Advanced issues in fMRI statistics: Nonparametric Inference, Power & Meta-Analysis

Thomas E. Nichols
University of Warwick

Zurich SPM Course

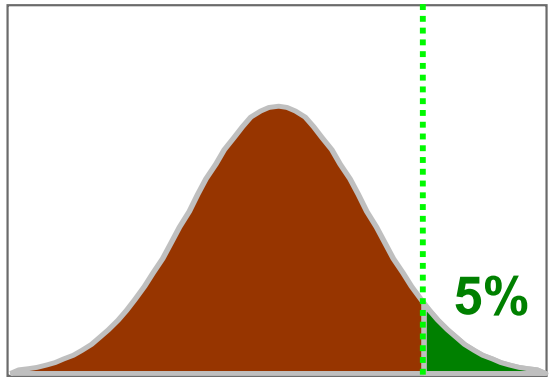
18 February, 2016

Advanced issues in fMRI statistics

- Nonparametric Inference
  - What if I don't trust my assumptions?
- Power
  - What's the chance of finding my effect? (pre-data)
- Meta-Analysis
  - What does the literature say?

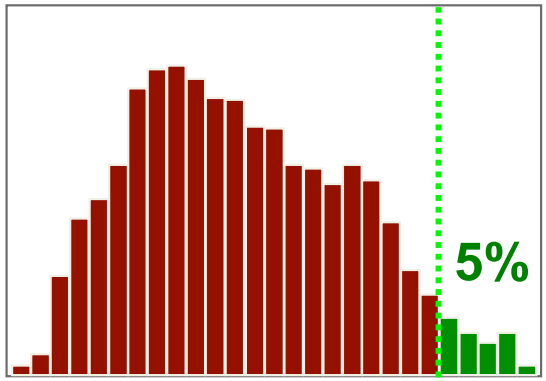
Nonparametric Inference

- Parametric methods
  - Assume distribution of statistic under null hypothesis
  - Needed to find P-values, u_α



Parametric Null Distribution

- Nonparametric methods
  - Use *data* to find distribution of statistic under null hypothesis
  - Any statistic!



Nonparametric Null Distribution

Permutation Test
Toy Example

- Data from tiny pharmaceutical trial
- 2 groups, 3 subjects each!
- Drug A & B; does A increase (BOLD) response?
  6 subjects, collected interleaved… ABABAB

A	B	A	B	A	B
103.00	90.48	99.93	87.83	99.76	96.06

- Null hypothesis H_o
  - No experimental effect, A & B labels arbitrary
- Statistic
  - Mean difference

Permutation Test
Toy Example

- Under H_o
  - Consider all equivalent relabelings

AAABBB	ABABAB	BAAABB	BABBAA
AABABB	ABABBA	BAABAB	BBAAAB
AABBAB	ABBAAB	BAABBA	BBAABA
AABBBA	ABBABA	BABAAB	BBABAA
ABAABB	ABBBAA	BABABA	BBBAAA

Permutation Test
Toy Example

- Under H_o
  - Consider all equivalent relabelings
  - Compute all possible statistic values

AAABBB 4.82	ABABAB 9.45	BAAABB -1.48	BABBAA -6.86
AABABB -3.25	ABABBA 6.97	BAABAB 1.10	BBAAAB 3.15
AABBAB -0.67	ABBAAB 1.38	BAABBA -1.38	BBAABA 0.67
AABBBA -3.15	ABBABA -1.10	BABAAB -6.97	BBABAA 3.25
ABAABB 6.86	ABBBAA 1.48	BABABA -9.45	BBBAAA -4.82

Permutation Test
Toy Example

- Under H_o
  - Consider all equivalent relabelings
  - Compute all possible statistic values
  - Find 95%ile of permutation distribution

AAABBB 4.82	ABABAB 9.45	BAAABB -1.48	BABBAA -6.86
AABABB -3.25	ABABBA 6.97	BAABAB 1.10	BBAAAB 3.15
AABBAB -0.67	ABBAAB 1.38	BAABBA -1.38	BBAABA 0.67
AABBBA -3.15	ABBABA -1.10	BABAAB -6.97	BBABAA 3.25
ABAABB 6.86	ABBBAA 1.48	BABABA -9.45	BBBAAA -4.82

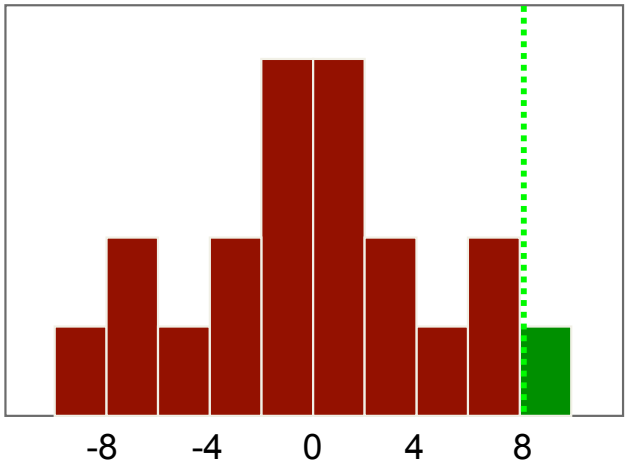
Permutation Test
Toy Example

- Under H_o
  - Consider all equivalent relabelings
  - Compute all possible statistic values
  - Find 95%ile of permutation distribution

AAABBB 4.82	ABABAB **9.45**	BAAABB -1.48	BABBAA -6.86
AABABB -3.25	ABABBA 6.97	BAABAB 1.10	BBAAAB 3.15
AABBAB -0.67	ABBAAB 1.38	BAABBA -1.38	BBAABA 0.67
AABBBA -3.15	ABBABA -1.10	BABAAB -6.97	BBABAA 3.25
ABAABB 6.86	ABBBAA 1.48	BABABA -9.45	BBBAAA -4.82

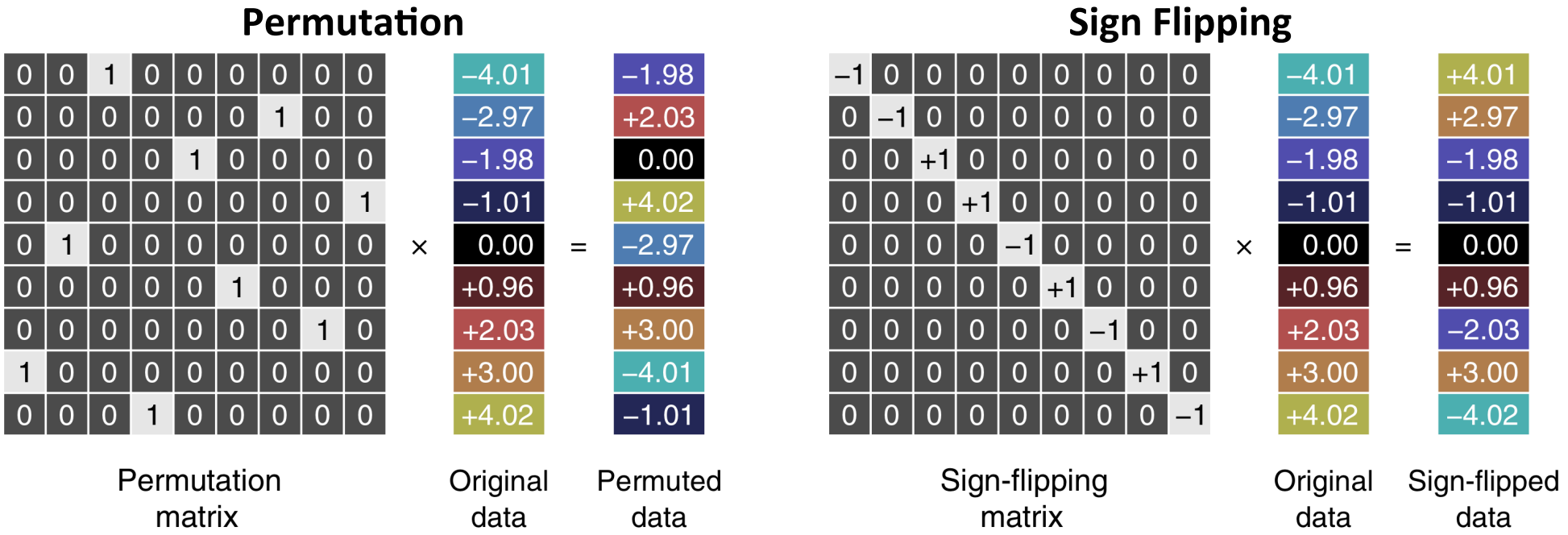
Permutation Test
Toy Example

- Under H_o
  - Consider all equivalent relabelings
  - Compute all possible statistic values
  - Find 95%ile of permutation distribution



Permutation vs. Sign Flipping

- For 2 groups, or for a correlation… permute
- For a 1-sample t-test… what can you permute?
- Sign flipping
  - Allows a ‘permutation’ test with 1 group data
  - Justified by symmetrically distributed errors

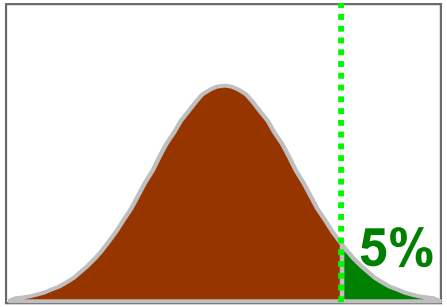


Multiple Tests: What is "A False Positive"?

- False Discovery Rate (FDR)
  - Expected proportion of false positives among detections
  - Compute from voxel- , peak- or cluster-wise uncorrected P-values
- Familywise Error Rate (FWE)
  - Chance of one or more false positives

$$\begin{aligned} \text{FWE}(u) &= \text{P}(\text{One or more false positives} \mid H_o) \\ &= \text{P}(\text{Max voxel above threshold } u \mid H_o) \\ &= \text{P}(\max T \geq u \mid H_o) \end{aligned}$$

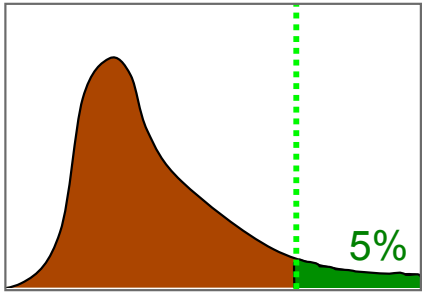
Uncorr: Tail area on 1-voxel null



H_o Distribution (1 voxel)

FWE-corrected thresholds / P-values just like uncorrected!

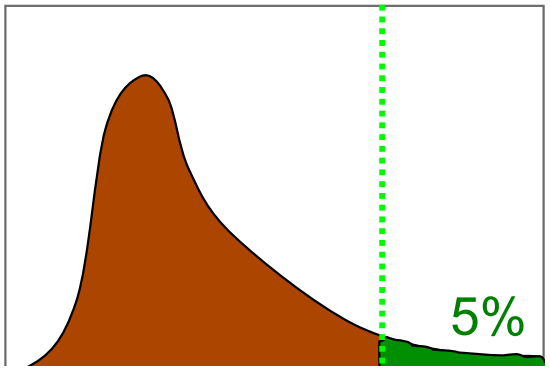
FWE-corr: Tail area on max. null



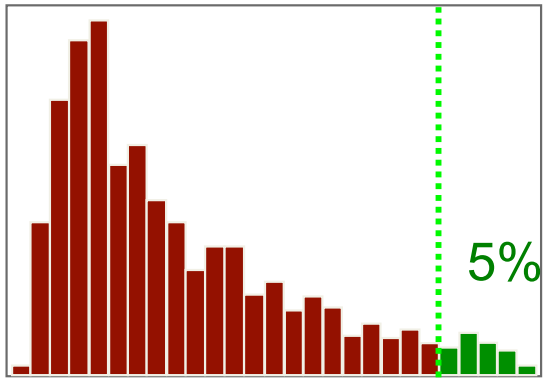
H_o Maximum Distribution

Controlling FWE: Permutation Test

- Parametric methods
  - Assume distribution of *max* statistic under null hypothesis
- Nonparametric methods
  - Use *data* to find distribution of *max* statistic under null hypothesis
  - Again, any max statistic!



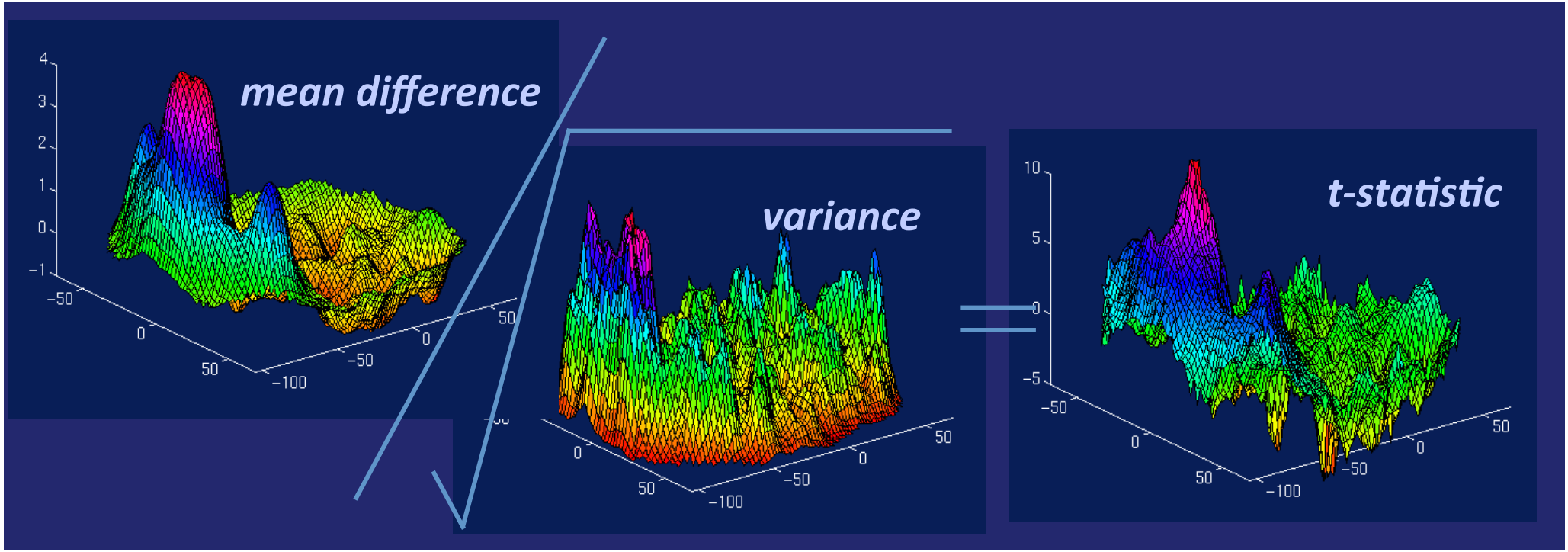
Parametric Null Max Distribution



Nonparametric Null Max Distribution

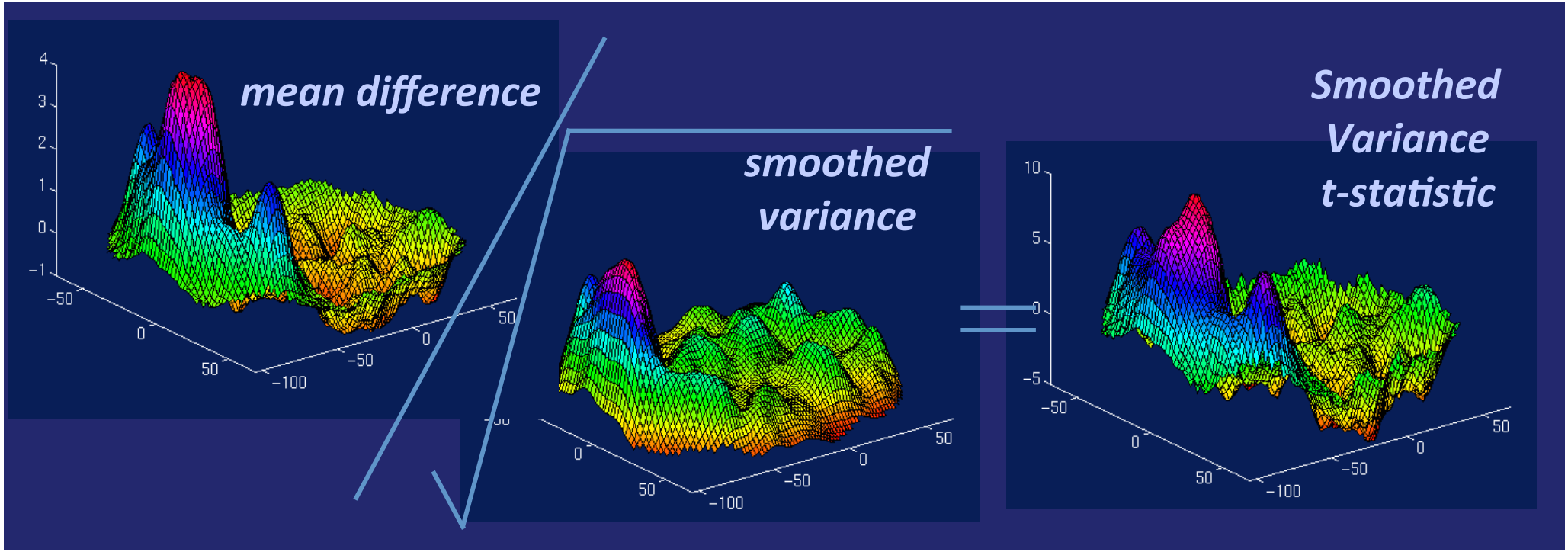
Permutation Test
Smoothed Variance *t*

- Collect max distribution
  - To find threshold that controls FWE
- Consider smoothed variance *t* statistic



Permutation Test
Smoothed Variance *t*

- Collect max distribution
  - To find threshold that controls FWE
- Consider smoothed variance *t* statistic



Permutation Test Strengths

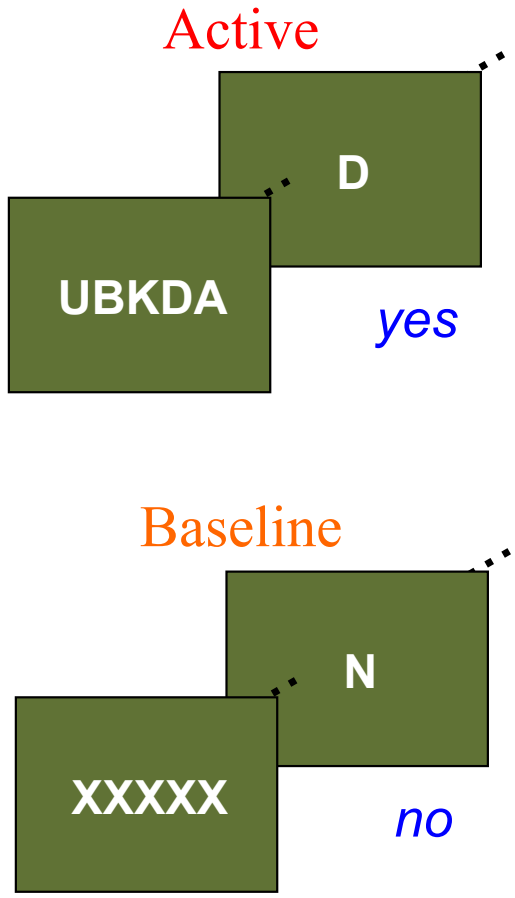
- Requires only assumption of exchangeability
  - Under Ho, distribution unperturbed by permutation
  - Allows us to build permutation distribution
- Subjects are exchangeable
  - Under Ho, each subject's A/B labels can be flipped
- fMRI scans not exchangeable under Ho
  - Due to temporal autocorrelation

Permutation Test Limitations

- Computational Intensity
  - Analysis repeated for each relabeling
  - Not so bad on modern hardware
    - No analysis discussed below took more than 2 minutes
- Implementation Generality
  - Each experimental design type needs unique code to generate permutations
    - Not so bad for population inference with t-tests

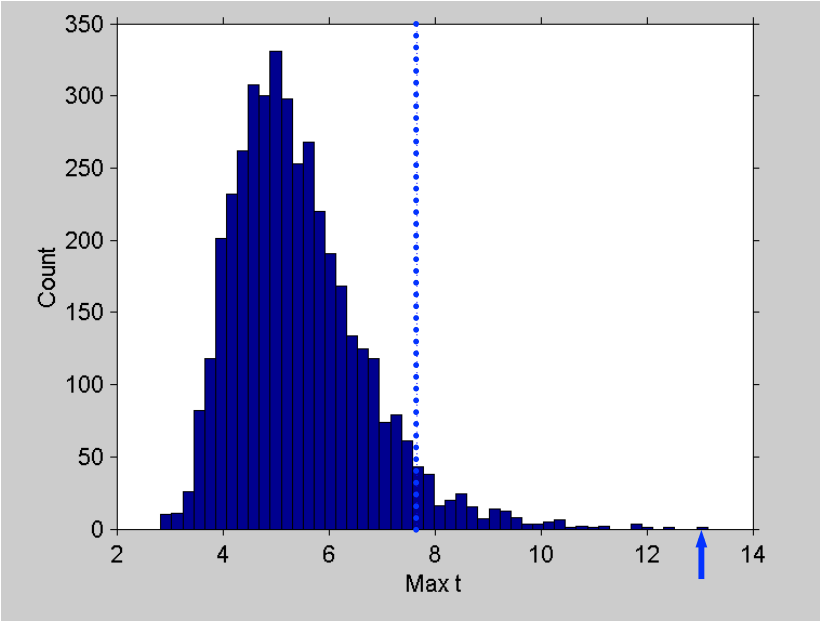
Permutation Test Example

- fMRI Study of Working Memory
  - 12 subjects, block design Marshuetz et al (2000)
  - Item Recognition
    - Active: View five letters, 2s pause, view probe letter, respond
    - Baseline: View XXXXX, 2s pause, view Y or N, respond
- Second Level RFX
  - Difference image, A-B constructed for each subject
  - One sample, smoothed variance *t* test

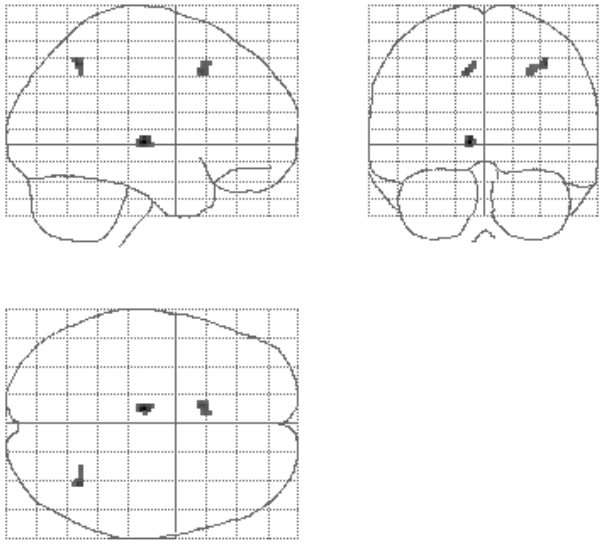


Permutation Test Example

- Permute!
  - $2^{12} = 4{,}096$ ways to flip 12 A/B labels
  - For each, note maximum of *t* image



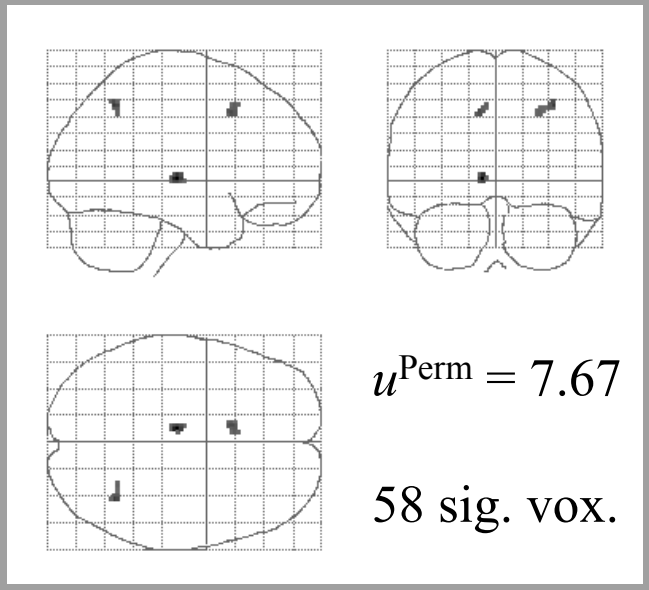
Permutation Distribution
Maximum *t*



Maximum Intensity Projection
Thresholded *t*

Permutation Test Example

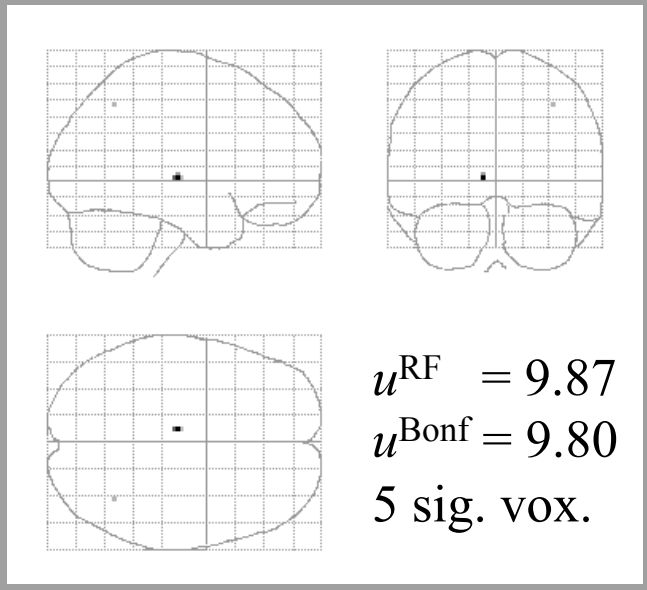
- Compare with Bonferroni
  - $\alpha = 0.05/110{,}776$
- Compare with parametric RFT
  - 110,776 2×2×2mm voxels
  - 5.1×5.8×6.9mm FWHM smoothness
  - 462.9 RESELs



$u^{\text{Perm}} = 7.67$

58 sig. vox.

t_{11} Statistic, Nonparametric Threshold

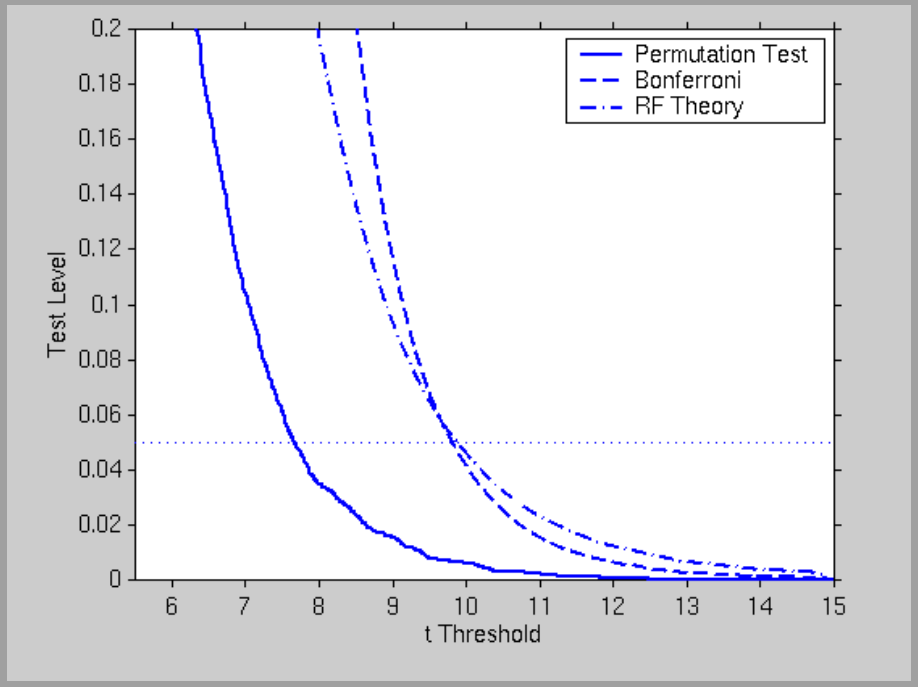


$u^{\text{RF}} = 9.87$

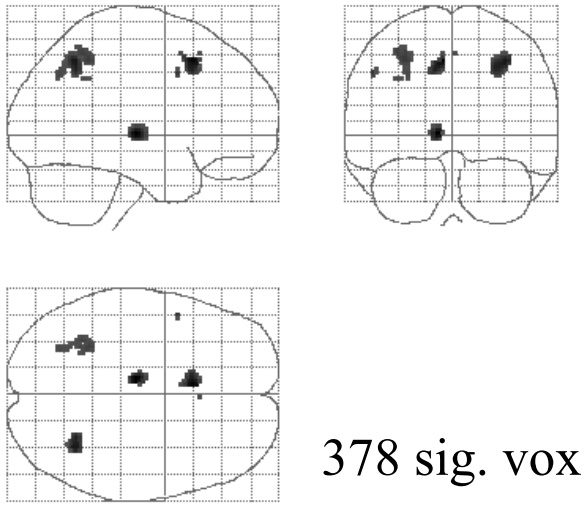
$u^{\text{Bonf}} = 9.80$

5 sig. vox.

t_{11} Statistic, RF & Bonf. Threshold



Test Level vs. t_{11} Threshold



378 sig. vox.

Smoothed Variance *t* Statistic, Nonparametric Threshold

Does this Generalize? RFT vs Bonf. vs Perm.

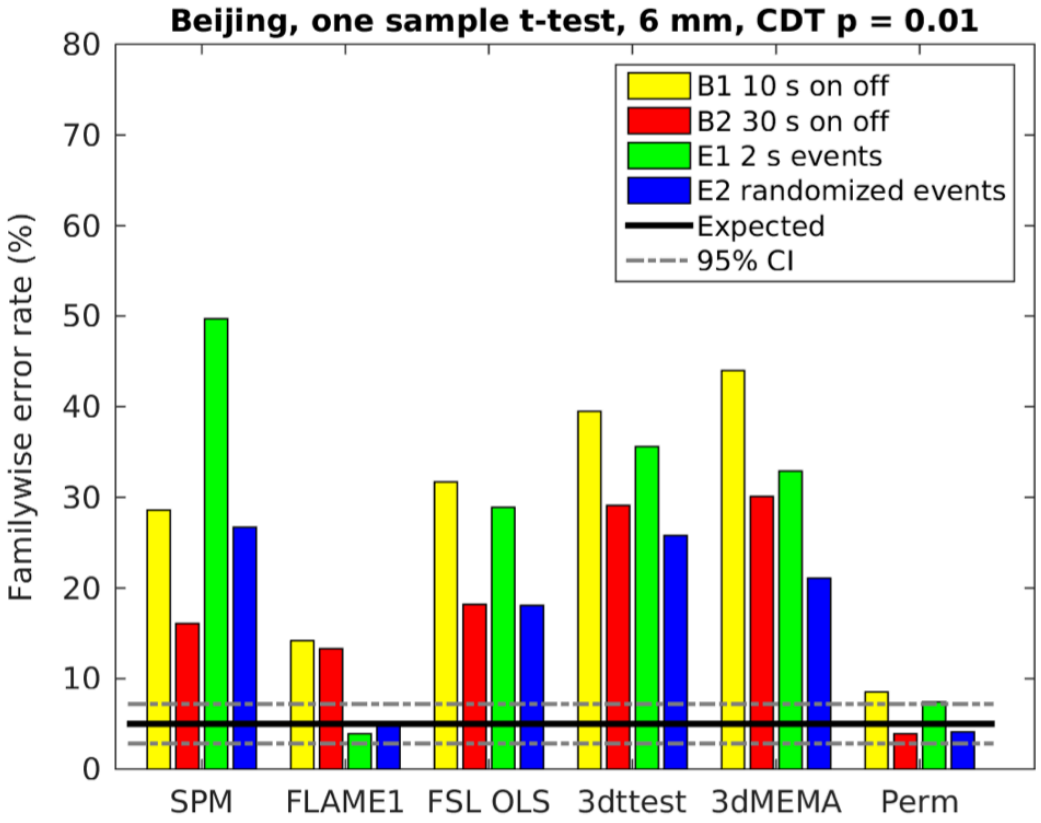
	df	*t* Threshold (0.05 Corrected) RF	Bonf	Perm
Verbal Fluency	4	4701.32	42.59	10.14
Location Switching	9	11.17	9.07	5.83
Task Switching	9	10.79	10.35	5.10
Faces: Main Effect	11	10.43	9.07	7.92
Faces: Interaction	11	10.70	9.07	8.26
Item Recognition	11	9.87	9.80	7.67
Visual Motion	11	11.07	8.92	8.40
Emotional Pictures	12	8.48	8.41	7.15
Pain: Warning	22	5.93	6.05	4.99
Pain: Anticipation	22	5.87	6.05	5.05

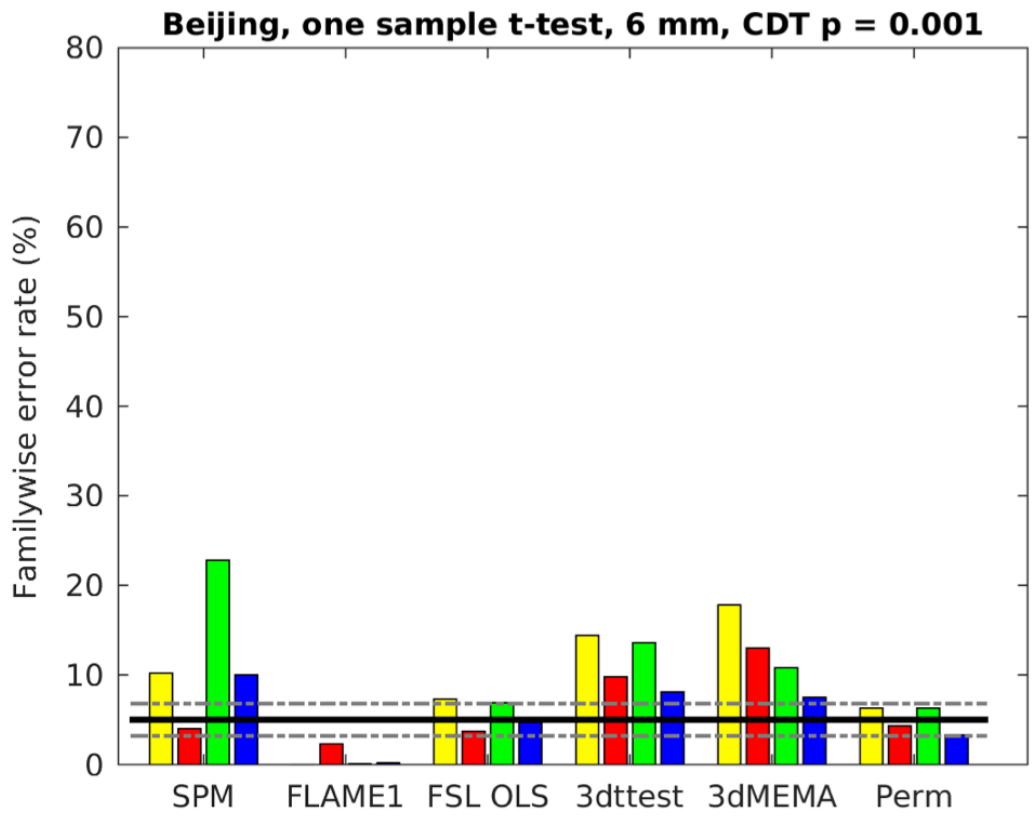
RFT vs Bonf. vs Perm.

		No. Significant Voxels (0.05 Corrected)			
		t			SmVar *t*
	df	RF	Bonf	Perm	Perm
Verbal Fluency	4	0	0	0	0
Location Switching	9	0	0	158	354
Task Switching	9	4	6	2241	3447
Faces: Main Effect	11	127	371	917	4088
Faces: Interaction	11	0	0	0	0
Item Recognition	11	5	5	58	378
Visual Motion	11	626	1260	1480	4064
Emotional Pictures	12	0	0	0	7
Pain: Warning	22	127	116	221	347
Pain: Anticipation	22	74	55	182	402

Null Data Evaluation of RFT

- Evaluation of RFT on 100's resting state datasets, w/ various fake designs
  - FWE not controlled for CFT P=0.01! (P=0.001 not so bad)
  - FWE OK for voxelwise RFT (not shown)





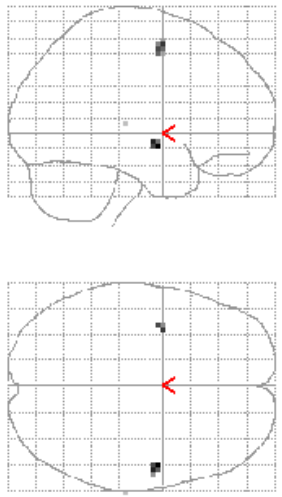
Eklund, Nichols, Knutsson (2015). Can parametric statistical methods be trusted for fMRI based group studies? **PDF**

Using SnPM to Assess Reliability with Small Groups

- Consider n=50 group study
  - Event-related Odd-Ball paradigm, Kiehl, et al.
- Analyze all 50
  - Analyze with SPM and SnPM, find FWE thresh.
- Randomly partition into 5 groups 10
  - Analyze each with SPM & SnPM, find FWE thresh
- Compare reliability of small groups with full
  - With and without variance smoothing

Skip

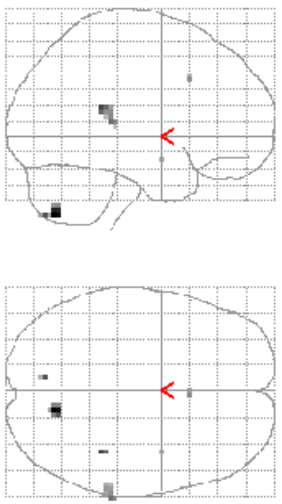
SPM t_{11}: 5 groups of 10 vs all 50
5% FWE Threshold



T>10.93

SPM{T_9}

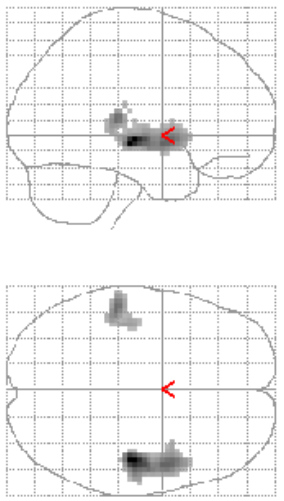
2 8 11 15 18 35 41 43 44 50



T>11.04

SPM{T_9}

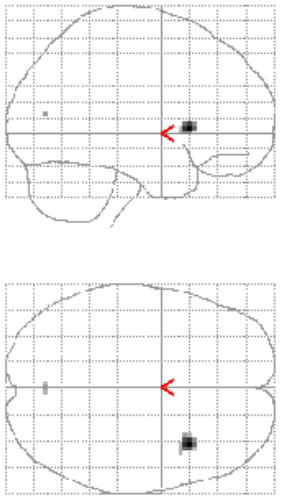
1 3 20 23 24 27 28 32 34 40



T>11.01

SPM{T_9}

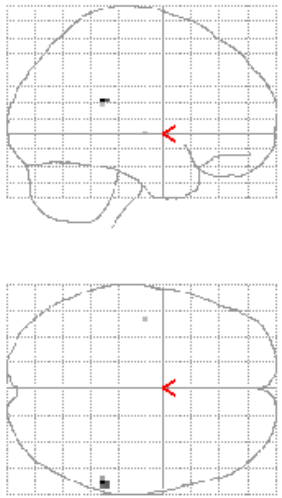
9 13 14 16 19 21 25 29 30 45



T>10.69

SPM{T_9}

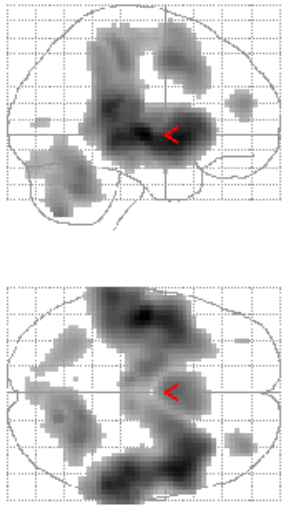
4 5 10 22 31 33 36 39 42 47



T>10.10

SPM{T_9}

6 7 12 17 26 37 38 46 48 49

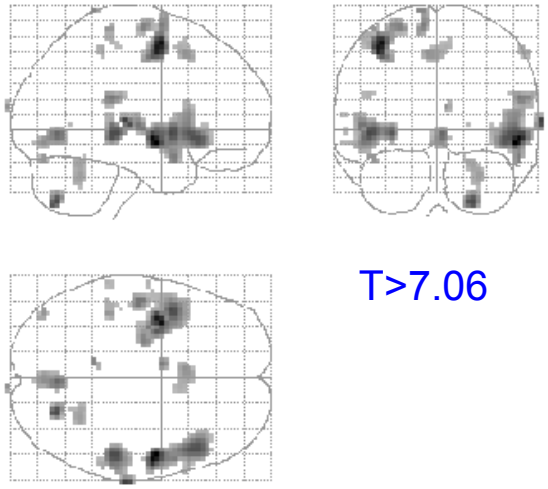


T>9.00

SPM{T_{49}}

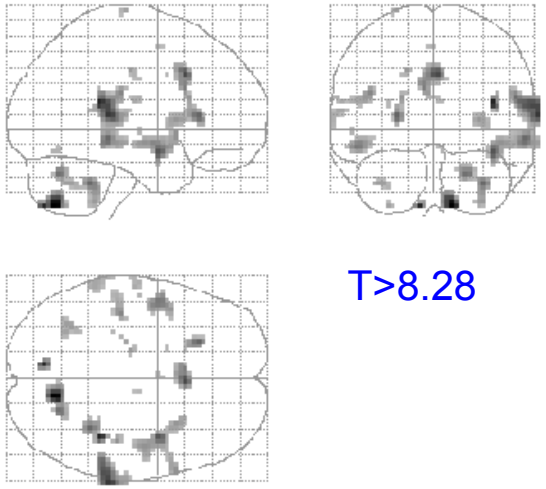
Arbitrary thresh of 9.0

SnPM *t*: 5 groups of 10 vs. all 50
5% FWE Threshold



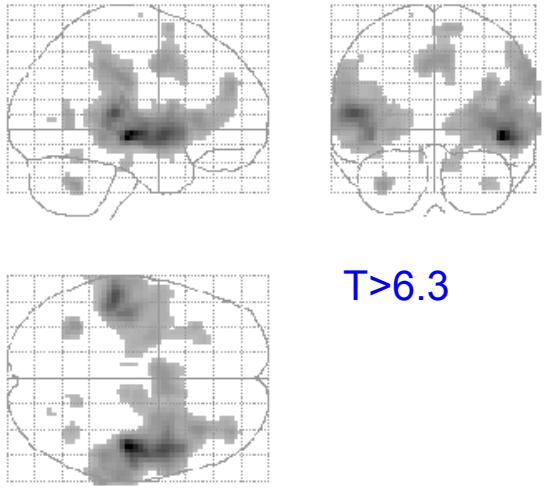
T>7.06

2 8 11 15 18 35 41 43 44 50



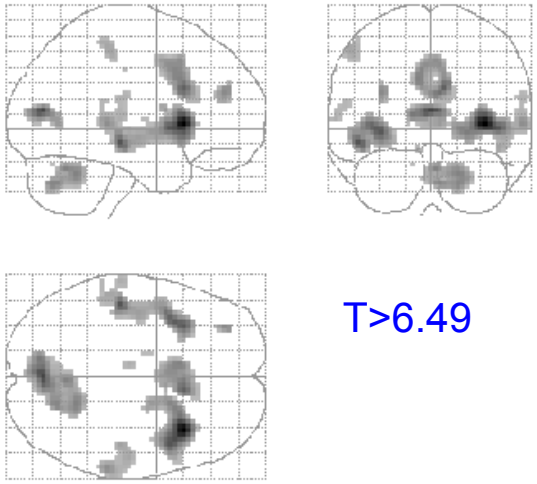
T>8.28

1 3 20 23 24 27 28 32 34 40



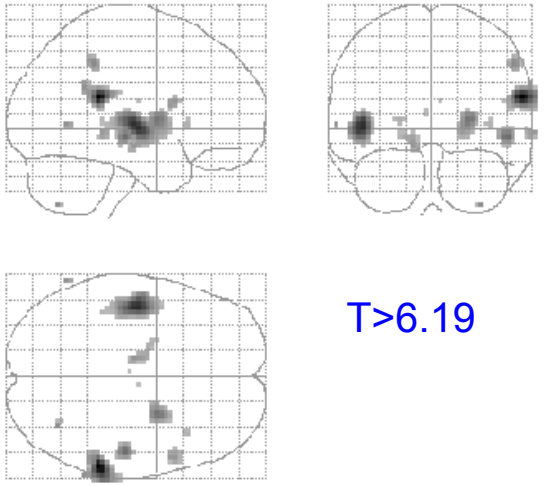
T>6.3

9 13 14 16 19 21 25 29 30 45



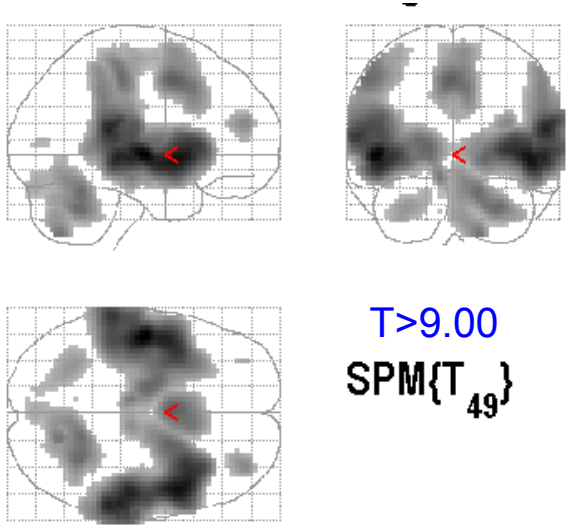
T>6.49

4 5 10 22 31 33 36 39 42 47



T>6.19

6 7 12 17 26 37 38 46 48 49

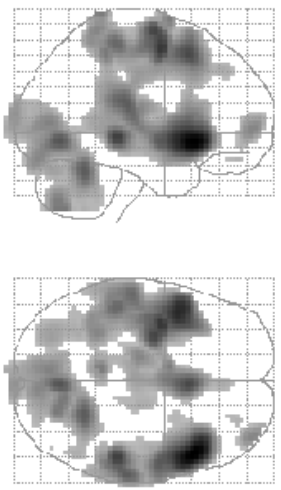


T>9.00

SPM{T_{49}}

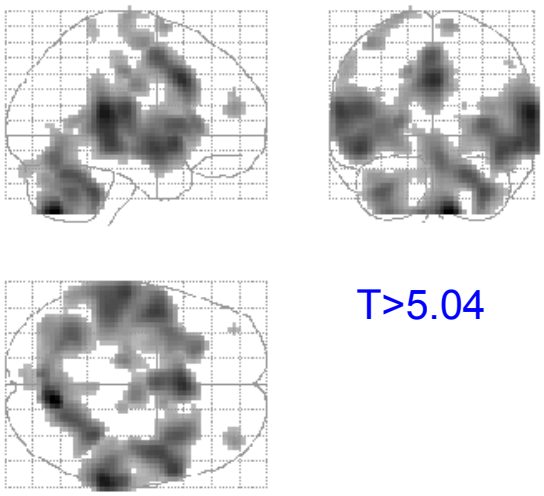
Arbitrary thresh of 9.0

SnPM SmVar *t*: 5 groups of 10 vs. all 50
5% FWE Threshold



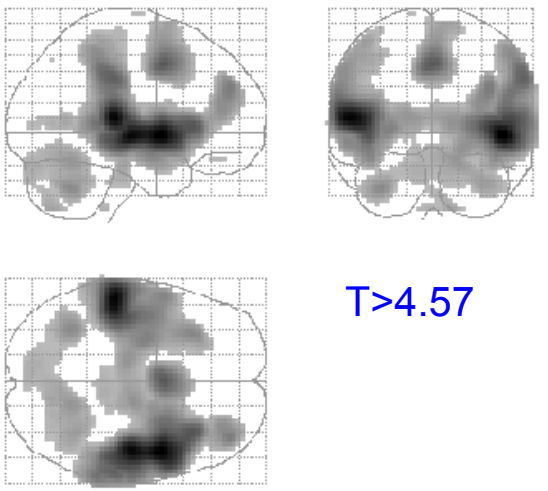
T>4.69

2 8 11 15 18 35 41 43 44 50



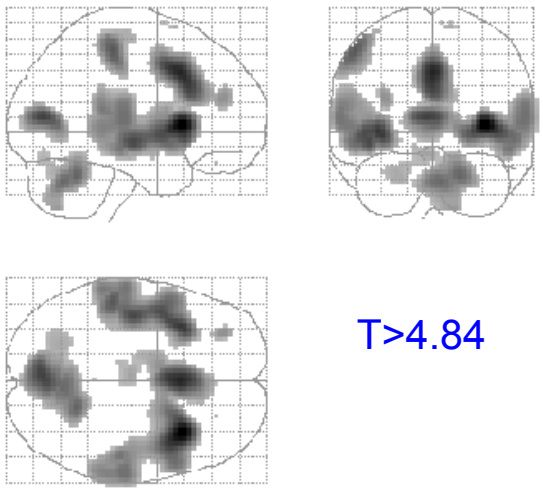
T>5.04

1 3 20 23 24 27 28 32 34 40



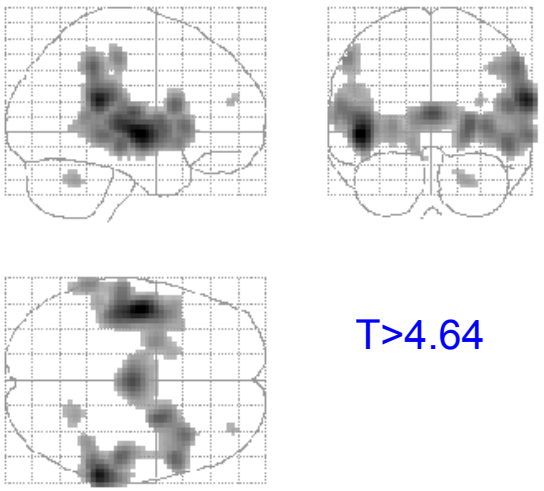
T>4.57

9 13 14 16 19 21 25 29 30 45



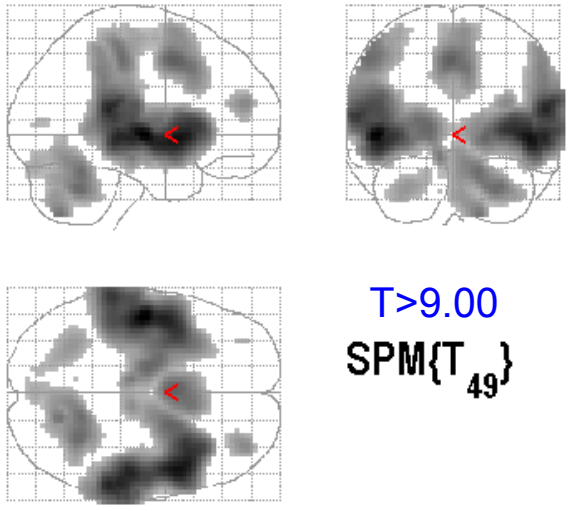
T>4.84

4 5 10 22 31 33 36 39 42 47



T>4.64

6 7 12 17 26 37 38 46 48 49



T>9.00

SPM{T_{49}}

Arbitrary thresh of 9.0

Nonparametric Conclusions

- Nonparametric Permutation
  - Good when Normality is question
  - Good with tiny group inference & variance smoothing

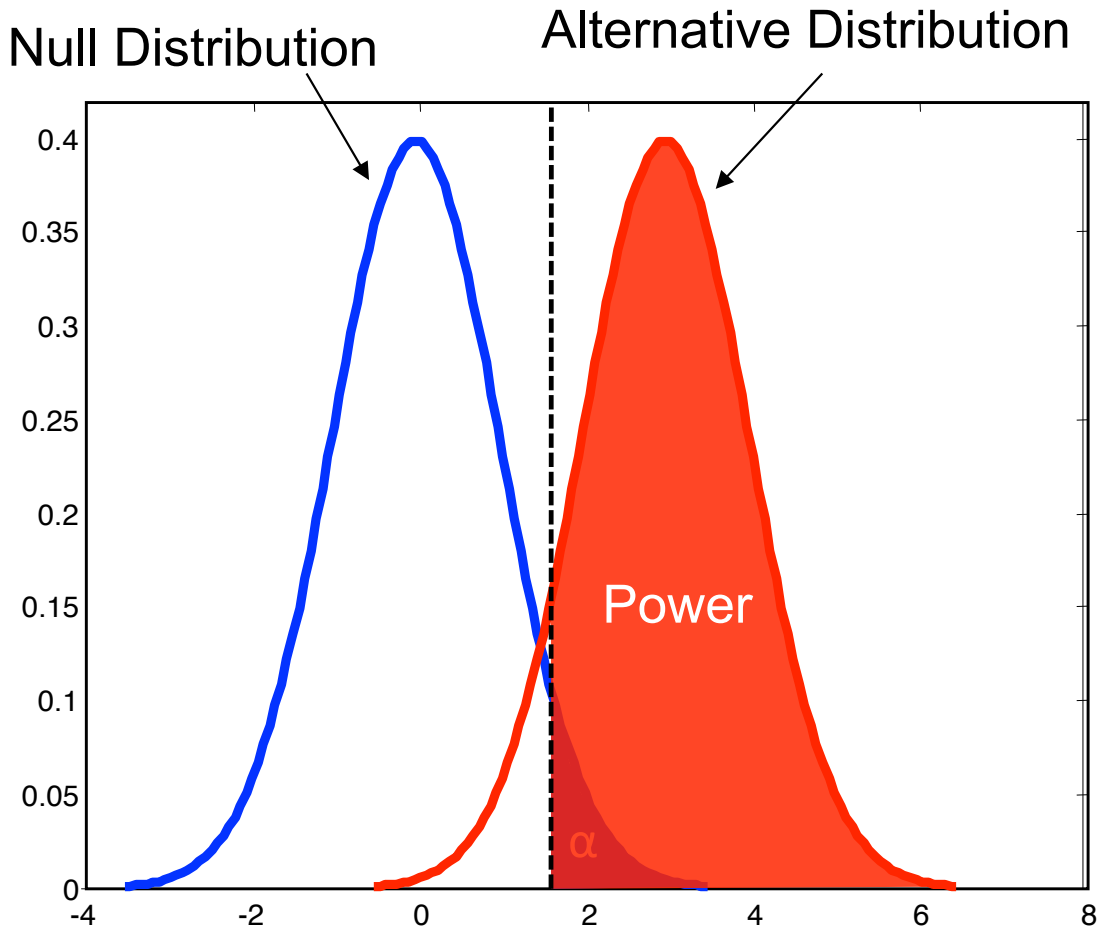
- Come to practical for more!

Advanced issues in fMRI statistics

- Nonparametric Inference
- Power
- Meta-Analysis

Power: 1 Test

- Power:
  Probability of rejecting H_0 when H_A is true
- Must specify:
  - Sample size *n*
  - Level α
    (allowed false positive rate)
  - Standard deviation σ
    (population variability; not StdErr)
  - Effect magnitude Δ
- Last two can be replaced with
  - Effect size δ = Δ/σ



Power: Statistic vs. Data Units

- 10 subjects
- % BOLD stdev σ = 0.5

One-Sample T-test…

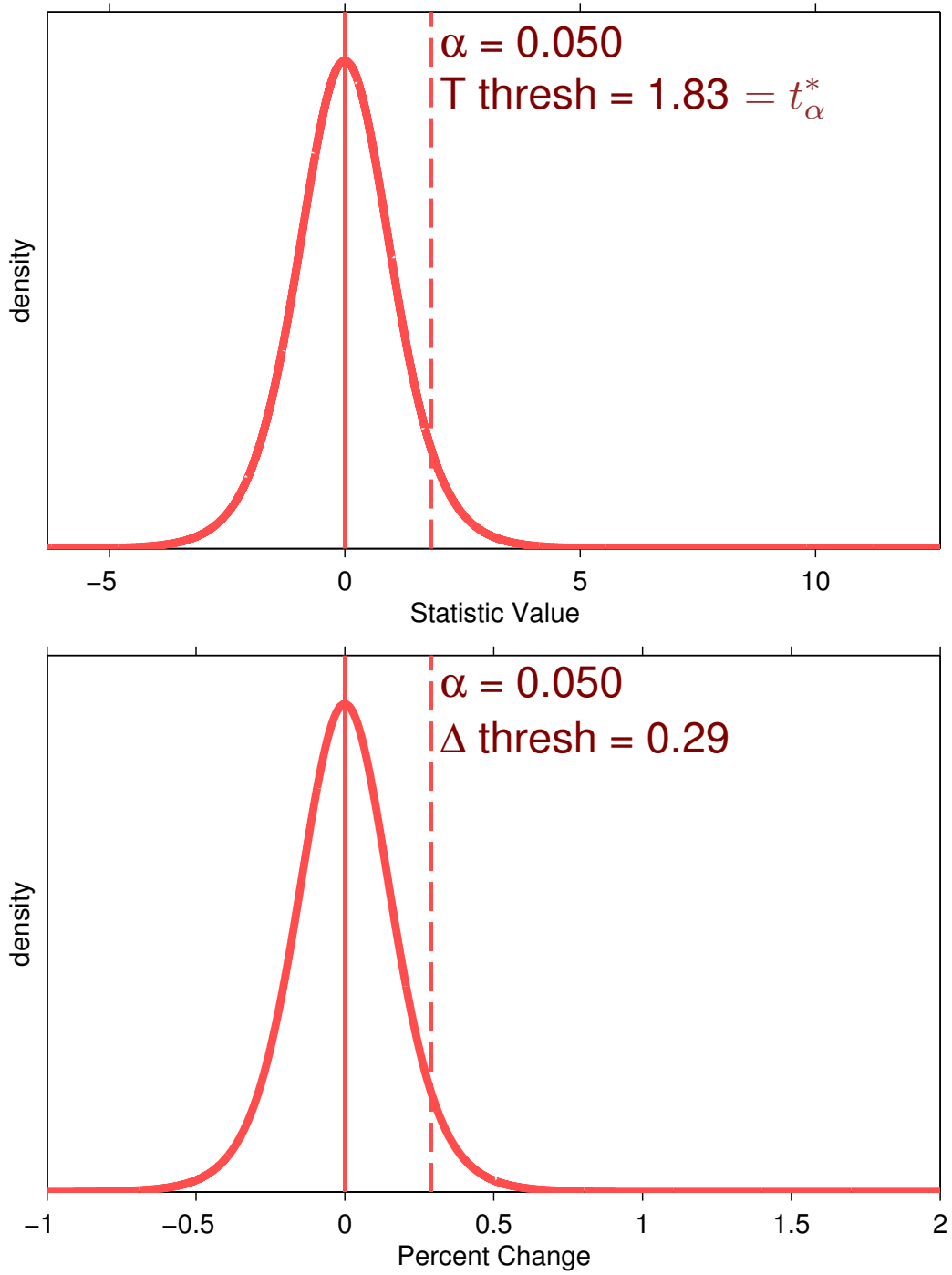
$$t = \frac{\bar{x}}{s/\sqrt{n}}$$

Reject H_0 if…

$$\frac{\bar{x}}{s/\sqrt{n}} \geq t^*_\alpha$$

Equivalently, reject H_0 if…

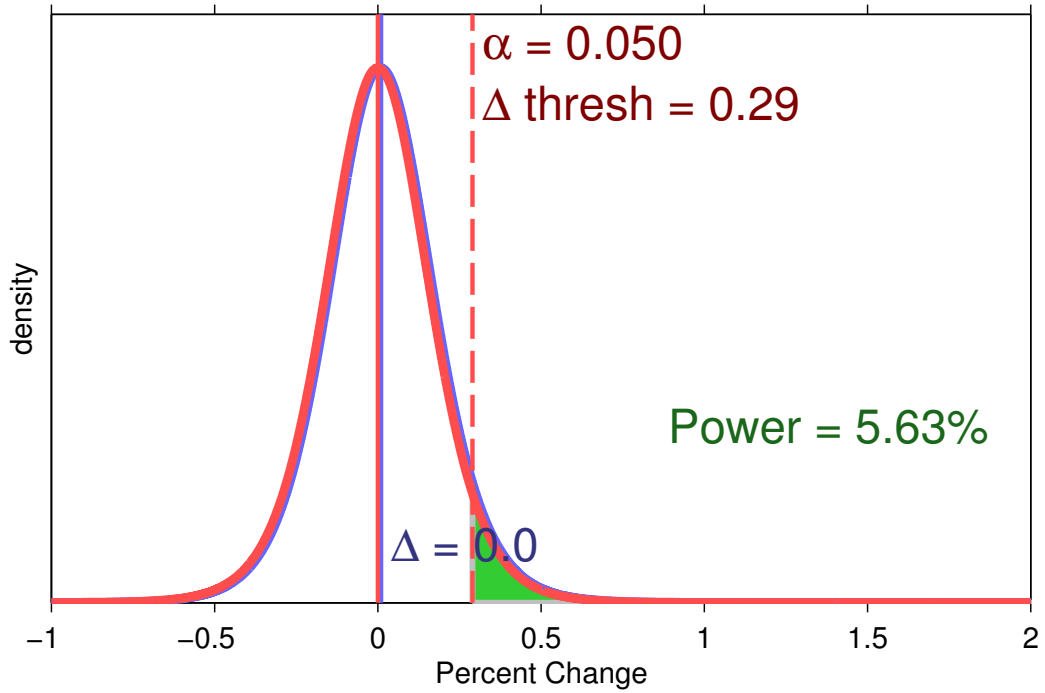
$$\bar{x} \geq t^*_\alpha \times s/\sqrt{n}$$

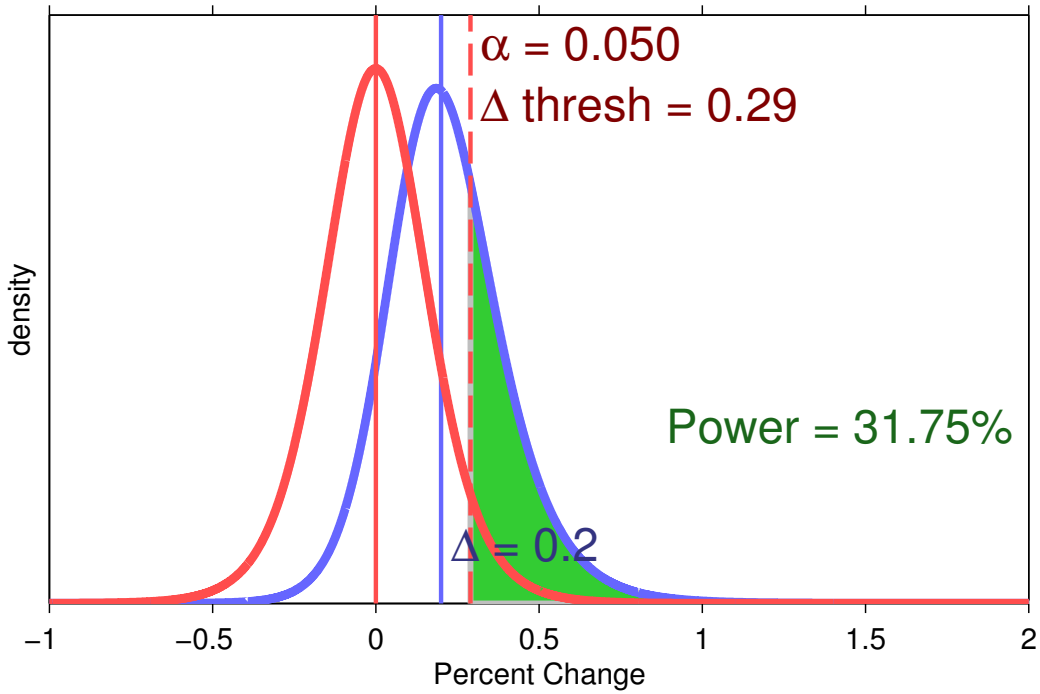


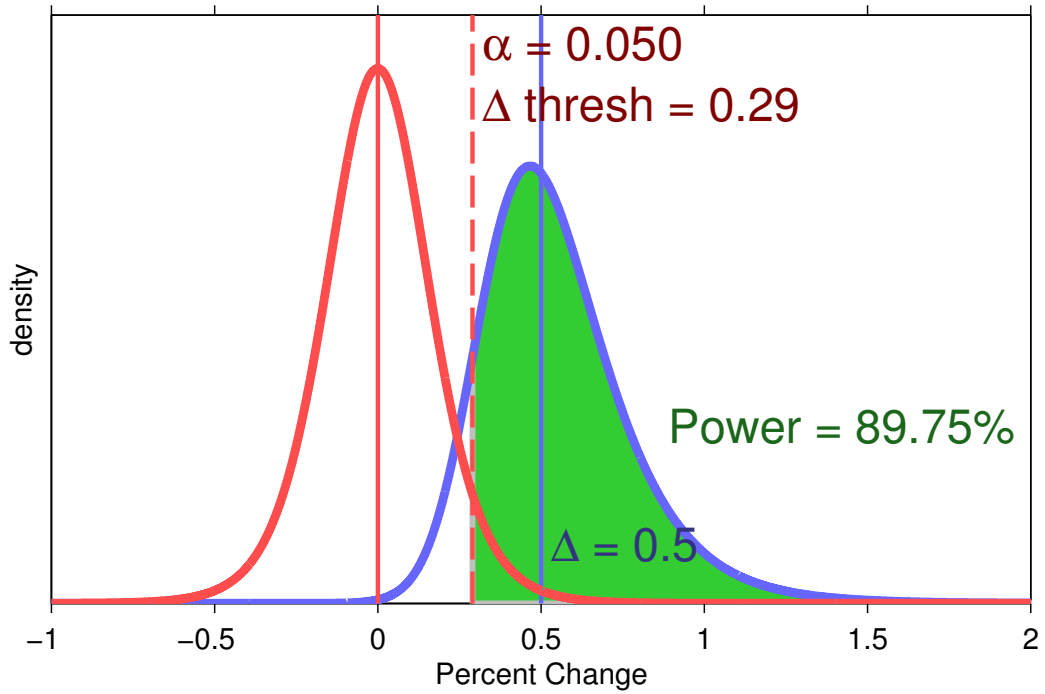
Power & Effect Magnitude

- 10 subjects
- % BOLD stdev σ = 0.5
- True %BOLD
  Δ = 0.01, 0.2, 0.5
- Effect Size δ = Δ/σ
  δ = 0.02, 0.4, 1.0

... assuming these are the right numbers!







Power: 100,000 Tests?

- Multiple testing (easy part)
  - Set α to reflect multiplicity
  - If FWE corrected is typically $t^*=5$, then $\alpha = 0.00036$
- Alternative: $\delta_1, \delta_2, \delta_3, \ldots, \delta_{99,999}, \delta_{100,000}$ (hard part)
  - Must consider all anticipated alternatives
  - These 10 voxels active, and those other 20, and…
  - Oh, and don't forget to specify $\sigma_1, \sigma_2, \sigma_3$ … too!

But see…

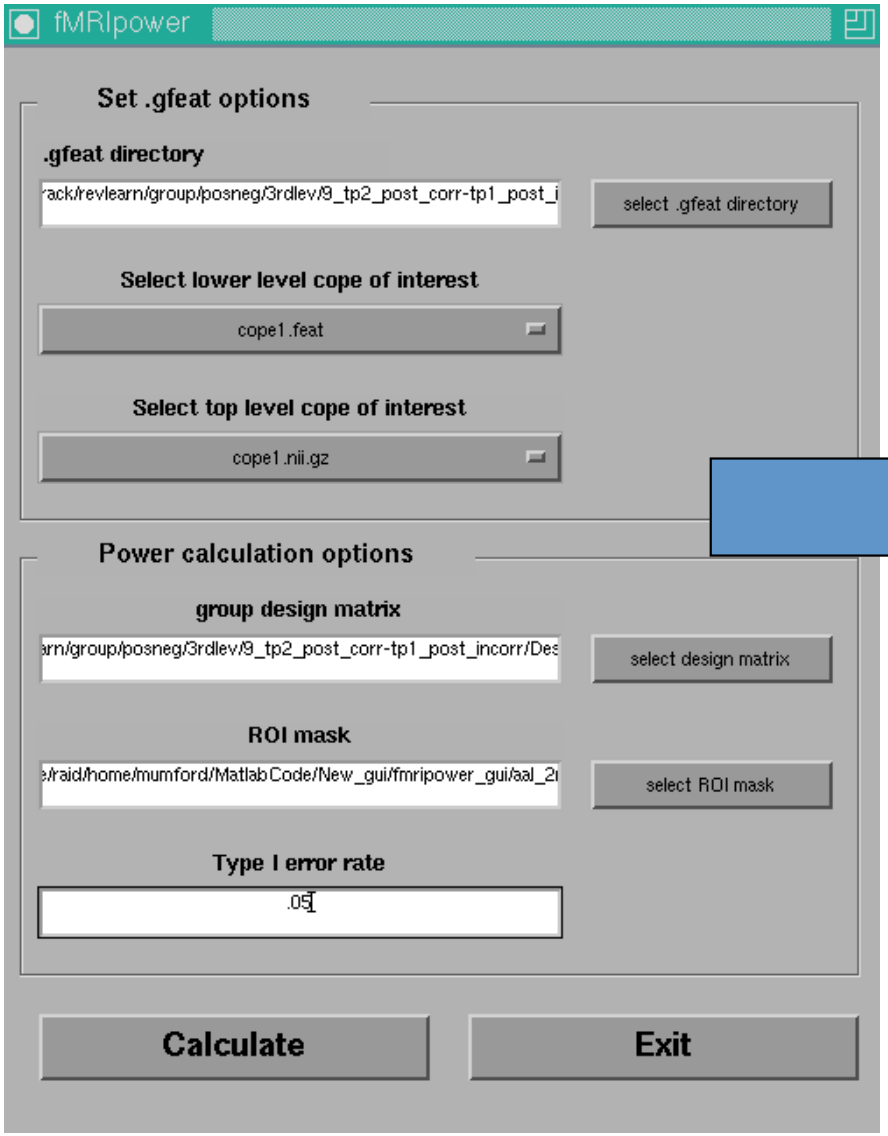
fMRIpower: http://fmripower.org
PowerMap: http://sourceforge.net/projects/powermap
NeuroPower: http://neuropower.shinyapps.io/neuropower

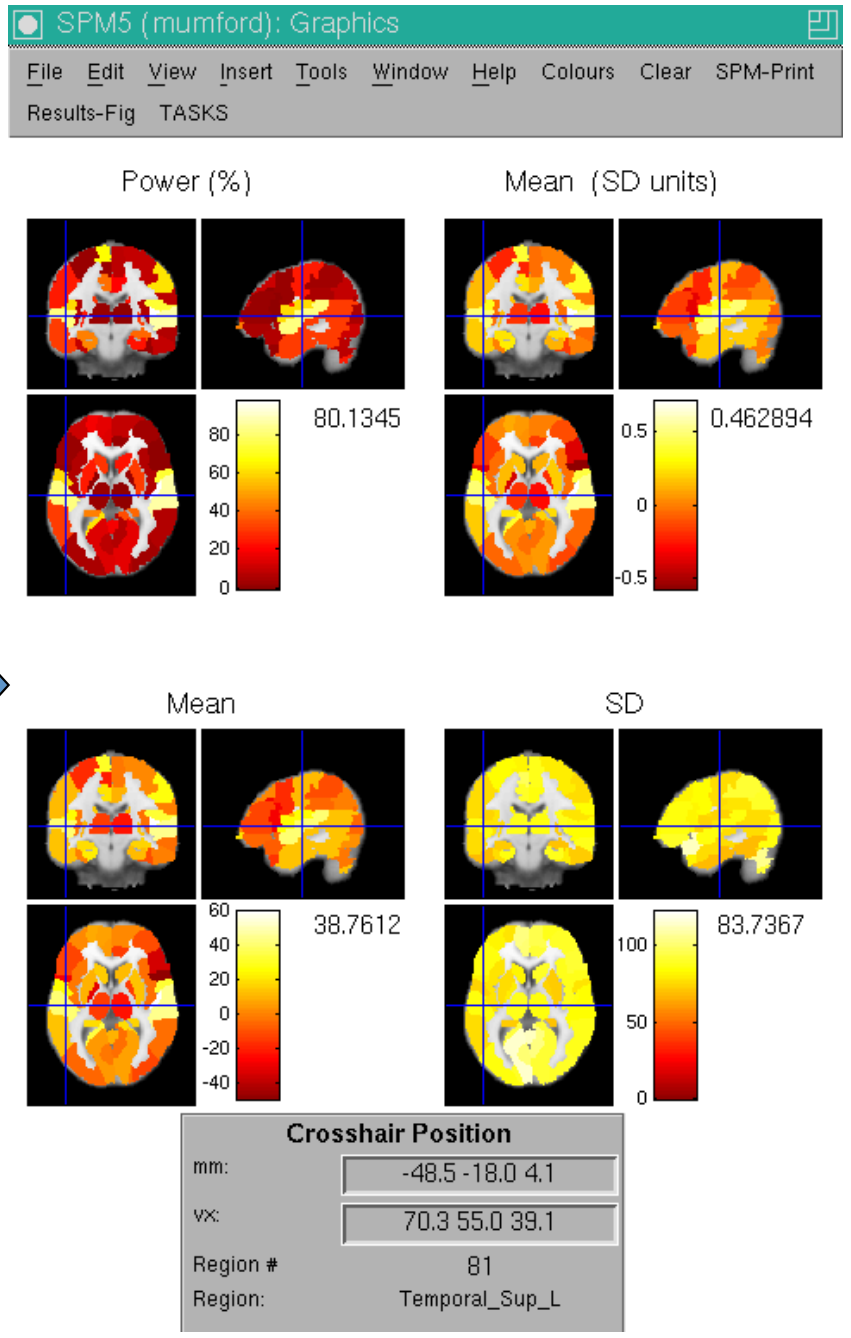
- In practice…
  - Base power on extracted summary values
  - Corresponds to a clinical trial's "primary outcome"
  - Come to practical to see the mechanics

fMRIpower tool

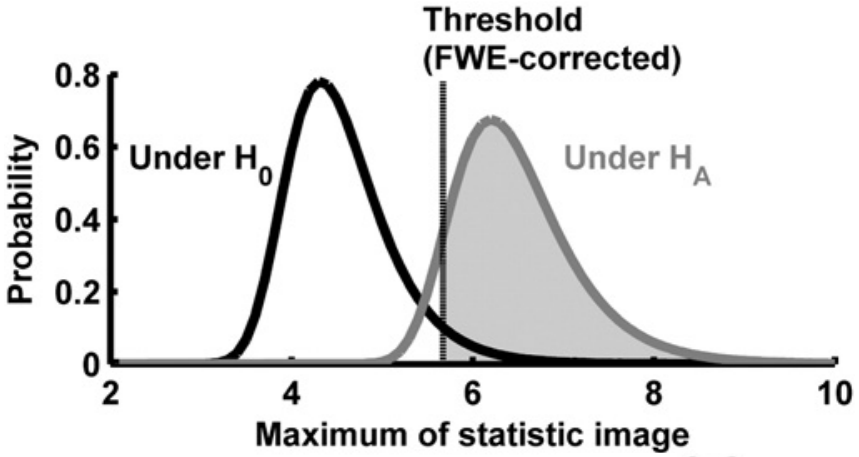
http://fmripower.org

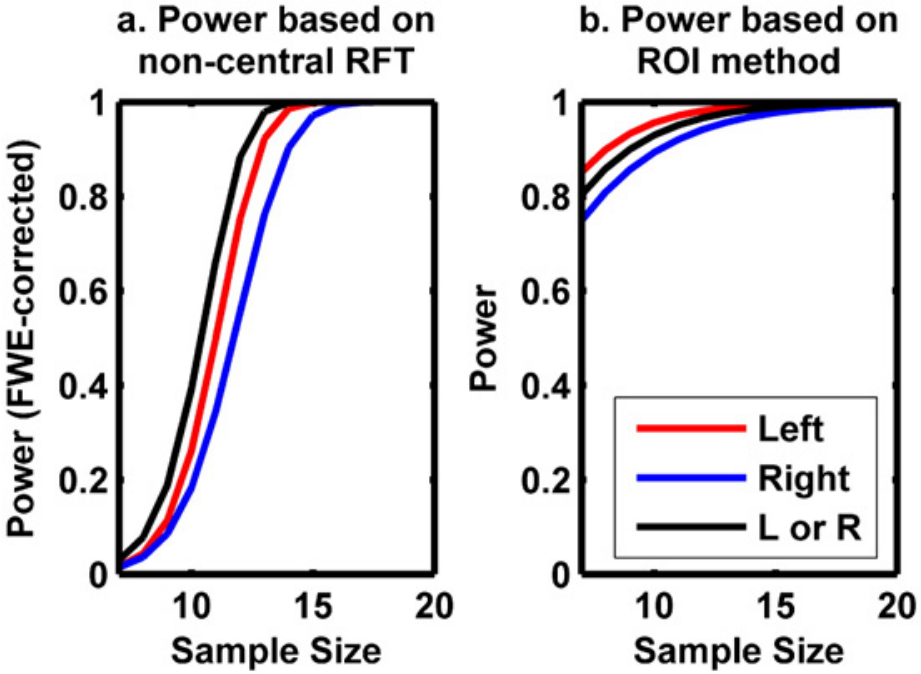
for both SPM & FSL





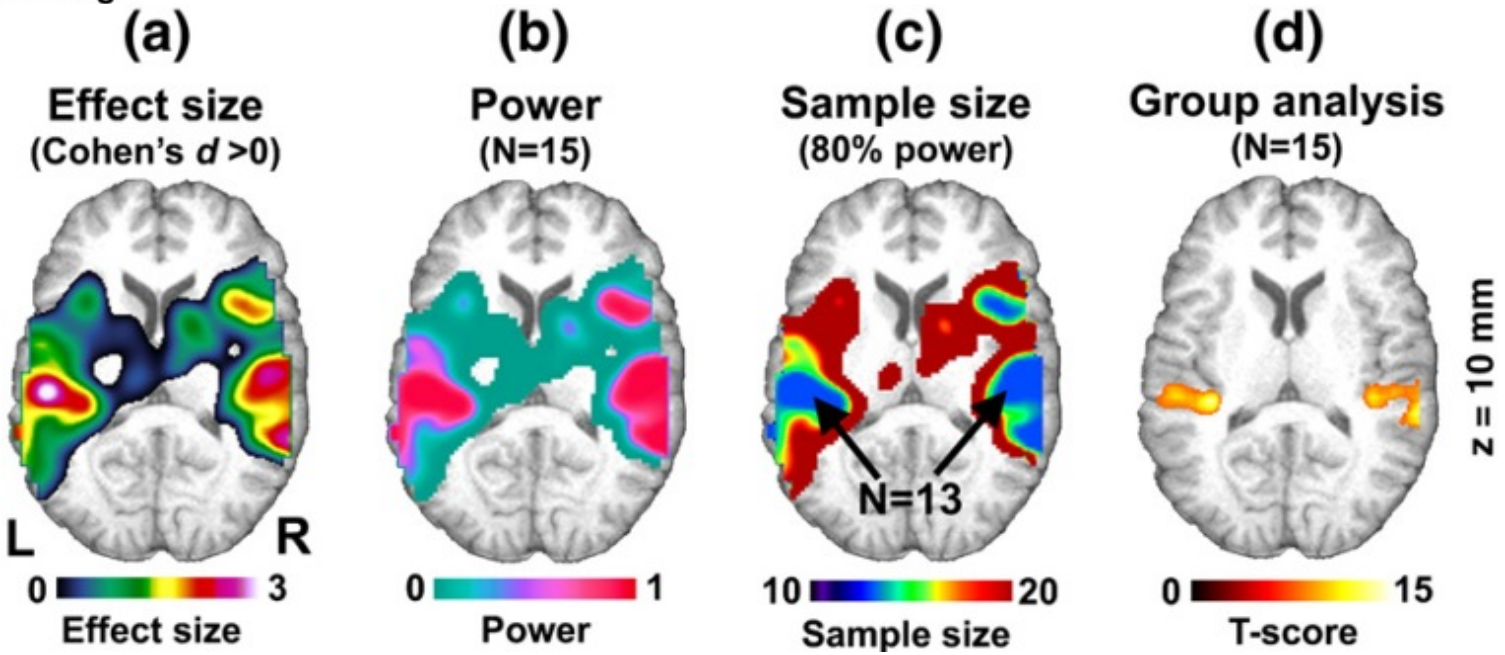
Voxel-wise Power Analyses (with RFT)





PowerMap tool

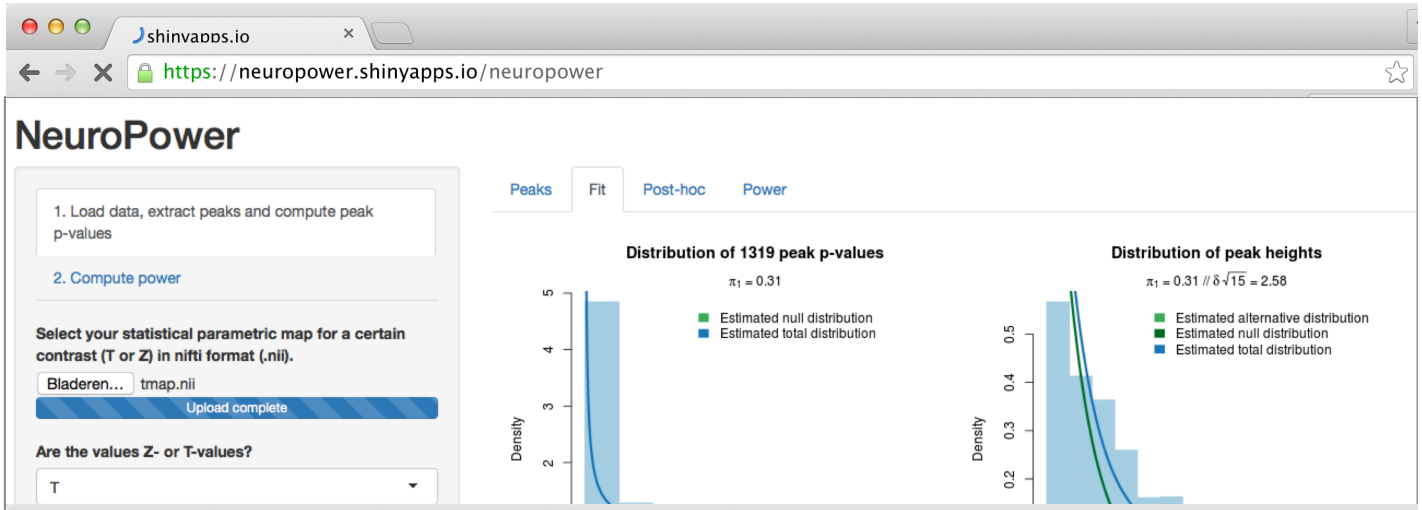
`http://sourceforge.net/projects/powermap`

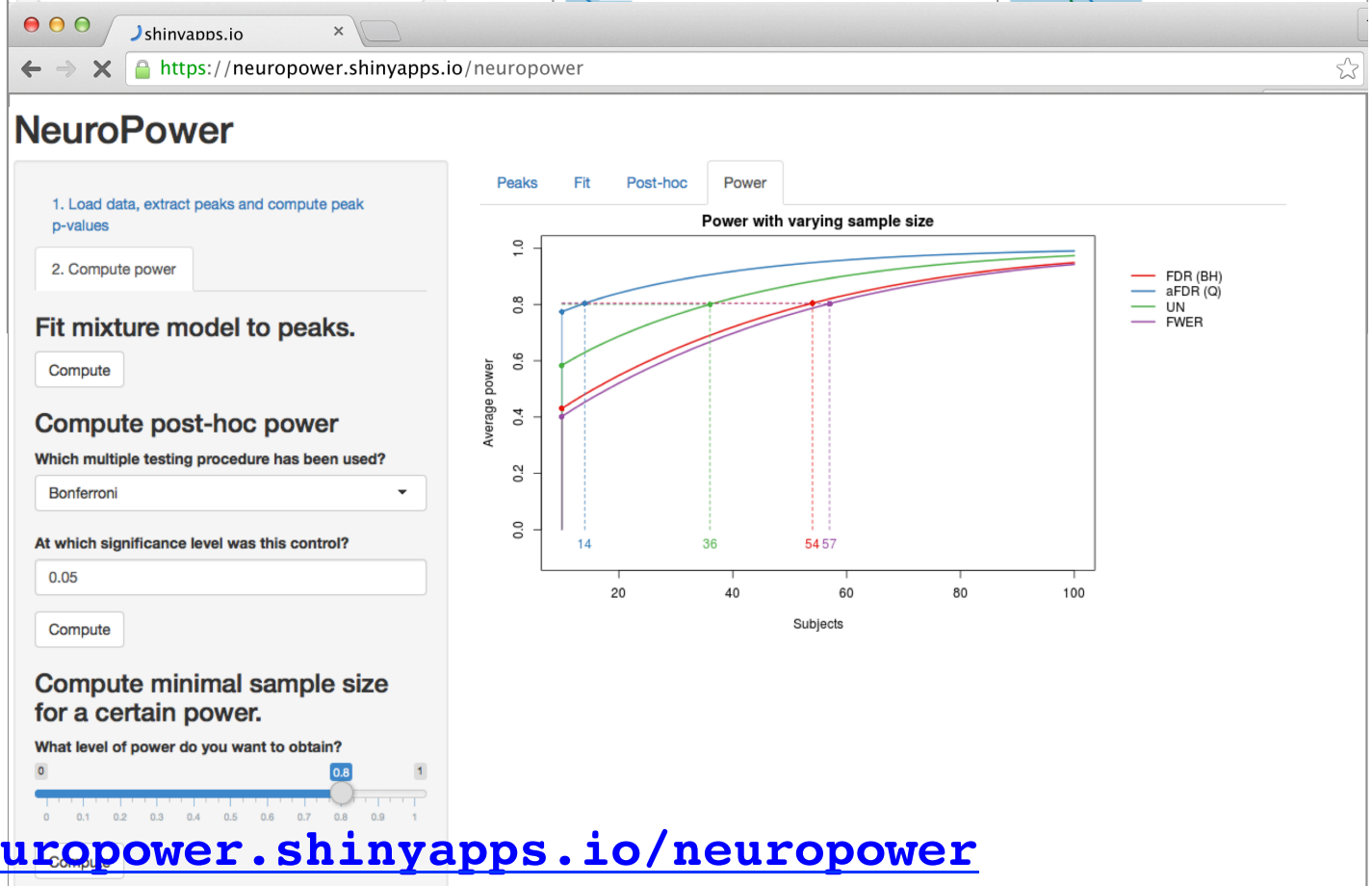


S Hayasaka, AM Peiffer, CE Hugenschmidt, PJ Laurienti. Power and sample size calculation for neuroimaging studies by non-central random field theory. NeuroImage 37 (2007) 721–730

NeuroPower

- Effect prevalance and effect size estimated from peaks *only*
- Then computes power for given number of subjects, peak threshold

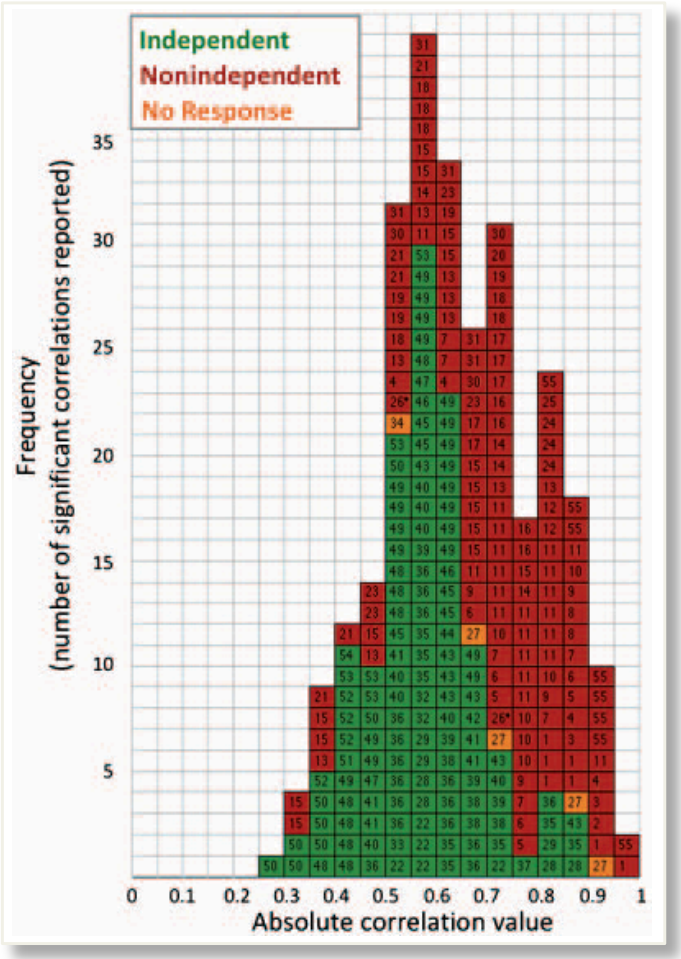




http://neuropower.shinyapps.io/neuropower

Power Dangers

- Retrospective Power
  - Power is a probability of a *future* true positive
  - Can't take current data (e.g. t=1.3) and say "What was my power for this result?"
- Estimating Effect Sizes
  - Voodoo correlations!
    - Effect size at peak is biased
      - Circularly defined as *the best* effect
  - Must use independent ROIs
    - Independent data, contrasts
    - Anatomical ROI



Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Persp. on Psych. Science*, 4, 274-290

Power & Replicability

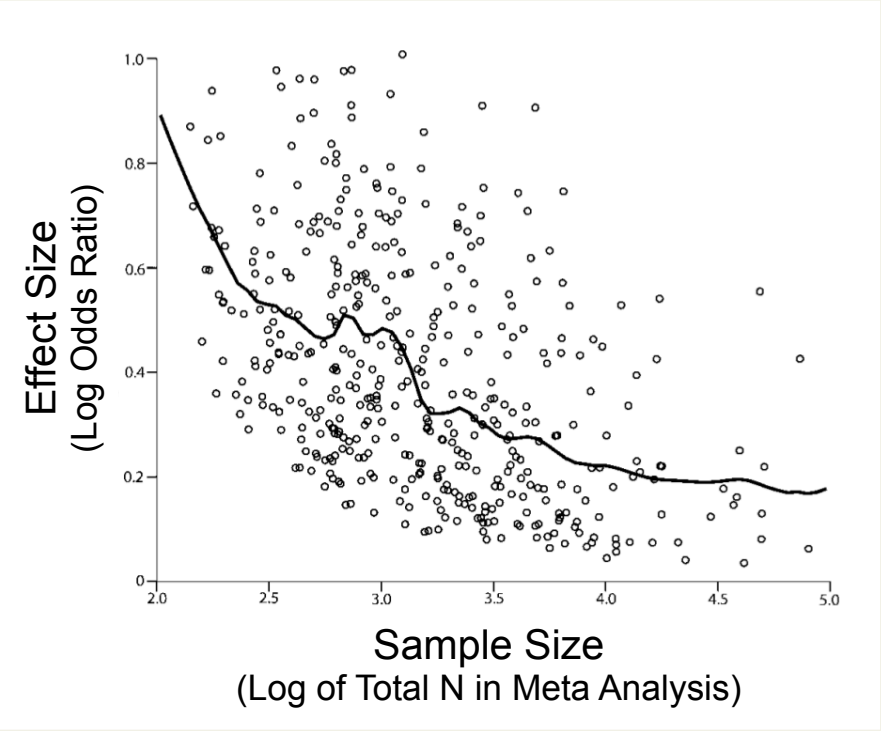
- I got a significant result, who cares about power!?
- Law of Small Numbers aka "Winner's Curse"
  - Small studies over-estimate effect size

☐ 256 meta analyses for a dichotomous effect (odds ratio) from Cochrane database

☐ Low N studies: At the best

- ☑ Have low power
  i.e. less likely to be positive
- ☑ But if <u>are positive</u>, likely due to
  - ☑ Randomly high effect or
  - ☑ Randomly small variance

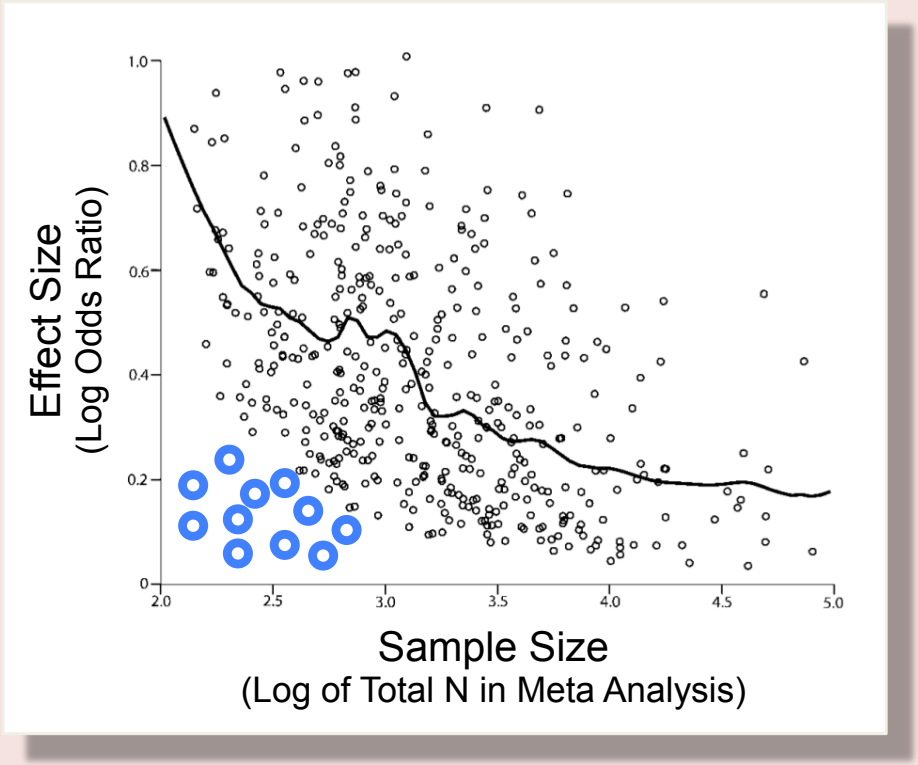
☐ Low power = hard to replicate!



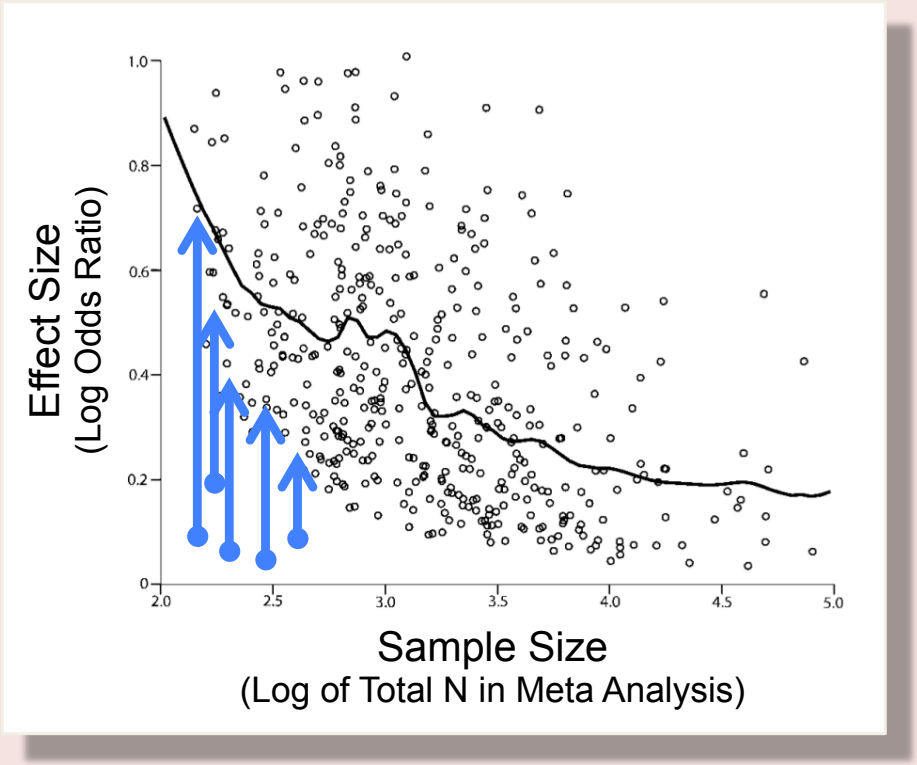
Ioannidis (2008). "Why most discovered true associations are inflated." *Epidemiology*, 19(5), 640-8.

Low N studies: At the worst

- Suppressed studies & Biased effects
  - P>0.05 not published
  - Biases that afflict small studies more than large studies



File drawer problem
(Unpublished non-significant studies)



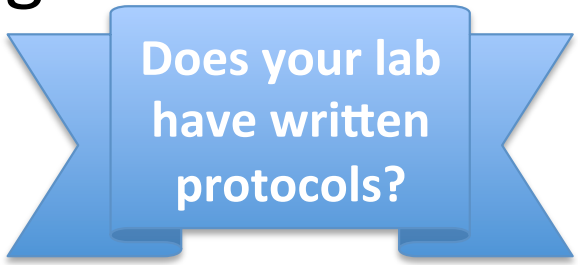
Bias
(Fishing or Vibration Effects)

Vibration Effects

- Sloppy or nonexistent analysis protocols

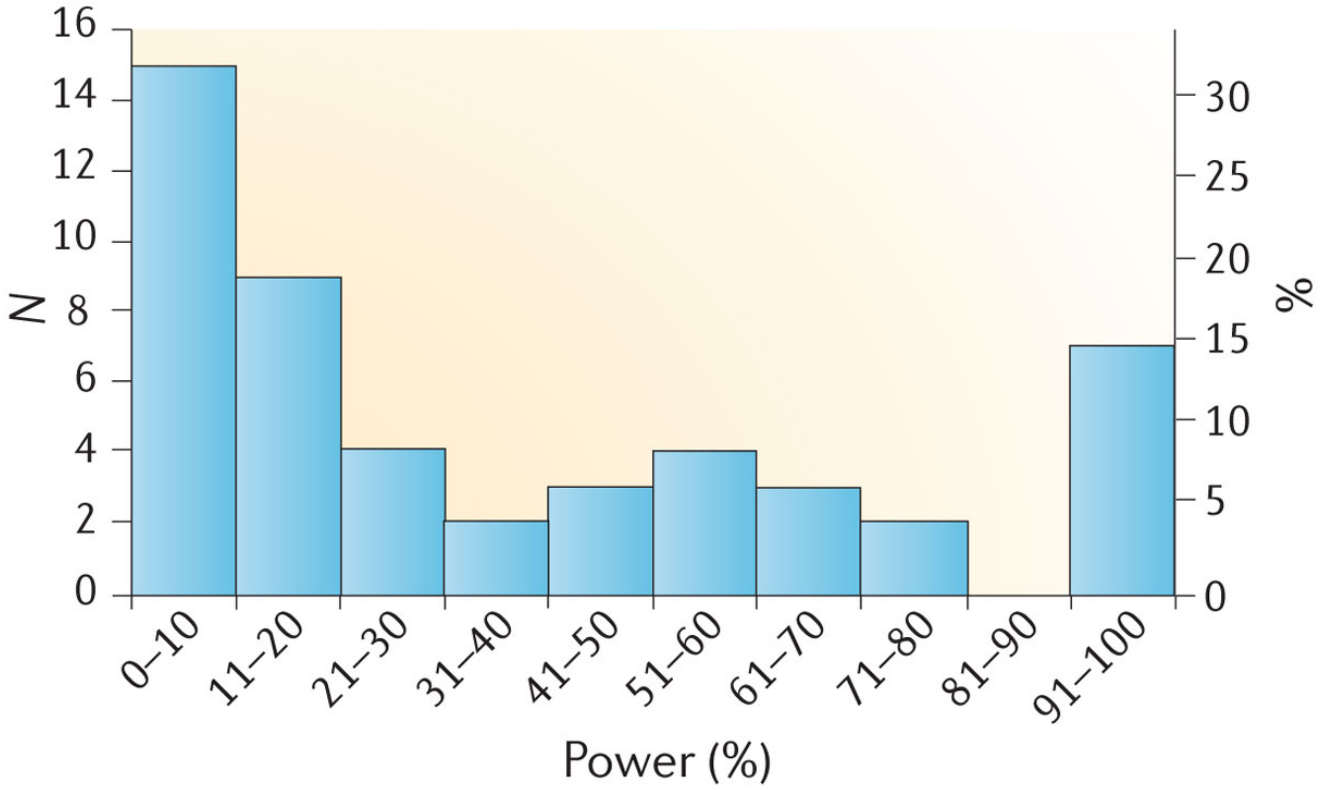
> "Try voxel-wise whole brain, then cluster-wise, then if not getting good results, look for subjects with bad movement, if still nothing, maybe try a global signal regressor; if still nothing do SVC for frontal lobe, if not, then try DLPFC (probably only right side), if still nothing, will look in literature for xyz coordinates near my activation, use spherical SVC… surely that'll work!"

  - You stop when you get the result you expect
  - These "vibrations" can only lead to inflated false positives
- Afflicts well-intended researchers
  - Multitude of preprocessing/modelling choices
    - Linear vs. non-linear alignment
    - Canonical HRF? Derivatives? FLOBS?



Power failure: Button et al.

- Meta-Analysis of (non-imaging) Neuroscience Meta-Analyses
- Recorded median power per meta-analysis
  - Median median power **21%**



50% of all neuroscience studies have **at most** a **1-in-5** chance of replicating!

Button, et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neuros*, 14(5), 365–76.

Button et al's Recommendations

- Do power calculations
- Disclose methods & findings transparently
- Pre-register your study protocol and analysis plan
- Make study materials and data available
  - Check out http://neurovault.org !
  - See also Brain Imaging Data Structure http://bids.neuroimaging.io
- Work collaboratively to increase power and replicate findings

Button, et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neuros*, 14(5), 365–76.

Power Conclusions

- Power = Replicability
  - Best gauge on whether you'll find the effect again
- Whole image-wise power possible
  - With either fMRIpower & powermap
- "Targeted outcome" power practical
  - Based on effect size at one location
  - But be aware of circularity issues

Advanced issues in fMRI statistics

- Nonparametric Inference
- Power
- Meta-Analysis

Overview

- Non-imaging meta-analysis
- Menu of meta-analysis methods
  - ROI's, IBMA, CBMA
- CBMA details
  - Kernel-based methods – What's in common
  - m/ALE, M/KDA – What's different
- Limitations & Thoughts

Stages of (non-imaging) Meta-Analysis


1. Define review's specific objectives.


2. Specify eligibility criteria.


3. Identify all eligible studies.


4. Collect and validate data rigorously.


5. Display effects for each study, with measures of precision.


6. Compute average effect, random effects std err


7. Check for publication bias, conduct sensitivity analyses.

Jones, D. R. (1995). Meta-analysis: weighing the evidence. *Statistics in Medicine*, 14(2), 137–49.

Methods for (non-imaging) Meta-Analysis (1)

- P-value (or Z-value) combining
  - Fishers (≈ average –log P)
  - Stouffers (≈ average Z)
  - Used only as method of last resort
    - Based on significance, not effects in real units
    - Differing *n* will induce heterogeneity (Cummings, 2004)
- Fixed effects model
  - Requires effect estimates and standard errors
    - E.g. Mean survival (days), and standard error of mean
  - Gives weighted average of effects
    - Weights based on per-study standard errors
  - Neglects inter-study variation

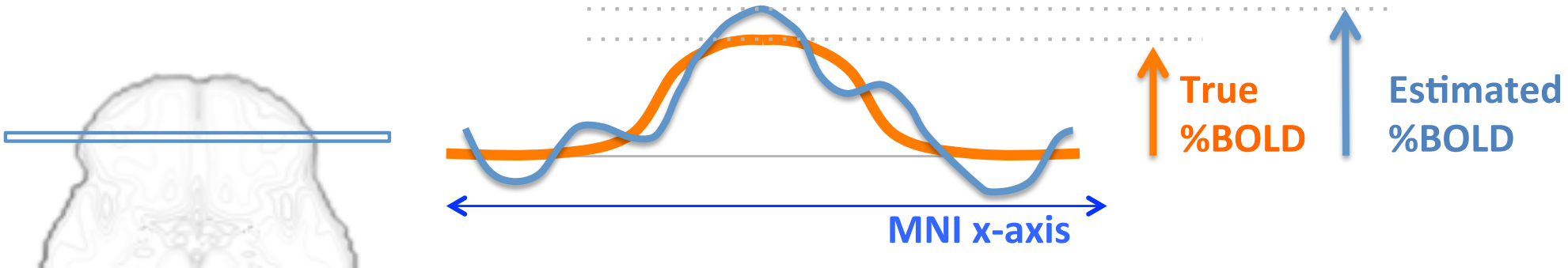
Cummings (2004). Meta-analysis based on standardized effects is unreliable. *Archives of Pediatrics & Adolescent Medicine*, 158(6), 595–7.

Methods for (non-imaging) Meta-Analysis (2)

- Random effects model
  - Requires effect estimates and standard errors
  - Gives weighted average of effect
    - Weights based on per-study standard errors *and* inter-study variation
  - Accounts for inter-study variation
- Meta regression
  - Account for study-level regressors
  - Fixed or random effects

Neuroimaging Meta-Analysis Approaches (1)

- Region of Interest
  - Traditional Meta-Analysis, on mean %BOLD & stderr
  - Almost impossible to do
    - ROI-based results rare (exception: PET)
    - Different ROIs used by different authors
    - Peak %BOLD useless, due to voodoo bias
      - Peak is overly-optimistic estimate of %BOLD in ROI



Neuroimaging Meta-Analysis Approaches (2)

- Intensity-Based Meta-Analysis (IBMA)
  - With P/T/Z Images only
    - Only allows Fishers/Stouffers *Not best practice* ☹
  - With contrast images only
    - Only allows random-effects model without weights
      - Can't weight by sample size! *Not best practice* ☹
  - With contrast and standard error images
    - SPM's spm_mfx and FSL's FEAT/FLAME:
      - 2nd-level : Combining subjects
      - 3rd-level : Combining studies
    - Allows meta-regression *Best practice* ☺
  - But image data rarely shared

Bad practice ☹

Neuroimaging Meta-Analysis Approaches (3)

- Coordinate-Based Meta-Analysis (CBMA)
  - x,y,z locations only
    - Activation Likelihood Estimation (ALE)

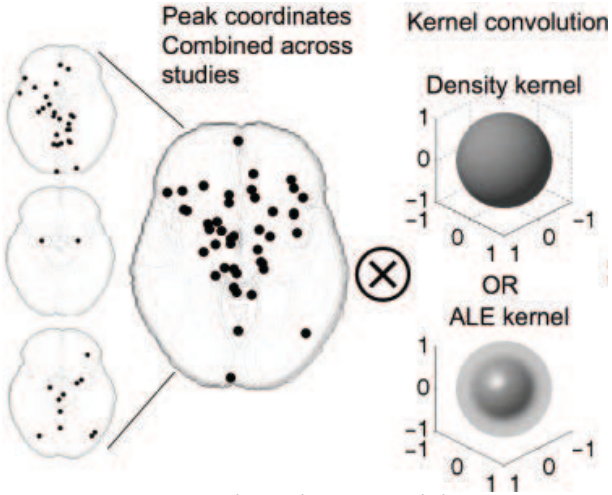
      Turkeltaub et al. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, 16(3), 765–780.
      Eickhoff et al. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, 30(9), 2907-26.
      Eickhoff et al. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, 59(3), 2349–61

    - Multilevel Kernel Density Analysis (MKDA)

      Wager et al. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage* 22 (4), 1679–1693.
      Kober et al. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *NeuroImage*, 42(2), 998–1031.
  - x,y,z and Z-value
    - Signed Difference Mapping (SDM)

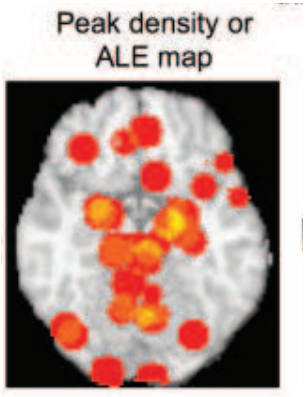
      Radua & Mataix-Cols (2009). Voxel-wise meta-analysis of grey matter changes in obsessive-compulsive disorder. *British Journal of Psychiatry*, 195:391-400.
      Costafreda et al. (2009). A parametric approach to voxel- based meta-analysis. *NeuroImage*, 46(1):115-122.

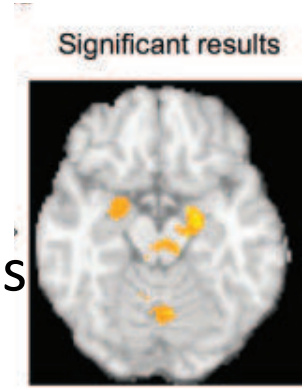
CMBA Kernel Methods



Wager et al. (2007). SCAN, 2(2), 150–8.

- Create study maps
  - Each focus is replaced with kernel
    - Important details on kernel overlap
- Create meta maps
  - Study maps combined
- Inference
  - Traditional voxel-wise or cluster-wise
    - Voxel-wise – FDR or FWE
    - Cluster-wise – FWE
  - Monte Carlo test
    - H_0: no consistency over studies
    - Randomly place each study's foci, recreate meta maps
    - Not actually a permutation test (see Besag & Diggle (1977))

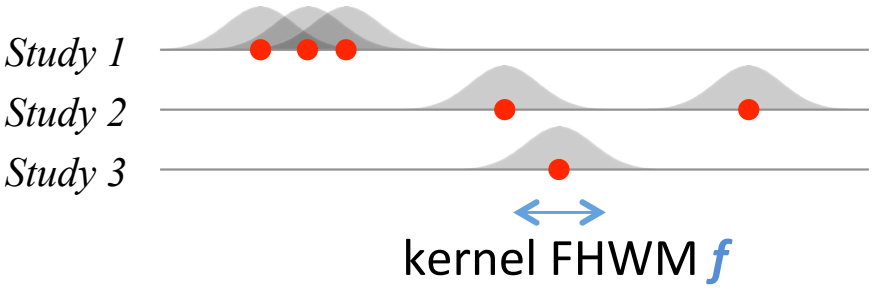




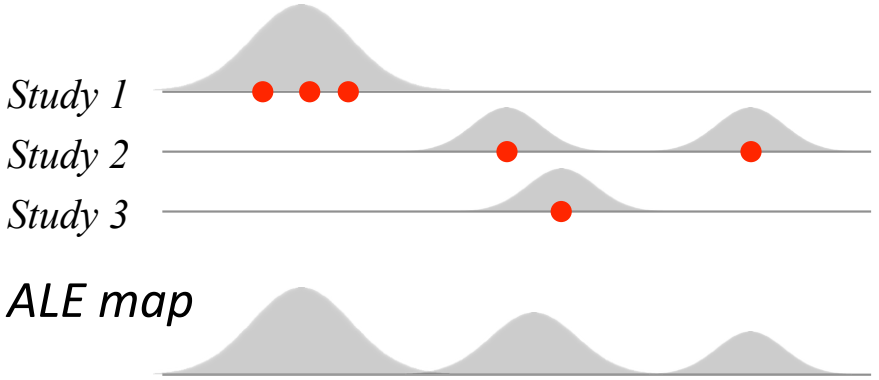
Besag & Diggle (1977). Simple Monte Carlo tests for spatial pattern. *JRSS C (Applied Statistics)*, 26(3), 327–333.

Kernel Methods History – m/ALE

ALE – Activation Likelihood Estimation
(Turkeltaub et al., 2002)

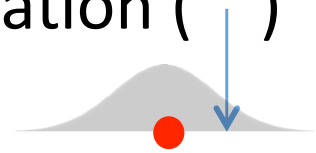


ALE per-study map



ALE interpretation for single focus (●)

Probability of observing a focus at that location ( )



ALE combining

Probability of union of independent events…

$ALE(p_1,p_2) = p_1 + p_2 - p_1 \times p_2$

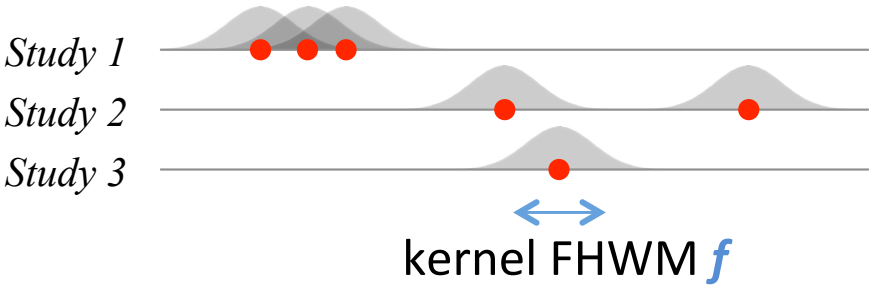
$ALE(p_1,p_2,p_3) = p_1 + p_2 + p_3 - p_1 \times p_2 - p_1 \times p_3 - p_2 \times p_3 + p_1 \times p_2 \times p_3$

ALE interpretation:

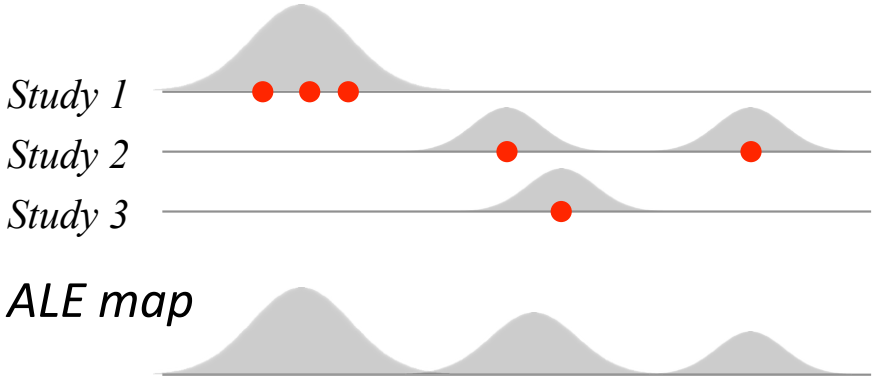
Probability of observing one or more foci at a given location *based* on a model of Gaussian spread with FWHM *f*

Kernel Methods History – m/ALE

ALE – Activation Likelihood Estimation
(Turkeltaub et al., 2002)



ALE per-study map



Problem with first ALE

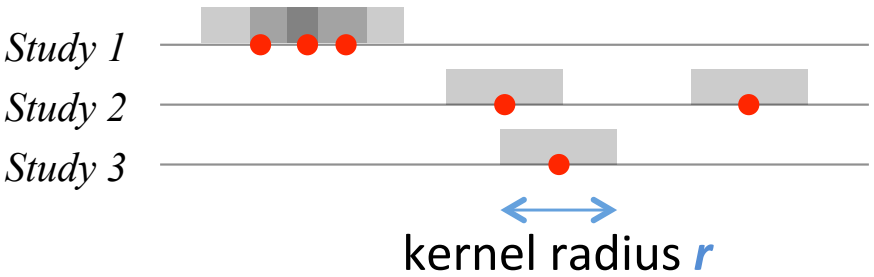
- Single study could dominate, if lots one has lots of points

Modified ALE (Eickhoff et al., 2009; Eickhoff et al., 2012)

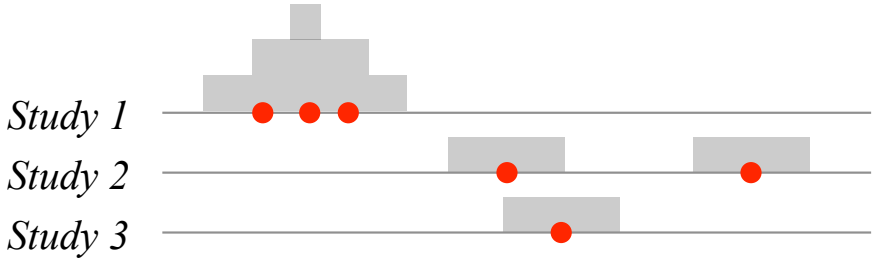
- Revised Monte Carlo test accounts for studies
  - Fix foci, randomly sample each map
- Adapt kernel size *f* to study sample size
- Voxel-wise test – no Monte Carlo!
- Cluster-wise test – still requires Monte Carlo

Kernel Methods History – M/KDA

KDA – Kernel Density Analysis
(Wager et al., 2004)



KDA per-study map



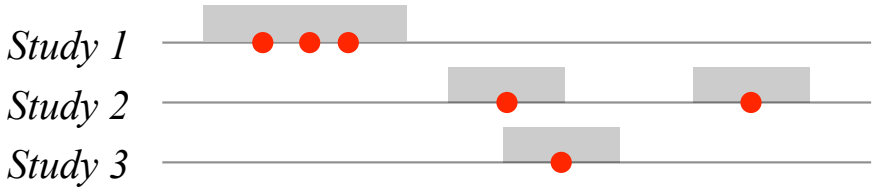
KDA map – average of study maps

Same problem with individual profligate studies

MKDA (Kober et al., 2008)

- Truncated study maps
- Monte Carlo test
  - Moves clusters, not individual foci

MKDA – Multilevel Kernel Density Analysis
per-study map



MKDA map – weighted average of study maps

MKDA (unweighted) interpretation:

Proportion of studies having one or more foci within distance *r*

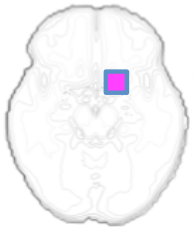
CBMA Limitations

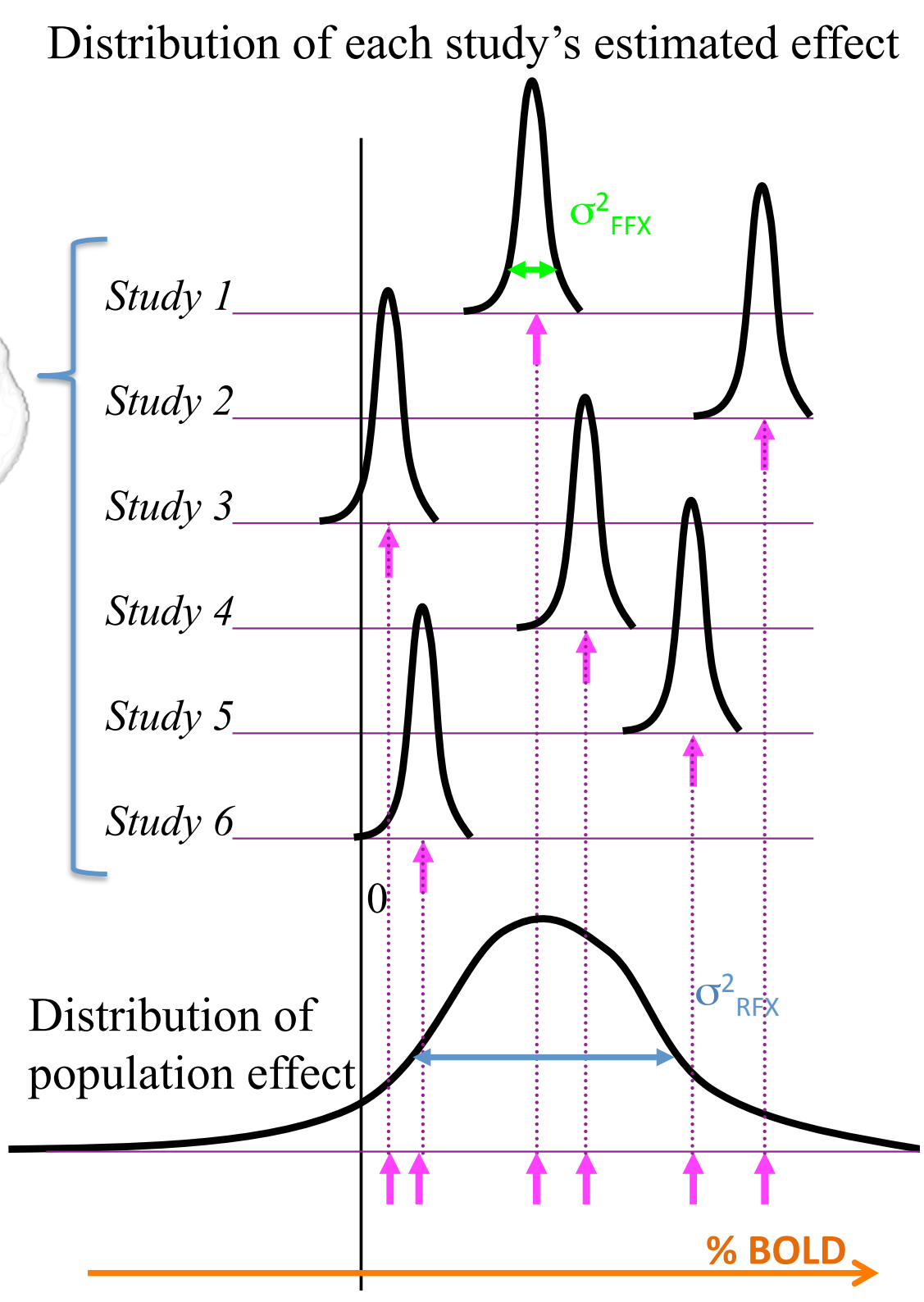
- Effect size
  - Non-imaging MA is all about effect size, CI's
  - What is the effect size?
    - MKDA – Proportion of study result in neighborhood
    - ALE – Probability at individual voxel one or foci
  - Standard errors? CI's?
  - Power/sensitivity
    - 5/10 studies – Great!
    - 5/100 studies – Not great? Or subtle evidence?

- Fixed vs. Random Effects?

IBMA Random Effects?

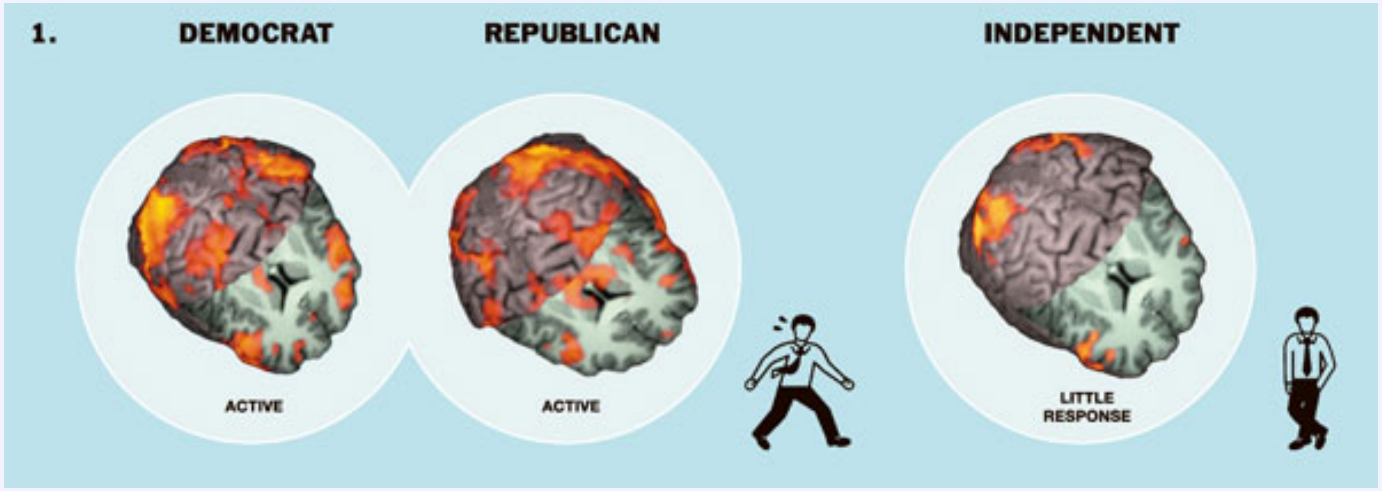
- An effect that generalizes to the population studied
- Significance relative to between-study variation

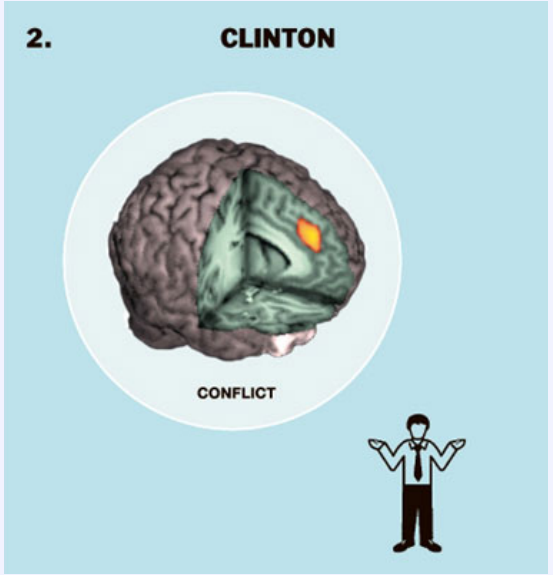




Reverse Inference & Brain Imaging

- Politics study from 2007
  - Voters viewed images of Democratic candidates (N=20)
  - Subset that disliked Clinton:
    - "…exhibited significant activity in the anterior cingulate cortex, an emotional center"…, activated when one "feels compelled to act in two different ways but must choose one."

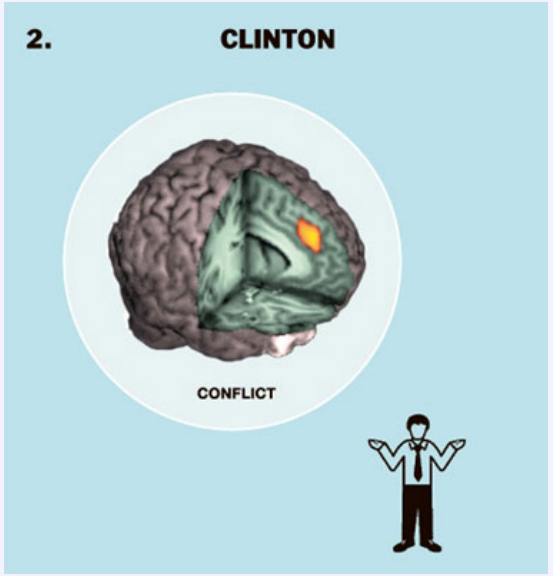




Iacoboni, et al., "This is your brain on politics". OP-ED, *The New York Times*, Nov. 11, 2007

Reverse Inference & Brain Imaging

- Logic
  - Emotion conflict resolution task
    ➔ Anterior Cingulate activation
    known from the literature
  - Hillary Clinton
    ➔ Anterior Cingulate activation
    observed in this experiment
  - Ergo
    ➔Hillary Clinton induces emotional conflict



Iacoboni, et al., *The New York Times*, Nov. 11, 2007

➔Faulty Reverse Inference
  - High P(A.C. Act. | Emot. Conf.) *doesn't imply* high P(Emot. Conf. | A.C. Act.) *!!!*

Reverse Inference: Correctly!

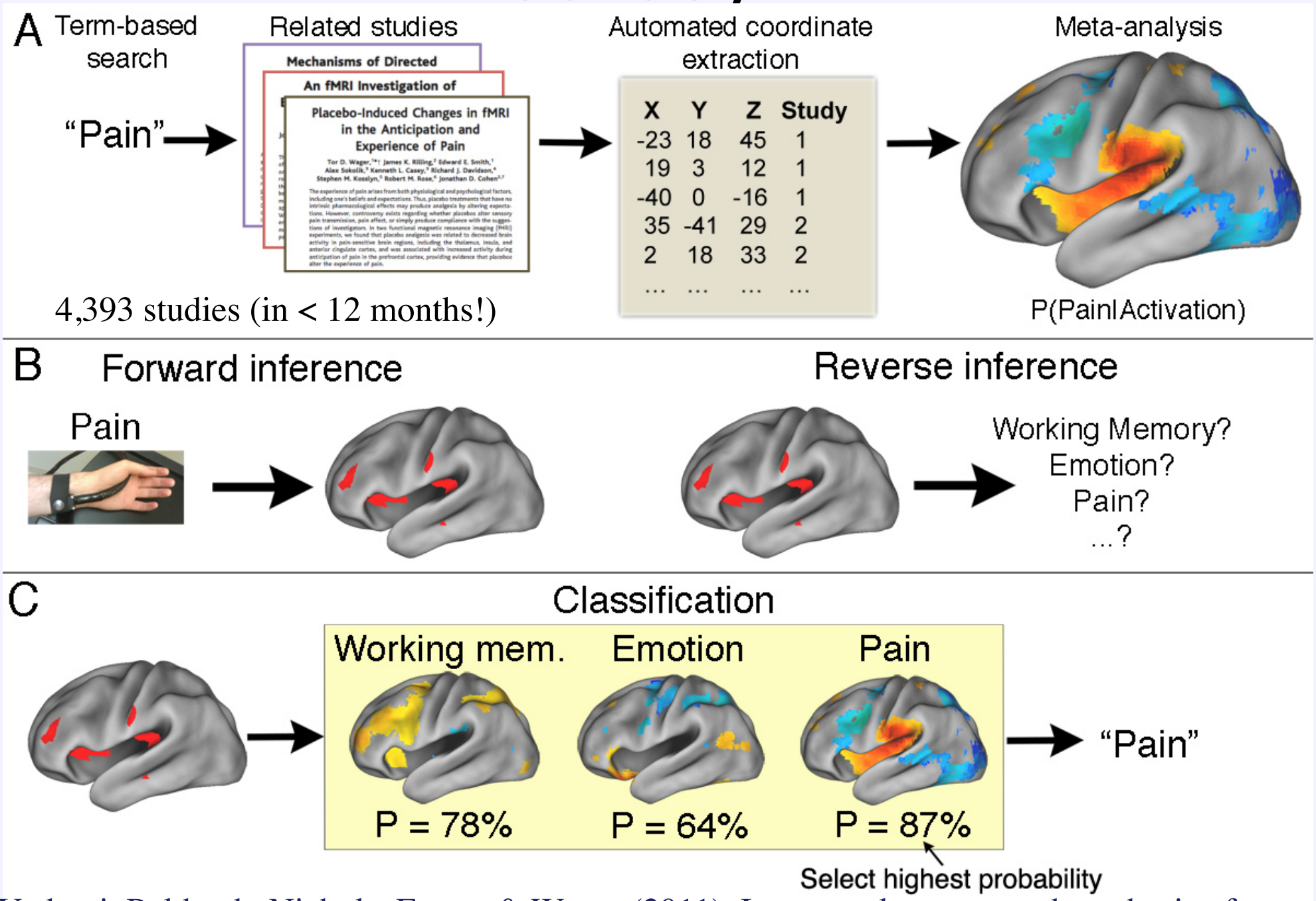
- Bayes Rule
  - Cognitive Domain C, Activation A

$$P(C=c|A) = \frac{P(A|C=c)\,P(C=c)}{\Sigma_{c^*} P(A|C=c^*)P(C=c^*)}$$

summation over *all cognitive domains!*

- Can we find "P(Emot. Conflict | ACC Act.)"?
  - Need to run 100's of experiments!
  - Or, use meta analysis!
  - But best Neuroimaging Meta Analysis databases are still limited
    - BrainMap.org has 2,355 studies (started in 1988)
    - Pubmed finds 21,017 refs "fMRI" in title/abstract

Neurosynth



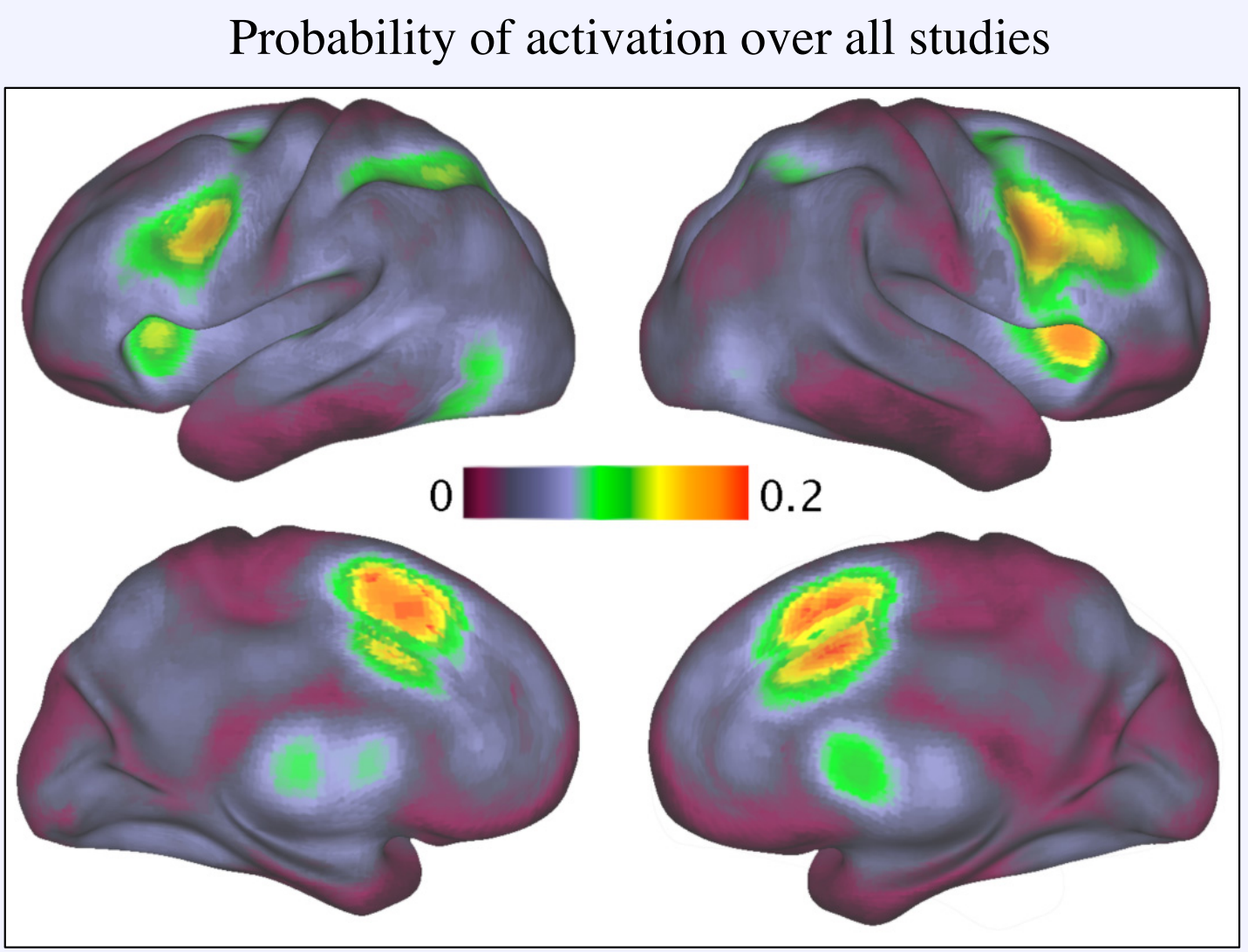
Yarkoni, Poldrack, Nichols, Essen, & Wager (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665-670. www.neurosynth.org

Neurosynth Methods

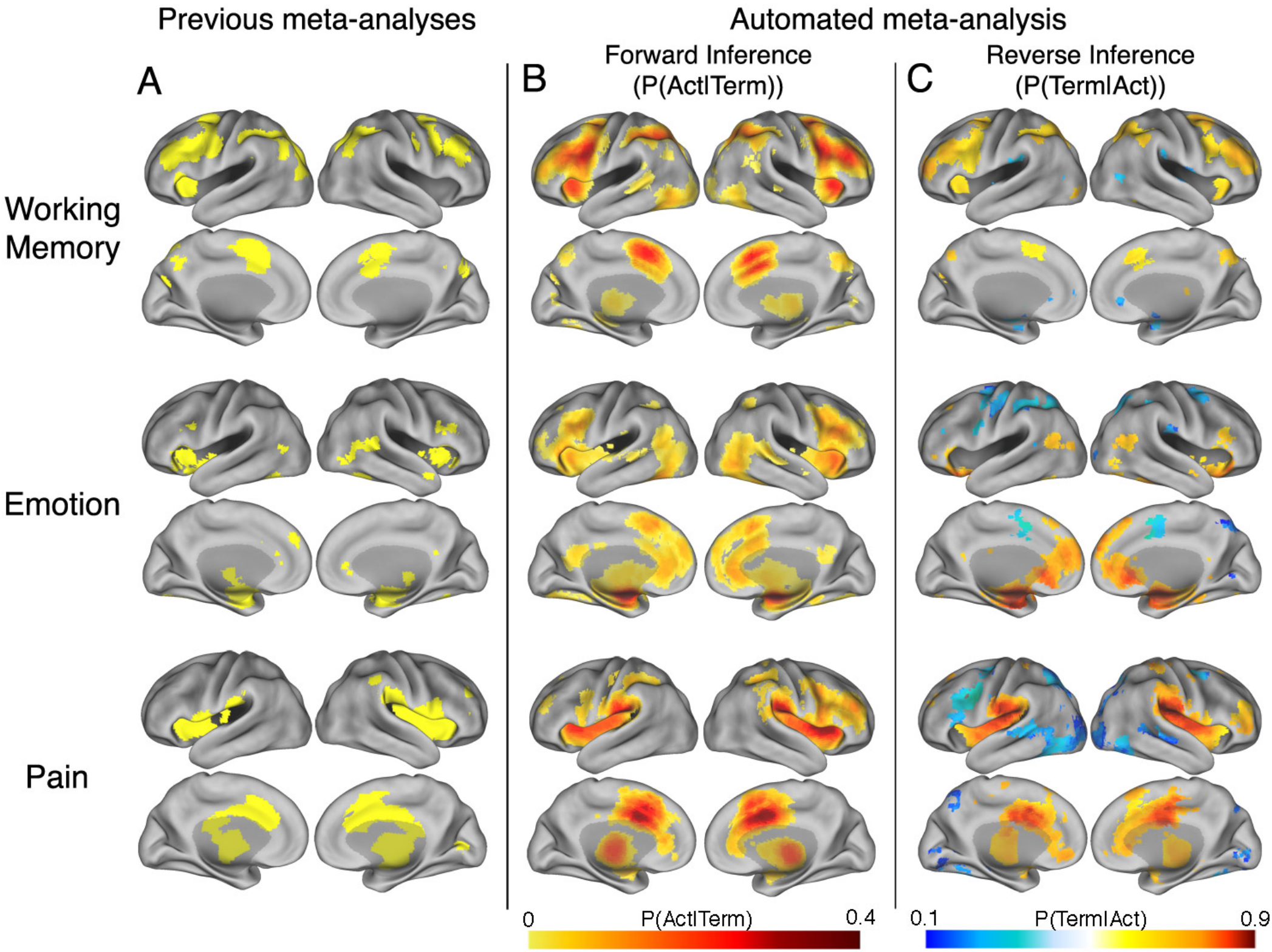
- 17 Neuroscience-focused journals used
  - *Biological Psychiatry, Brain, Brain and Cognition, Brain and Language, Brain Research, Cerebral Cortex, Cognitive Brain Research, Cortex, European Journal of Neuroscience, Human Brain Mapping, Journal of Neurophysiology, Journal of Neuroscience, NeuroImage, NeuroLetters, Neuron, Neuropsychologia*, & *Pain*.
- Tagging
  - Each article 'tagged' with psychological terms
  - Scored as high frequency (>1/1000 words) or not
- Coordinate harvesting
  - Tables parsed for x,y,z coordinates
- Not exhaustive, but already massive
  - 4,400+ studies, 145,000+ foci

What about Anterior Cingulate?

- It's always there!

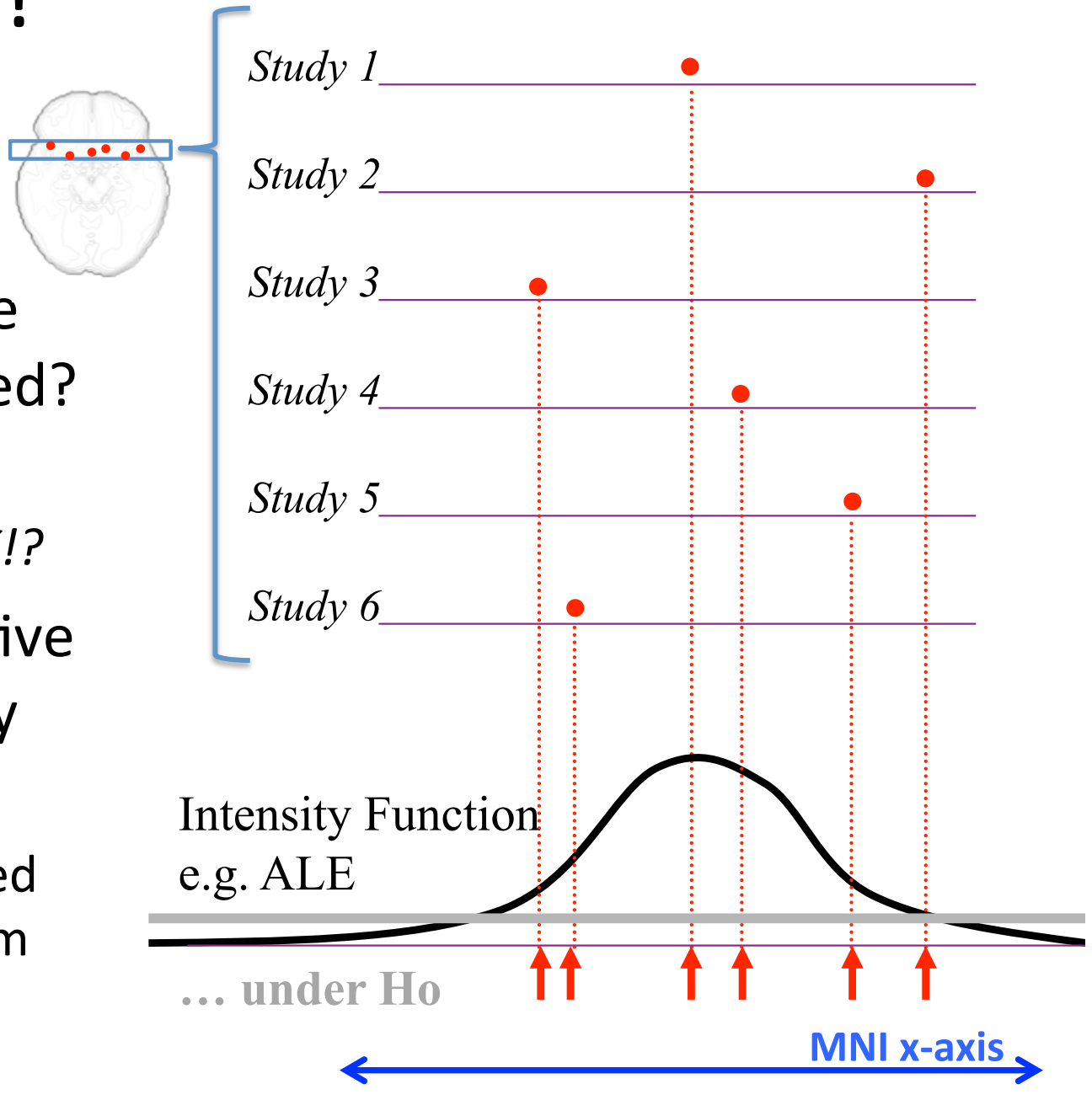


- Finally, can do real reverse inference…


Previous meta-analyses
Automated meta-analysis
A
B
Forward Inference
(P(ActlTerm))
C
Reverse Inference
(P(TermlAct))
Working Memory
Emotion
Pain
0
P(ActlTerm)
0.4
0.1
P(TermlAct)
0.9

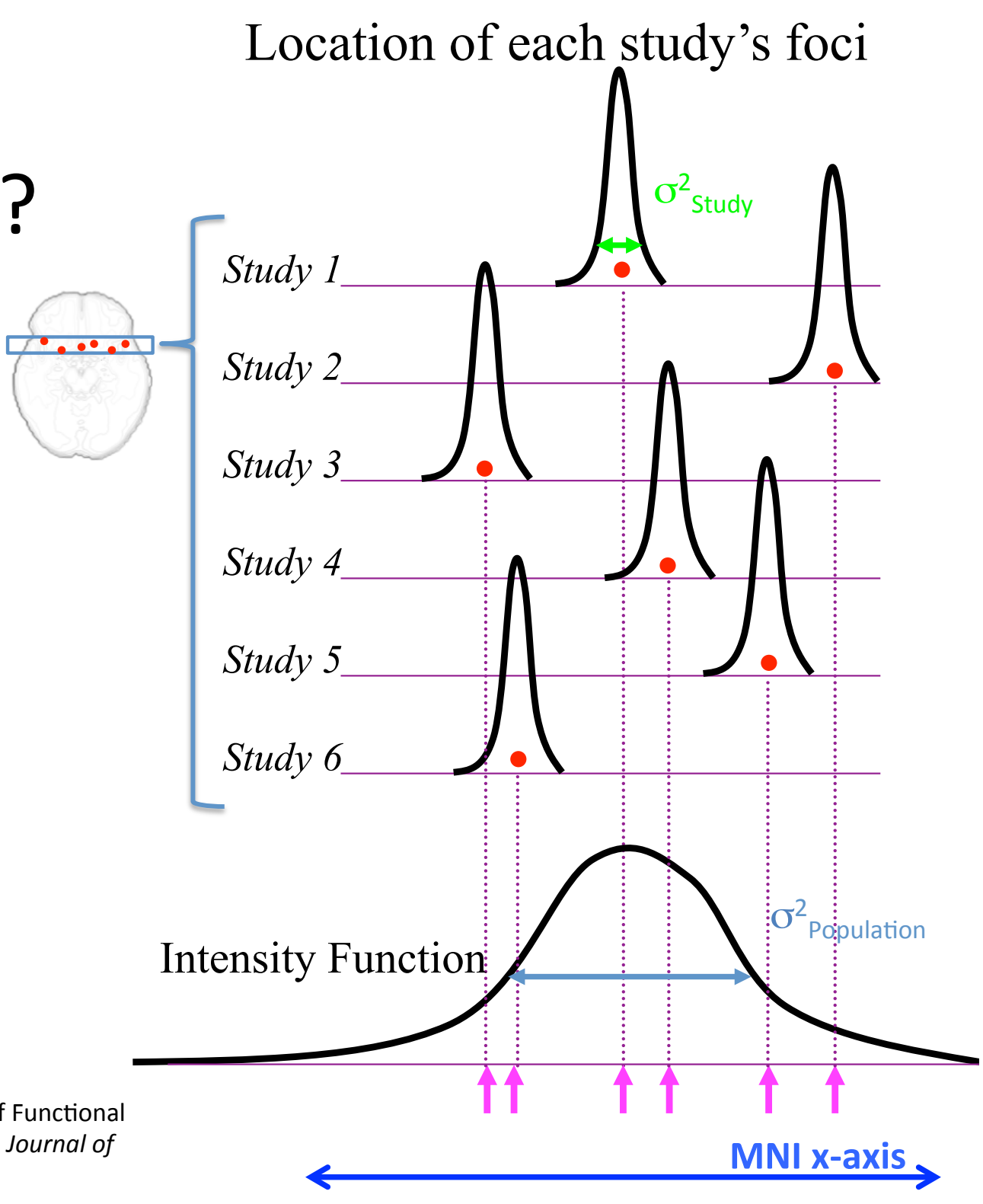
What is a Random Effect?

- CBMA
  - An effect that generalizes to the population studied?
    - *5/10 signif.: OK?*
    - *5/100 signif.: OK!?*
  - Significance relative to between-study variation?
    - Significance based on null of random distribution



What is a Random Effect?

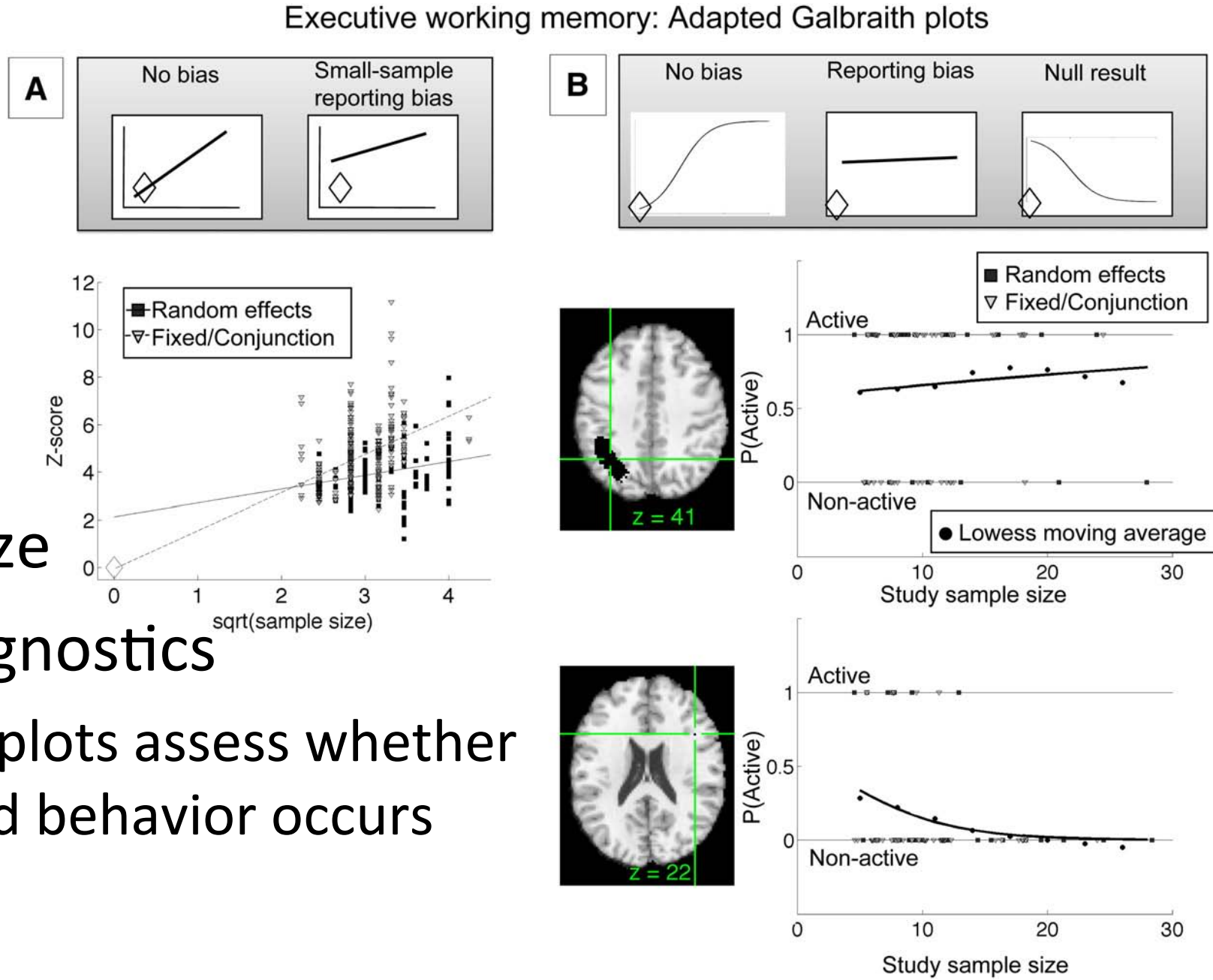
- Bayesian Hierarchical Marked Spatial independent Cluster Process
  - Explicitly parameterizes intra- and inter-study variation



Kang, Johnson, Nichols, & Wage (2011). Meta Analysis of Functional Neuroimaging Data via Bayesian Spatial Point Processes. *Journal of the American Statistical Association*, 106(493), 124–134.

CBMA Sensitivity analyses

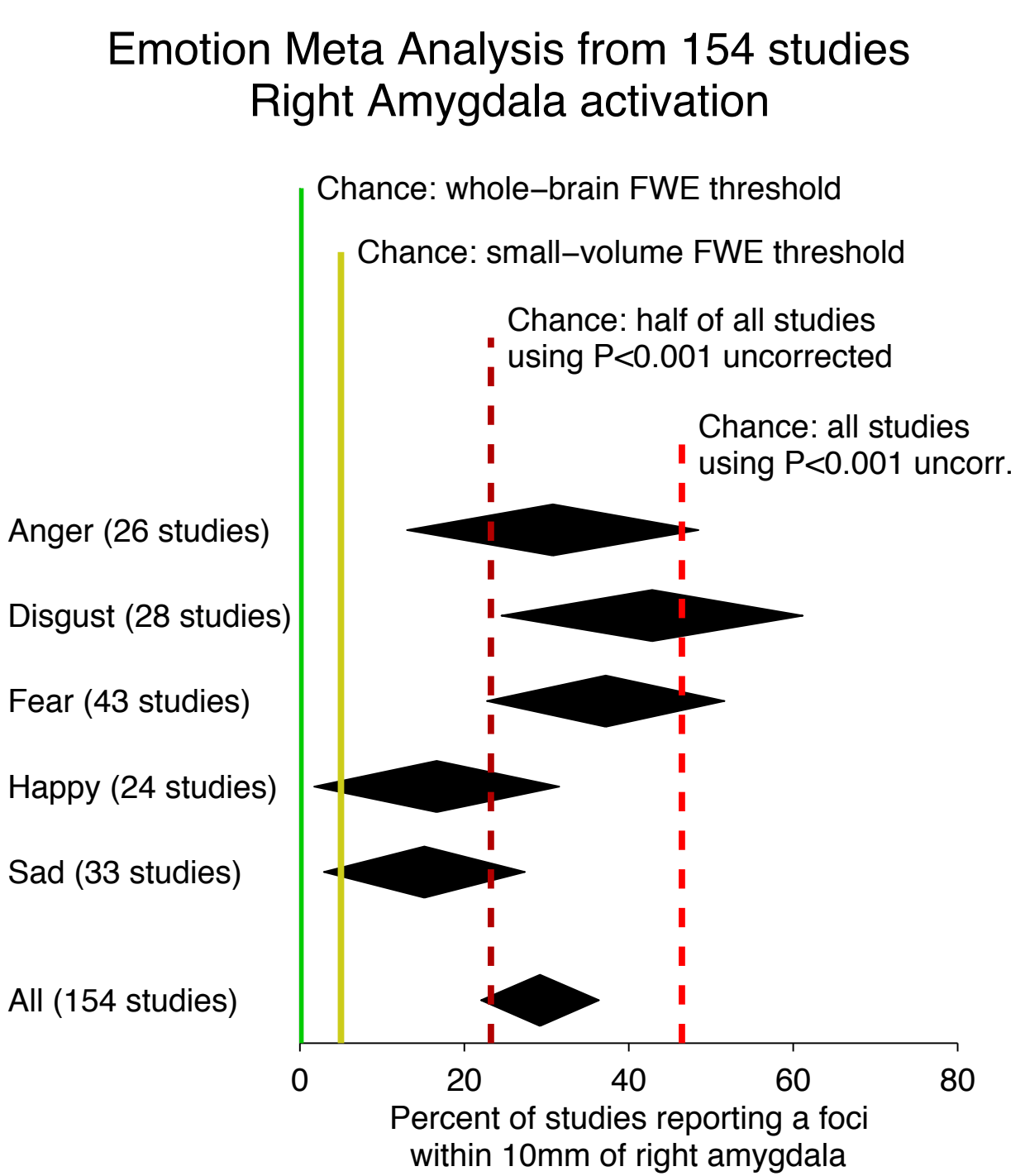
- Z-scores should fall to zero with sample size
- Meta Diagnostics
  - Various plots assess whether expected behavior occurs



Wager et al. (2009). Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *NeuroImage*, 45(1S1), 210–221.

CBMA File Drawer Bias?

- What about "P<0.001 uncorrected" bias?
- Forrest plot
  - MKDA values for right amygdala
  - Can explore different explanations for the effect



T. Nichols

Meta-Analysis Conclusions

- IBMA
  - Would be great, rich tools available
- CBMA
  - 2+ tools available
  - Still lots of work to deliver best (statistical) practice to inferences