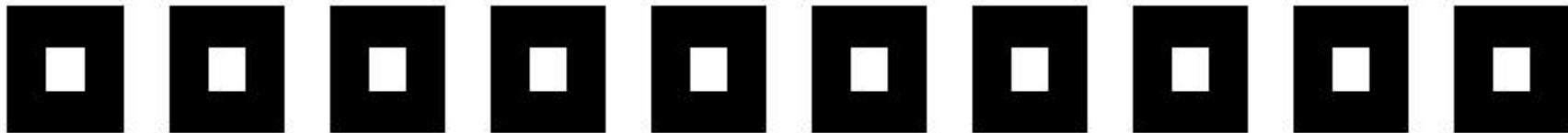


Multiple Comparisons II

Example

Signal



+

Noise

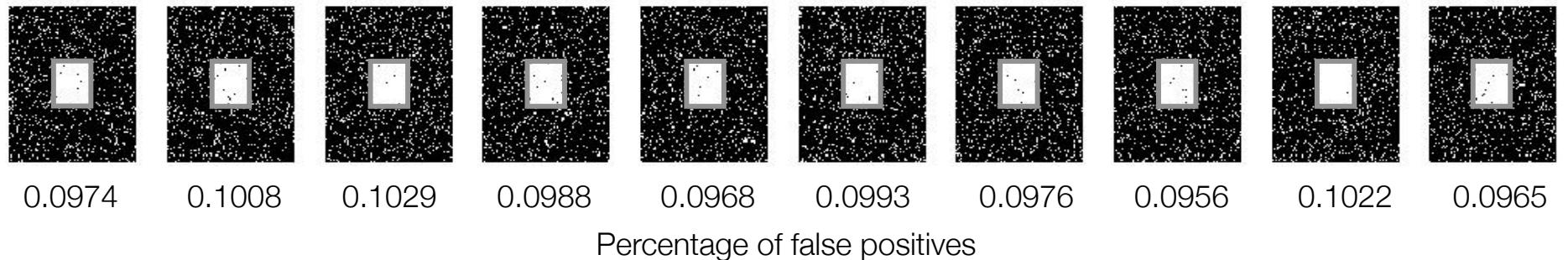


=

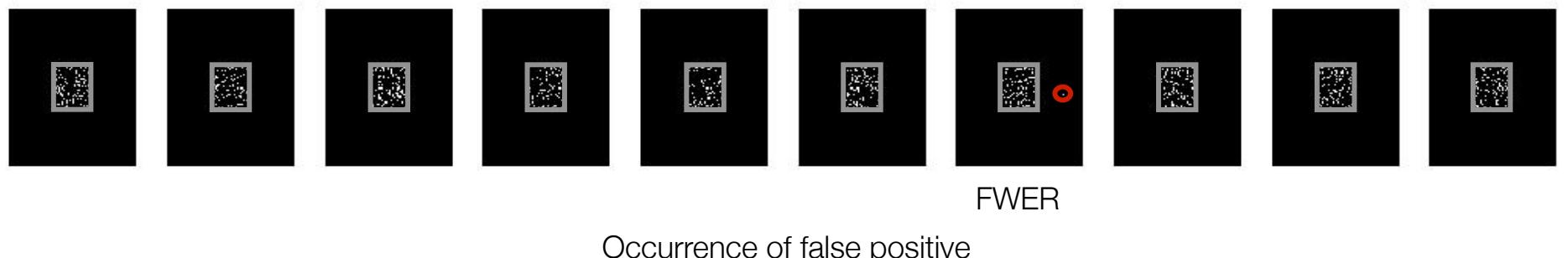
Signal + Noise



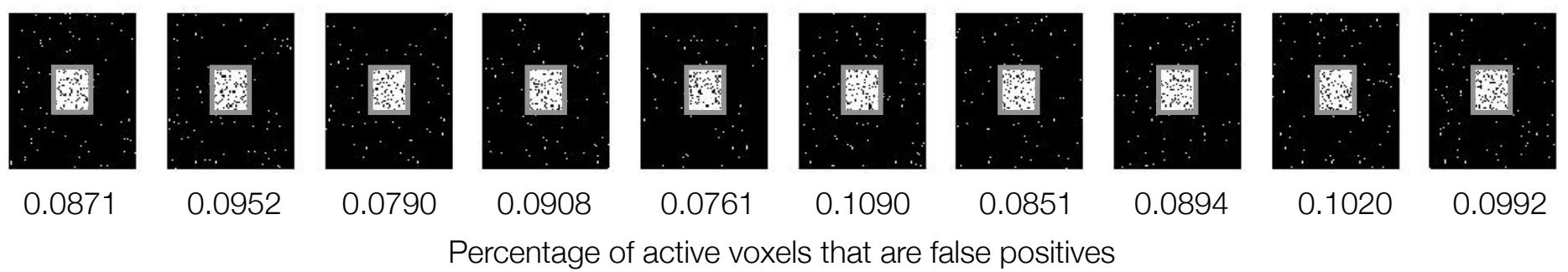
$\alpha=0.10$, No correction



FWER control at 10%



FDR control at 10%



Cluster-level Inference

- Two step-process
 - Define clusters by arbitrary threshold u_{clus}
 - Retain clusters larger than α -level threshold k_α

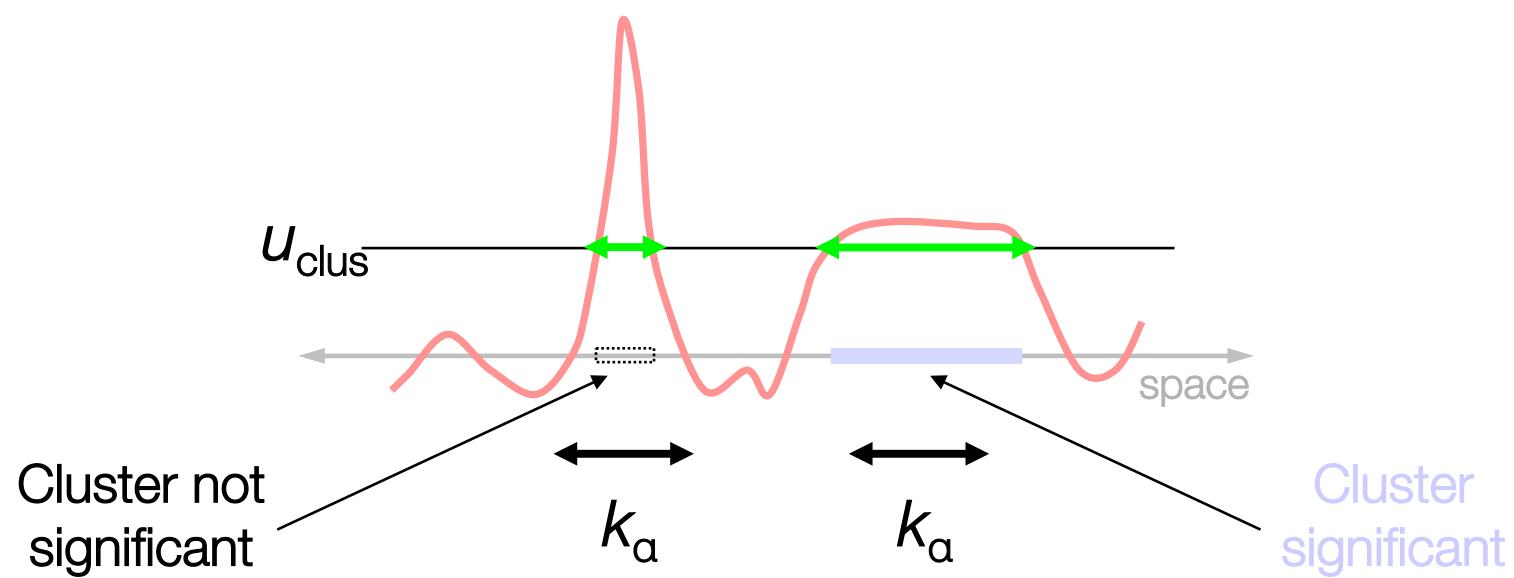


Diagram: Tom Nichols

Cluster-level Inference

- Typically better sensitivity
- Worse spatial specificity
 - The null hyp. of entire cluster is rejected
 - Only means that *one or more* voxels in cluster active

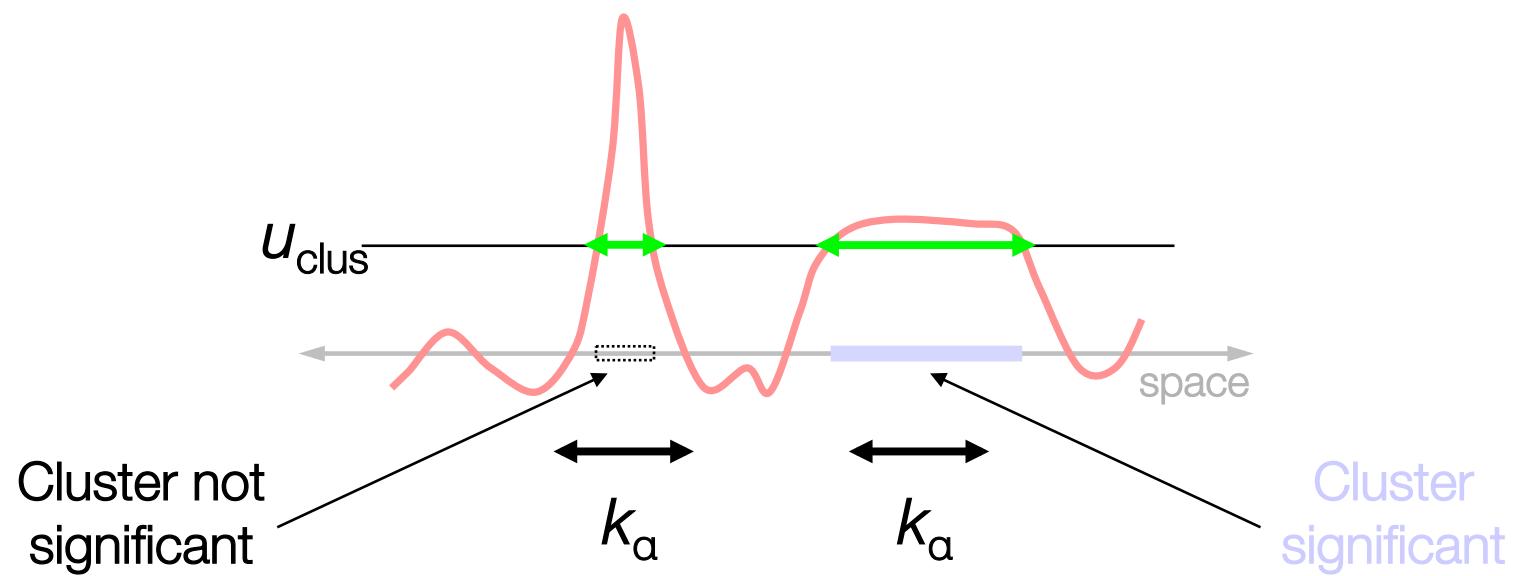
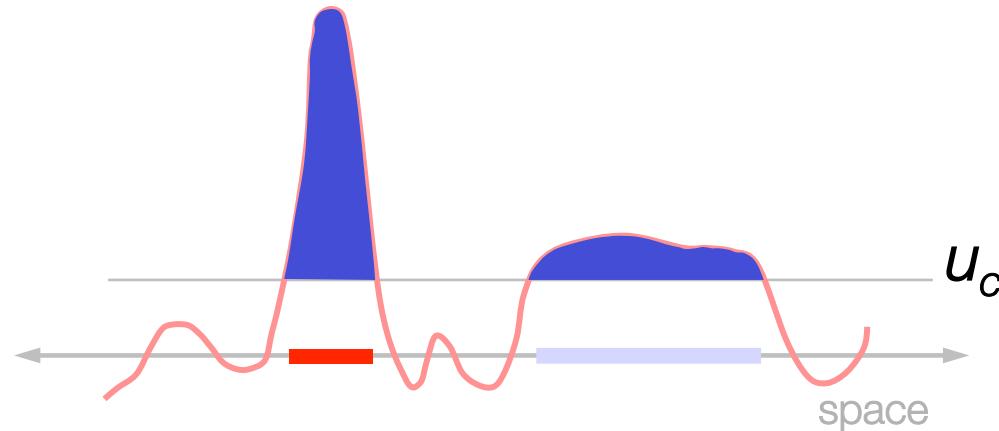


Diagram: Tom Nichols

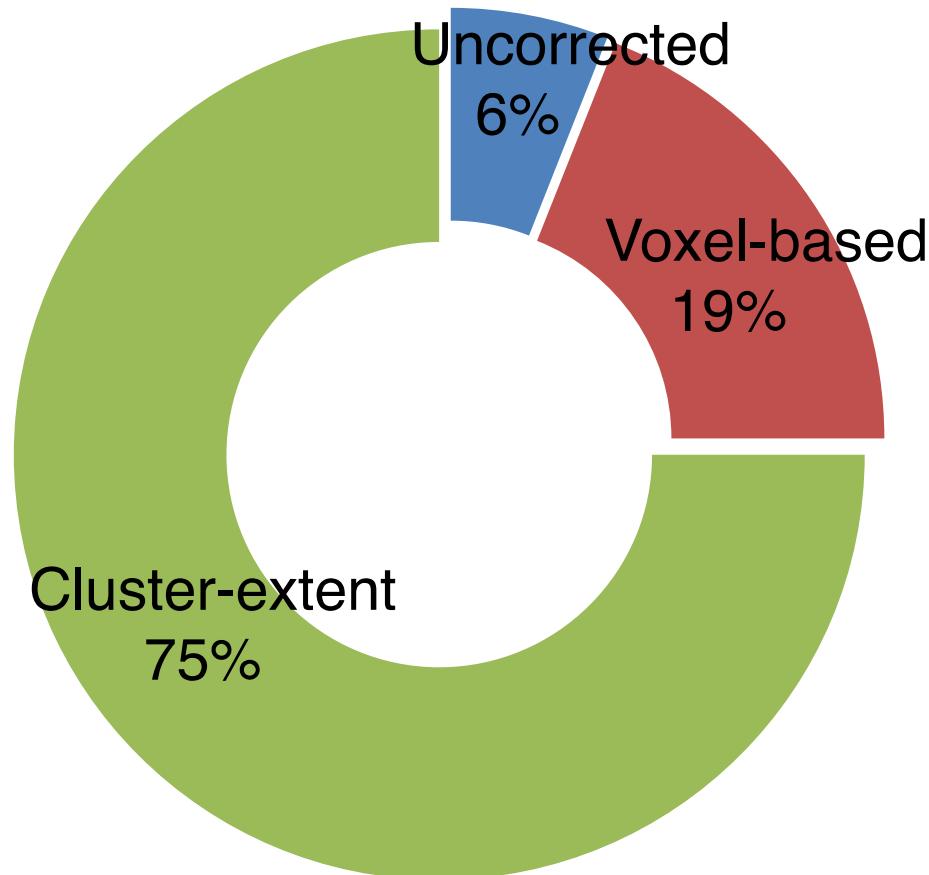
Combining Cluster Size with Intensity Information

- ‘Threshold-free cluster enhancement’ (TFCE)
 - Integral M above threshold
 - More powerfully combines peak & height (Hayasaka & Nichols, NI 2004)



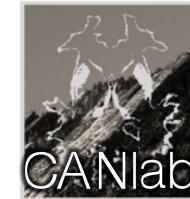
TFCE in FSL’s *Randomize*

Diagram: Tom Nichols





University of Colorado **Boulder**



PITFALLS OF USING UNCORRECTED THRESHOLDS

Uncorrected Thresholds

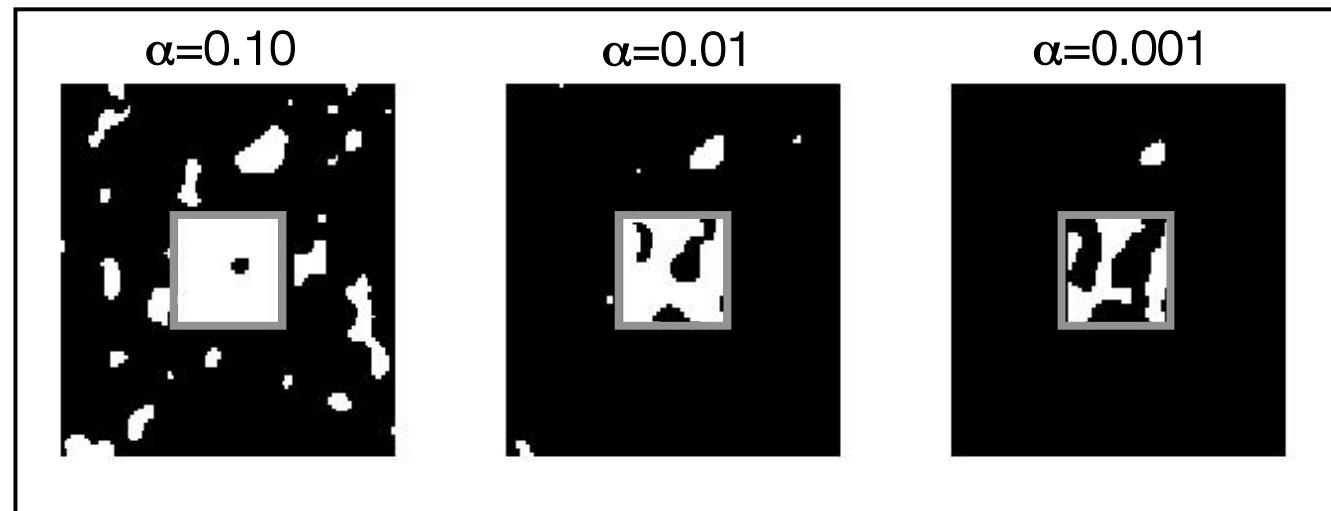
- Many published PET and fMRI studies use arbitrary uncorrected thresholds (e.g., $p < 0.001$).
 - A likely reason is that with available sample sizes, corrected thresholds are so stringent that power is extremely low.
- Using uncorrected thresholds is problematic when interpreting conclusions from individual studies, as many activated regions may be false positives.
- Null findings are hard to disseminate, hence it is difficult to refute false positives established in the literature.

Extent Threshold

- Sometimes an arbitrary **extent threshold** is used when reporting results.
- Here a voxel is only deemed truly active if it belongs to a cluster of k contiguous active voxels (e.g., $p < 0.001$, 10 contingent voxels).
- Unfortunately, this does not necessarily correct the problem because imaging data are spatially smooth and therefore false positives may appear in clusters.

Example

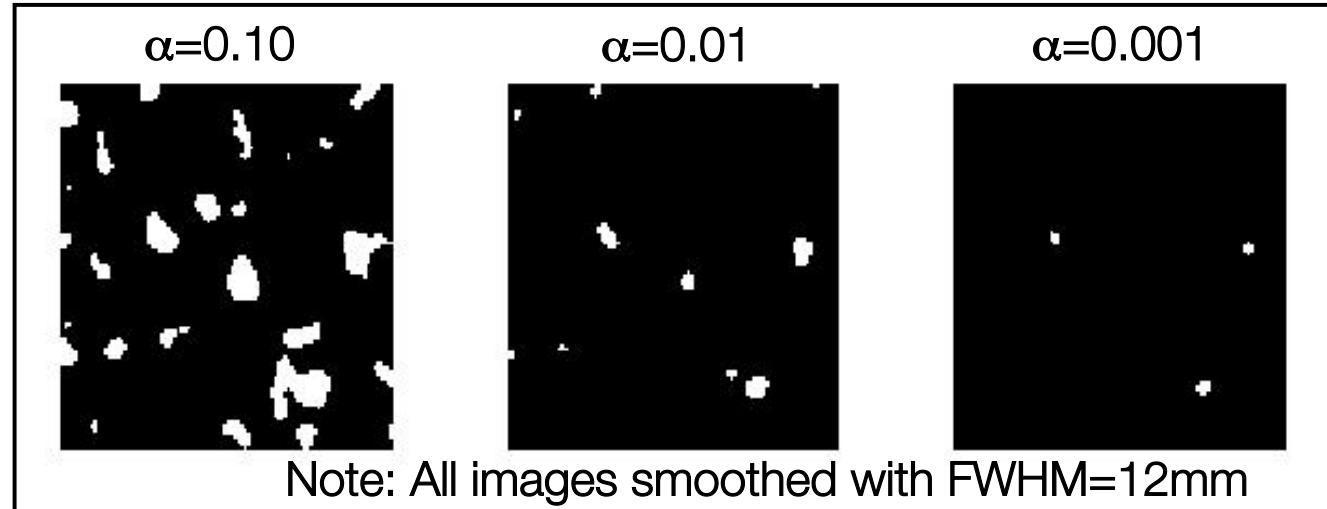
- Activation maps with spatially correlated noise thresholded at three different significance levels. Due to the smoothness, the false-positive activation are contiguous regions of multiple voxels.



Note: All images smoothed with FWHM=12mm

Example

- Similar activation maps using null data.

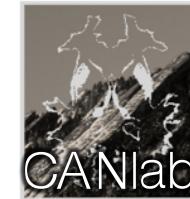


- Using typical arbitrary height and extent threshold is too liberal!
(e.g., $p < .001$ and 10 voxels; Forman et al., 1995).
- Why? Data is spatially smooth, and false positives tend to be “blobs” of many voxels

e.g. Wager, Lindquist, & Hernandez, 2009. “Essentials of functional neuroimaging.”
In: Handbook of Neuroscience for the Behavioral Sciences



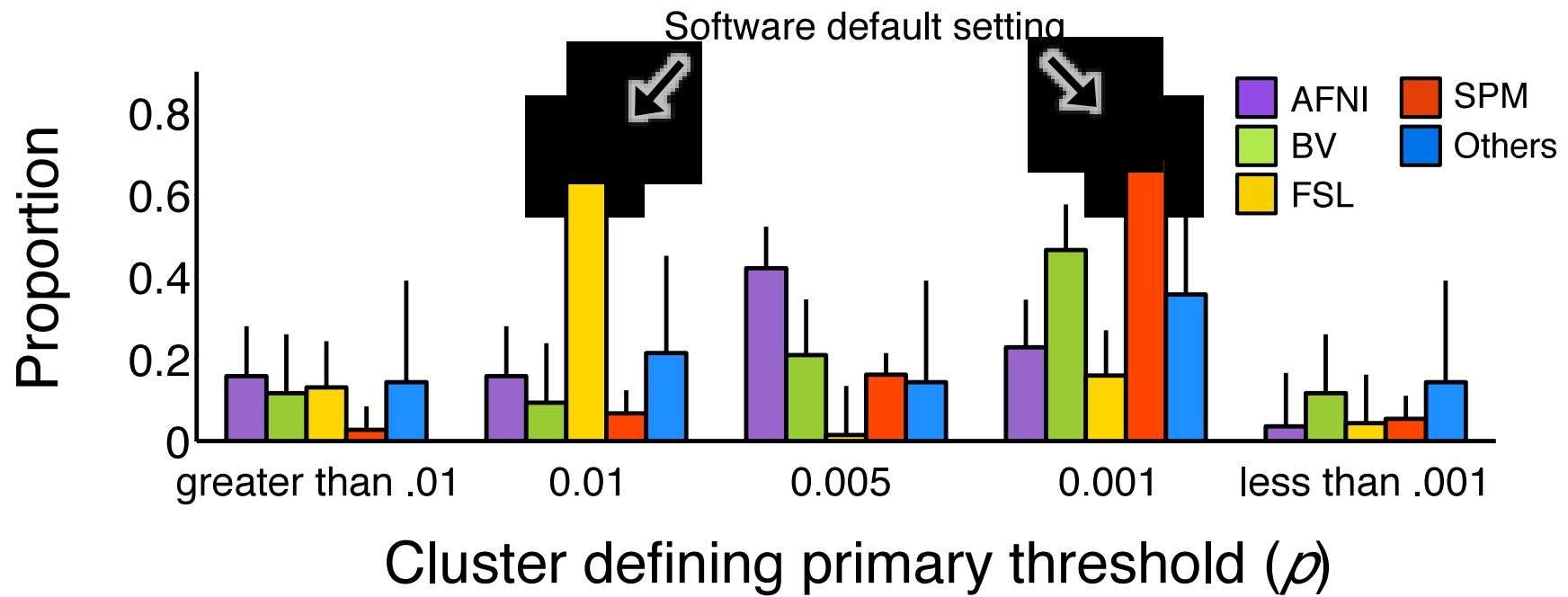
University of Colorado **Boulder**



PITFALLS OF CLUSTER EXTENT-BASED* CORRECTION

* Also applies to TFCE

What are people doing?



fMRI studies from Cerebral Cortex, Nature, Nature Neuroscience, NeuroImage, Neuron, PNAS, and Science ($N = 814$).

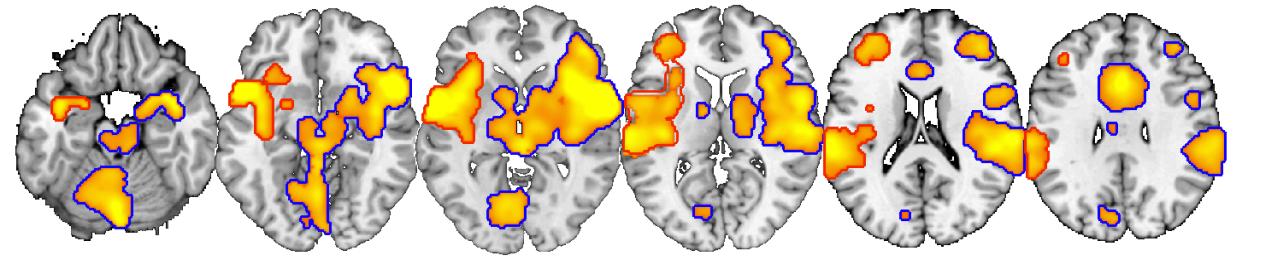


Woo et al. 2014

Example cluster-level inference results

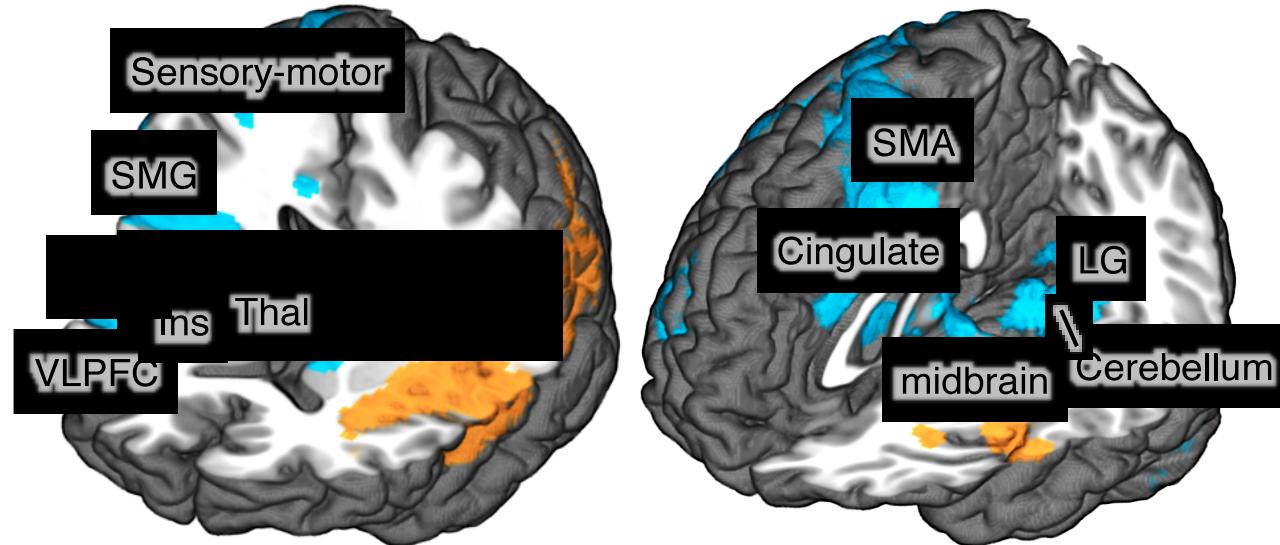
Cluster $p < .05$, FWER corrected with primary $p < .01$

Thermal
pain, N =
20



○ Cluster ○ Cluster 2

Cluster 1
Cluster 2

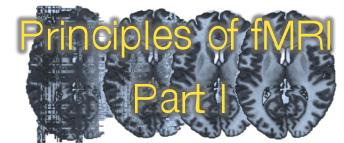


Neuroanatomy test: Which regions are activated?

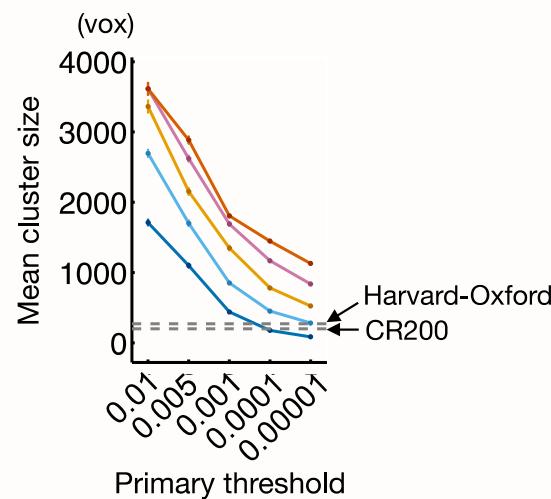


Woo et al. 2014

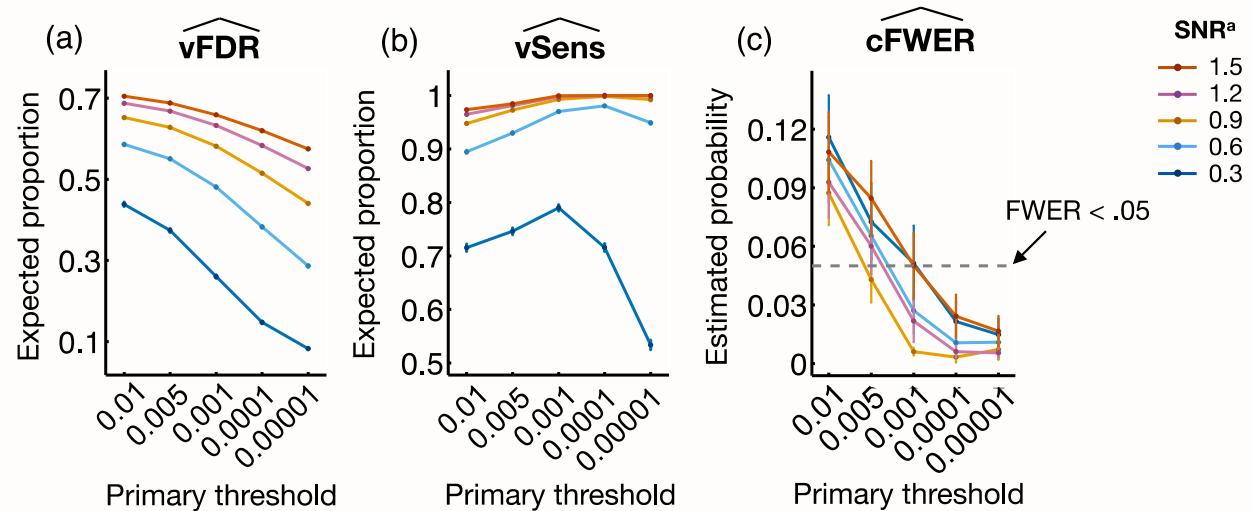
Issues with cluster-extent based thresholding



A. Mean size of significant clusters



B. Evaluation measures at cluster $p < .05$, FWER corrected
(Low vFDR, high vSens, and lower cFWER than .05 are desirable)



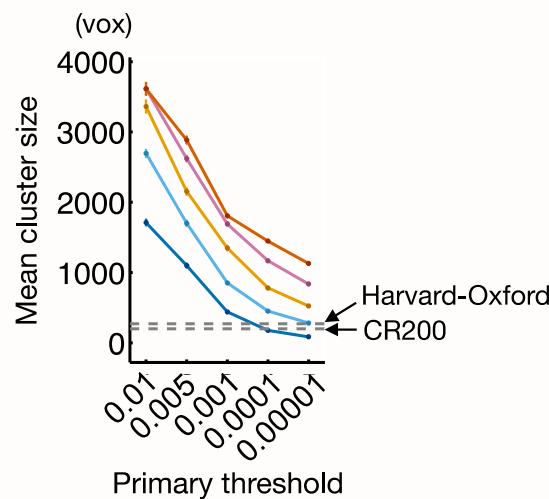
Non-specificity: Most findings are larger than useful anatomical areas



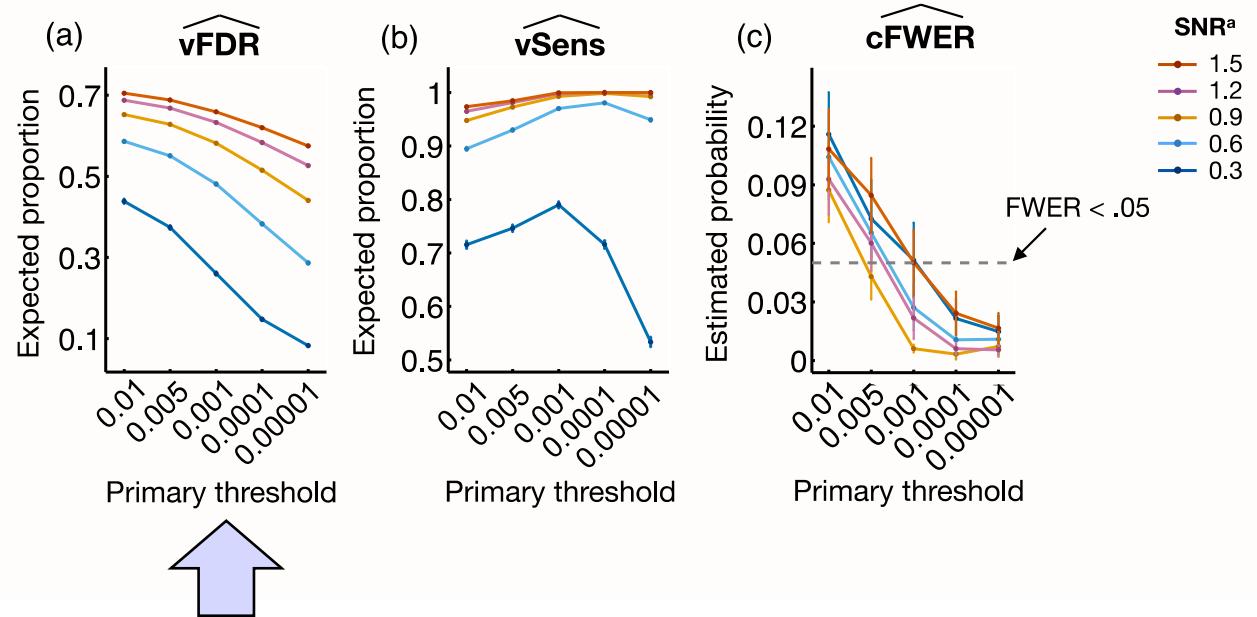
Issues with cluster-extent based thresholding



A. Mean size of significant clusters



B. Evaluation measures at cluster $p < .05$, FWER corrected
(Low vFDR, high vSens, and lower cFWER than .05 are desirable)

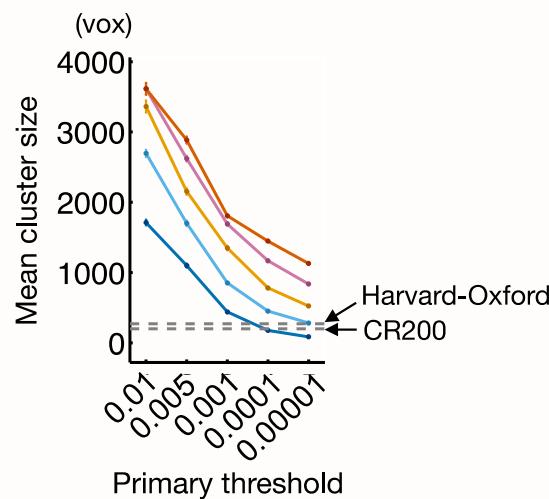


False discoveries: Most (45-70%) activated voxels in map are not truly active

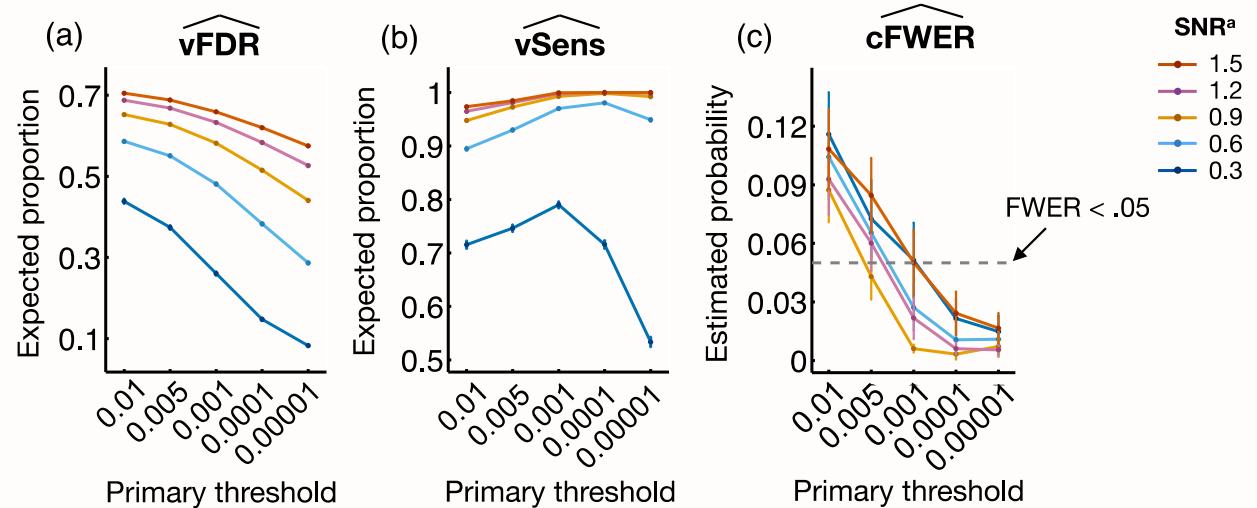


Issues with cluster-extent based thresholding

A. Mean size of significant clusters



B. Evaluation measures at cluster $p < .05$, FWER corrected
(Low vFDR, high vSens, and lower cFWER than .05 are desirable)



False discoveries: Family-wise error rate is not controlled with low thresholds
-- You are not actually getting $p < .05$ FWER corrected!!



Wrap-up

- Multiple comparison methods
 - Uncorrected
 - FWE
 - FDR
 - Cluster extent-based (FWE, FDR)

- Pitfalls
 - Uncorrected thresholds
 - Cluster extent-based thresholds



End of Module



@fMRIstats