To Estimate or Infer?

Lessons from Genetics (& elsewhere)

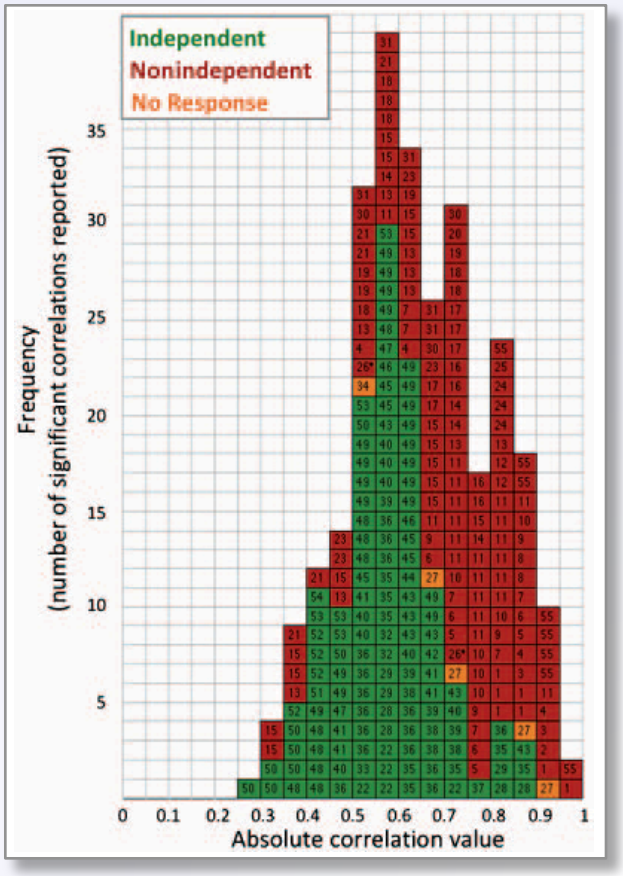
Thomas Nichols, PhD
Department of Statistics &
Warwick Manufacturing Group
University of Warwick

Outline

- Circularity in Imaging & Elsewhere
- Distinguishing Estimation vs Inference
- Understanding Circularity Bias
- Action!

Perilous P-values! Corrupt Correlations!

- Vul, et al. (2009).
  - Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. (aka Voodoo Correlations). *Perspectives on Psychological Science*, 4, 274-290.
  - fMRI-behavioral correlations impossibly large
    - Theoretically limited due to imperfect reliability of BOLD & behavioral measures
  - Due to "circularity"
    - Using same data to select voxels *and* to measure the effect
      - Literature review finds independently derived ROIs yield smaller correlations

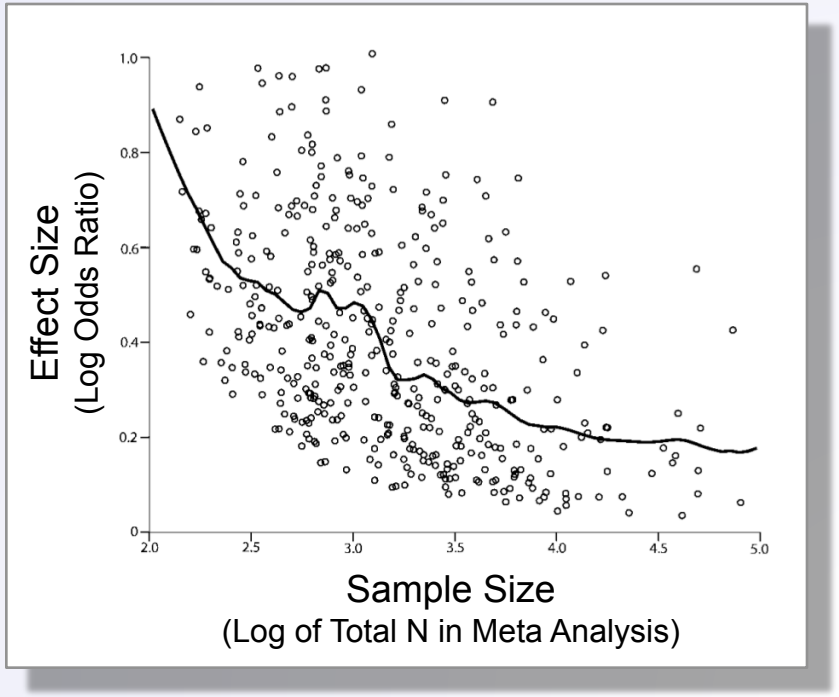


See also: Kriegeskorte, et al. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, 12(5), 535-540.

Perilous P-values! Corrupt Correlations!

- Ioannidis (2008).
  - "Why most discovered true associations are inflated" *Epidemiology*, 19(5), 640-8.
  - "Winner's Curse" – studies over-estimate effect size, due to low power

- ☐ 256 meta analyses for a dichotomous effect (odds ratio)
- ☐ Studies with smallest N have biggest effect size!
  - ☑ Low N studies have low power
  - ☑ Low-power studies rarely succeed (i.e. get published)
  - ☑ But when they do, is result of randomly high effect or randomly small variance, biasing effect size
- ☐ Explains difficulty with replication



‘Circularity’ in Genetics

Am. J. Hum. Genet. 69:1357–1369, 2001

Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans

Harald H. H. Göring,[1] Joseph D. Terwilliger,[2,3,4] and John Blangero[1]

[1]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio; [2]Department of Psychiatry and [3]Columbia Genome Center, Columbia University, and [4]New York State Psychiatric Institute, New York

Am. J. Hum. Genet. 70:575–585, 2002

Bias in Estimates of Quantitative-Trait–Locus Effect in Genome Scans: Demonstration of the Phenomenon and a Method-of-Moments Procedure for Reducing Bias

David B. Allison,[1] Jose R. Fernandez,[2] Moonseong Heo,[2] Shankuan Zhu,[2] Carol Etzel,[3] T. Mark Beasley,[1] and Christopher I. Amos[3]

[1]Department of Biostatistics and Center for Research on Clinical Nutrition, University of Alabama at Birmingham, Birmingham; [2]Obesity Research Center, Saint Luke’s/Roosevelt Hospital, Institute of Human Nutrition, Columbia University College of Physicians and Surgeons, New York; and [3]University of Texas, M. D. Anderson Cancer Center, Houston

The p
It is s
show
Simul
estim
virtua
How
the lo
and t
throu
that a
set m
condi
respo
the in

An attractive feature of variance-components methods (including the Haseman-Elston tests) for the detection of quantitative-trait loci (QTL) is that these methods provide estimates of the QTL effect. However, estimates that

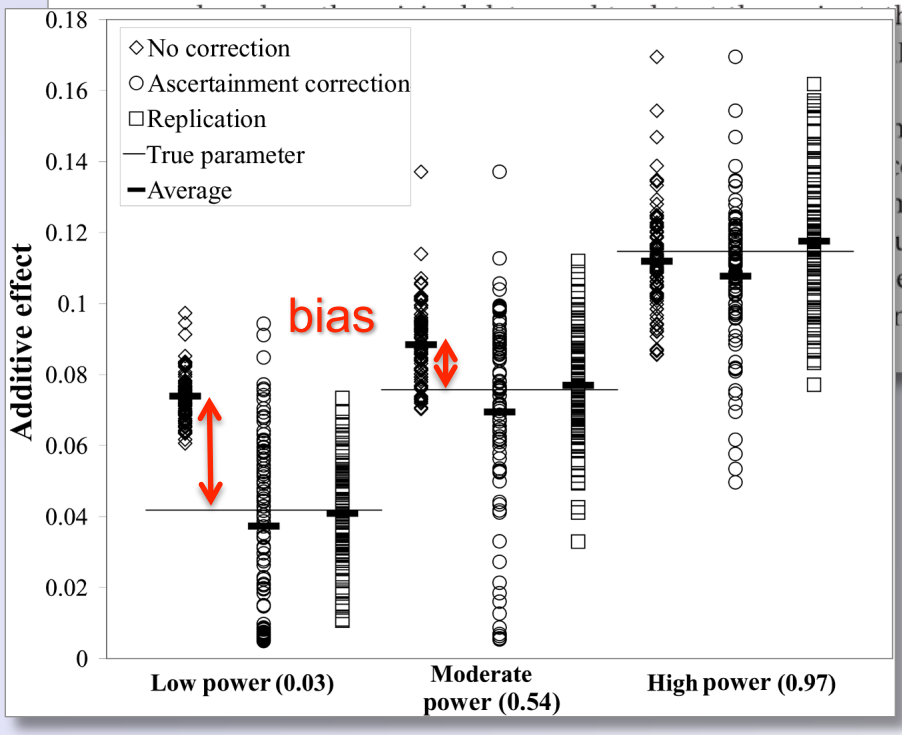
www.ajhg.org The American Journal of Human Genetics Volume 80 April 2007 605

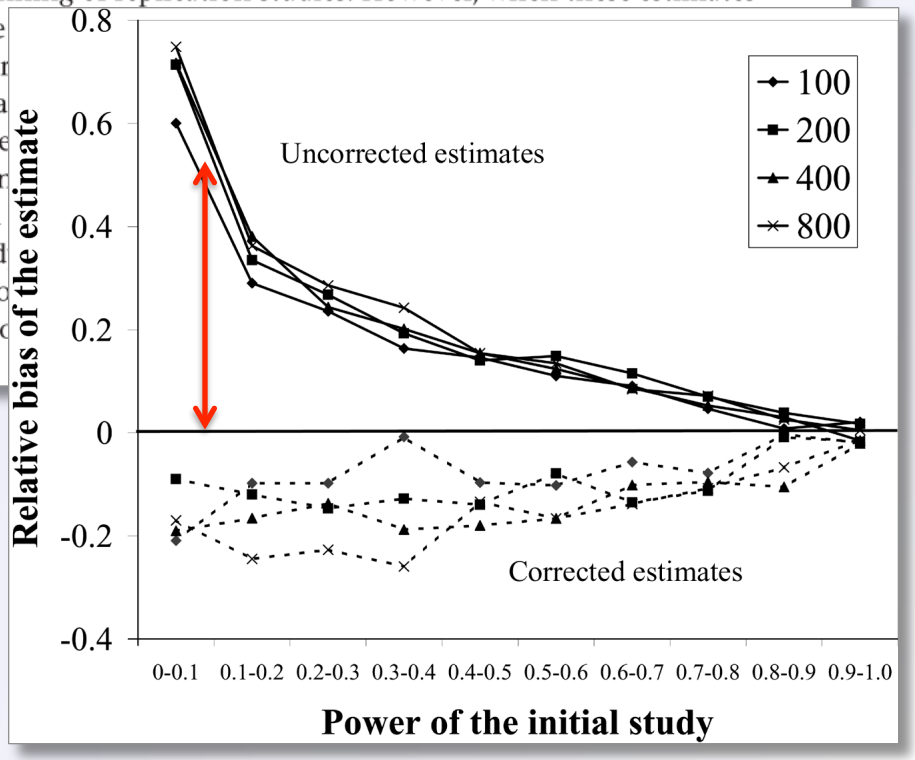
ARTICLE

Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data

Sebastian Zöllner and Jonathan K. Pritchard

Genomewide association studies are now a widely used approach in the search for loci that affect complex traits. After detection of significant association, estimates of penetrance and allele-frequency parameters for the associated variant indicate the importance of that variant and facilitate the planning of replication studies. However, when these estimates





Bias is worst for low power, diminishes with higher power

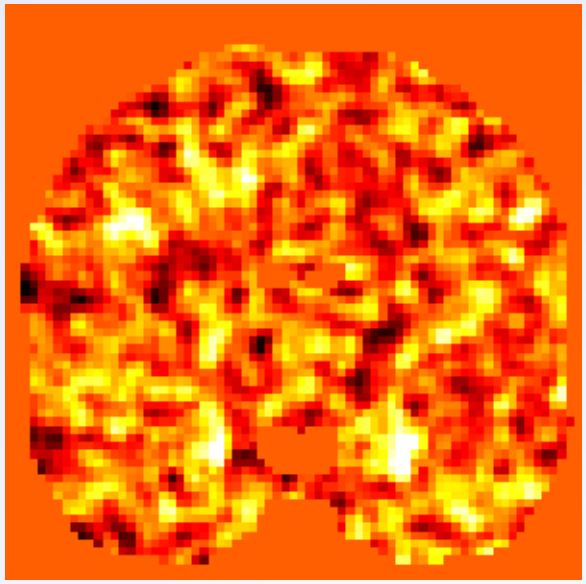
Inference or Estimation?

- Statistics delivers on
  - Inference
    - Deciding on presence or absence of signal
  - Estimation (as usually implemented)
    - Computing a good guess of a continuous quantity
- Estimation assumes no selection
  - Requiring successful detection biases estimation
    - Only looking at OR of peak GWAS-significant SNPs
    - Only measuring %BOLD in significant voxels

Brain Mapping *Inference*

(on where any signal is)

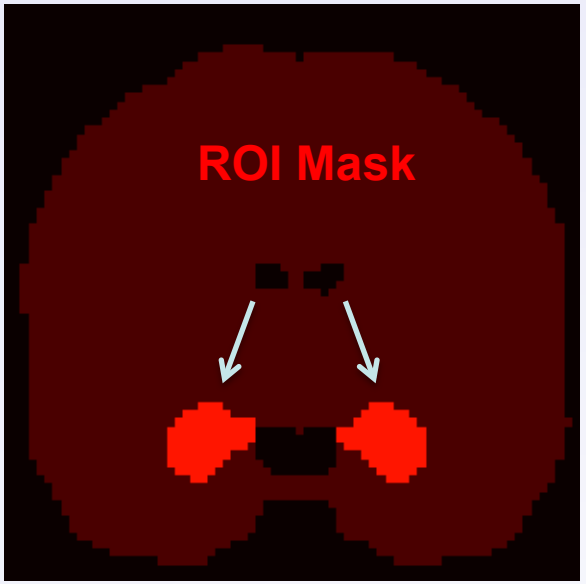
- Perform t-test at 100,000 voxels
- Threshold, mark significant
  - FWE 0.05
    - 95% confident all true positives
  - FDR 0.05
    - 95% true positives on average
- Estimation of effect magnitude?
  - *None*!
  - Only ‘estimation’ of set of signal voxels

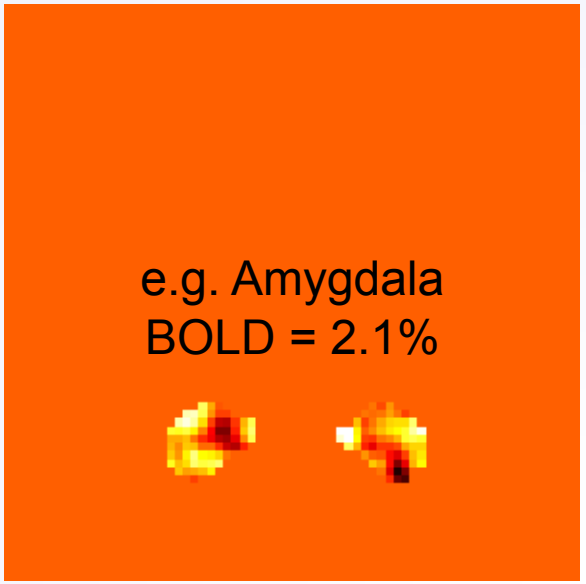


Brain Mapping *Estimation*

(on signal in a given location)

- Define ROI Mask
  - Average voxel-wise %BOLD within mask
- Inference on location?
  - *None!*
  - Location assumed!
  - Only inference is
    - H_0: zero BOLD in ROI
    - H_A: non-zero BOLD in ROI

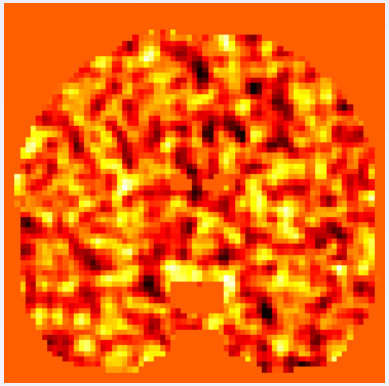




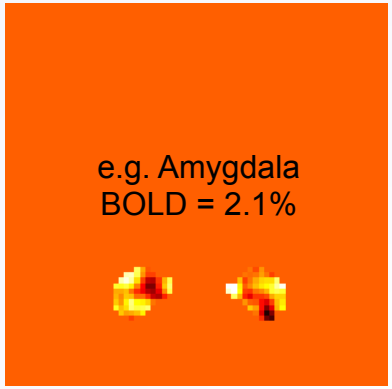
Estimation-after-Inference: Why is there circularity bias?

- Only measure ‘winner voxels’
  - Those with
    - Randomly high sample mean
    - Randomly low standard error

How bad is circularity bias?



- Basic one-sample fMRI analysis
  - N subjects, i=1,…,N
  - X_i BOLD response for subject i, $X_i \sim N(\mu, \sigma)$
  - X (no subscript) = $\sum_i X_i / N$, sample mean
  - $Z = X/(\sigma/\sqrt{N})$, test statistic & Z-score
- No circularity – No Bias
  - Anatomical ROI, or ROI from independent fMRI data
  - $E(X - \mu) = 0$



Estimation Bias: With Circularity

- Conditional inference
  - Only measure X in voxels with $Z \geq u$

- Bias – Conditional on a detection
  - $E(X - \mu \mid Z \geq u) = \phi(u^*)/[1-\Phi(u^*)] \ \sigma/\sqrt{N}$

$$u^* = u - \mu / (\sigma/\sqrt{N})$$

  - Assume no null voxels in ROI

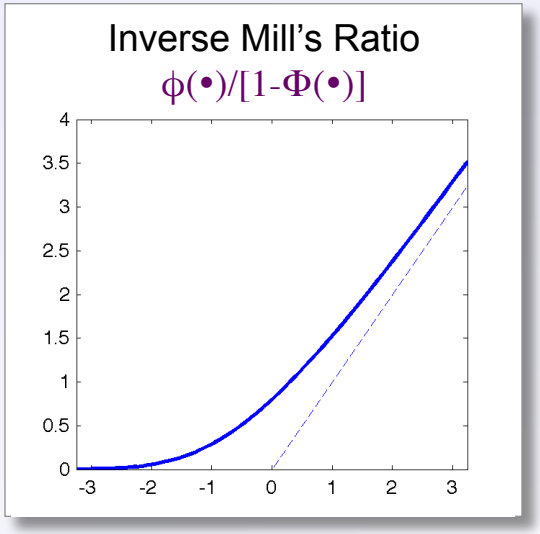
ϕ – CDF of Standard Normal
Φ – PDF of Standard Normal

- Biased by term that depends on
  - Standard Error $\sigma/\sqrt{N}$
  - Shifted threshold u^*
    - Shifted by non-centrality parameter (NCP) $\mu / (\sigma/\sqrt{N})$

Estimation Bias: Under Circularity

$$E(X - \mu \mid Z \geq u) = \phi(u^*)/[1-\Phi(u^*)]\ \sigma/\sqrt{N},$$
$$u^* = u - \mu / (\sigma/\sqrt{N})$$

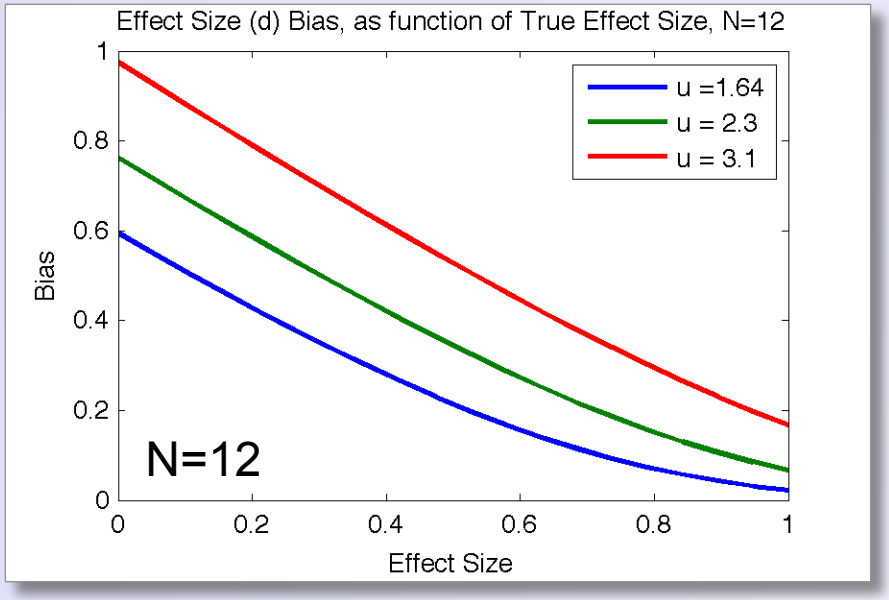
- Components of bias
  - Transformed relative threshold $\phi(u^*)/[1-\Phi(u^*)]$
  - Standard error $\sigma/\sqrt{N}$
- In Z-score units

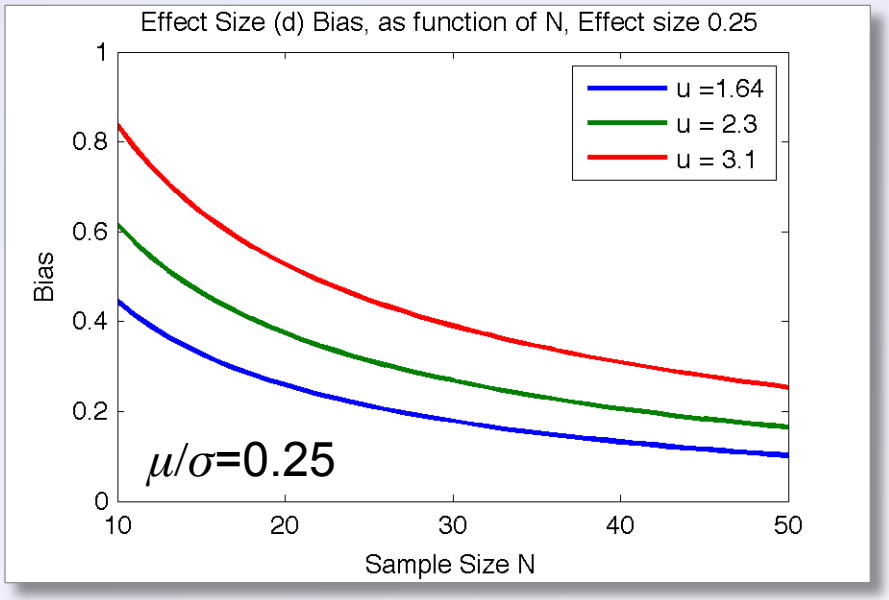


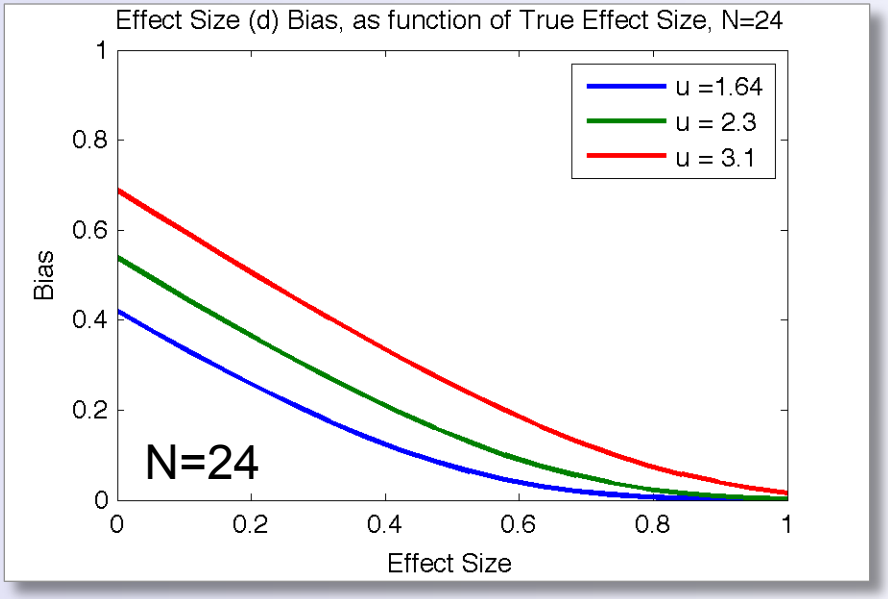
$$E((X - \mu)/(\sigma/\sqrt{N}) \mid Z \geq u) = \phi(u^*)/[1-\Phi(u^*)]$$

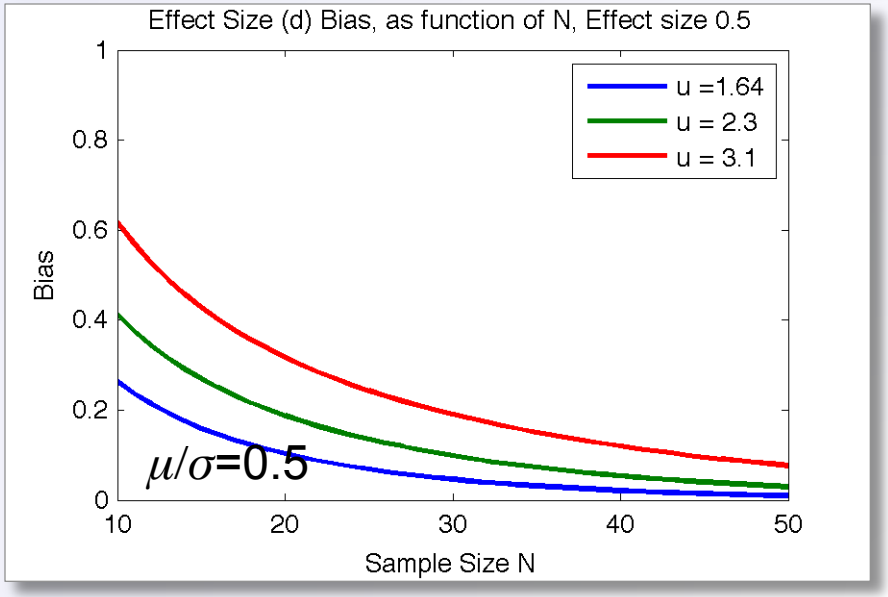
  - So if $\mu \approx 0$, Z-score bias $\approx u$
    - Very approximate correction: $Z^* \approx Z - u$

Bias Reduces with Signal & N









Estimation Bias: What if ROI misses signal?

- Conditional expectation now a mixture
  - $E(X - \mu \mid Z \geq u) =$

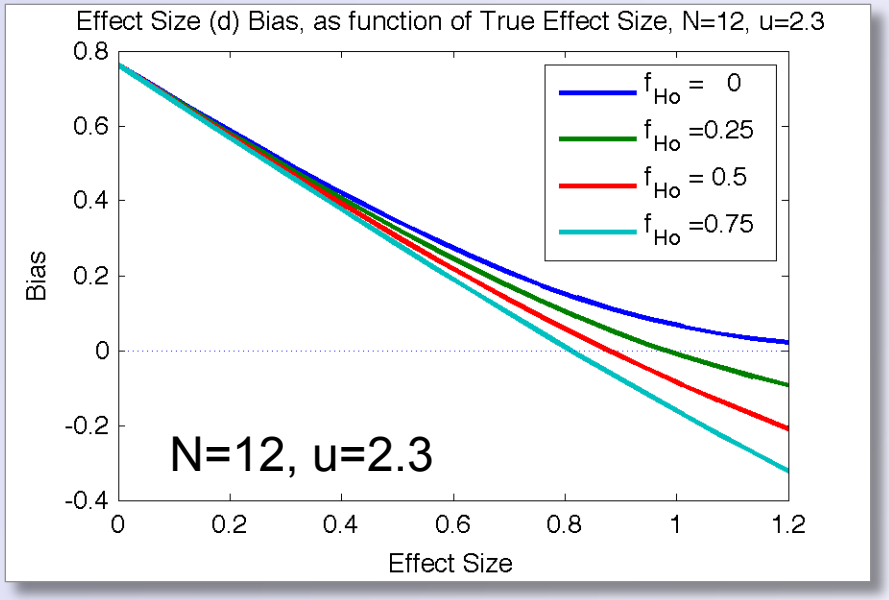
$$f_{Ho} \quad \phi(u)/[1-\Phi(u)] \quad \sigma/\sqrt{N} \quad +$$
$$-f_{Ho}\,\mu \quad + (1-f_{Ho})\ \phi(u^*)/[1-\Phi(u^*)] \quad \sigma/\sqrt{N}$$

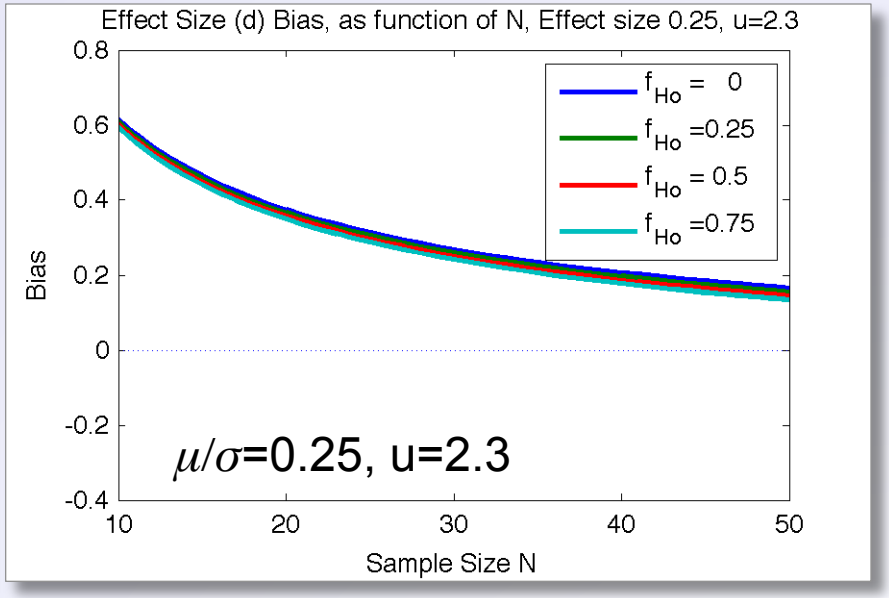
  where

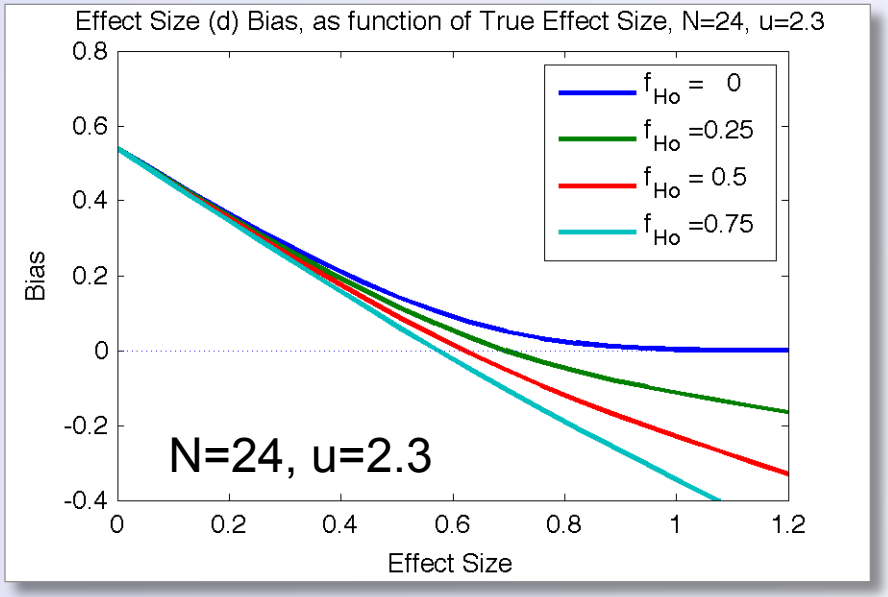
  f_{Ho} is fraction of ROI that is null

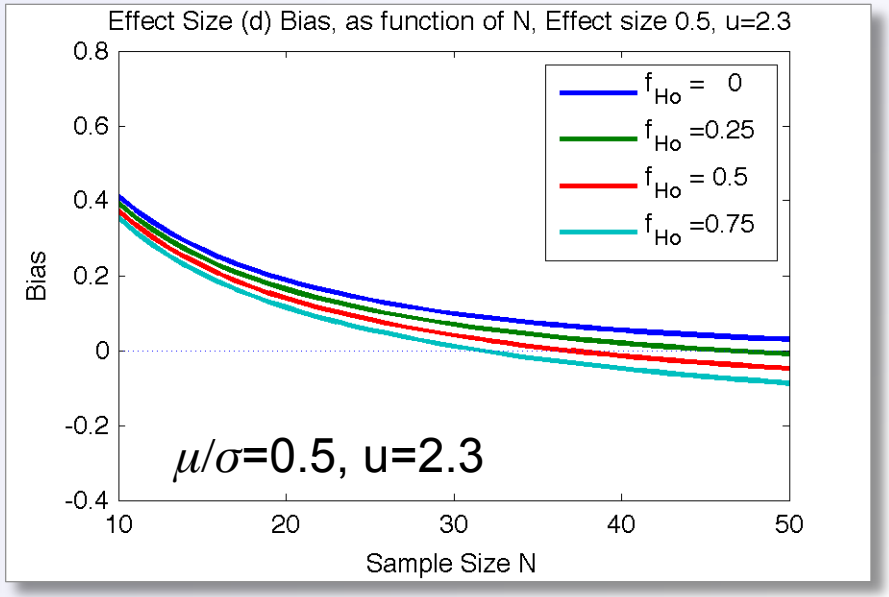
- So now have different directions of bias
  - "Winner's Curse" biases up
  - False positive voxels biases up
  - Diluting true positives biases down

Bias: Effect of Null voxels in ROI









Estimation Bias: Test-retest is Circular!

- Vul et al. suggest intrasubject split-halves
  - Split each subject's data into two
    - Use half A to define a ROI, half B to estimate
  - This is still circular!
- Conditional expectation
  - $\mathrm{E}(X_B \mid Z_A \geq u) = \mu + \phi(u^*)/[1-\Phi(u^*)] \ \sigma/\sqrt{N} \ \rho$ where

$$\rho = \sigma_G^2 / (\sigma_G^2 + \sigma_\varepsilon^2)$$

  - Same, except scaled by A/B correlation
    - Depends on RFX (σ_G^2) and FFX (σ_ε^2) variance

Conclusions: Bias Exploration

- Bias is variable
- Depends on
  - True signal magnitude
    - Worst positive bias with small signal (low power)
  - Proportion of true signal in ROI
    - Null voxels add positive bias
    - Dilution of signal gives negative bias
- Practical suggestions?
  - Use smallest ROIs practical

Conclusions: Circularity

- Circularity well described in other fields
  - If not well-appreciated
- Estimation
  - Assume location → Measure signal
- Inference
  - Search over space → Localize

Voodoo Correlations Redux: Is the sky falling? *Yes!*

- Multiple comparisons
  - Authors still use "P<0.001 uncorrected"
  - Reviewers still accept it!!!
- Sloppy or nonexistent 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!"

  - These "vibrations" can only lead to inflated false positives