# Selective peak inference: Unbiased estimation of the effect size at local maxima

Samuel Davenport and Thomas E. Nichols

University of Oxford

April 6, 2021

# Double Dipping

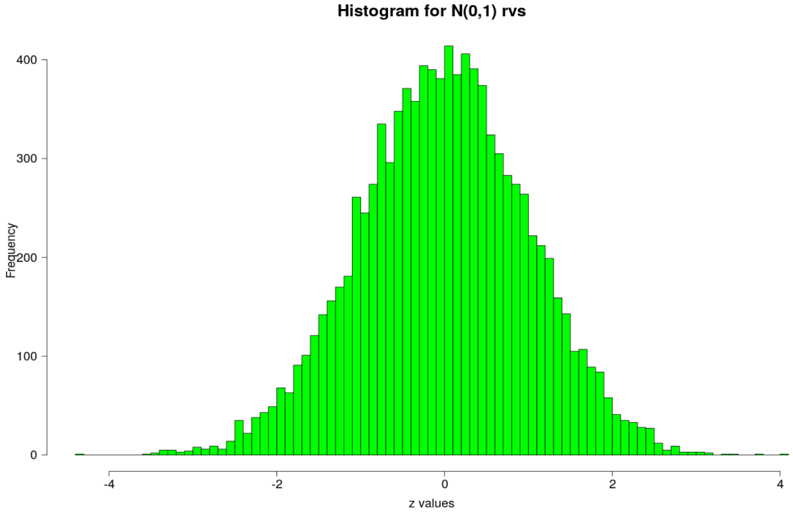
# Examples

- Dice Example: Imagine you roll 10 fair dice and at random some of them show a 6. If you rolled them again would you expect them still to be 6?

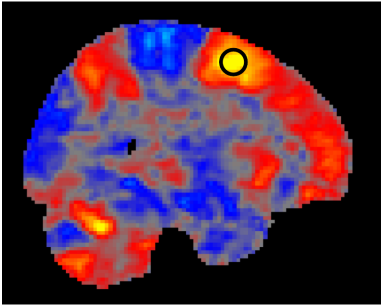
Figure 1: Some Dice

# Mean 0 example

Suppose for now that we have 10000 independent N(0,1) random variables. Then the largest are biased estimates for the true mean.



# The Winner's Curse in fMRI



- Choose significant voxels based on some statistic and its maxima.
- Report uncorrected values at peaks

# Data-Splitting Approach

- Split your subjects into two groups.

# Data-Splitting Approach

- Split your subjects into two groups.

# Data-Splitting Approach

- Split your subjects into two groups.

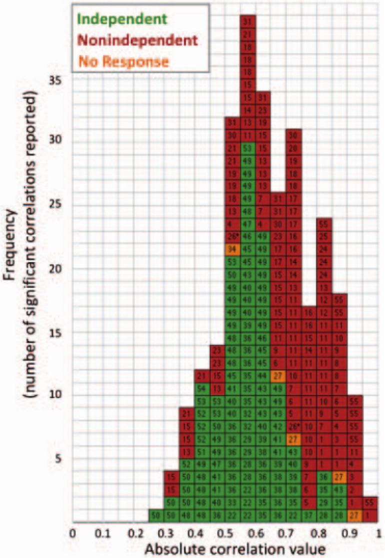
- Use half for significance and half for estimation of the effect size.
- Solves the bias problem as have independence across subjects.

# Data-Splitting Approach

- Split your subjects into two groups.

- Use half for significance and half for estimation of the effect size.
- Solves the bias problem as have independence across subjects.
- Issues: Less data to estimate so higher variance.

# Vul et al 2009



# Methods

# General Definitions

- $\mathcal{V}$: set of voxel locations
- Define an **image** to be a map $Z : \mathcal{V} \rightarrow \mathbb{R}$.
- Define a **local maxima** or **peak** of $Z$ to be a voxel $v \in \mathcal{V}$ such that the value that $Z$ takes at that location is larger than the value $Z$ takes at neighbouring voxels

# One-Sample Model

Suppose that we have $N$ subjects and for each $n = 1, \ldots, N$ a corresponding random image $Y_n$ on $\mathcal{V}$ such that for every voxel $v \in \mathcal{V}$,

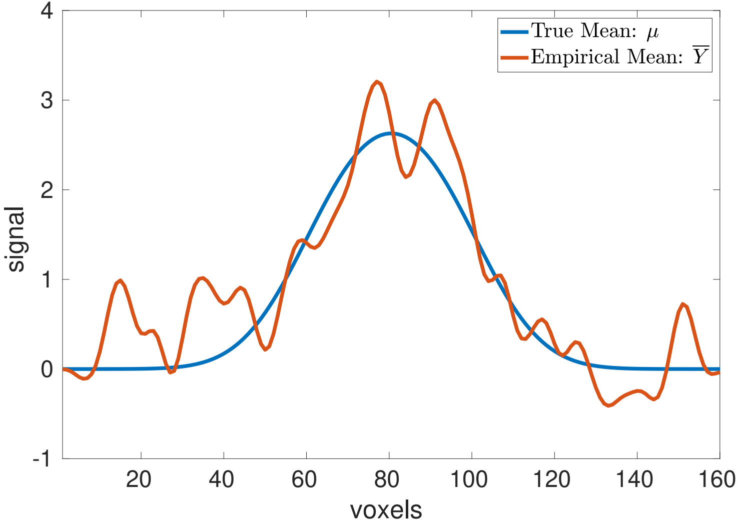
$$Y_n(v) = \mu(v) + \epsilon_n(v).$$

- $\mu(v)$ is the common mean intensity
- $\epsilon_1, \ldots, \epsilon_n$ are iid mean zero random images from some unknown multivariate distribution on $\mathcal{V}$
- Let $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} Y_n$
- let $\hat{v}_k$ be the location of the $k$th largest local maximum of $\hat{\mu}$

We want to know $\mu(\hat{v}_k)$, but we have $\hat{\mu}(\hat{v}_k)$.

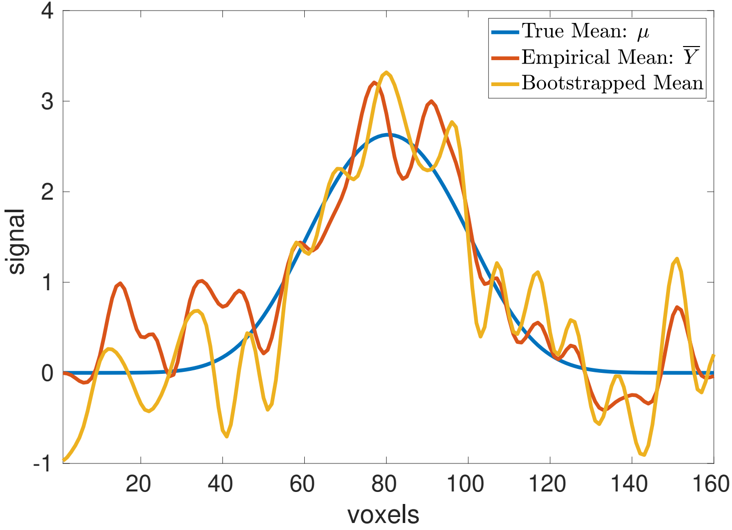
# 1D Example

20 subjects, $Y_n(t) = \mu(t) + \epsilon_n(t)$, $\hat{\mu} = \overline{Y} = \frac{1}{20}\sum_{n=1}^{20} Y_n$

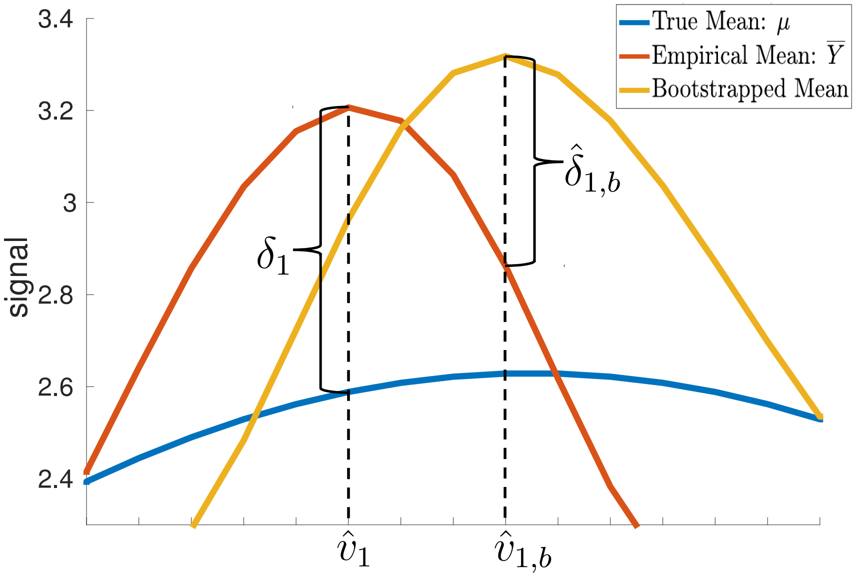


# 1D Example - Bootstrap Method

20 subjects, $Y_n(t) = \mu(t) + \epsilon_n(t)$, $\hat{\mu} = \overline{Y} = \frac{1}{20}\sum_{n=1}^{20} Y_n$



# 1D Example - Bootstrap Method



**Algorithm 1** Non-Parametric Bootstrap Bias Calculation

1: **Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and screening threshold $u$.

2: Let $\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} Y_n$ and let $K$ be the number of peaks of $\hat{\mu}$ above $u$, and for $k = 1, \dots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{\mu}$.

**Algorithm 2** Non-Parametric Bootstrap Bias Calculation

**Input**: Images $Y_1, \ldots, Y_N$, the number of bootstrap samples $B$ and screening threshold $u$.
Let $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} Y_n$ and let $K$ be the number of peaks of $\hat{\mu}$ above $u$, and for $k = 1, \ldots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{\mu}$.
**for** $b = 1, \ldots, B$ **do**
    Sample $Y^*_{1,b}, \ldots, Y^*_{N,b}$ independently with replacement from $Y_1, \ldots, Y_N$.
    Let $\hat{\mu}_b = \frac{1}{N}\sum_{n=1}^{N} Y^*_{N,b}$ and for $k = 1, \ldots, K$, let $\hat{v}_{k,b}$ be the location of the $k$th largest local maxima of $\hat{\mu}_b$.
    For $k = 1, \ldots, K$, let $\hat{\delta}_{k,b} = \hat{\mu}_b(\hat{v}_{k,b}) - \hat{\mu}(\hat{v}_{k,b})$ be an estimate of the bias at the $k$th largest local maxima.
**end for**

**Algorithm 3** Non-Parametric Bootstrap Bias Calculation

**Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and screening threshold $u$.

Let $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} Y_n$ and let $K$ be the number of peaks of $\hat{\mu}$ above $u$, and for $k = 1, \dots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{\mu}$.

**for** $b = 1, \dots, B$ **do**

  Sample $Y_{1,b}^*, \dots, Y_{N,b}^*$ independently with replacement from $Y_1, \dots, Y_N$.

  Let $\hat{\mu}_b = \frac{1}{N}\sum_{n=1}^{N} Y_{N,b}^*$ and for $k = 1, \dots, K$, let $\hat{v}_{k,b}$ be the location of the $k$th largest local maxima of $\hat{\mu}_b$.

  For $k = 1, \dots, K$, let $\hat{\delta}_{k,b} = \hat{\mu}_b(\hat{v}_{k,b}) - \hat{\mu}(\hat{v}_{k,b})$ be an estimate of the bias at the $k$th largest local maxima.

**end for**

For $k = 1, \dots, K$, let $\hat{\delta}_k = \frac{1}{B}\sum_{b=1}^{B} \hat{\delta}_{k,b}$.

**return** $(\hat{\mu}(\hat{v}_1) - \hat{\delta}_1, \dots, \hat{\mu}(\hat{v}_K) - \hat{\delta}_K)$.

# One-Sample t-statistics/Cohen's $d$

In neuroimaging we are interested in testing

$$H_0(v) : \mu(v) = 0 \text{ versus } H_1(v) : \mu(v) \neq 0$$

using the one-sample $t$-statistic:

$$t = \frac{\hat{\mu}\sqrt{N}}{\hat{\sigma}}$$

where

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} Y_n, \quad \hat{\sigma}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \hat{\mu})^2.$$

Effect size is measured via

$$\hat{d}(v) = \frac{\hat{\mu}}{\hat{\sigma}}$$

but this is a biased estimator for the population Cohen's $d$:

$$d(v) = \frac{\mu}{\sigma}.$$

# Unbiased Cohen's $d$ Estimation

This $t$-statistic $\hat{\mu}\sqrt{N}/\hat{\sigma}$ has a non-central $t$-distribution with non-centrality parameter $\mu\sqrt{N}/\sigma$ and $N-1$ degrees of freedom. Thus

$$\mathbb{E}\left[\frac{\hat{\mu}\sqrt{N}}{\hat{\sigma}}\right] = \frac{\mu}{\sigma}\sqrt{\frac{N-1}{2}}\frac{\Gamma((N-2)/2)}{\Gamma((N-1)/2)} = C_N\frac{\mu\sqrt{N}}{\sigma}$$

for $N > 2$, where $\Gamma$ is the gamma function and $C_N$ is a bias correction factor (?, ?). So we can use

$$\frac{\hat{\mu}}{\hat{\sigma}C_N}$$

as an unbiased of the population Cohen's $d$.

**Algorithm 4** Non-Parametric Bootstrap Bias Calculation

1: **Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and threshold $u$.

2: Let $K$ be the number of peaks of $t$ above $u$ and for $k = 1, \dots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{d} = \hat{\mu}/\hat{\sigma}$.

**Algorithm 5** Non-Parametric Bootstrap Bias Calculation

**Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and threshold $u$.

Let $K$ be the number of peaks of $t$ above $u$ and for $k = 1, \dots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{d} = \hat{\mu}/\hat{\sigma}$.

**for** $b = 1, \dots, B$ **do**

    Sample $Y^*_{1,b}, \dots, Y^*_{N,b}$ independently with replacement from $Y_1, \dots, Y_N$.

    Let $\hat{\mu}_b = \frac{1}{N}\sum_{n=1}^N Y^*_{n,b}$ and let $\hat{\sigma}^2_b(v) = \frac{1}{N-1}\sum_{n=1}^N (Y^*_{n,b}(v) - \hat{\mu}_b(v))^2$ for each $v \in \mathcal{V}$.

**Algorithm 6** Non-Parametric Bootstrap Bias Calculation

**Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and threshold $u$.
Let $K$ be the number of peaks of $t$ above $u$ and for $k = 1, \dots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{d} = \hat{\mu}/\hat{\sigma}$.
**for** $b = 1, \dots, B$ **do**
    Sample $Y^*_{1,b}, \dots, Y^*_{N,b}$ independently with replacement from $Y_1, \dots, Y_N$.
    Let $\hat{\mu}_b = \frac{1}{N}\sum_{n=1}^N Y^*_{n,b}$ and let $\hat{\sigma}^2_b(v) = \frac{1}{N-1}\sum_{n=1}^N (Y^*_{n,b}(v) - \hat{\mu}_b(v))^2$ for each $v \in \mathcal{V}$.
    For $k = 1, \dots, K$, let $\hat{v}_{k,b}$ be the location of the $k$th largest local maxima of $\hat{d}_b = \hat{\mu}_b/\hat{\sigma}_b$.
    Let $\hat{\delta}_{k,b} = (\hat{d}_b(\hat{v}_{k,b}) - \hat{d}(\hat{v}_{k,b}))/C_N$ be an estimate of the bias.
**end for**

**Algorithm 7** Non-Parametric Bootstrap Bias Calculation

1: **Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and threshold $u$.

2: Let $K$ be the number of peaks of $t$ above $u$ and for $k = 1, \dots, K$, let $\hat{v}_k$ be the location of the $k$th largest maxima of $\hat{d} = \hat{\mu}/\hat{\sigma}$.

3: **for** $b = 1, \dots, B$ **do**

4: Sample $Y^*_{1,b}, \dots, Y^*_{N,b}$ independently with replacement from $Y_1, \dots, Y_N$.

5: Let $\hat{\mu}_b = \frac{1}{N}\sum_{n=1}^N Y^*_{n,b}$ and let $\hat{\sigma}^2_b(v) = \frac{1}{N-1}\sum_{n=1}^N (Y^*_{n,b}(v) - \hat{\mu}_b(v))^2$ for each $v \in \mathcal{V}$.

6: For $k = 1, \dots, K$, let $\hat{v}_{k,b}$ be the location of the $k$th largest local maxima of $\hat{d}_b = \hat{\mu}_b/\hat{\sigma}_b$.

7: Let $\hat{\delta}_{k,b} = (\hat{d}_b(\hat{v}_{k,b}) - \hat{d}(\hat{v}_{k,b}))/C_N$ be an estimate of the bias.

8: **end for**

9: For $k = 1, \dots, K$, let $\hat{\delta}_k = \frac{1}{B}\sum_{b=1}^B \hat{\delta}_{k,b}$

10: **return** $(\hat{d}(\hat{v}_1)/C_N - \hat{\delta}_1, \dots, \hat{d}(\hat{v}_K)/C_N - \hat{\delta}_K)$.

# Estimation of the mean

To infer on $\mu$ instead of $\mu/\sigma$ can just use

$$\hat{\delta}_{k,b} = \hat{\mu}_b(\hat{v}_{k,b}) - \hat{\mu}(\hat{v}_{k,b})$$

# Circular Inference and Data-Splitting

- Circular inference estimates are: $\hat{d}(\hat{v}_1)/C_N, \dots, \hat{d}(\hat{v}_K)/C_N$.
- For data-splitting, we first divide the images into two groups: $Y_1, \dots, Y_{N/2}$ and $Y_{N/2+1}, \dots, Y_N$. Then find the peaks using the first half of the subjects and estimate the values at those peaks using the second half of the subjects.

# GLM

Let $Y$ be an $N$-dimensional random image such that for each $v \in \mathcal{V}$

$$Y(v) = X\beta(v) + \epsilon(v)$$

- $N \times p$ design matrix $X$
- parameter vector $\beta(v) \in \mathbb{R}^p$
- $\epsilon(v) = (\epsilon_1(v), \ldots, \epsilon_N(v))^T$ is the random image of the noise

# GLM

Let $Y$ be an $N$-dimensional random image such that for each $v \in \mathcal{V}$

$$Y(v) = X\beta(v) + \epsilon(v)$$

- $N \times p$ design matrix $X$
- parameter vector $\beta(v) \in \mathbb{R}^p$
- $\epsilon(v) = (\epsilon_1(v), \ldots, \epsilon_N(v))^T$ is the random image of the noise

We are interested in testing

$$H_0(v) : C\beta(v) = 0 \text{ versus } H_1(v) : C\beta(v) \neq 0$$

for some contrast matrix $C \in \mathbb{R}^{m \times p}$. We can test this at each voxel with the usual $F$-test,

$$F(v) = \frac{(C\hat{\beta}(v))^T (C(X^T X)^{-1} C^T)^{-1} (C\hat{\beta}(v))/m}{\hat{\sigma}(v)^2} \quad (1)$$

where $\hat{\beta}(v) = (X^T X)^{-1} X^T Y$ and $\hat{\sigma}^2(v)$ is the error variance. Under the alternative has a non-central $F$-distribution.

**Algorithm 8** Non-Parametric Bootstrap Bias Calculation

1: **Input**: Images $Y_1, \dots, Y_N$, the number of bootstrap samples $B$ and threshold $u$.

2: Let $\hat{\beta} = \hat{\beta}(X, Y) = (X^T X)^{-1} X^T Y$ and let $\hat{\epsilon} = Y - X\hat{\beta}$ be the residuals.

3: For each $n = 1, \dots, N$, let $r_n = \hat{\epsilon}_n / \sqrt{1 - p_n}$ be the modified residuals, where $p_n = (X(X^T X)^{-1} X^T)_{nn}$. Let $\overline{r} = \frac{1}{N} \sum_{n=1}^{N} r_i$ be their mean.

4: **for** $b = 1, \dots, B$ **do**

5: Sample $\epsilon^*_{1,b}, \dots, \epsilon^*_{N,b}$ independently with replacement from $r_1 - \overline{r}, \dots, r_N - \overline{r}$ and let $\epsilon^*_b = (\epsilon^*_{1,b}, \dots, \epsilon^*_{N,b})^T$ and set $Y^*_b = X\hat{\beta} + \epsilon^*$.

6: Let $F^*_b$ be the bootstrapped $F$-statistic image computed using $Y^*_b$. Let $R^2_b$ be the bootstrapped partial $R^2$ image and set $\hat{\delta}_{k,b} = R^2_b(\hat{v}_{k,b}) - R^2(\hat{v}_{k,b})$ to be the estimate of the bias.

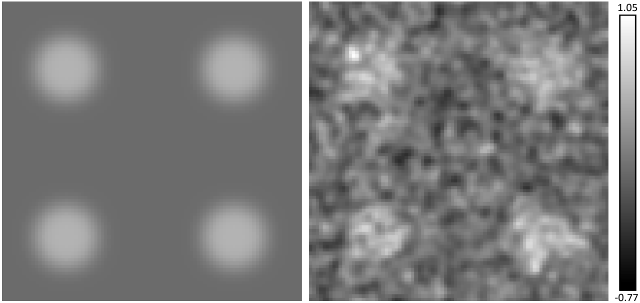
7: **end for**

8: For $k = 1, \dots, K$, let $\hat{\delta}_k = \frac{1}{B} \sum_{b=1}^{B} \hat{\delta}_{k,b}$.

9: **return** $(R^2(\hat{v}_1) - \hat{\delta}_1, \dots, R^2(\hat{v}_K) - \hat{\delta}_K)$.

# Simulations - Cohen's $d$

All simulations generated using code from the RFTtoolbox `https://github.com/BrainStatsSam/RFTtoolbox` (avoiding edge problems)




(a) True Signal (b) Sample Cohen's $d$

- Panel (a) illustrates a slice through the true signal (actually 9 peaks only 4 shown).
- Panel (b) illustrates the same slice through the one sample Cohen's $d$ for 50 subjects. Noise: Gaussian random field with FWHM 6.

# Bias, RMSE and standard deviation

Traditionally, one estimates a common $\theta$ with estimators $\hat{\theta}_1, ..., \hat{\theta}_K$ however we have estimators $\hat{\theta}_1, \dots, \hat{\theta}_K$ of parameters $\theta_1, \dots, \theta_K$ where $K$ is the number of significant peaks that are found over all realizations. As such we instead define

$$\tilde{\theta}_k = \hat{\theta}_k - \theta_k$$

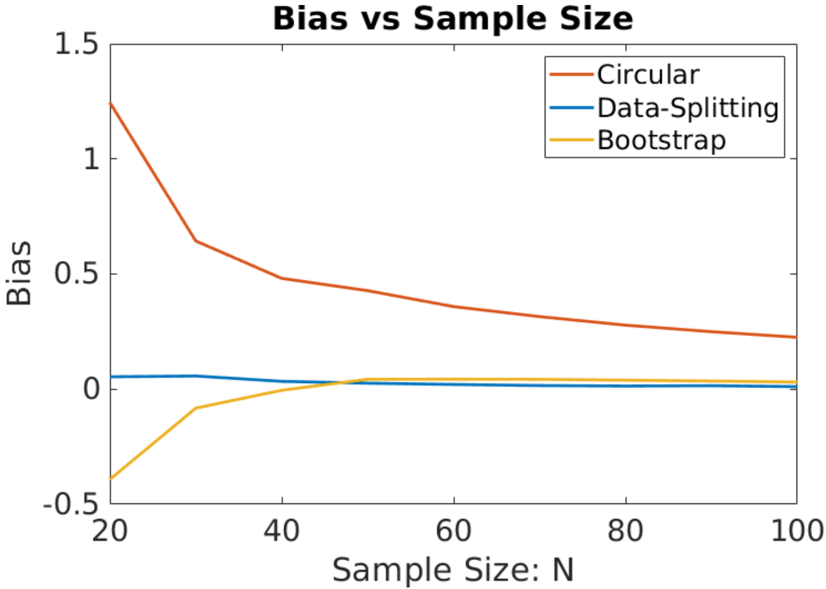
and use the fact that the noise-free value of $\tilde{\theta}_k$ is 0 for each $k$.

$$\begin{aligned}\text{MSE} &= \frac{1}{K}\sum_{k=1}^{K}(\tilde{\theta}_k - 0)^2 \\ &= \frac{1}{K}\sum_{k=1}^{K}(\tilde{\theta}_k - \frac{1}{K}\sum_{k=1}^{K}\tilde{\theta}_k)^2 + \left(\frac{1}{K}\sum_{k=1}^{K}\tilde{\theta}_k\right)^2\end{aligned}$$

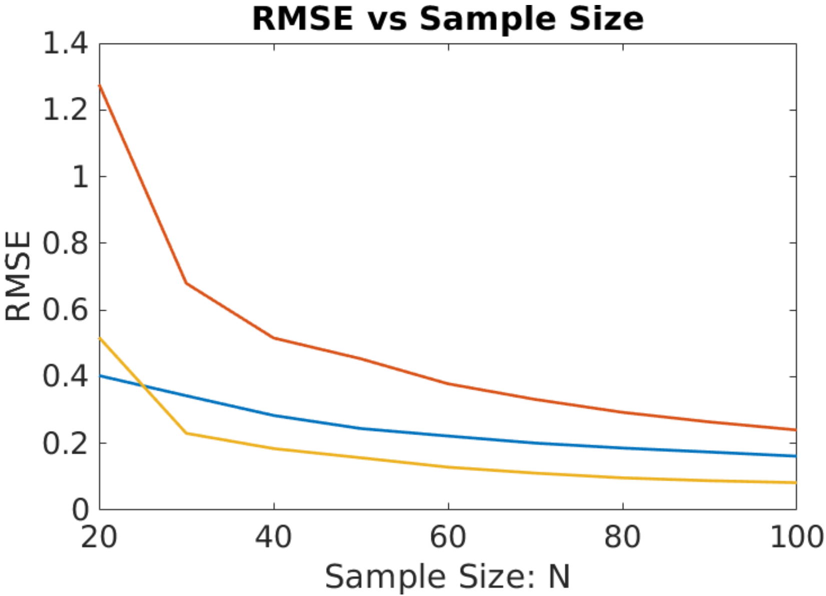
# Simulation Evaluation and Thresholding

- We evaluate our methods for $N = \{20, 30, \ldots, 100\}$.
- For each $N$ we generate 1,000 realizations and compare the performance of the three methods across realizations.

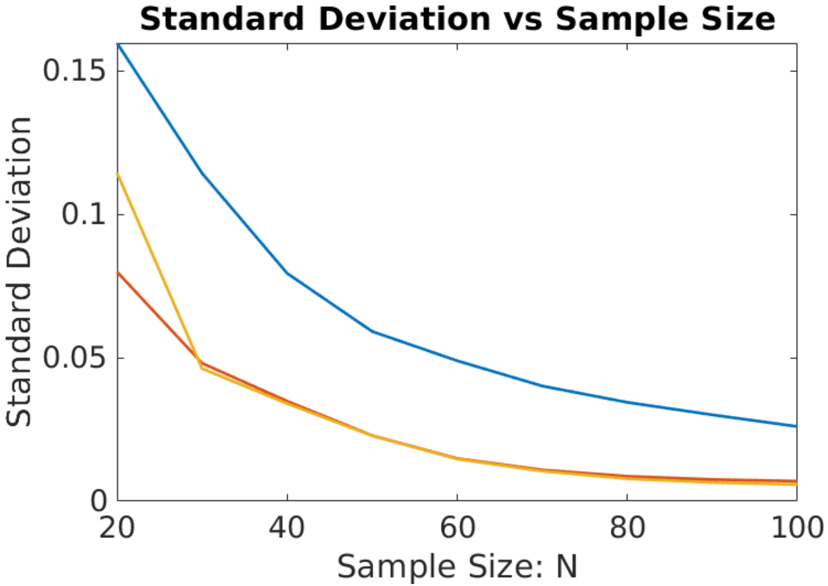
# Results - One Sample Cohen's $d$ simulations Bias



# Results - One Sample Cohen's $d$ simulations RMSE



# Results - One Sample Cohen's $d$ simulations STD



# Big Data Validation

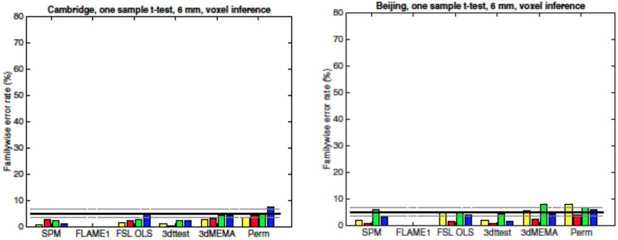
# Data and Testing

- 8940 total subjects from the UK biobank. We have task fMRI and VBM data from all subjects
- We test the one-sample methods using the task fMRI data and the GLM methods using the VBM data (as the $R^2$ effect sizes are very small for the task fMRI data sets)
- For the task-fMRI data we estimate Cohen's $d$ or $\mu$.
- For the VBM data we regress against age, sex and an intercept and compute the partial $R^2$ for age.
- Set aside 4000 subjects to compute a ground truth and divide the rest into $G_N = 4940/N$ groups of size $N = 20, 50, 100$.
- Actually for the VBM data we take $N = 50, 100, 150$ as the effect size is lower

We recommend this type of testing framework for all statistical methods.

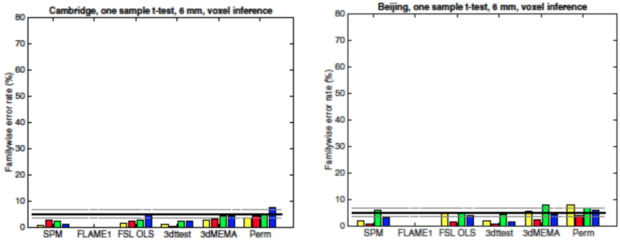
# Thresholding

- We threshold using voxelwise RFT. This doesn't have the same problems as clusterwise inference as it doesn't make the same assumptions.



# Thresholding

- We threshold using voxelwise RFT. This doesn't have the same problems as clusterwise inference as it doesn't make the same assumptions.




- Our method independent of the threshold.
- For the big data analysis we do permutation is very computational so is not practical.
- But permutation testing can be used to compute the voxelwise threshold when doing a general analysis.

# Cohen's $d$ ground truth

Computing the ground truth is difficult due to memory constraints. So you have load images sequentially. Let $\mathcal{D}$ be the set of all possible voxels. Typically $\mathcal{D}$ is a $91 \times 109 \times 91$ grid. Define

$$M_n(v) = \begin{cases} 1 & \text{if subject } n \text{ has data at } v \\ 0 & \text{otherwise} \end{cases}$$

# Cohen's $d$ ground truth

Computing the ground truth is difficult due to memory constraints. So you have load images sequentially. Let $\mathcal{D}$ be the set of all possible voxels. Typically $\mathcal{D}$ is a $91 \times 109 \times 91$ grid. Define

$$M_n(v) = \begin{cases} 1 & \text{if subject } n \text{ has data at } v \\ 0 & \text{otherwise} \end{cases}$$

Take $\mathcal{S} \subset \{1, \dots, 8940\}$ of size 4000 and let

$$\mu(v) = \frac{\sum_{n \in \mathcal{S}} Y_n(v) M_n(v)}{\sum_{n \in \mathcal{S}} M_n(v)} \times 1\!\!1(M_n(v) = 1 \text{ for at least } 100\ n \in \mathcal{S})$$

$$\sigma^2(v) = \frac{\sum_{n \in \mathcal{S}} (Y_n - \mu(v))^2 M_n(v)}{\sum_{n \in \mathcal{S}} M_n(v) - 1} \times 1\!\!1(M_n(v) = 1 \text{ for at least } 100\ n \in \mathcal{S}),$$

# Cohen's $d$ ground truth

$$\mu(v) = \frac{\sum_{n \in \mathcal{S}} Y_n(v) M_n(v)}{\sum_{n \in \mathcal{S}} M_n(v)} \times \mathbb{1}(M_n(v) = 1 \text{ for at least } 100 \; n \in \mathcal{S})$$

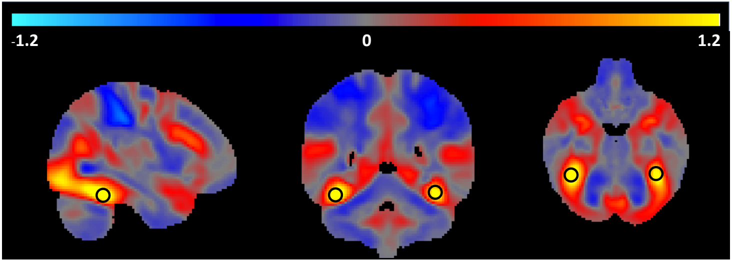
$$\sigma^2(v) = \frac{\sum_{n \in \mathcal{S}} (Y_n - \mu(v))^2 M_n(v)}{\sum_{n \in \mathcal{S}} M_n(v) - 1} \times \mathbb{1}(M_n(v) = 1 \text{ for at least } 100 \; n \in \mathcal{S}),$$

and the **ground truth Cohen's** $d$ estimate as

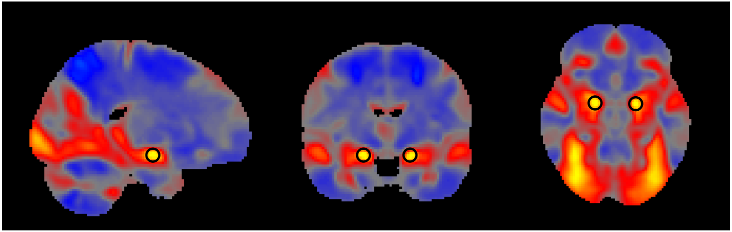
$$d(v) = \frac{\mu(v)}{\sigma(v)}.$$

Finally each of these are additionally masked with the 2mm MNI brain mask.

# Cohen's $d$ Ground Truth Slices




(a) Top 2 peaks



(b) 3rd and 4th Highest Peaks

# Illustrating the Winner's Curse

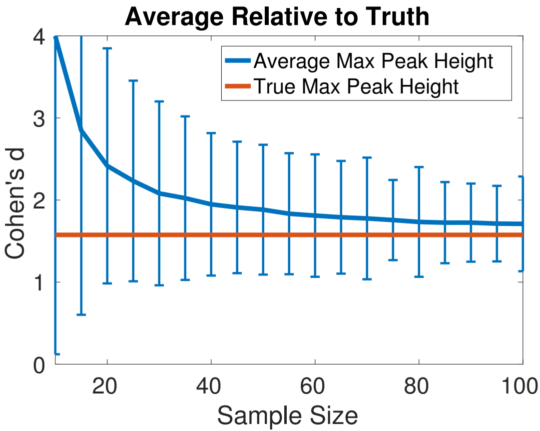



Figure 4: Comparing the maximum values at small sample Cohen's $d$ (over the $G_N$ groups) to the max ground truth value.

# GLM ground truth

For now assume that no data is missing and that we have

- $N_{\text{all}} = 4000$ subjects
- an $N_{\text{all}} \times p$ design matrix $X = (x_1, \dots, x_{N_{\text{all}}})^T$
- $V$ is the number of voxels in each subject image $Y_n$
- $Y$ be the $N_{\text{all}} \times V$ matrix of all the subject images

# GLM ground truth

For now assume that no data is missing and that we have

- $N_{\text{all}} = 4000$ subjects
- an $N_{\text{all}} \times p$ design matrix $X = (x_1, \dots, x_{N_{\text{all}}})^T$
- $V$ is the number of voxels in each subject image $Y_n$
- $Y$ be the $N_{\text{all}} \times V$ matrix of all the subject images

For $Y = X\beta + \epsilon$, we want to compute

$$\hat{\beta} = (X^T X)^{-1} X^T Y,$$

at each voxel. For each $v \in \mathcal{V}$,

$$X^T Y(v) = \left(x_1, \dots, x_{N_{\text{all}}}\right) \begin{pmatrix} Y_1(v) \\ \vdots \\ Y_{N_{\text{all}}}(v) \end{pmatrix} = \sum_{n=1}^{N_{all}} Y_n(v) x_n,$$

$$\hat{\sigma}^2 = (N_{\text{all}} - p)^{-1} \sum_{n=1}^{N_{all}} (Y_n - x_n^T \hat{\beta})^2.$$

and this allows $F$ and $R^2$ to be calculated

# GLM ground truth

For each $v \in \mathcal{V}$,

$$X^T Y(v) = \left(x_1, \dots, x_{N_{\text{all}}}\right) \begin{pmatrix} Y_1(v) \\ \vdots \\ Y_{N_{\text{all}}}(v) \end{pmatrix} = \sum_{n=1}^{N_{all}} Y_n(v) x_n,$$

Can compute $\hat{\beta} = (X^T X)^{-1} X^T Y$ from this and estimate

$$\hat{\sigma}^2 = (N_{\text{all}} - p)^{-1} \sum_{n=1}^{N_{all}} (Y_n - x_n^T \hat{\beta})^2.$$

and this allows $F$ and $R^2$ to be calculated.

# GLM ground truth with missingness

Let $C(v) := \{n : M_n(v) = 1\}$. Then for each voxel $v$ we need to compute the complete case estimate

$$\hat{\beta}(v) = (X_{C(v)}^T X_{C(v)})^{-1} X_{C(v)}^T Y_{C(v)}.$$

The first and second parts of this expression can be computed as

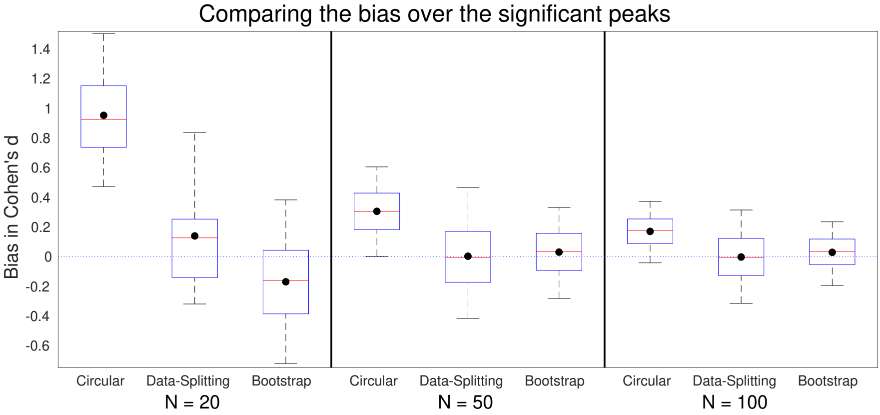
$$(X_{C(v)}^T X_{C(v)})^{-1} = \left( \sum_{n=1}^{N_{\text{all}}} M_n(v) x_n x_n^T \right)^{-1}$$

and

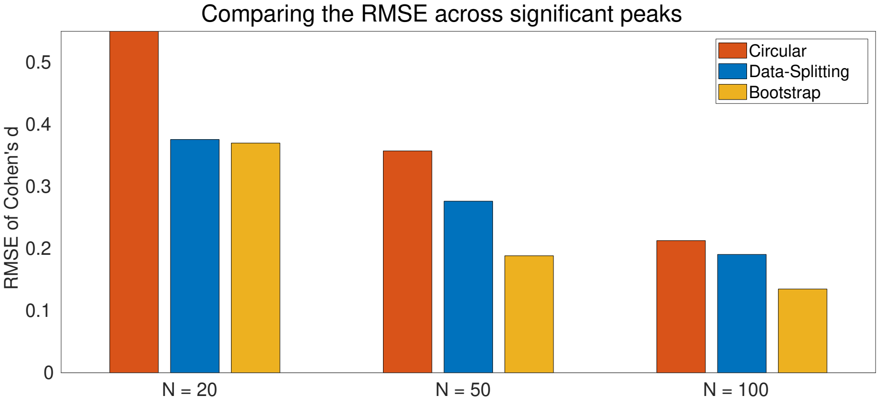
$$X_{C(v)}^T Y_{C(v)} = \sum_{n=1}^{N_{\text{all}}} M_n(v) Y_n(v) x_n$$

$\hat{\sigma}^2, F$ and $R^2$ can similarly be computed.

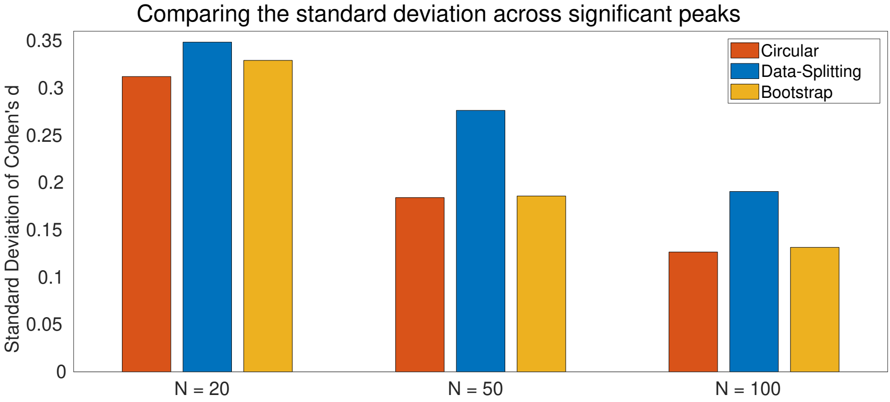
# One Sample Cohen's $d$ - Bias



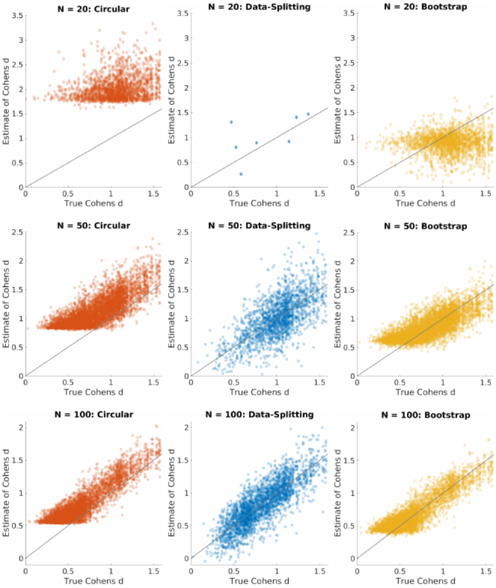
# One Sample Cohen's $d$ - RMSE



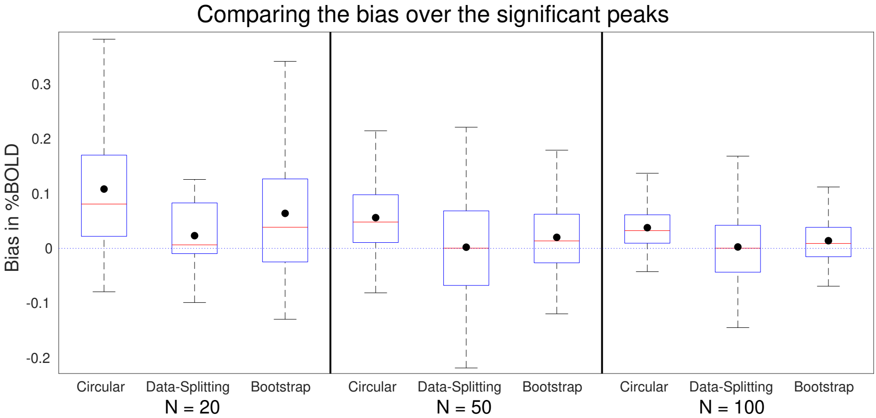
# One Sample Cohen's $d$ - Standard Deviation



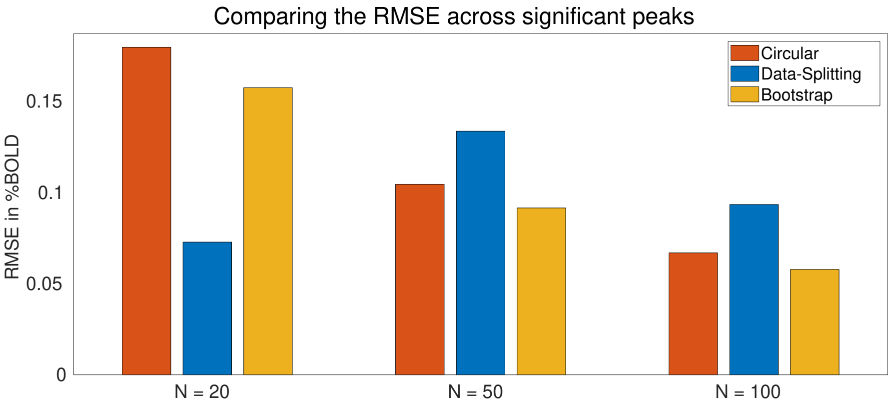
# One Sample Cohen's $d$ - Estimates vs Ground truth



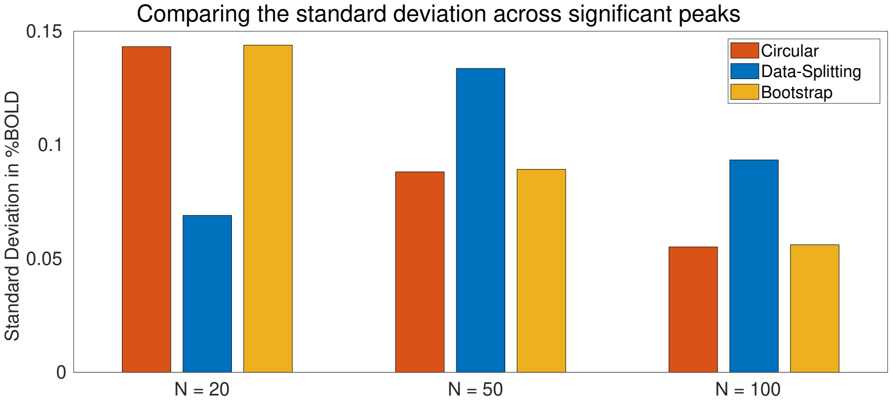
# Mean estimation - Bias



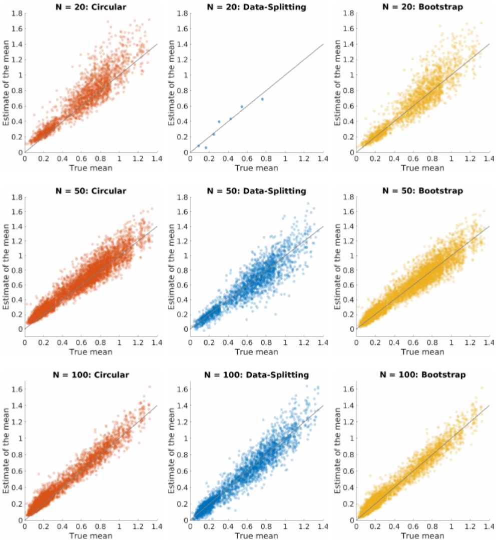
# Mean estimation - RMSE



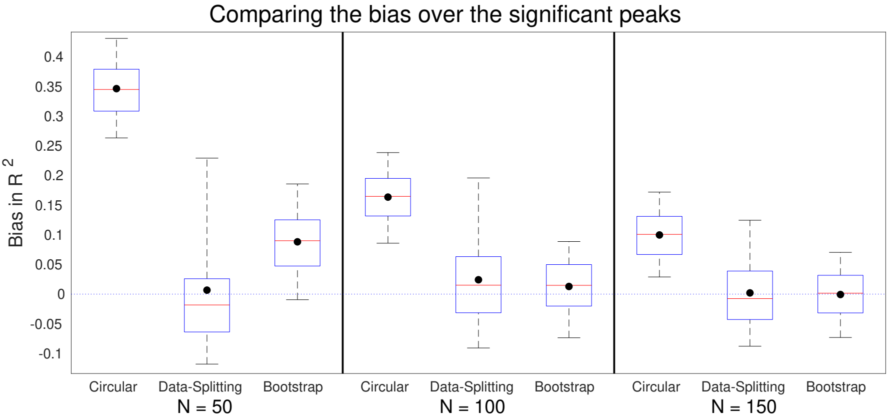
# Mean estimation - Standard Deviation



# Mean estimation - Estimates versus Ground truth

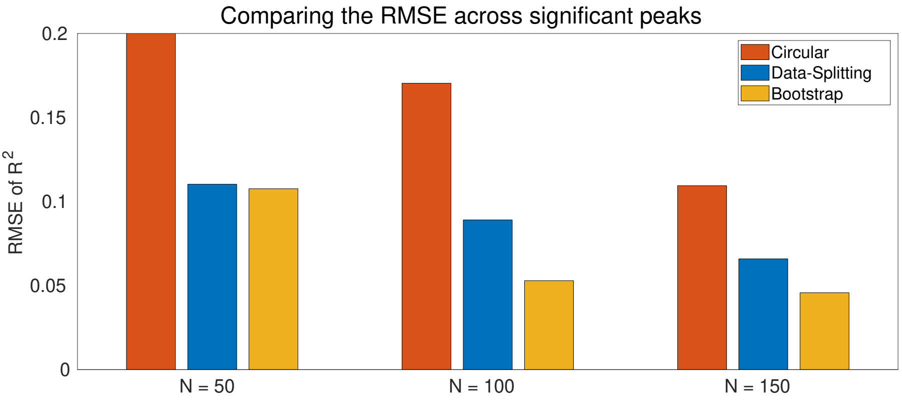


# $R^2$ - Bias

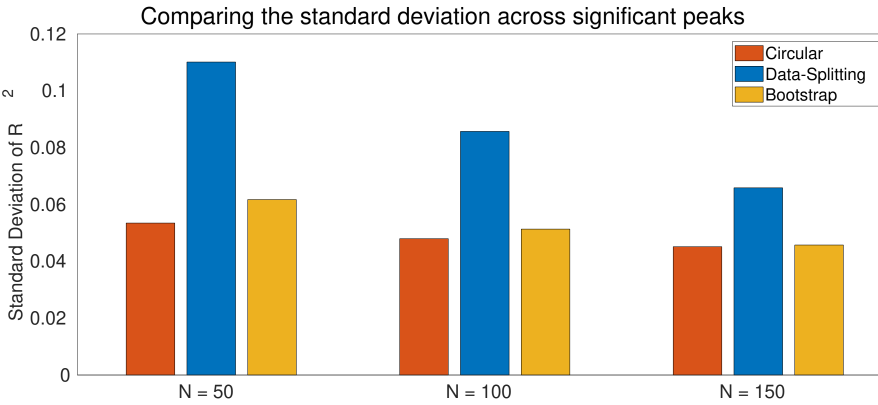


Comparing the bias over the significant peaks
Bias in R 2
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
Circular
Data-Splitting
Bootstrap
N = 50
Circular
Data-Splitting
Bootstrap
N = 100
Circular
Data-Splitting
Bootstrap
N = 150

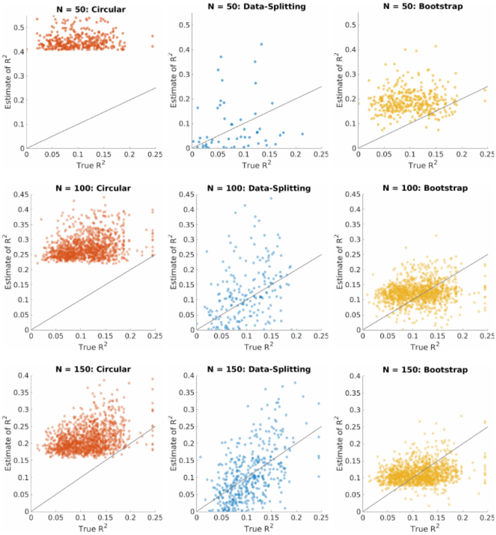
# $R^2$ - RMSE


Comparing the RMSE across significant peaks
RMSE of R 2
0.2
0.15
0.1
0.05
0
N = 50
N = 100
N = 150
Circular
Data-Splitting
Bootstrap

# $R^2$ - Standard Deviation



# $R^2$ - Estimates versus Ground truth



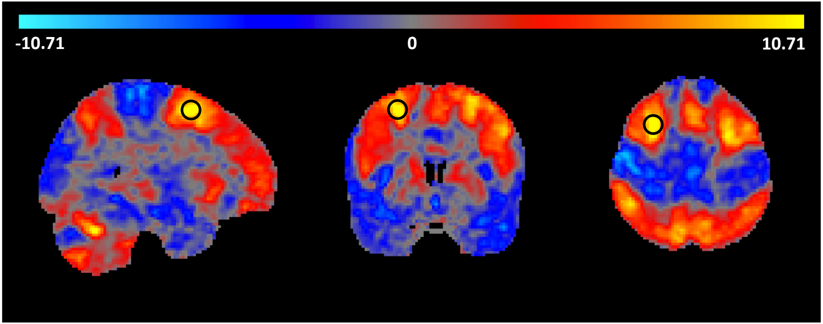
# One-Sample t-statistic power

Given a potential future sample size $N'$ and an estimate of Csohen's $d : \hat{d}$, the power is:

$$\mathbb{P}(T_{N'-1,\hat{d}} > t_{1-\alpha,N'-1})$$

where $t_{1-\alpha,N'-1}$ is chosen such that $\mathbb{P}(T_{N'-1,0} > t_{1-\alpha,N'-1}) = \alpha$ and $T_{N'-1,\lambda}$ has a non-central $T$ distribution with $N' - 1$ degrees of freedom.

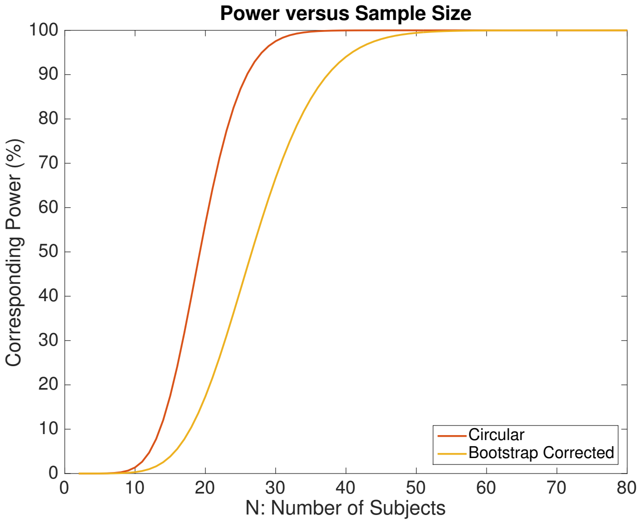
# Working Memory Example




- One-sample $t$-statistic for 80 subjects from the HCP.
- Activation in the Medial Frontal Gyrus.
- At the maximum the observed (circular) Cohen's $d$ is 1.519, while the bootstrap-corrected value is 1.161
- The observed %BOLD change there is %0.450 and corrected estimate is %0.433.

# Power graph

At the maximum the observed (circular) Cohen's $d$ is 1.519, while the bootstrap-corrected value is 1.161. So we can generate a power graph:



# Conclusion and Future Work

- We provide a method for dealing with the winner's curse which outperforms existing methods in terms of RMSE.
- Can also be used to estimate the maximum rather than the maximum at a given location.
- So far have mainly considered voxelwise inference but it would be interesting to extend this to other types of inference but
- Would be cool to develop an RFT method to do this but this is probably quite difficult!
- Worth noting once again that it's important (especially in light of clusterfailure) that existing and emerging statistical methods are tested using this type of big data validation.

# Software Availability

- Paper available online.
- Code and scripts to reproduce figures available in SIbootstrap toolbox (github.com/sjdavenport/SIbootstrap). Simulations and thresholding were performed using RFTtoolbox available at github.com/sjdavenport/RFTtoolbox/.
- Slides available on my website.

# Bibliography