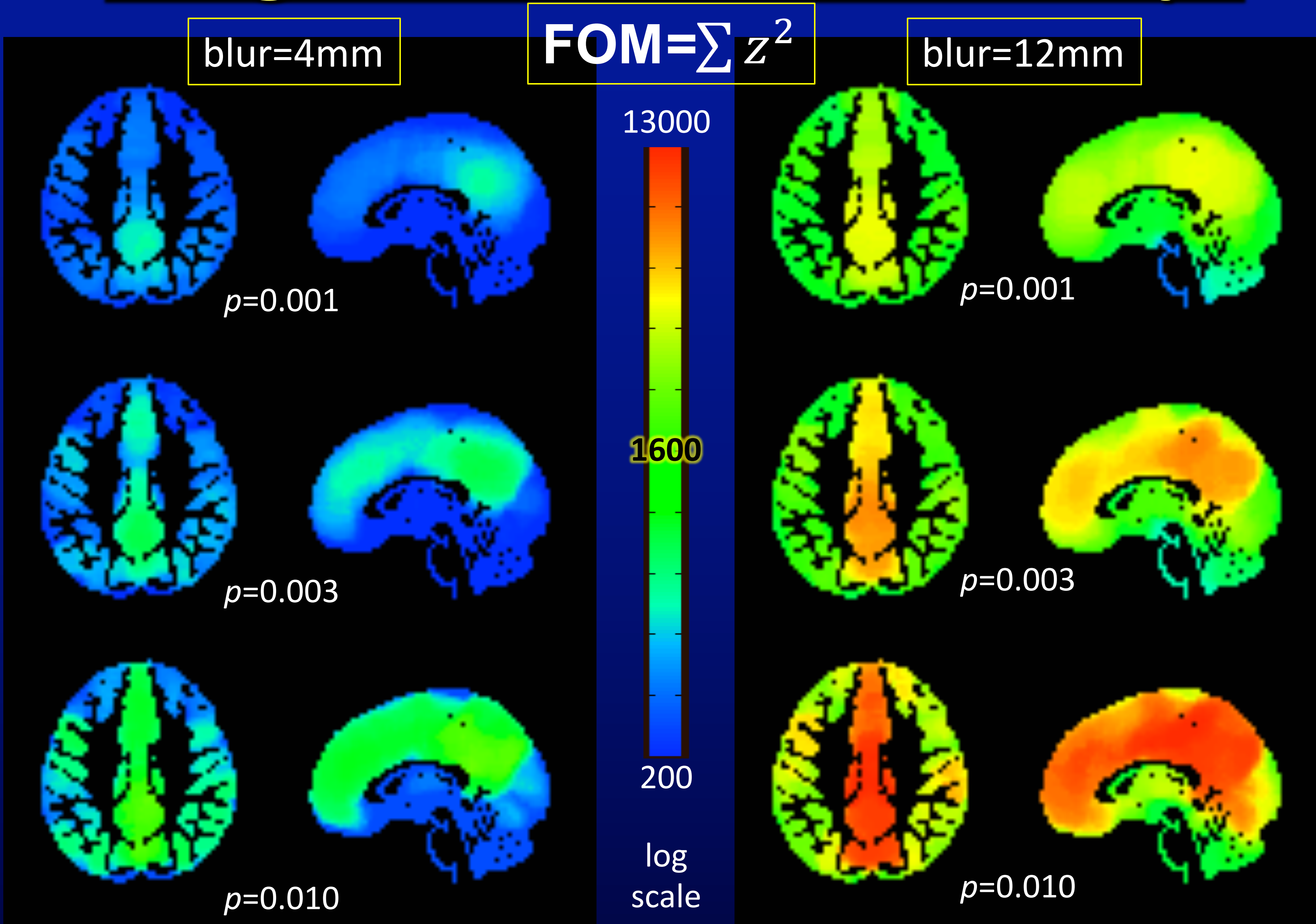
- **ETAC** algorithm: program **`3dXClustSim`**
- User adds options to **`3dttest++`** to run **ETAC** after the group *t*-tests are done
  - **`-ETAC`** to enable the algorithm
  - **`-ETAC_blur`** to specify blur cases to use
  - **`-ETAC_opt`** to specify thresholding options
    - To change from default per-voxel *p*-values of
      0.0100 0.0056 0.0031 0.0018 0.0010
    - To change default clustering parameters
      NN=2 FOM=$\sum z^2$ 2-sided tests goal=$\alpha_{Goal}$=5%

# ETAC Sample Command

```
3dttest++
 -setA datasets
 -setB datasets { other options here ... }
 -prefix Gtest.nii
 -prefix_clustsim GtestX
 -ETAC
 -ETAC_blur 6 12
 -ETAC_opt
 sid=2:pthr=0.01,0.003,0.001:name=TestA
 -ETAC_opt
 sid=1:pthr=0.01,0.003,0.001:name=TestB
```

Combines with any other blurring (applies to -ETAC_blur 6 12)

# Images of Multi-Thresh Maps



# How ETAC Works

- More complex than older ClustSim
- Must keep cluster-FOM tables for each sub-method and for each voxel (local ETAC)
  - Some voxels don't get many "hits"
  - Clusters are dilated to get brain coverage
    - But FOM for cluster is based on original size

→How to apply spatially variable cluster-FOM to a given cluster in real data?

  - Sort thresholds for all voxels in real cluster
  - Use the 80% point (100% = maximum)

# ETAC: Things to be Done - 1

- Single-subject via mixed-model ACF
  - Spatially non-stationary? A little complex.

✓ **ETAC** algorithm *without* voxel equity
  - Multi-method with global cluster thresholds

- Implementation details (short term):
  - ✓ Different $\alpha_{Goal}$s in same run (*e.g.*, 2% 3% 4% 5%)
  - Apply multi-thresholds to other *t*-volumes in `3dttest++` output
    - *e.g.*, 1-sample results in 2-sample tests
  - Other cluster-FOMs (*e.g.*, ~TFCE's)?

# ETAC: Things to be Done - 2

- Test more null cases for FPR
  - `3dttest++` options, such as covariates
    - Do multi-threshold maps from the main effect apply to the extra *t*-tests, such as covariates and 1-sample results in 2-sample tests?
      - And give approximately the desired FPR?
    - Or does **ETAC** need to be run separately for each *t*-test included in the output? ☹☹
  - Resting state FMRI seed-based correlation maps (all tests up to now are task-based)
  - Other scenarios?

# ETAC: Things to be Done - 3

- Test more *positive* cases for power
  - Task-based and resting state
  - Need large number of subjects for this work
    - So can test subsets of different sizes
    - And draw lots of random sub-collections
  - For task cases, need a variety of conditions
    - So can cover large parts of brain
    - Including conditions with small (focal) activations, such as amygdala
      - Will **ETAC** work well for such cases?

# ETAC: Things to be Done - 4

- Extend method to work on surface domains, not just 3D volumes
  - Will need a *lot* of work ☹ ☹ ☹ ☹ ☹ ☹ ☹ ☹
  - Need to write ClustSim for surfaces
  - Need to write **ETAC** (multi-thresholding and FPR solving) for surfaces
  - *Or* for mixed 2D+3D domains, as in the CIFTI-format data (*e.g.*, HCP)
    - Cortical surfaces plus basal ganglia volumes
    - ETAC is based on topology *not* on geometry

# ETAC: Things to be Done - 5

- ✓ Should **ETAC** output show you *which* sub-methods a voxel passed?
  - *e.g.*, which *p*-values, which blur cases?
- Need experience with actual users/actual studies to find things out:
  - What other outputs would be interesting?
  - How useful is **ETAC** *now*, compared to other methods for global thresholding?
- These 5 slides are just *part* of *the list* …

# Other Thoughts

- With many subjects in a study, does cluster-FOM thresholding continue to make sense?
  - More and more of brain will pass test
    - Unless looking at a restricted hypothesis, such as brain regions correlated with some subject behavior/condition
  - How to interpret such results?
- At what point does voxel-wise *only* thresholding become ”reasonable”?

# Conclusions (At Long Last!)

- If `3dttest++` can do your group analysis, **ETAC** might be your new friend
  - Fewer arbitrary thresholding choices ☺
  - No loss of power ☺
  - Not fully tested yet ☹
  - No publication to cite yet ☹☹
- If you need **`3dLME`**, **`3dMVM`**, *etc.*, then the mixed model ACF method is decent
  - With per-voxel $p \leq 0.002$
  - Publication you can cite ☺

# AFNI Clustering Papers

- Somewhere over the rainbow – **ETAC** paper
  - https://www.biorxiv.org/content/early/2018/04/05/295931
- FMRI Clustering and False Positive Rates. PNAS 114: E3370–E3371, 2017.
  - https://arxiv.org/abs/1702.04846
  - https://doi.org/10.1073/pnas.1614961114
- FMRI Clustering in **AFNI**: False Positive Rates Redux. Brain Connectivity 7:152-171, 2017.
  - https://arxiv.org/abs/1702.04845
  - https://doi.org/10.1089/brain.2016.0475

# Where It Started

# Finally … Thanks

- The list of people I should thank is not *quite* as large as Skewes' number★ …

**MM Klosek. JS Hyde. JR Binder. EA DeYoe. SM Rao.**
**EA Stein. A Jesmanowicz. MS Beauchamp. BD Ward.**
**KM Donahue. PA Bandettini. AS Bloom. T Ross. B Biswal.**
**M Huerta. ZS Saad. K Ropella. B Knutson. J Bobholz.**
**G Chen. RM Birn. J Ratke. PSF Bellgowan. J Frost.**
**K Bove-Bettis. R Doucette. RC Reynolds. PP Christidis.**
**LR Frank. R Desimone. L Ungerleider. KR Hammett.**
**DS Cohen. DA Jacobson. EC Wong. J Gonzalez-Castillo. D Glen.**
**P Kundu (AKA IMoM). E Raab. A Martin. S Gotts. PA Taylor.**

***And YOU, the suffering audience …***

★Currently thought to be about $1.4 \times 10^{316}$