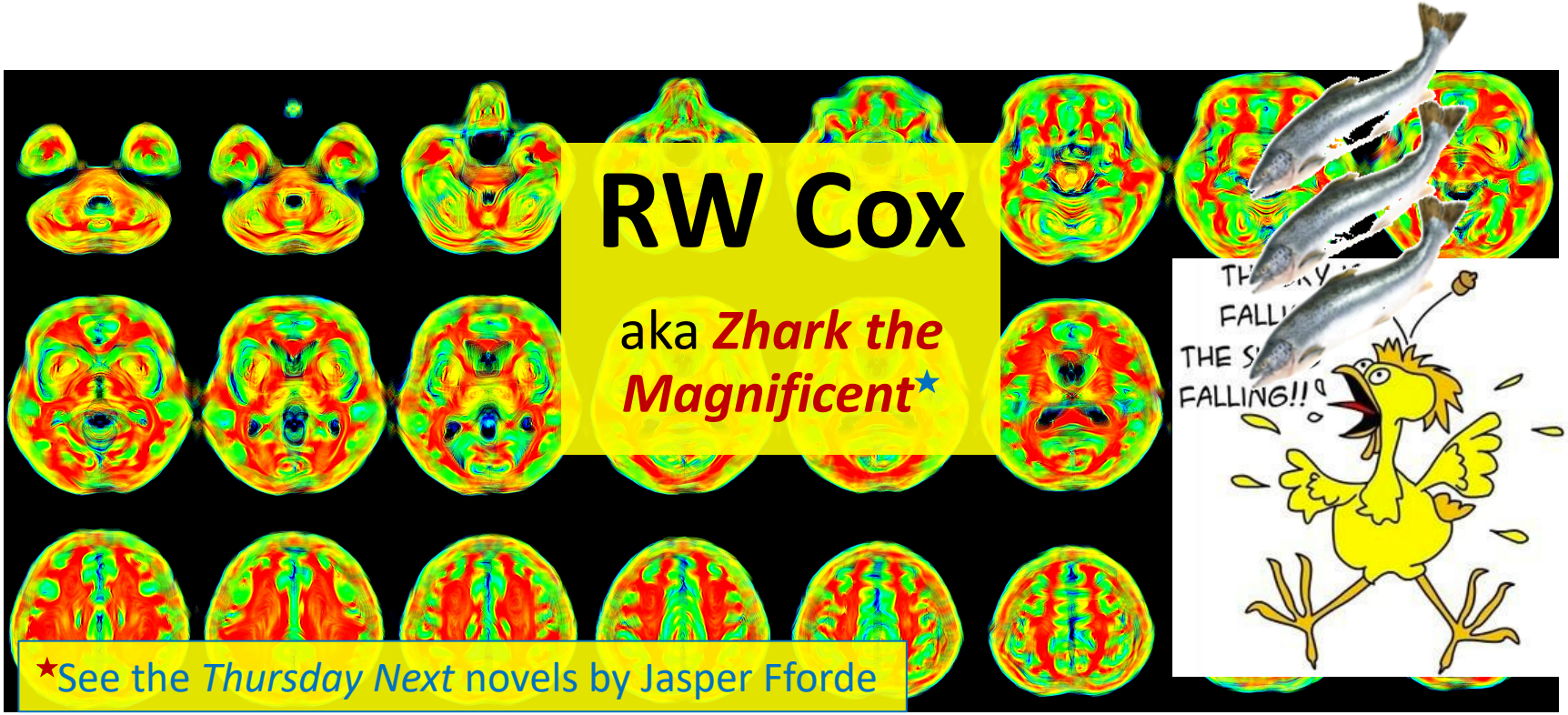
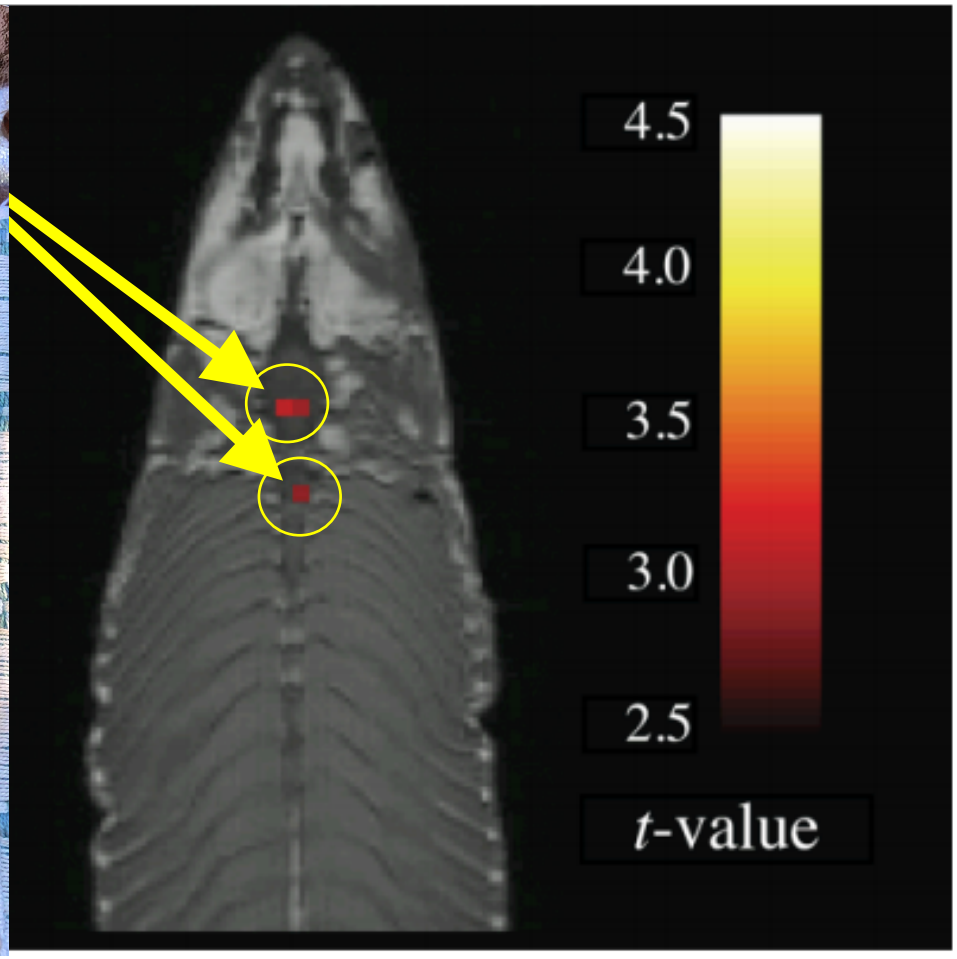
Functional MRI:
Dead Salmon (2009) +
the Great Cluster Panic of 2016



Dead Salmon: 2009

- ”Task” activation: scanning a deceased Atlantic salmon
  - “Shown” a series of photographs of humans
  - “Analyzed” with SPM2 – 16 voxels with voxelwise $p < 0.001$
  - http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf



Dead Salmon: *Popular* Perception

The New York Times

But when you divide the brain into bitty bits and make millions of calculations according to a bunch of inferences, there are abundant opportunities for error, particularly when you are relying on software to do much of the work. This was made glaringly apparent back in 2009, when a graduate student conducted an fM.R.I. scan of a dead salmon and found neural activity in its brain when it was shown photographs of humans in social situations. Again, it was a salmon. And it was dead.

- The NYT got it almost exactly wrong.

Dead Salmon: *Popular* Perception

WIRED

"By complete, random chance, we found some voxels that were significant that just happened to be in the fish's brain," Bennett said. "And if I were a ridiculous researcher, I'd say, 'A dead salmon perceiving humans can tell their emotional state.'"

The result is completely nuts — but that's actually exactly the point. Bennett, who is now a post-doc at the University of California, Santa Barbara, and his adviser, George Wolford, wrote up the work as a warning about the dangers of false positives in fMRI data. They wanted to call attention to ways the field could improve its statistical methods.

- Wired is closer, but still obfuscates the reality.

Dead Salmon: *Popular* Perception

- Lazy people *still* bring up the dead salmon paper as a way to slam FMRI – that NY Times article is from 2016
- Not every popular publication is off base:

SCIENTIFIC AMERICAN®

The original poster almost didn't make it to a conference, but when it did, it made a major splash, and reactions were very positive. Some people like to use the salmon study as proof that fMRI is woo, but this isn't the case, it's actually a study to show the importance of correcting your stats★.

★What we mean by that is coming up later 🙂

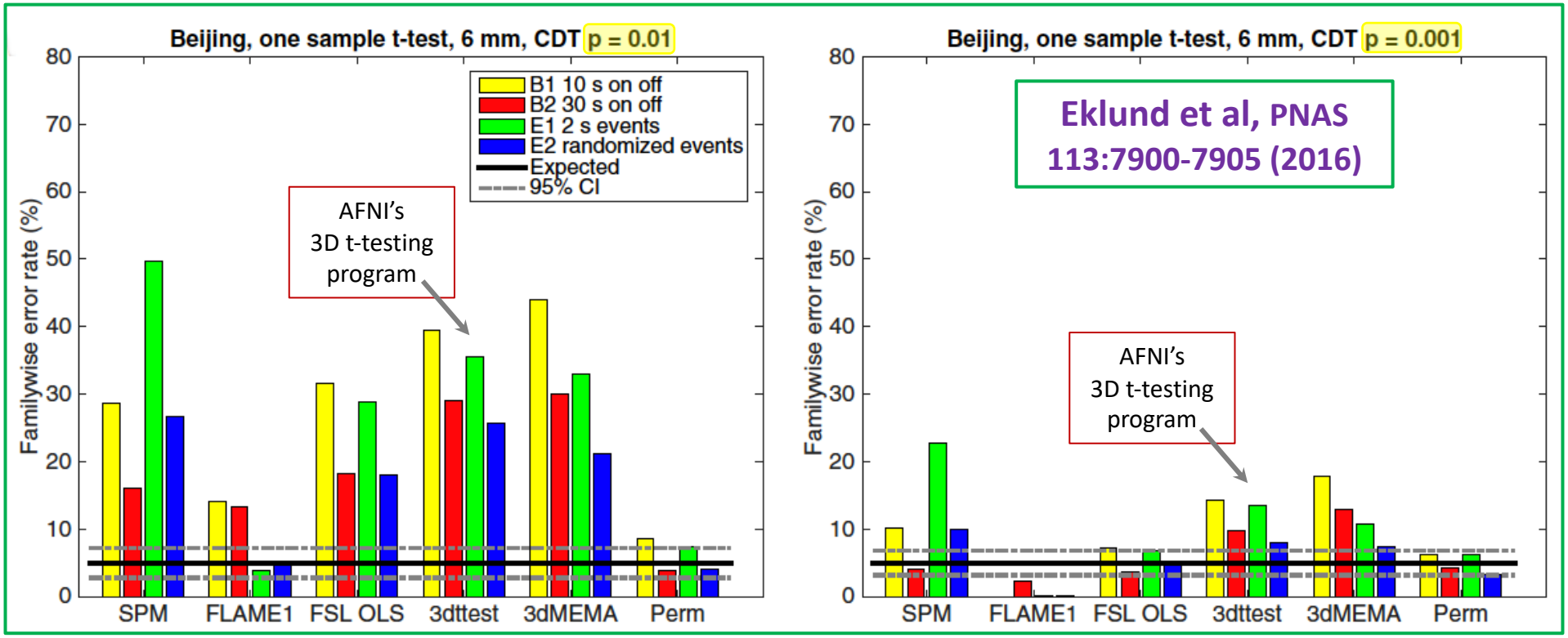
Dead Salmon: Authors' *True* Points

- Statistics controlling for the familywise error rate (FWER) and false discovery rate (FDR) both indicated that **no active voxels were present**, even at relaxed statistical thresholds.
- We argue that relying on **standard**★ statistical thresholds ($p < 0.001$) and low minimum cluster sizes ($k > 8$) is an ineffective control for multiple comparisons. [★RWC: Refers to *ad hoc* bad methods]
- We further argue that the vast majority of fMRI studies should be utilizing proper multiple comparisons correction as standard practice when thresholding their data.
  - RWC: Many papers using FMRI have not used *any* reasonable attempt to control for multiple comparisons
  - Especially those published in journals not used to imaging papers – for the most part, this doesn't apply any more

The Great Cluster Panic: 2016

- This one is more serious [https://doi.org/10.1073/pnas.1602413113]
  - But *also* over-hyped and poorly popularized
- Methods: Take many (living human) resting FMRI datasets
  - Analyze as if they were task (make up some timings)
  - Take collection of individual subject analysis results, and analyze *those* as if they were a group analysis
  - Anything found surviving the thresholding algorithm is a false positive (since there is no actual task in the data)
  - Run 1000 random such pseudo-task pseudo-group tests to find the *false positive rate* (FPR, aka FWER) – when the settings are for a nominal 5%
  - So should find about 50 false positives out of 1000 such collective simulations – ***if all is well in FMRI-land***
  - Analyses carried out with **AFNI**, **FSL**, and **SPM** software

The Great Cluster Panic: 2016



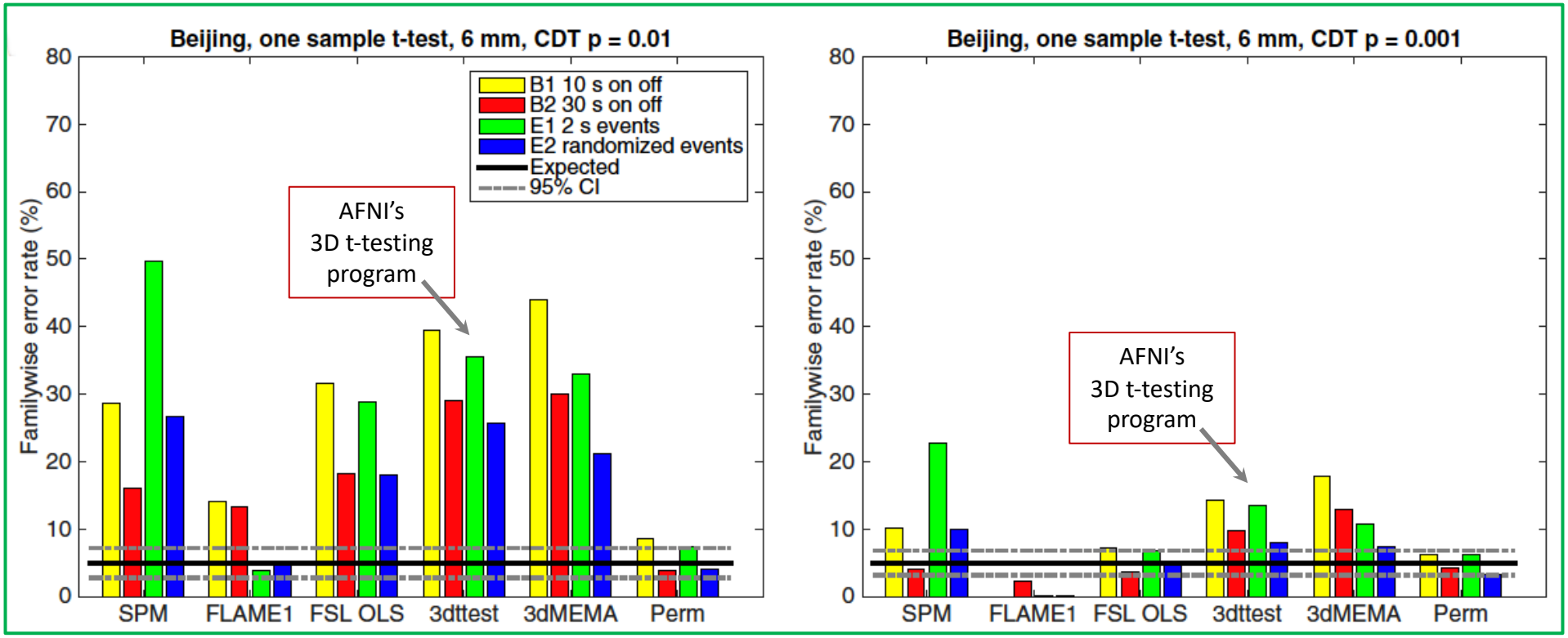
- FPR ≫ 5%: notably for voxelwise p=0.01
- A lot of doom-crying about this in 2016: 😱

Original paper

"These results question the validity of some 40,000 fMRI studies"

"Could Invalidate 15 Years of Brain Research" ← ScienceAlert

The Great Cluster Panic: 2016



- Response by journals: require *p*-threshold of 0.001 (or smaller)
- Response by SPM: *p*=0.001, nothing to see here; we've got *finesse*
- Response by FSL: we're fine with *p*=0.001 (after a little harrumphing)
- Response by **AFNI** (i.e., *moi*): some changes – a few slides ahead*!*
  - But first: a little background

Background: Voxelwise 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`)
- Second level group analysis on voxel β values = % signal change (*not* ROIs)
  - Can be as simple as *t*-tests (`3dttest++`)
  - Or a complicated model such as Linear Mixed Effects (`3dLME`), *etc.*

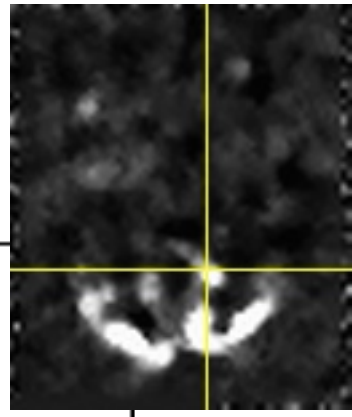
Background: Group Spatial Inference - 1

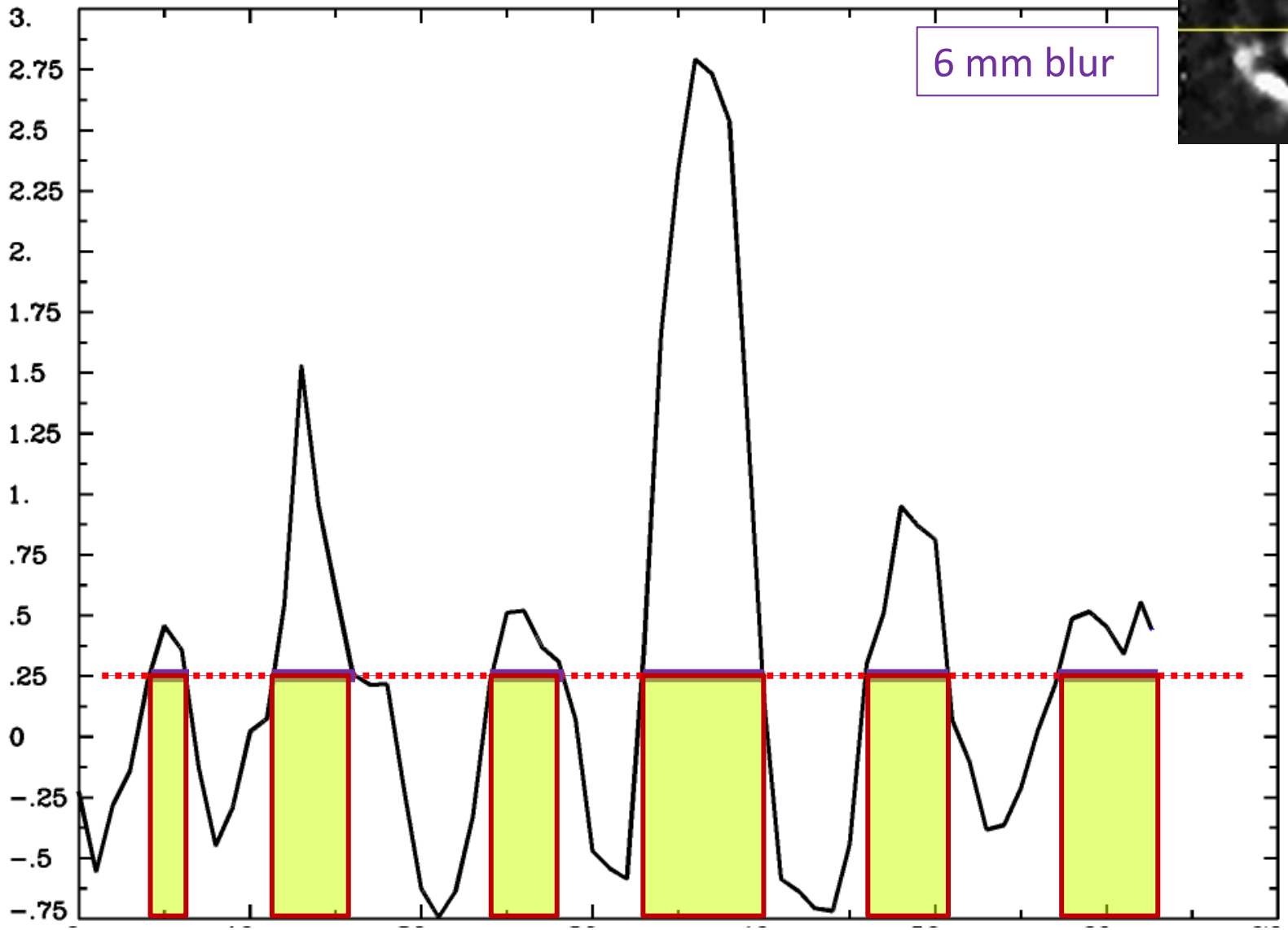
- Voxelwise thresholding on group *t*-statistic is usually super conservative (to get global FPR≈5%)
- ***A Solution***: form clusters of neighboring voxels, each above lower (less strict) voxelwise *t*-statistic (or *z*-statistic)
  - With larger voxelwise *p*-value (=smaller *t*) – e.g., *p*=0.01 vs 0.001
  - Clustering is (a) **correcting** for the voxelwise statistics giving too many false positives, and (b) building a *spatial* model for activation
- ***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 voxelwise *p*, adjust cluster-FOM threshold to get desired global FPR ⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛⇛…

Background: Group Spatial Inference - 2

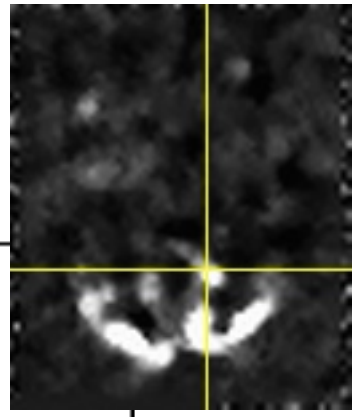
- Dual threshold method (voxel then cluster) can be weak (low power to detect)
- ***A Solution***: use spatial blurring ≈ average nearby voxel β ("Coef ") values together, in each subject, *before* group statistics
  - To reduce spatial noise and reinforce commonality
  - To reduce effective number of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

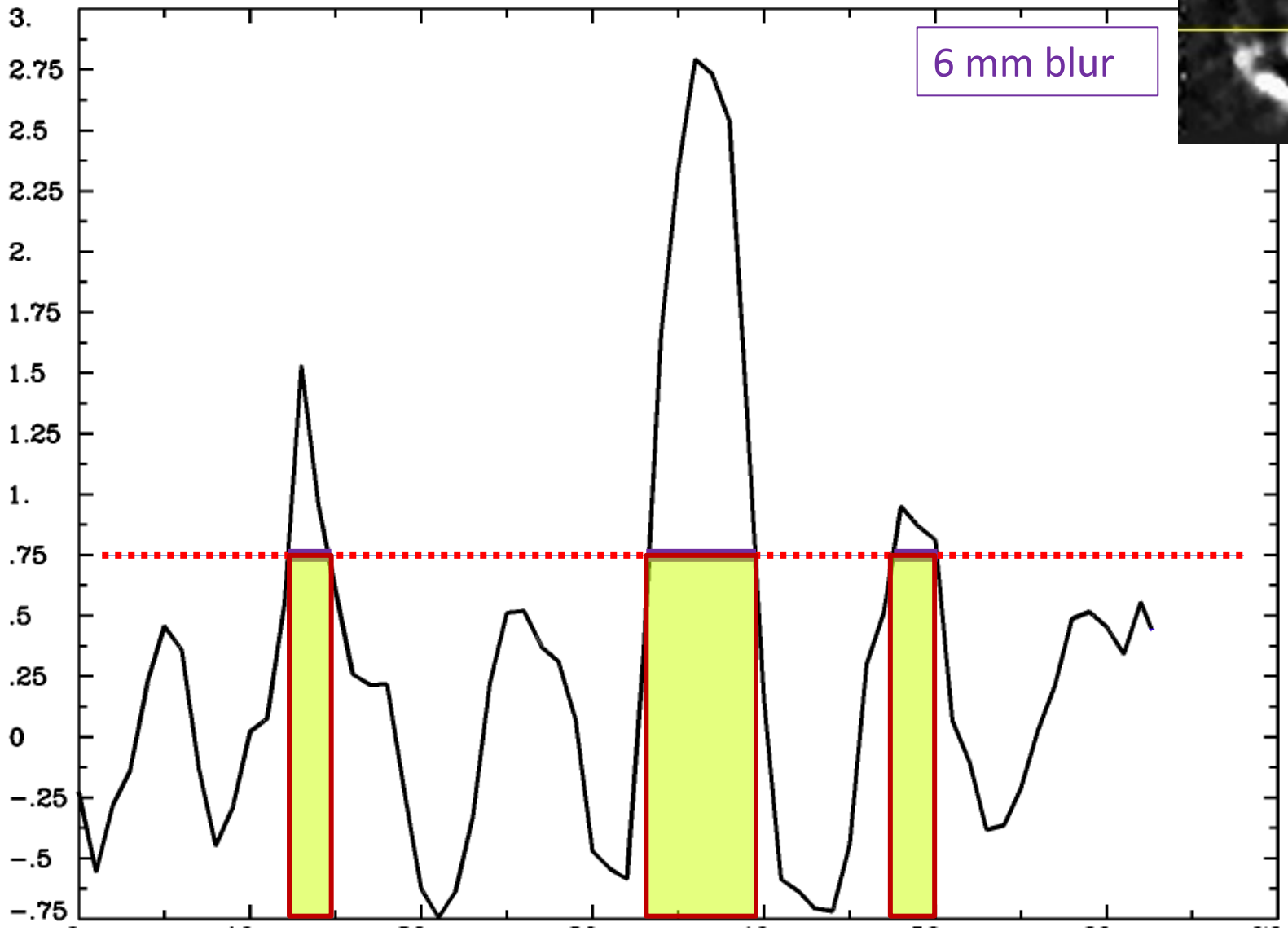
1D Dual Thresholding (real data)



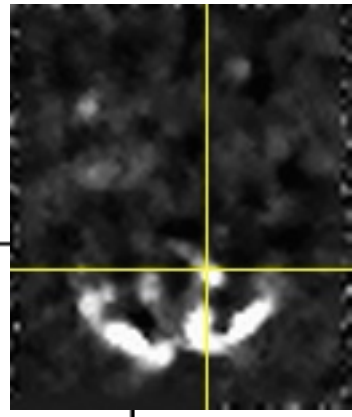


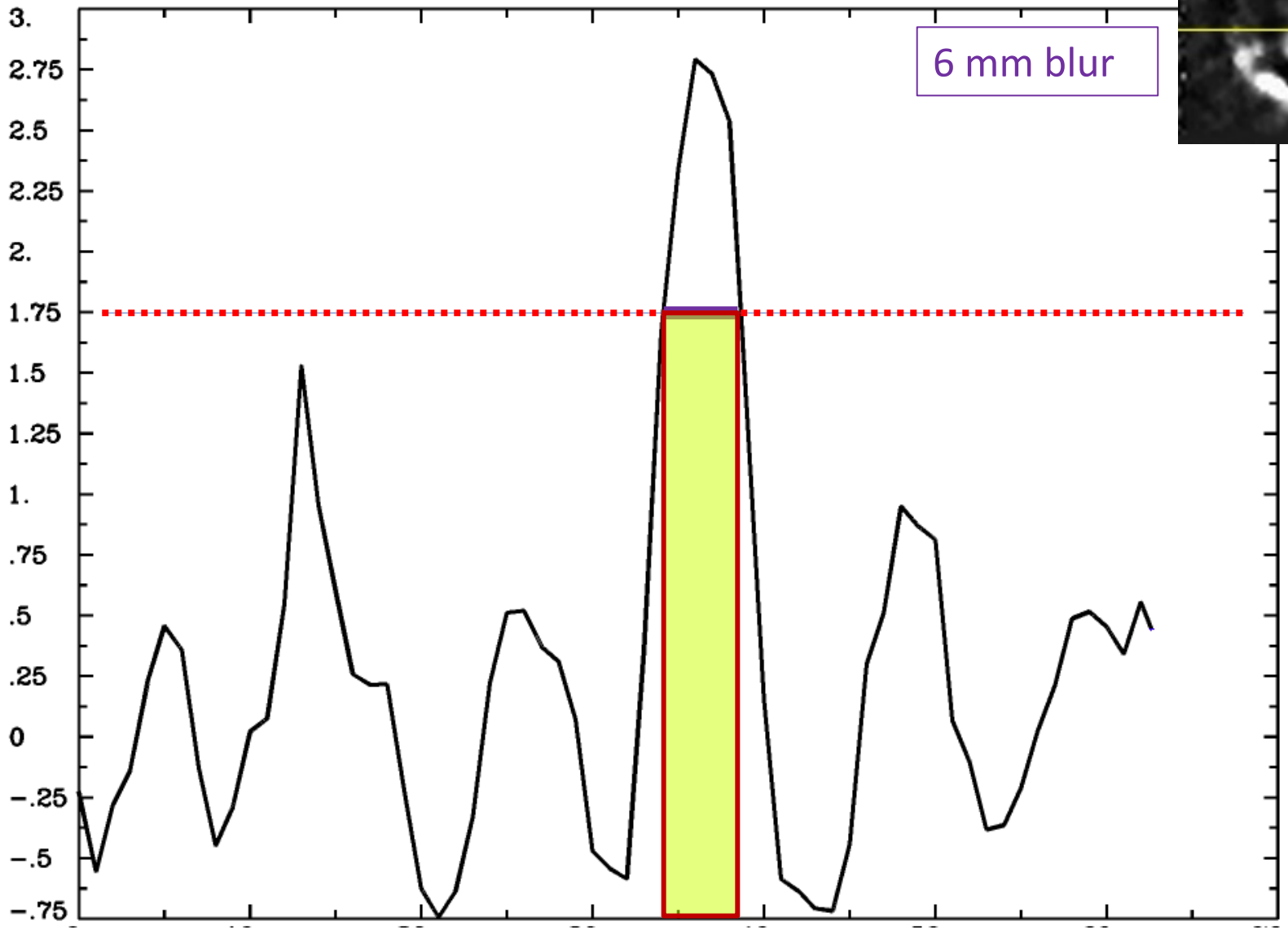
1D Dual Thresholding (real data)





1D Dual Thresholding (real data)

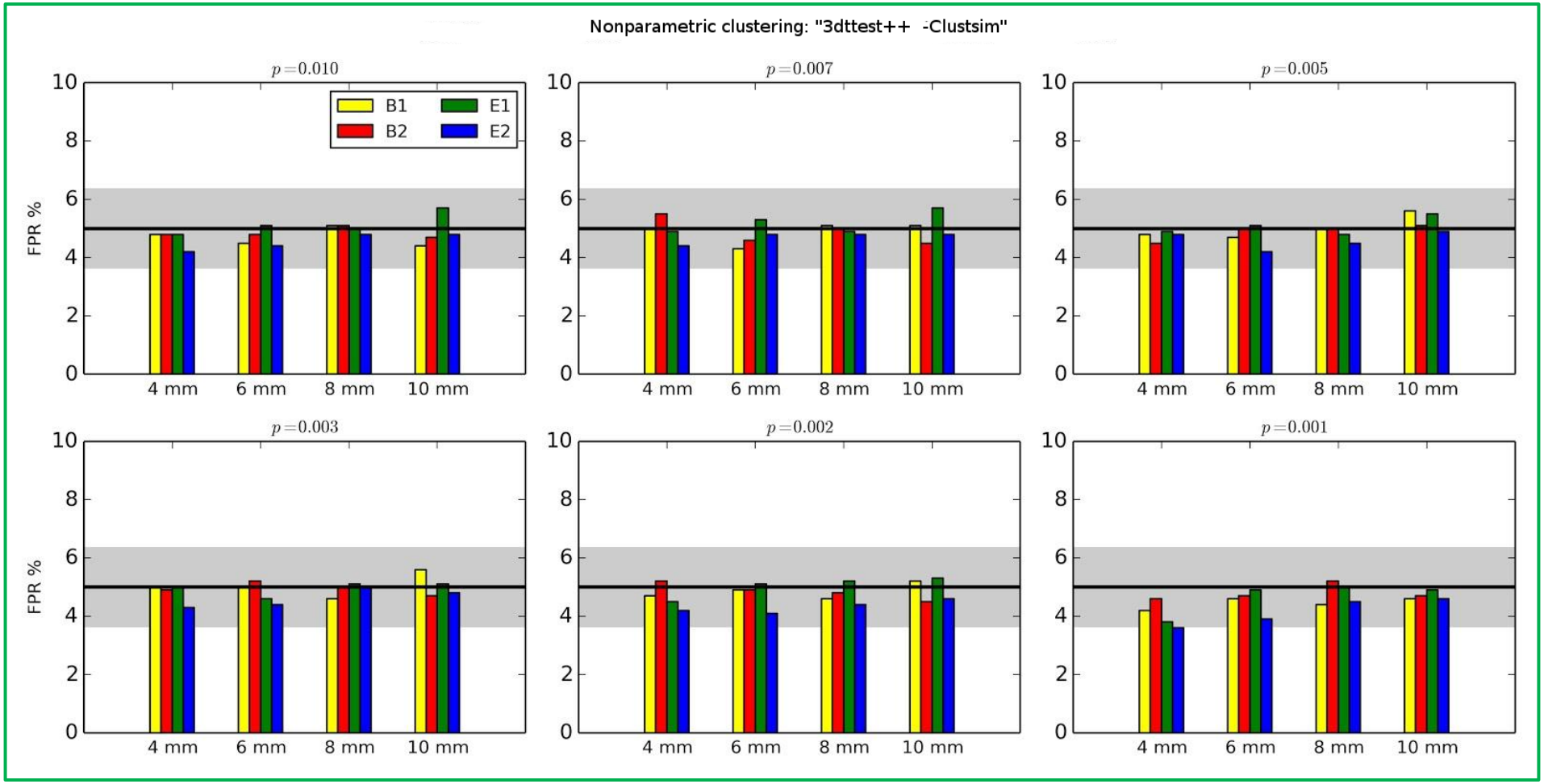




AFNI-land: Took Cluster Panic Seriously

- Problems arise in how cluster-threshold is calculated
  - Cluster-threshold that gives 5% FPR, as a function of per-voxel *p*-threshold, has traditionally been based on a mathematical model for the spatial distribution of the FMRI noise
    - To determine chance of finding anything in noise-only volumes
  - Difficulty with this approach: a good model is hard to get
- Alternative approach
  - Randomization of actual data to produce synthetic examples of ”noise-only” volumes, and then analyze *those* volumes to find cluster-threshold that gives a decent FPR (e.g., 5%)
  - 1 group analysis = randomize the signs of the data
  - 2 group analysis = also permute subjects between groups
  - **Brute force computation** to avoid relying on unreliable mathematical models to create the “noise-only” volumes
    - e.g., repeat randomization 10,000 times

Nonparametric Clustering in **AFNI**

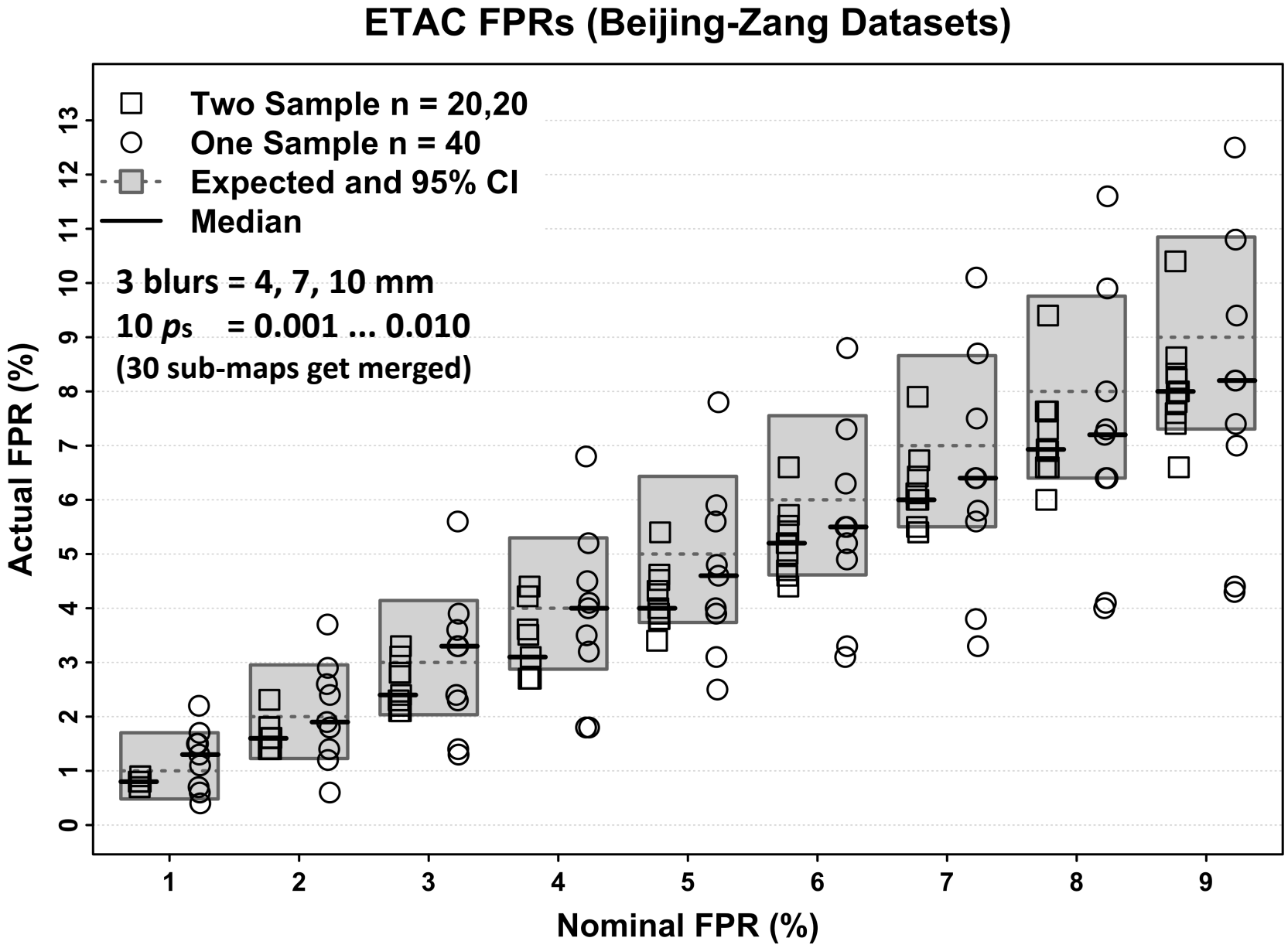


- *t*-test residuals are permuted/randomized (10,000 times)
- 10,000 re-*t*-tests computed from residuals, then fed to `3dClustSim`

AFNI: Beyond Mono-*p*-Thresholding

- Instead of using just a single *p*-threshold and then finding a cluster-threshold to give the desired FPR
- Using many *p*-thresholds (e.g., *p*=0.001, 0.002, …, 0.010)
  - Giving many supra-threshold brain maps
  - Cluster-threshold each of those separately
  - Merge the results
  - Adjust the multiple cluster-thresholds for the multiple *p*-thresholds to give the final desired FPR (say 5%)
  - **Eliminates the choice of an arbitrary *p*-threshold**
- Can do the same for amount of blurring
  - Use multiple blurring cases (e.g., 4mm, 6mm, 8mm) and merge results across blurs × *p*-thresholds
- Goal: find smaller intense clusters *and* larger weak clusters in the same group analysis

AFNI: Beyond Mono-Thresholding



Finally …

- Panics come around every few years in FMRI data analyses
- Earlier panics:
  - Voodoo Correlations
  - Head motion artifacts (at least twice)

- In every case, there was a real point deserving response
- In every case, the response was over-hyped, over-sold, and over-anxious (especially if it got into popular press)
- **Whence this anxiety***?*
  - **IMHO**, it is rooted in the fact we don't really know what we are measuring with FMRI and how it relates to neural "computation", and so don't really know how to interpret our results
  - But: we know that it is just a crude proxy for "real" brain activity
  - The worst functional brain mapping method, **except for all others**

Where My Clustering Ideas Started