Correcting for Circular Inference without data splitting

Samuel J. Davenport[12], Thomas E. Nichols[2]

[1]Department of Statistics, University of Oxford [2] Big Data Institute, University of Oxford

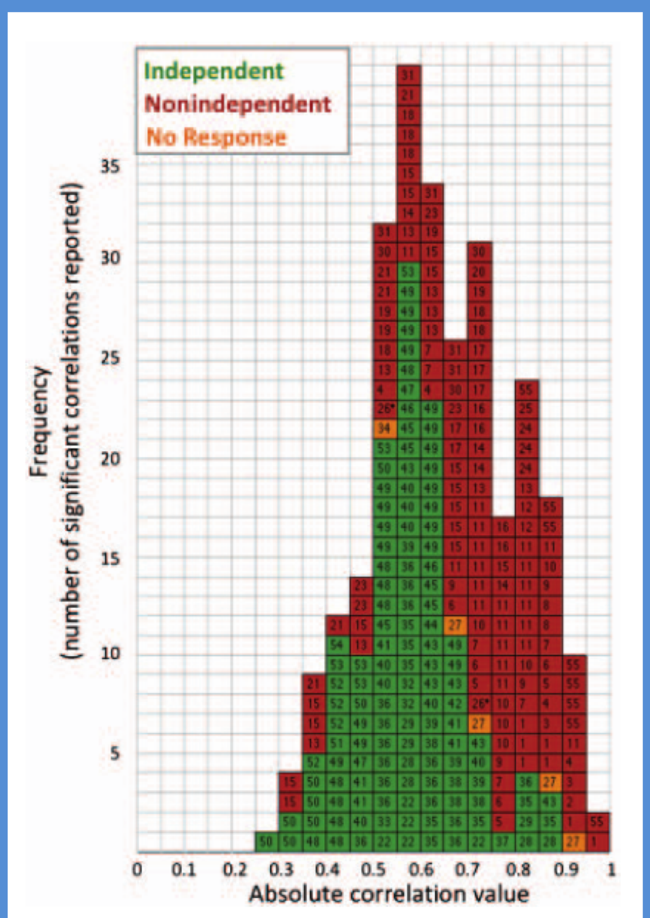
[3]Wellcome Centre for Integrative Neuroimaging, University of Oxford and [4]Department of Statistics, University of Warwick

UNIVERSITY OF OXFORD

BIG DATA INSTITUTE

Come see the talk! Tuesday 13.45 at the Open Science Room

1. Introduction

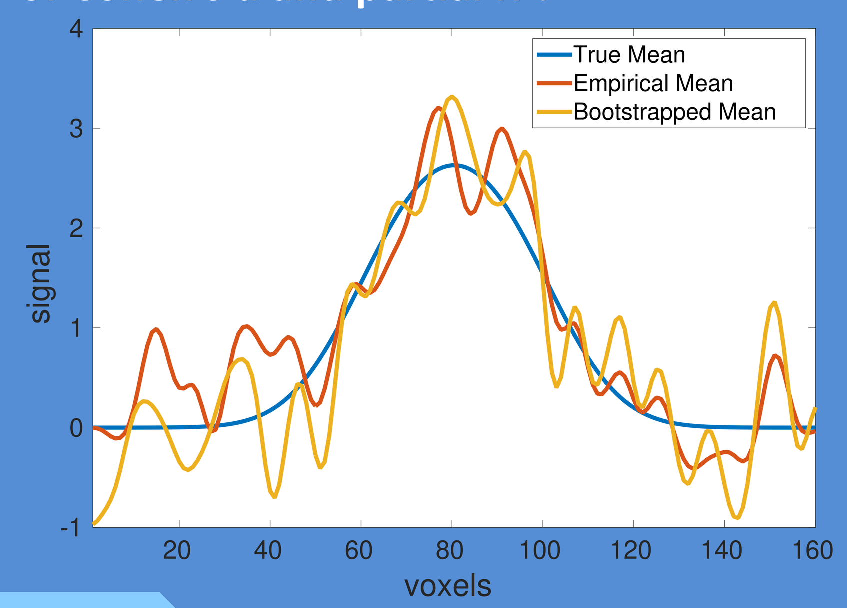


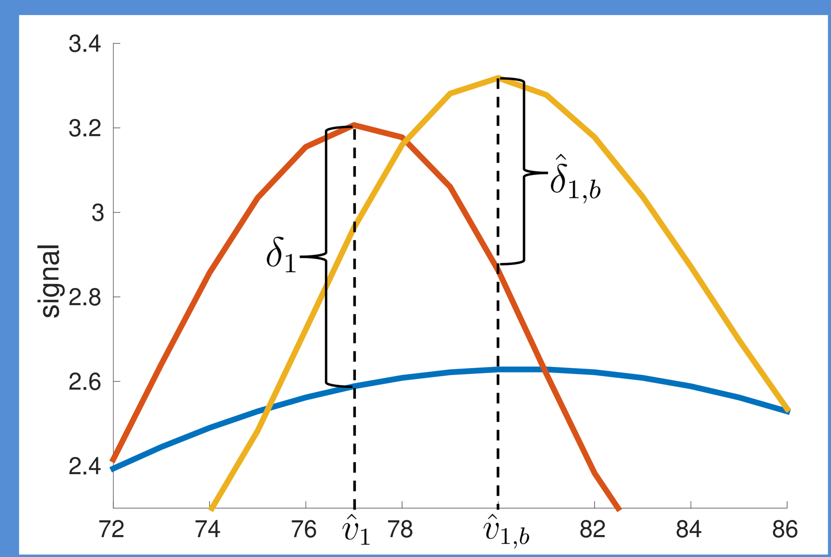
Reporting peak effect sizes following a search for significance is an example of the winner's curse problem. Even if an effect is truly present, the magnitude of the effect is over-estimated. This problem was described by the "Voodoo Correlations" paper: Vul et al (2009), that found that "circular" correlations (computed at locations determined from the data, red in the figure to the left) usually exceeded non-circular correlations (green). If this problem is addressed at all, the typical solution is Data-Splitting, using the first half of the data to find significant regions and the second half of the data to calculate the effect sizes causing a loss of power and spatial accuracy.

2. Methods – Bootstrap Bias Correction

We estimate the bias using the non-parametric bootstrap (building on a method from Tan et al (2014)) and then subtract this bias estimate from the peak effect size to yield a lower bootstrap-corrected estimate of the effect. To do so we resample the data and compare the peaks of each resample to observed peaks to obtain an estimate of the bias. In practice we compute peak height via Cohen's d, namely $\mathrm{d} = \mathrm{T}\sqrt{N}$ and partial R^2 to provide an N-independent measures of the effect.

The figures below provide an illustrative 1D simulation on a grid of 160 voxels where we consider correcting the global maximum of the mean. The bias above the noise-free signal (δ_1) is evident, and is estimated by comparing a bootstrap sample to the original, yielding an estimate of $\delta_{1,b}$. The method works similarly for 3D brain images and extends to estimation of Cohen's d and partial R^2.





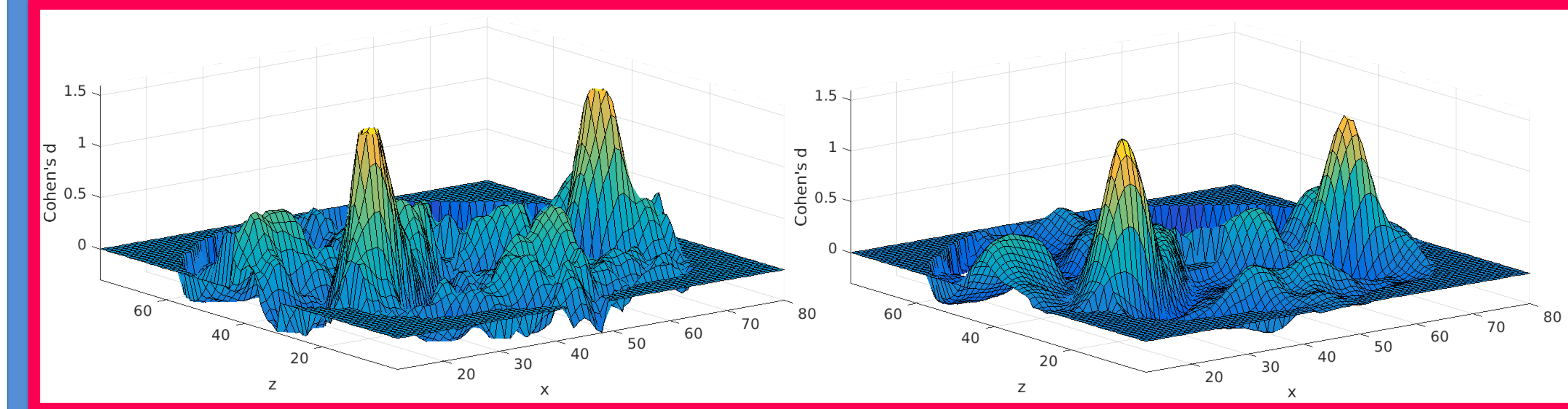
3. Methods - Big Data Evaluation

UK Biobank

We have used an empirical validation using real data from the UK Biobank with fMRI (faces-shapes contrast) and VBM data from 8940 subjects.

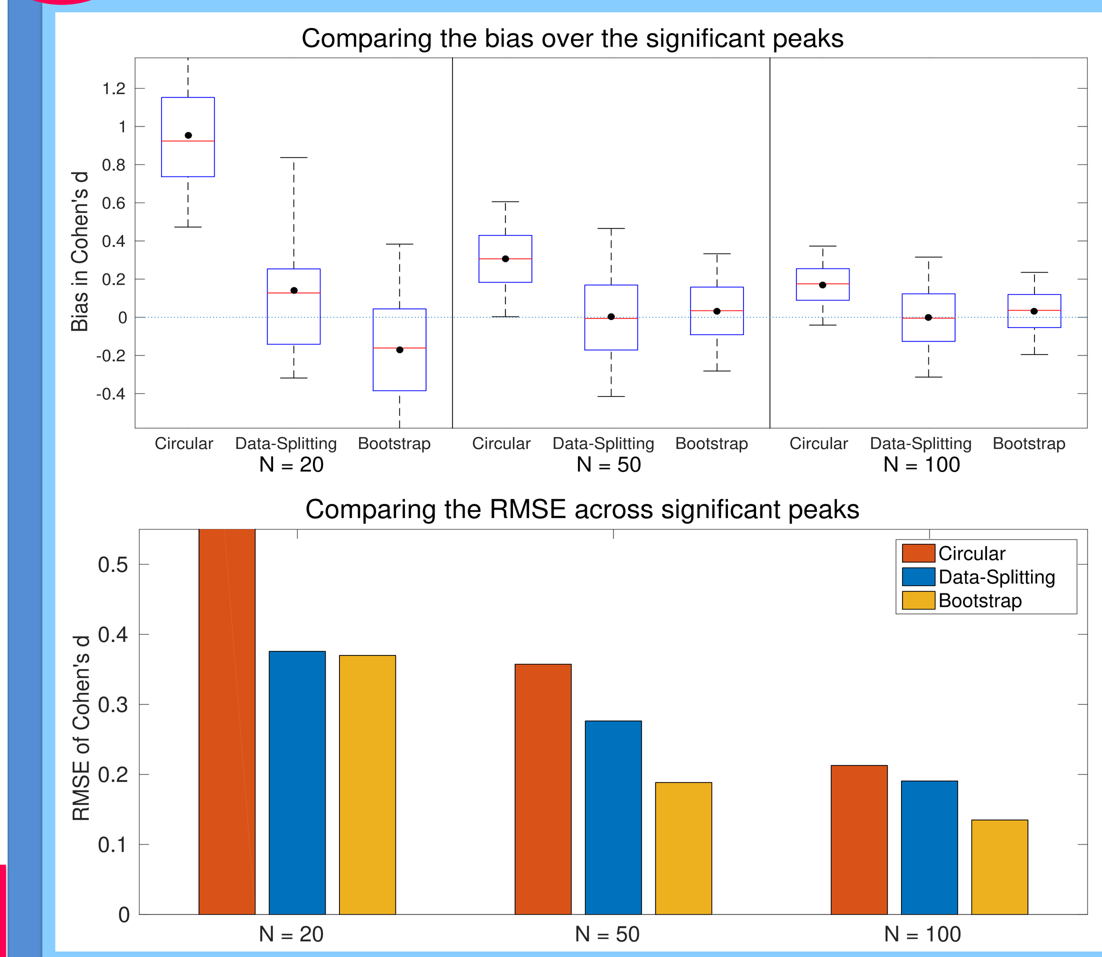
4000 subject ground truth

We set aside 4000 subjects to compute an accurate ground truth and divided the remaining data into groups with sizes typical of those used in neuroimaging in order to test the analyses. We recommend that all emerging methods be tested in this manner.

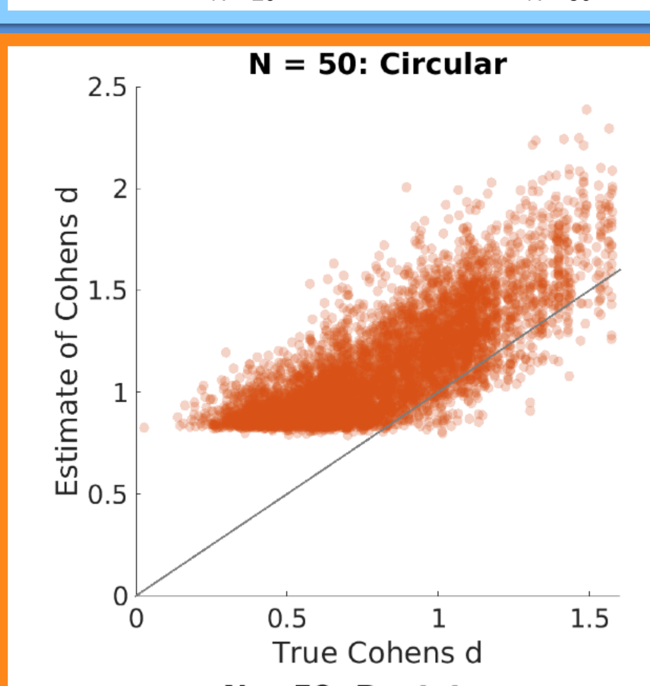


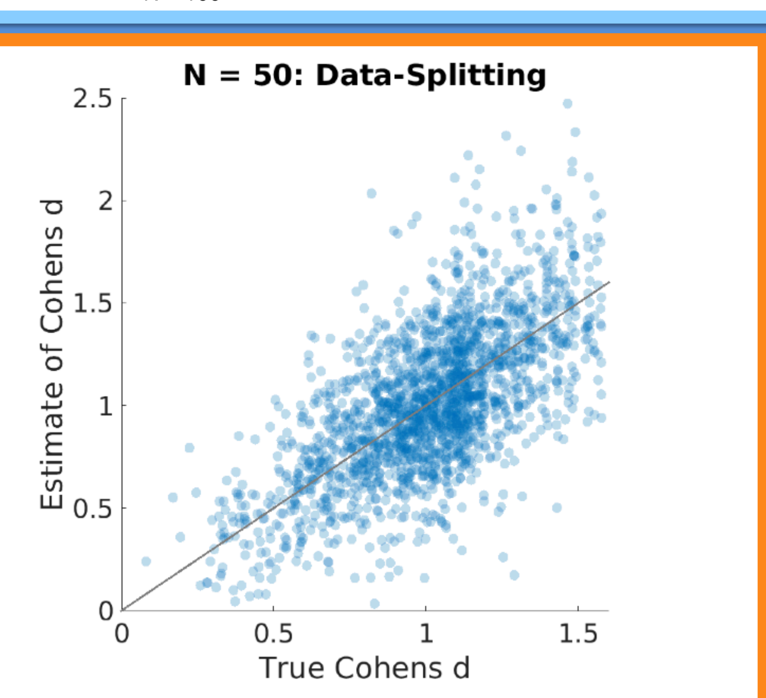
Left: Slice through a 50 subject Cohen's d estimate. Right: Slice through 4000 subject Cohen's d ground truth.

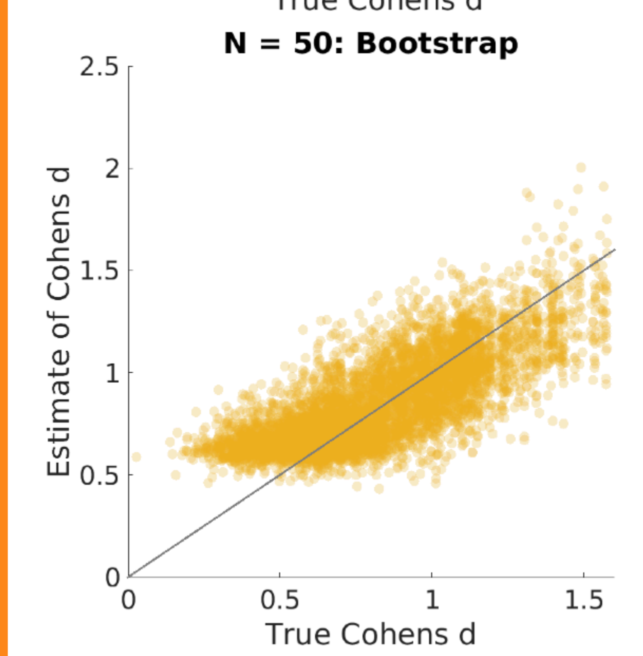
4. Results – One sample Cohen's d



Comparison of one sample Cohen's d estimation for task fMRI. Estimates are compared relative to the 4000 subject ground truth. Bias (top), and RMSE (bottom) are shown for sample sizes of N = 20, 50 and 100, based on 4940/N samples.

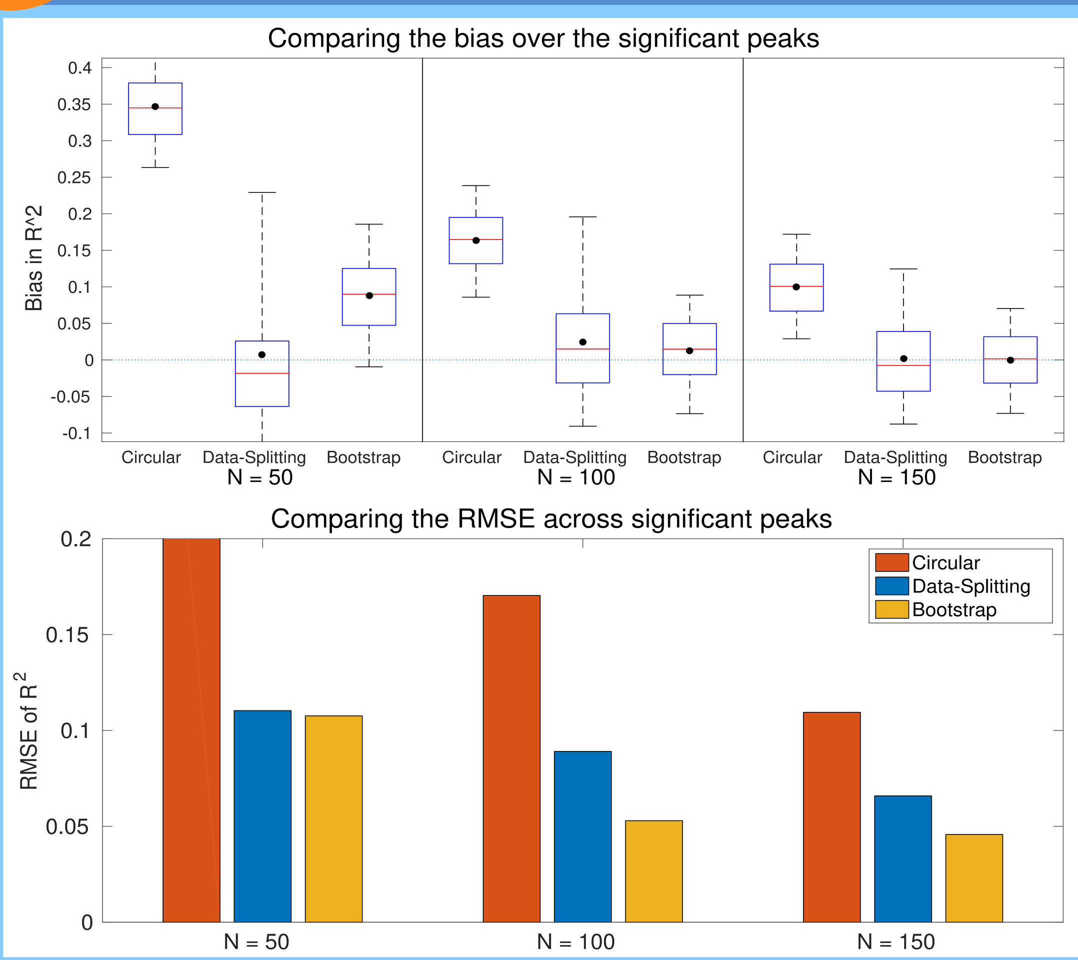




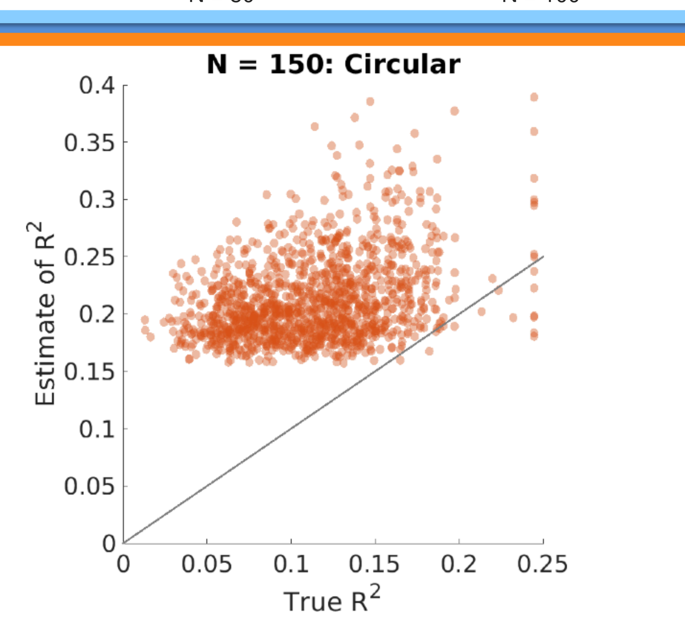


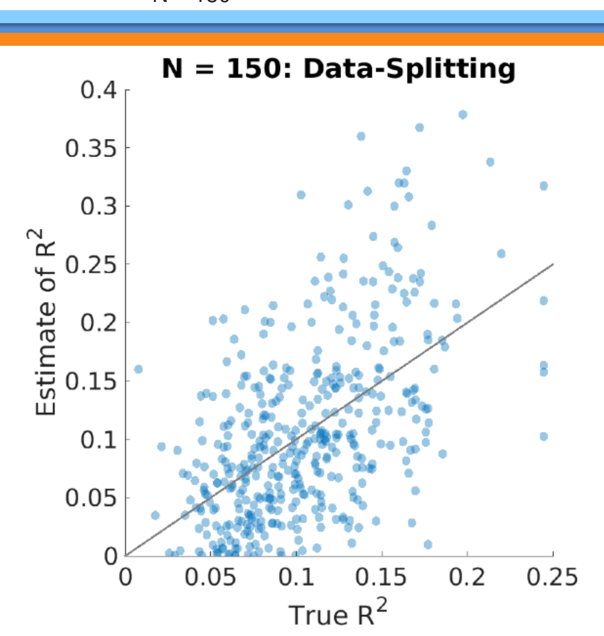
Plots of estimated versus true value of one-sample Cohen's d (calculated using 50 subjects) for task fMRI images, for circular (top-left), data-splitting (top-right), and bootstrap (bottom-left). These show all peaks found over the 4940/50 samples. For each peak the true Cohen's d is obtained at that location from the held-out 4,000 subject Cohen's d image. We can see that bootstrap estimates have low bias and variance while data-splitting estimates are unbiased but are more variable.

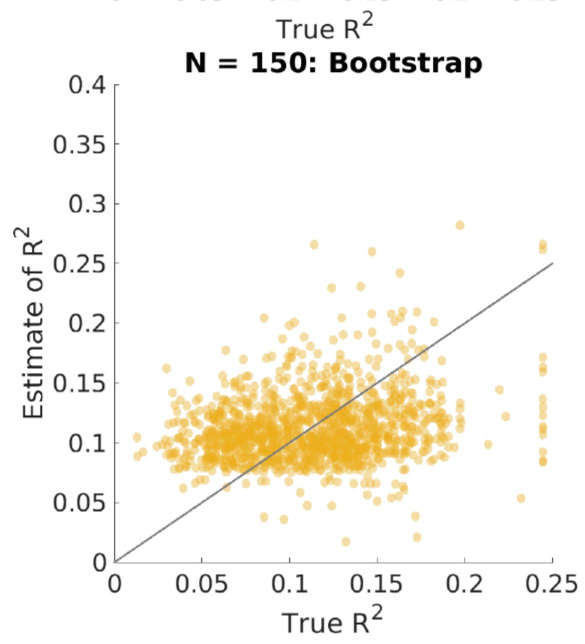
5. Results – partial R^2 for age



Comparison of estimates partial R^2 for age VBM data (regressed against age and sex). Bias (top), and RMSE (bottom) are shown for sample sizes of N = 50, 100 and 150, based on peaks in 4940/N samples.

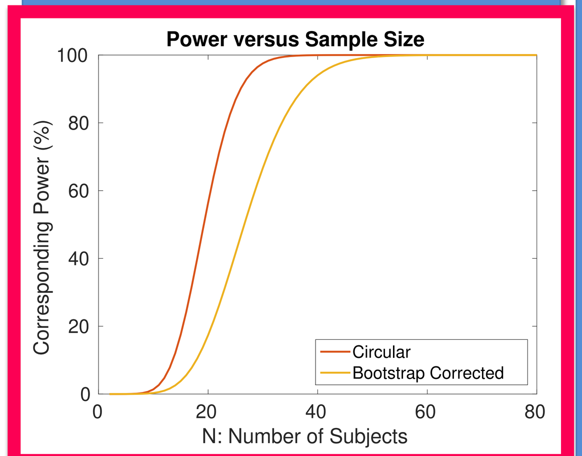






Plots of estimates (using 150 subjects) versus ground truth of the partial R^2 for age, obtained using a GLM regression on VBM data, for circular (top-left), data-splitting (top-right), and bootstrap (bottom-left). These show all peaks found over the 4940/50 samples. For each peak the true partial R^2 is obtained at that location from the held-out 4,000 subject partial R^2 image. Here the SNR is low and so the observed rank order can differ substantially from the noise-free rank order, reducing the accuracy of the bootstrap. However, in terms of RMSE, the bootstrap still outperforms circular inference and data-splitting. As the sample size increases the bootstrap estimates lie closer to the identity line and perform better.

6. Power



Implementing the method on a working memory dataset with 80 subjects yields a circular Cohen's d of 1.52 and a corrected one of 1.16.

7. Discussion

Unbiased estimation of effect size is essential yet absent from most neuroimaging studies. We have evaluated three methods for assessing the signal magnitude at peaks in neuroimaging analyses. The bootstrap method that we have introduced provides circularity corrected estimates from an analysis using all of the data enabling very accurate inference on peak locations relative to data-splitting. Compared to uncorrected, circular inference our method has dramatically less bias and lower RMSE. While data-splitting is unbiased by construction, our method has lower variance and RMSE in most settings.

8. Pre-Print

BioArxiv preprint available with more detailed methods and results: Davenport and Nichols (2019). See also the SIbootstrap package available at: sjdavenport@github.io/software

9. References

- Davenport and Nichols, Selective peak inference: Unbiased estimation of raw and standardized effect size at local maxima 2019.
- Tan et al, Selection Bias Correction and Effect Size Estimation under Dependence, 2014.
- Vul et al . Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition, 2009

- Thanks to Alexander Bowring at the University of Oxford for the poster template.