TRANSFORMING HEAVY TAILED DATA IMPROVES THE POWER OF INFERENCE

SAMUEL DAVENPORT – UNIVERSITY OF CALIFORNIA SAN DIEGO

JOINT WORK WITH THOMAS MAULLIN-SAPEY, THOMAS E. NICHOLS AND ARMIN SCHWARTZMAN

EXISTING TRANSFORMATION APPROACHS

- Data transformations are rather common however are typically designed to maximize Gaussianity.

An Analysis of Transformations

THE usual techniques for the analysis of linear models as exemplified by the analysis of variance and by multiple regression analysis are usually justified by assuming

(i) simplicity of structure for $E(y)$;
(ii) constancy of error variance;
(iii) normality of distributions;
(iv) independence of observations.

EXISTING TRANSFORMATION APPROACHS

- Data transformations are rather common however are typically designed to maximize Gaussianity.

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & (\lambda \neq 0), \\ \log y & (\lambda = 0), \end{cases}$$

$$\frac{1}{(2\pi)^{\frac{1}{2}n} \sigma^{n}} \exp\left\{ - \frac{(\mathbf{y}^{(\lambda)} - \mathbf{a}\boldsymbol{\theta})' (\mathbf{y}^{(\lambda)} - \mathbf{a}\boldsymbol{\theta})}{2\sigma^{2}} \right\} J(\lambda; \mathbf{y}),$$

- However its not clear why transforming to maximize Gaussianity is a good idea, especially in the context of testing as the test-statistic is Gaussian in the limit.
- Instead it may make more sense to transform to try to improve power.

MODEL

Consider the following signal plus noise model with a mean $\mu : \mathcal{V} \to \mathbb{R}$, for some finite set of indices $\mathcal{V}$, and random observations

$$Y_i(v) = \mu(v) + \epsilon_i(v), \text{ for all } v \in \mathcal{V}.$$

Here $\epsilon_i : \mathcal{V} \to \mathbb{R}$, for $1 \leq i \leq n$, are i.i.d. mean-zero random images (or vectors indexed by $\mathcal{V}$), for some number of observations $n \in \mathbb{N}$. Let $\sigma^2(v) := \text{var}(\epsilon_1(v))$ and assume that $\min_{v \in \mathcal{V}} \sigma^2(v) > 0$.

Remark 2.1. *In our applications, $\mathcal{V}$ will correspond to the set of vertices/voxels for the 2D surface and 3D volume data respectively. However the framework is fully general. In particular, given $m \in \mathbb{N}$, taking $\mathcal{V} = \{1, 2, \ldots, m\}$ we recover vectors in $\mathbb{R}^m$ and taking $\mathcal{V} = \{1\}$ we recover random variables in $\mathbb{R}$.*

NULL HYPOTHESES TO TEST

at each $v \in \mathcal{V}$, we shall be interested in testing the point-null:

$$H_0^P(v) : \mu(v) = 0 \quad \text{v.s.} \quad H_1^P(v) : \mu(v) \neq 0.$$

null set $\mathcal{N} \subseteq \mathcal{V}$ $\quad \mathcal{N} = \{v \in \mathcal{V} : \mu(v) = 0\}$

We shall also be interested in testing the directional null

$$H_0^D(v) : \mu(v) \leq 0 \quad \text{v.s.} \quad H_1^D(v) : \mu(v) > 0$$

$H_0^D(v)$ is true on $\mathcal{D} = \{v \in \mathcal{V} : \mu(v) \leq 0\}$

TEST-STATISTICS

$$T_n(v) = \frac{\sqrt{n}\hat{\mu}_n(v)}{\hat{\sigma}_n(v)}$$

for each $v \in \mathcal{V}$

where $\hat{\mu}_n(v) = \sum_{i=1}^{n} Y_i(v)$

$$\hat{\sigma}_n(v) = \left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i(v) - \sum_{i=1}^{n} Y_i(v))^2\right)$$

This test statistic is widely used despite not being necessarily being optimal

TRANSFORMED TEST-STATISTICS

$$T_n^*(v) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n f_v(Y_i(v))}{\left(\frac{1}{n-1} \sum_{j=1}^n (f_v(Y_j(v)) - \sum_{i=1}^n f_v(Y_i(v))^2 \right)^{1/2}}$$

for each $v \in \mathcal{V}$, where $f_v : \mathbb{R} \to \mathbb{R}$

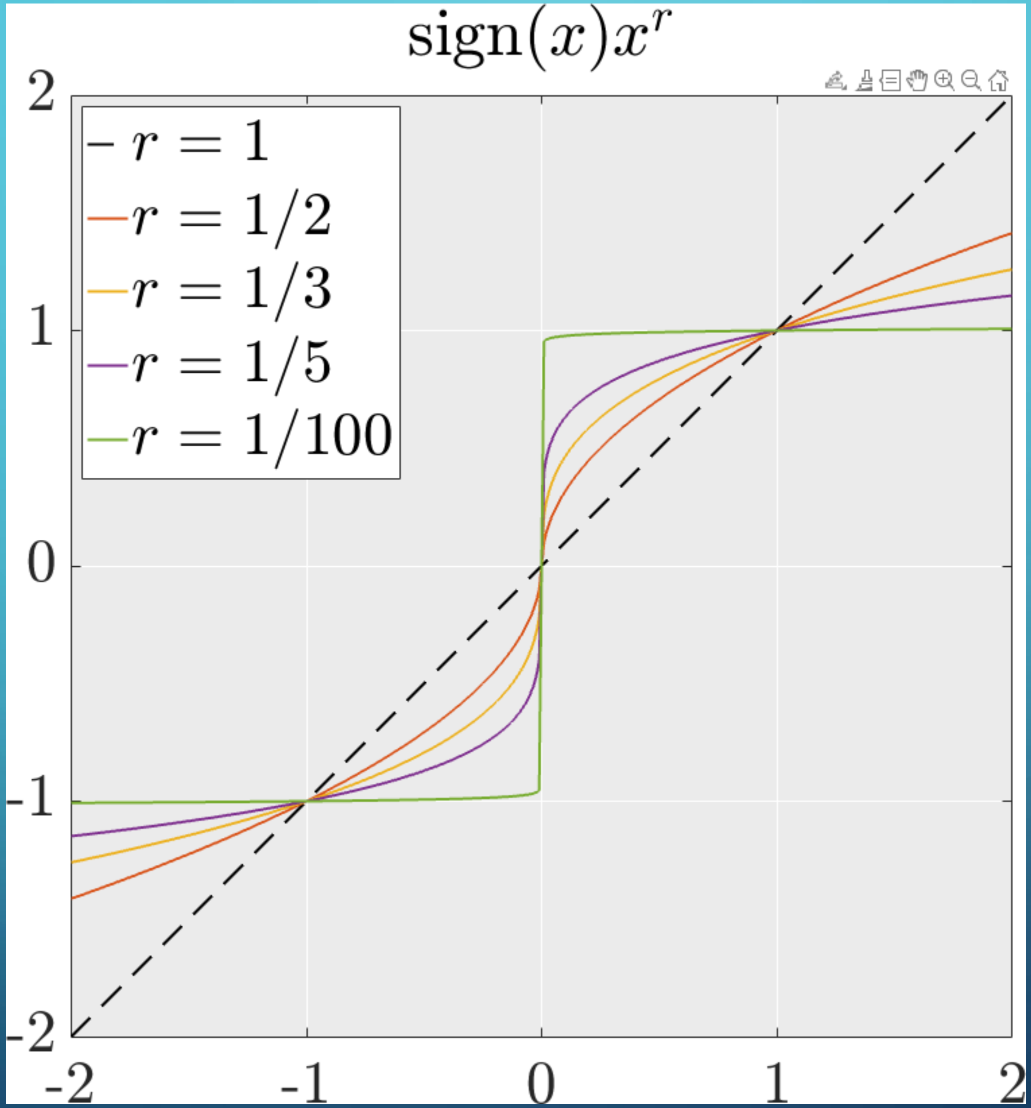
NATURAL FAMILY OF TRANSFORMATIONS

- Natural to choose the same function at each v.

$$f(x) = \begin{cases} x^r & \text{if } x \geq 0 \\ -(-x)^r & \text{if } x < 0 \end{cases}, \text{ for } r > 0,$$

- This family is the signed power family has the special property that it makes the test-statistics robust to the pointwise variance of the data
- Other transformation families such as those basd on the arcsinh family etc can be used but do not have this property.

ANTISYMMETRIC TRANSFORMATIONS



INVARIANCE UNDER SIGN-FLIPPING

We will develop theory for testing using sign-flipping. To do so we shall requrie the following assumption on the errors

Assumption 1. (Null invariance under sign-flipping) For all $g_1, \ldots, g_n \in \{\pm 1\}$,

$$\{g_1\epsilon_1(v), \ldots, g_n\epsilon_n(v)\}_{v\in\mathcal{N}} =_d \{\epsilon_1(v), \ldots, \epsilon_n(v)\}_{v\in\mathcal{N}}.$$

TRANSFORMING PRESERVES THE NULL

Definition 2.2. Given $f : \mathbb{R} \to \mathbb{R}$ we say that f is antisymmetric if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$.

Theorem 2.4. *Suppose that $\epsilon_1, \ldots, \epsilon_n$ each have a symmetric distribution and let f be an antisymmetric function. Then for each $v \in \mathcal{V}$ we have $\mathbb{E}[f(\epsilon_i(v))] = 0$ for all $1 \le i \le n$.*

Proof. To establish the first result note that the fact that f is antisymmetric and $\epsilon_1(v)$ is symmetric means that the problem is completely symmetric. In particular,

$$\mathbb{E}(f(\epsilon_1(v))) = \mathbb{E}(f(-\epsilon_1(v))) = -\mathbb{E}(f(\epsilon_1(v)))$$

So $\mu(v) = 0$ implies $\mathbb{E}[f(Y_i(v))] = 0$

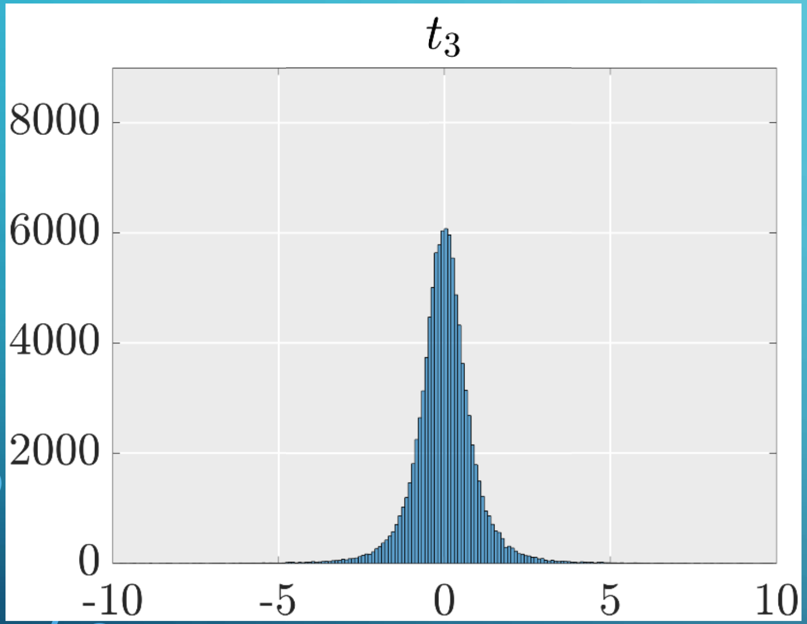
PRESERVING THE DIRECTIONAL NULL

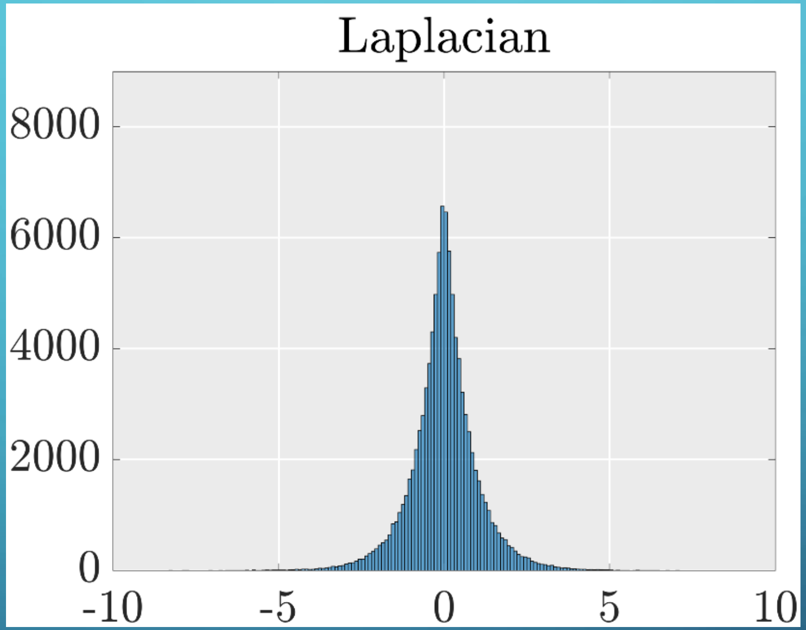
Theorem 2.5. *Suppose that* $\epsilon_1, \ldots, \epsilon_n$ *each have a symmetric distribution and let* f *be an increasing antisymmetric function, then* $\mathbb{E}[f(Y_i(v))] \leq 0$ *for all* $v \in \mathcal{V}$ *such that* $\mu(v) \leq 0$ *and all* $1 \leq i \leq n$.

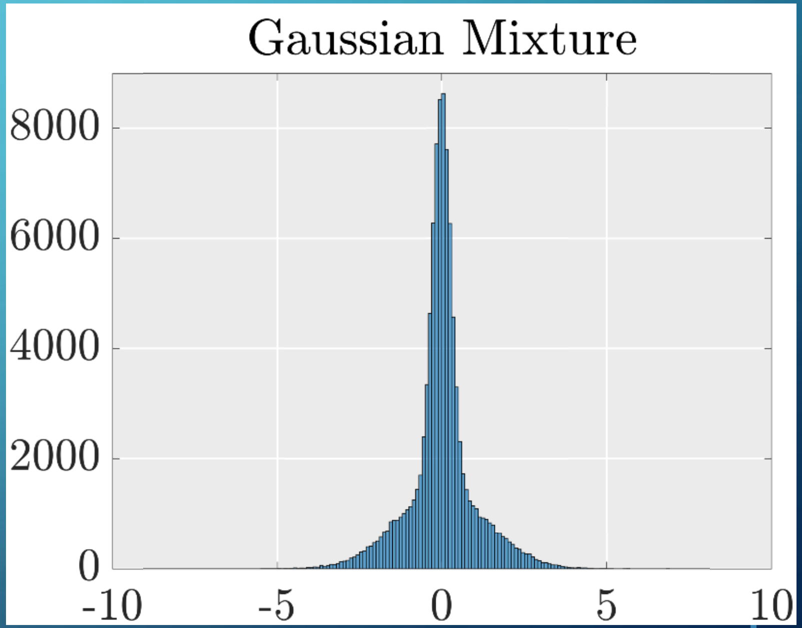
Proof. Since f is increasing, applying Theorem 2.4, it follows that $\mu(v) \leq 0$ implies

$$\mathbb{E}[f(Y_i(v))] = \mathbb{E}[f(\mu(v) + \epsilon_i(v))] \leq \mathbb{E}[f(\epsilon_i(v))] = 0.$$

SIMULATING DATA FROM HEAVY TAILED DISTRIBUTIONS



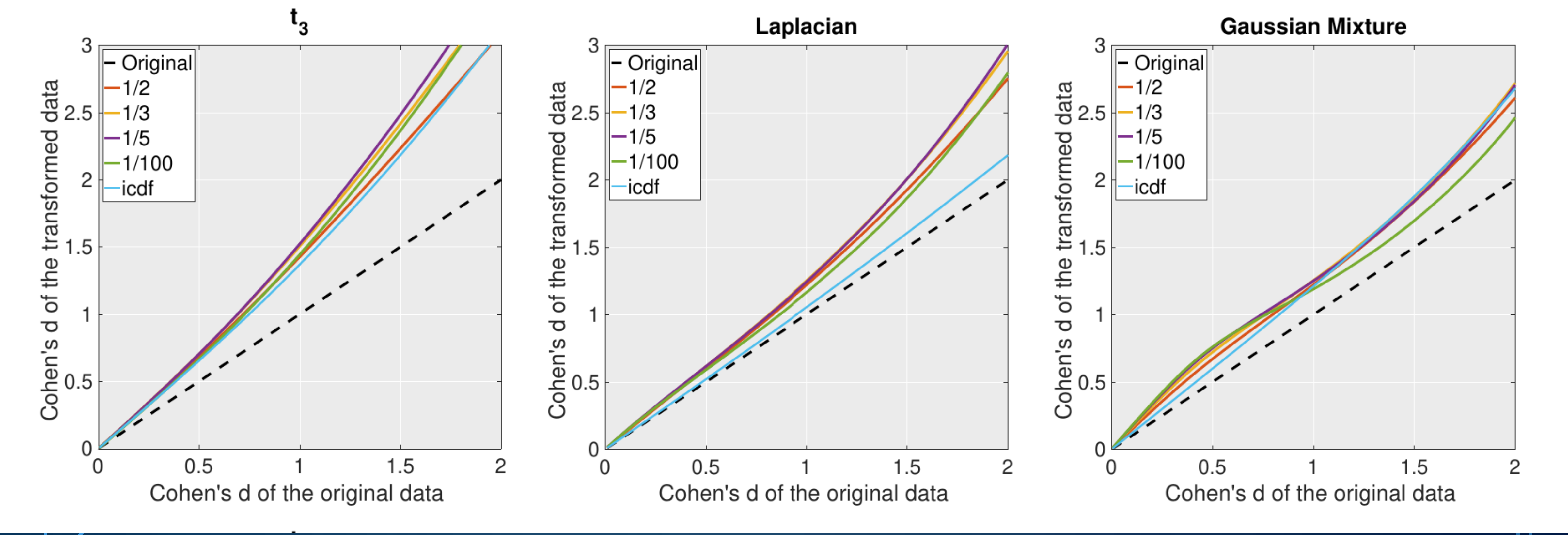




TRANSFORMATIONS CAN BE USED TO INCREASE COHEN'S D

$$\frac{\mathbb{E}[f(Y)]}{\mathrm{std}(f(Y))} \quad \text{vs} \quad \frac{\mathbb{E}[Y]}{\mathrm{std}(Y)}$$

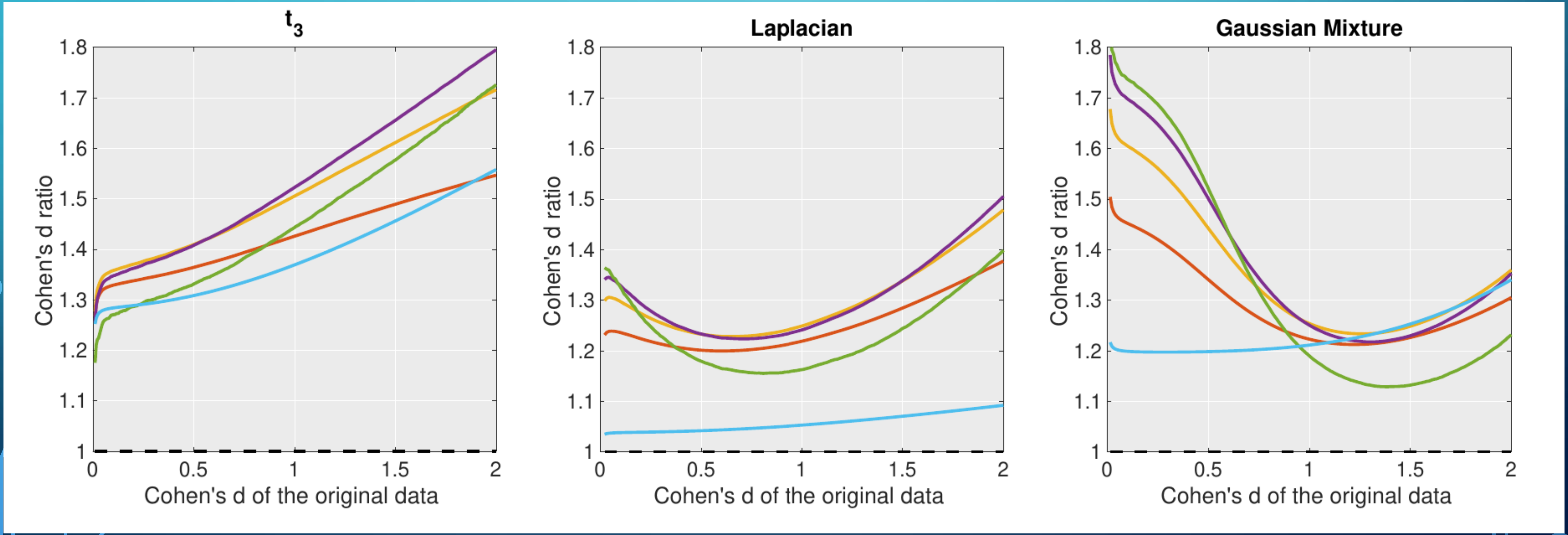
- Cohen's d is equivalent to asymptotic relative efficiency



TRANSFORMATIONS CAN BE USED TO INCREASE COHEN'S D

- Cohen's d is equivalent to asymptotic relative efficiency

$$\frac{\mathbb{E}[f(Y)]}{\mathrm{std}(f(Y))} \Big/ \frac{\mathbb{E}[Y]}{\mathrm{std}(Y)}$$



INFERENCE USING A CLT

Theorem 4.1. *Suppose that* $(f_v)_{v\in\mathcal{V}} : \mathbb{R} \to \mathbb{R}$ *is a collection of real valued anti-symmetric functions such that* $\max_{v\in\mathcal{V}} \mathbb{E}(f_v(Y_1(v))^2) < \infty$. *Then as* $n \to \infty$,

$$T_n^*|_{\mathcal{N}} \xrightarrow{d} \mathcal{G}|_{\mathcal{N}}(0, \rho)$$

where $\rho(u, v) = corr(f_u(\epsilon_1(u)), f_v(\epsilon_1(v)))$ *for* $u, v \in \mathcal{V}$. *Moreover* $\min_{v\in\mathcal{D}\setminus\mathcal{N}} T_n^*|_{\mathcal{N}}(v)$ *converges almost surely to* $-\infty$.

INFERENCE USING THE MULTIPLER BOOTSTRAP

Given a number of bootstraps $B \in \mathbb{N}$ for $2 \leq b \leq B$, define bootstrapped test-statistics, $T^*_{n,b}(v) = \frac{\sqrt{n}\hat{\mu}_{n,b}(v)}{\hat{\sigma}_{n,b}}$, at each $v \in \mathcal{V}$, where

$$\hat{\mu}_{n,b}(v) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_{bi} \left(f_v(Y_i(v)) - \frac{1}{n} \sum_{j=1}^{n} f_v(Y_j(v)) \right), \text{ and}$$

$$\hat{\sigma}_{n,b}(v) = \left(\frac{1}{n-1} \sum_{j=1}^{n} \left(g_{bi} \left(f_v(Y_i(v)) - \frac{1}{n} \sum_{j=1}^{n} f_v(Y_j(v)) \right) - \hat{\mu}_{n,b}(v) \right)^2 \right)^{1/2},$$

We take $T^*_{n,1} = T^*_n$

INFERENCE USING THE MULTIPLIER BOOTSTRAP

Theorem 4.2. *Suppose that* $(f_v)_{v\in\mathcal{V}} : \mathbb{R} \to \mathbb{R}$ *is a collection of real valued anti-symmetric functions such that* $\max_{v\in\mathcal{V}} \mathbb{E}(f_v(Y_1(v))^2) < \infty$. *Then*

$$\left(T^*_{n,2}, \ldots, T^*_{n,B}\right)^T \xrightarrow{d} \mathcal{G}(0, \rho I_{(B-1)\times(B-1)}), \text{ and moreover,}$$

$$\left(T^*_{n,1}|_{\mathcal{N}}, T^*_{n,2}|_{\mathcal{N}}, \ldots, T^*_{n,B}|_{\mathcal{N}}\right)^T \xrightarrow{d} \mathcal{G}|_{\mathcal{N}}(0, \rho I_{B\times B}),$$

where $\rho(u, v) = corr(f_u(\epsilon_1(u)), f_v(\epsilon_1(v)))$ *for each* $u, v \in \mathcal{V}$.

TESTING USING THE BOOTSTRAPPED DATA

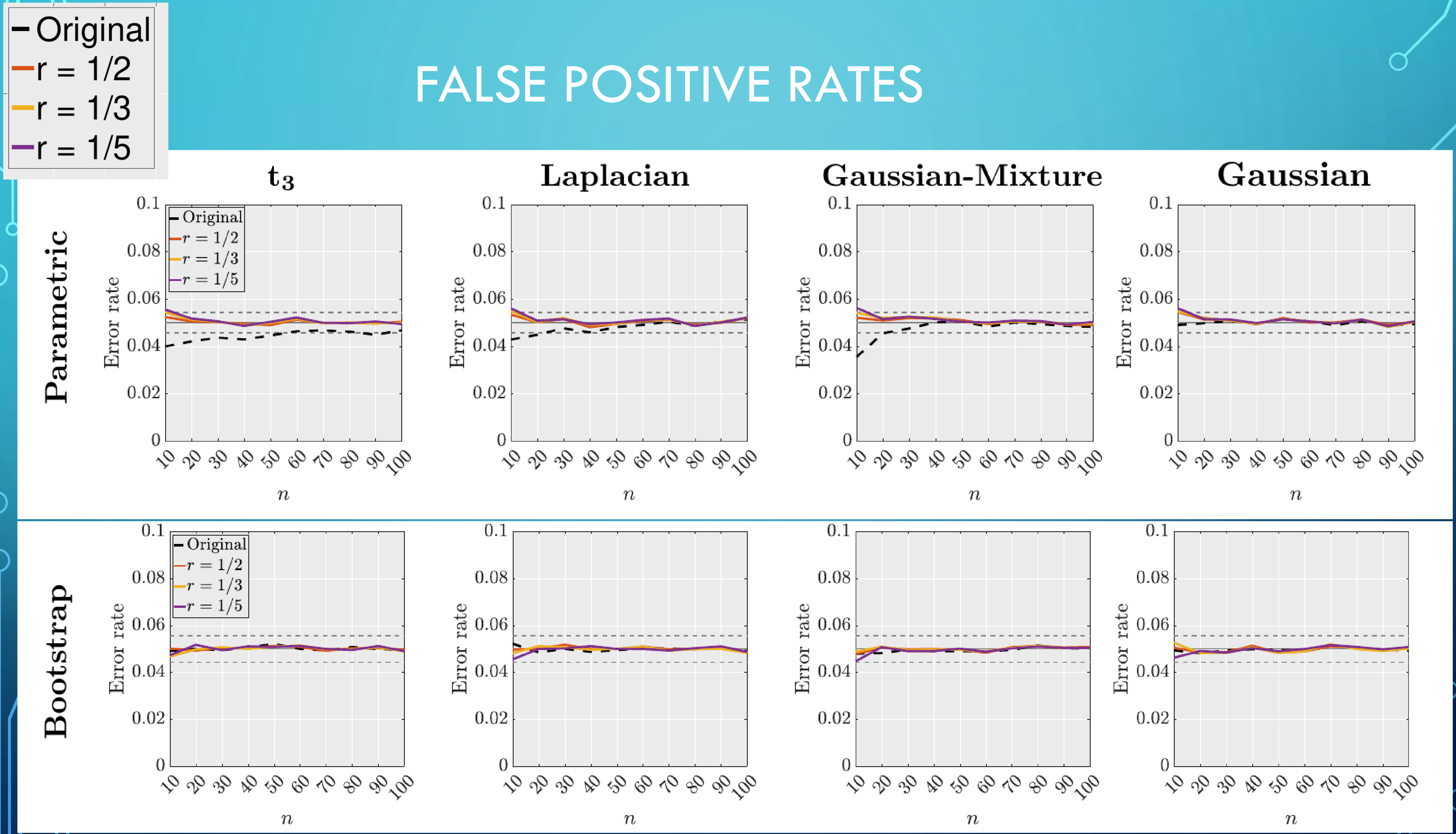
$$\lambda_{\alpha,n,B} = \inf\left\{\lambda : \frac{1}{B}\sum_{b=1}^{B} 1\left[\max_{v\in\mathcal{V}} T^*_{n,b}(v) \le \lambda\right] \ge \alpha\right\}$$

Consider the test which rejects $H_0^D(v)$ at each voxel $v \in \mathcal{V}$ if $T_n^*(v) > \lambda_{\alpha,n,B}$.

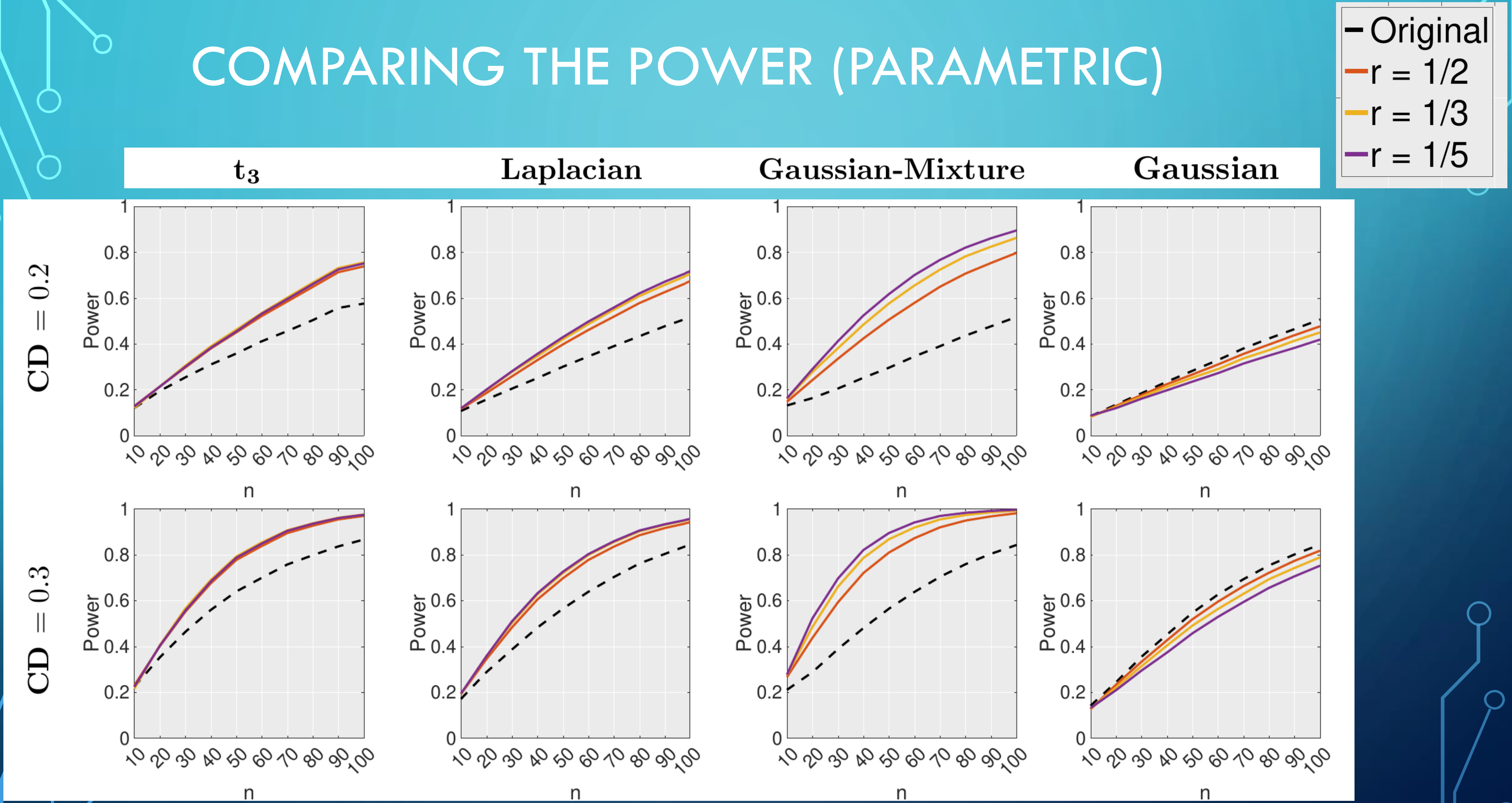
Theorem 4.3. *(Directional FWER control) Let* $\mathcal{R}_n = \{v \in \mathcal{V} : T_n^*(v) > \lambda_{\alpha,n,B}\}$ *and let* $\mathcal{D} = \{v \in \mathcal{V} : \mu(v) \le 0\}$. *Then, for* $\alpha \in \left(\frac{1}{B}, 1\right)$ *we have*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{R}_n \cap \mathcal{D} \ne \emptyset) \le \frac{\lfloor \alpha B \rfloor}{B} \le \alpha.$$

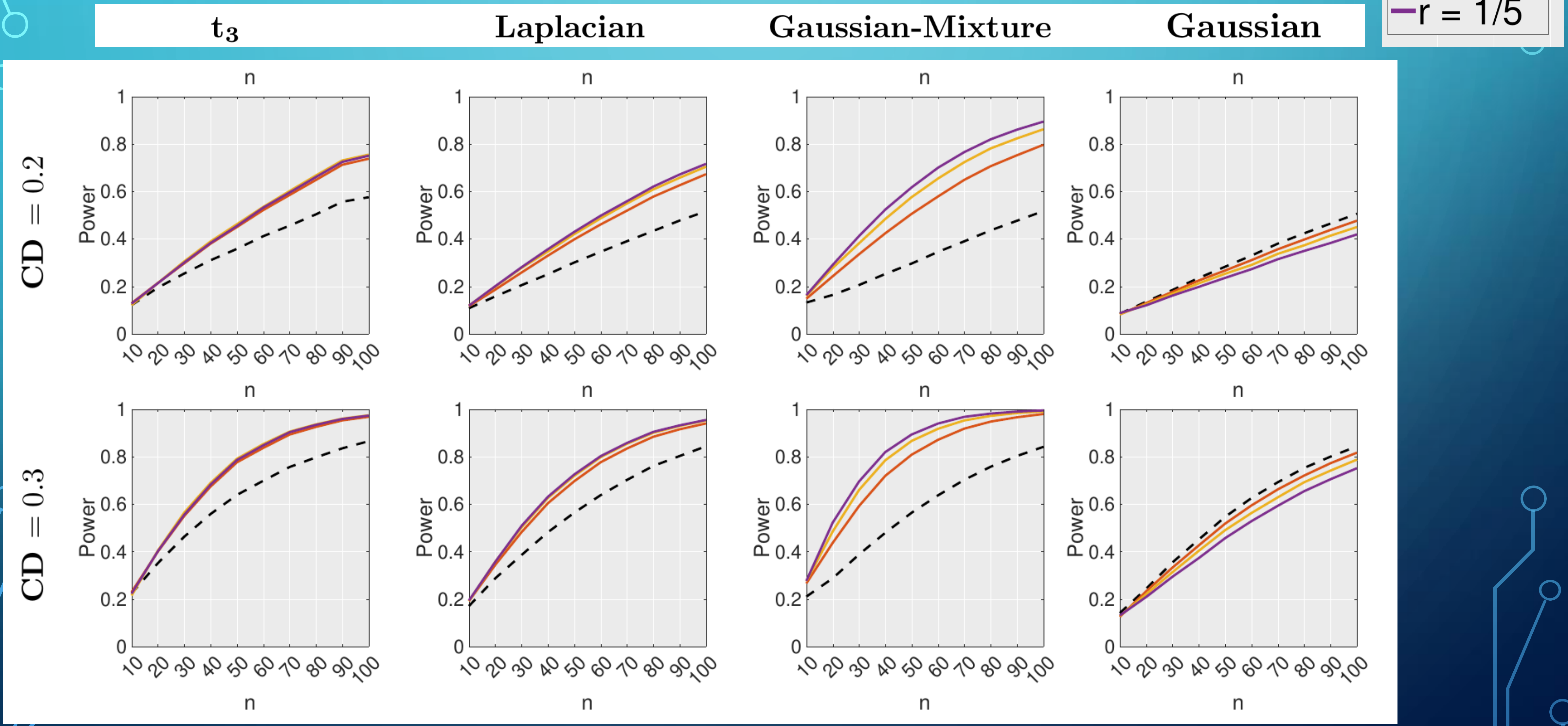
FALSE POSITIVE RATES



COMPARING THE POWER (PARAMETRIC)

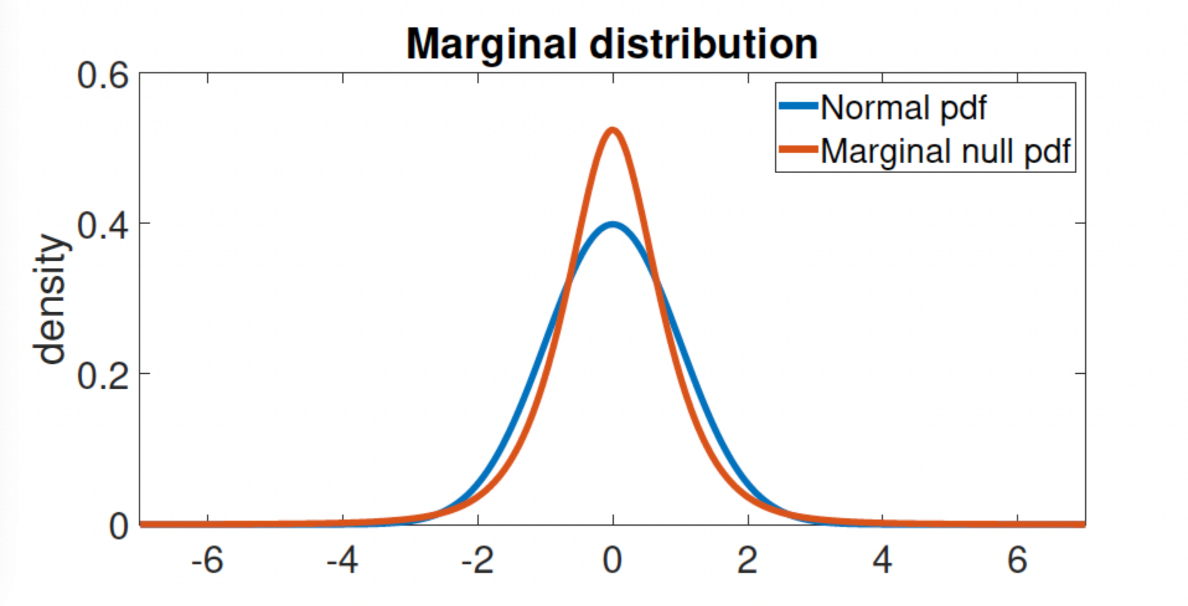


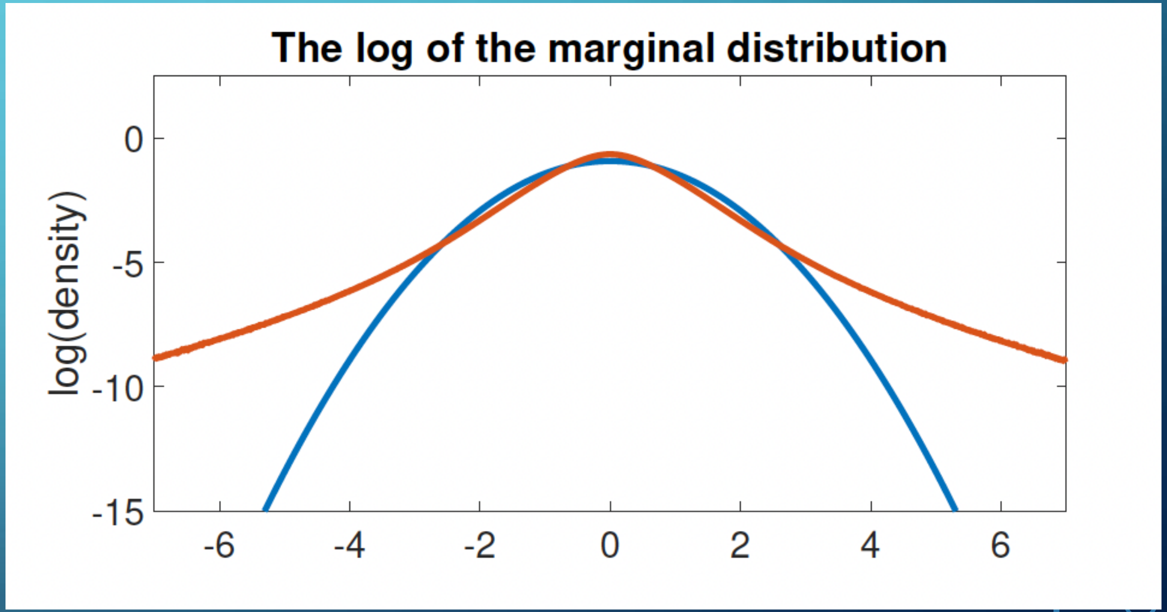
COMPARING THE POWER (BOOTSTRAP)



NULL DISTRIBUTION OF FMRI DATA IS HEAVY TAILED

Examine the null distribution using resting state data processed with fake task designs (from 7000 subjects from the UK Biobank).



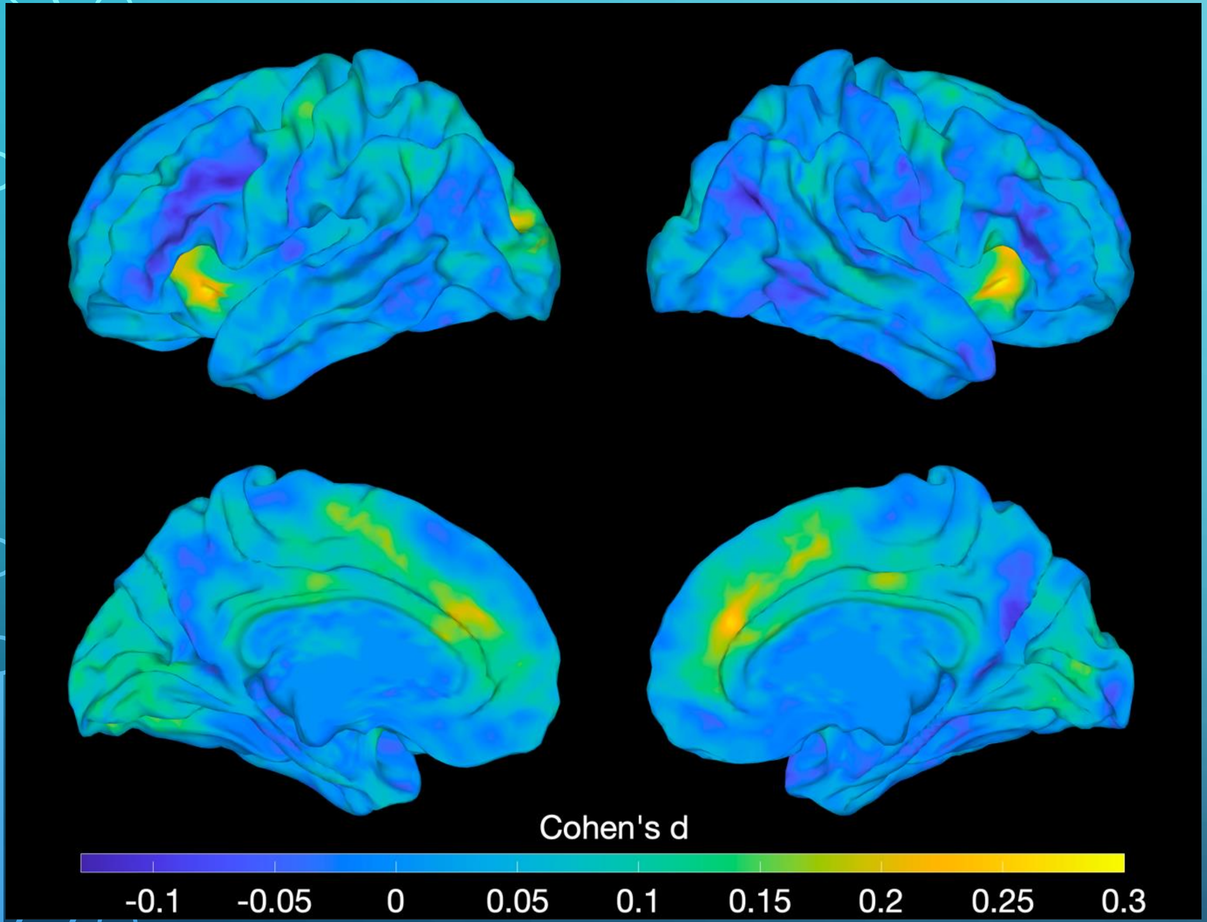


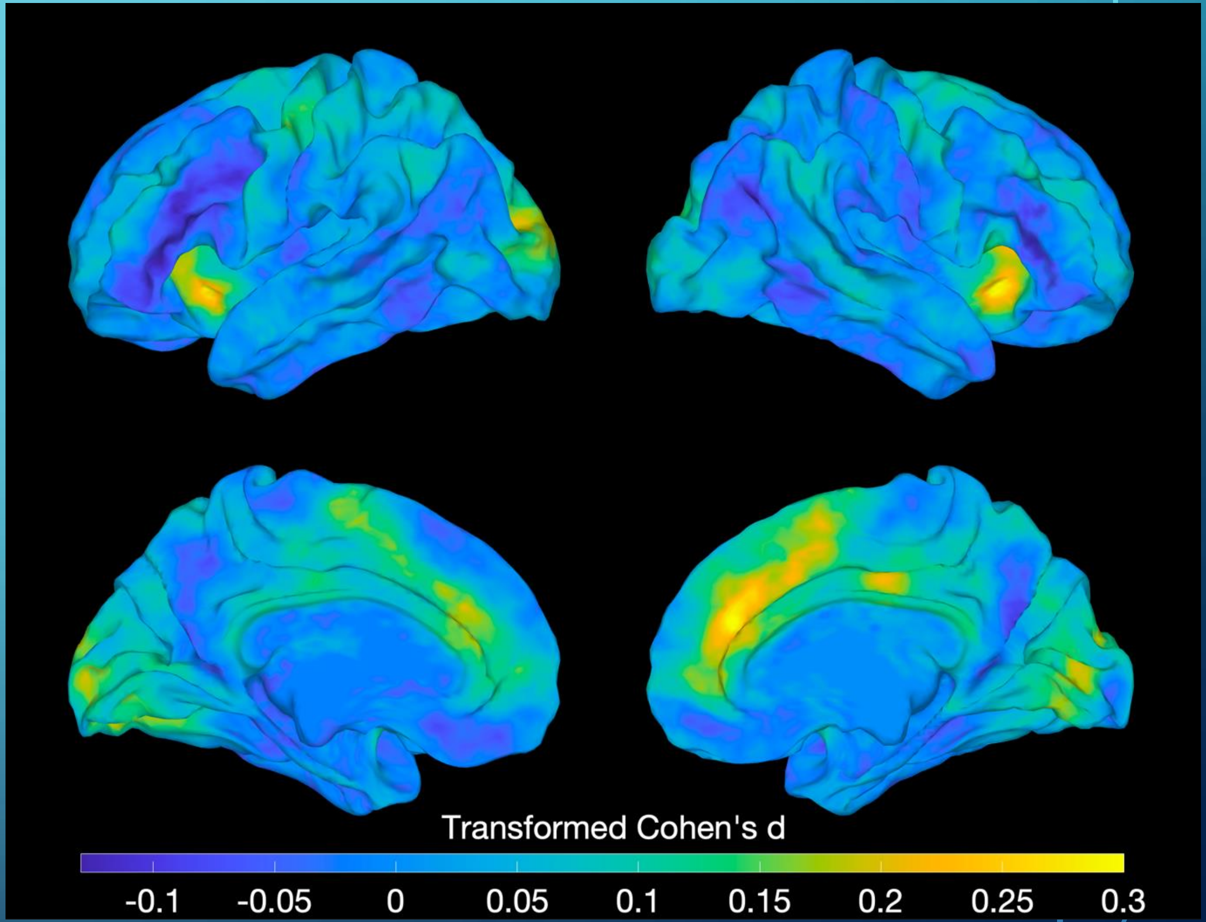
See Davenport et al (2023) for further details.

APPLICATION TO ABCD DATA

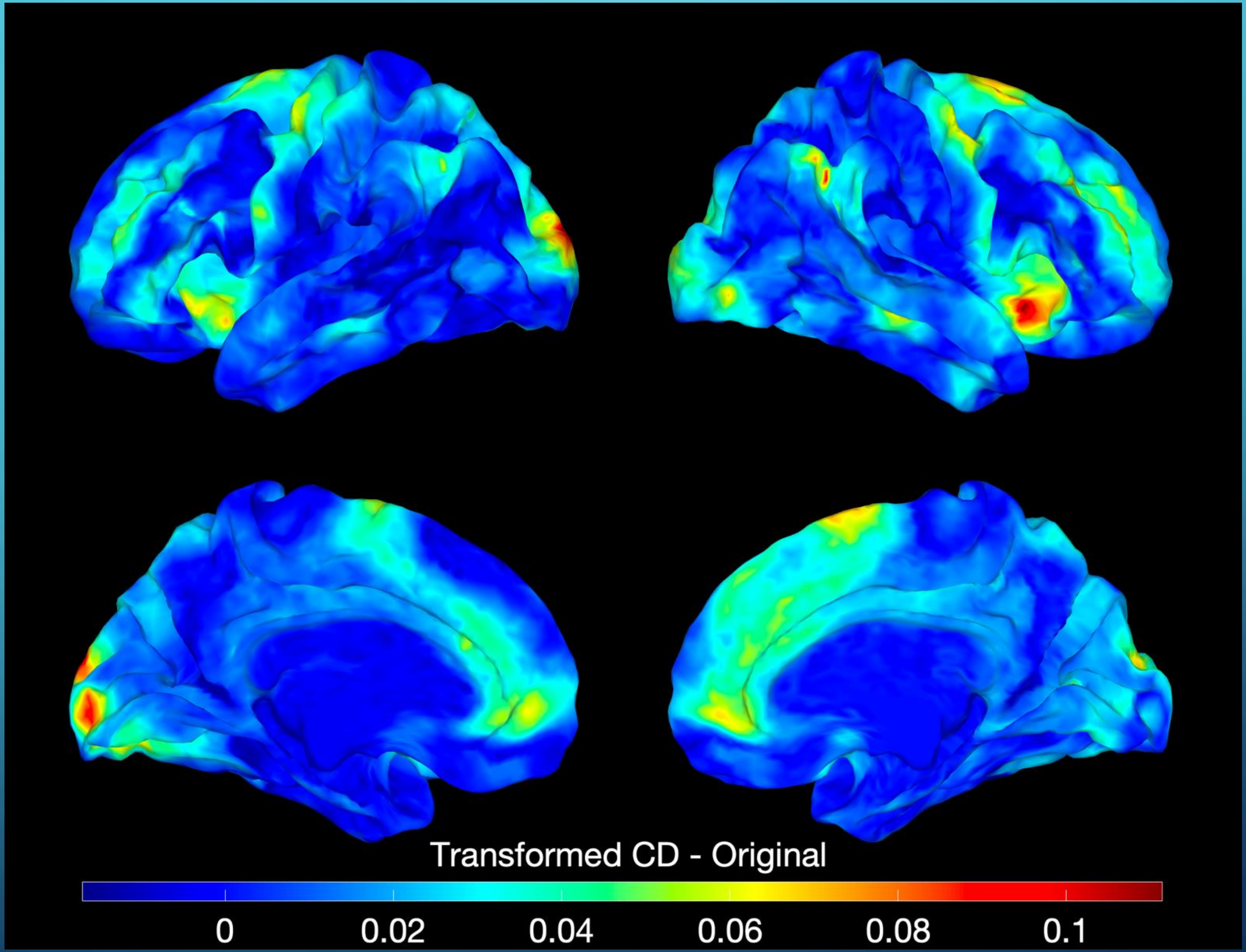
- We have n = 15000 subjects from the ABCD – MID task. We consider the constrast for antipating large vs small rewards.
- We have data at 18000 brain imaging vertices for each of these subjects.
- We consider inference with and without transformations.
- We use a transformation of r = 1/5 as this performed best in the simulations.

COHENS D BEFORE AND AFTER

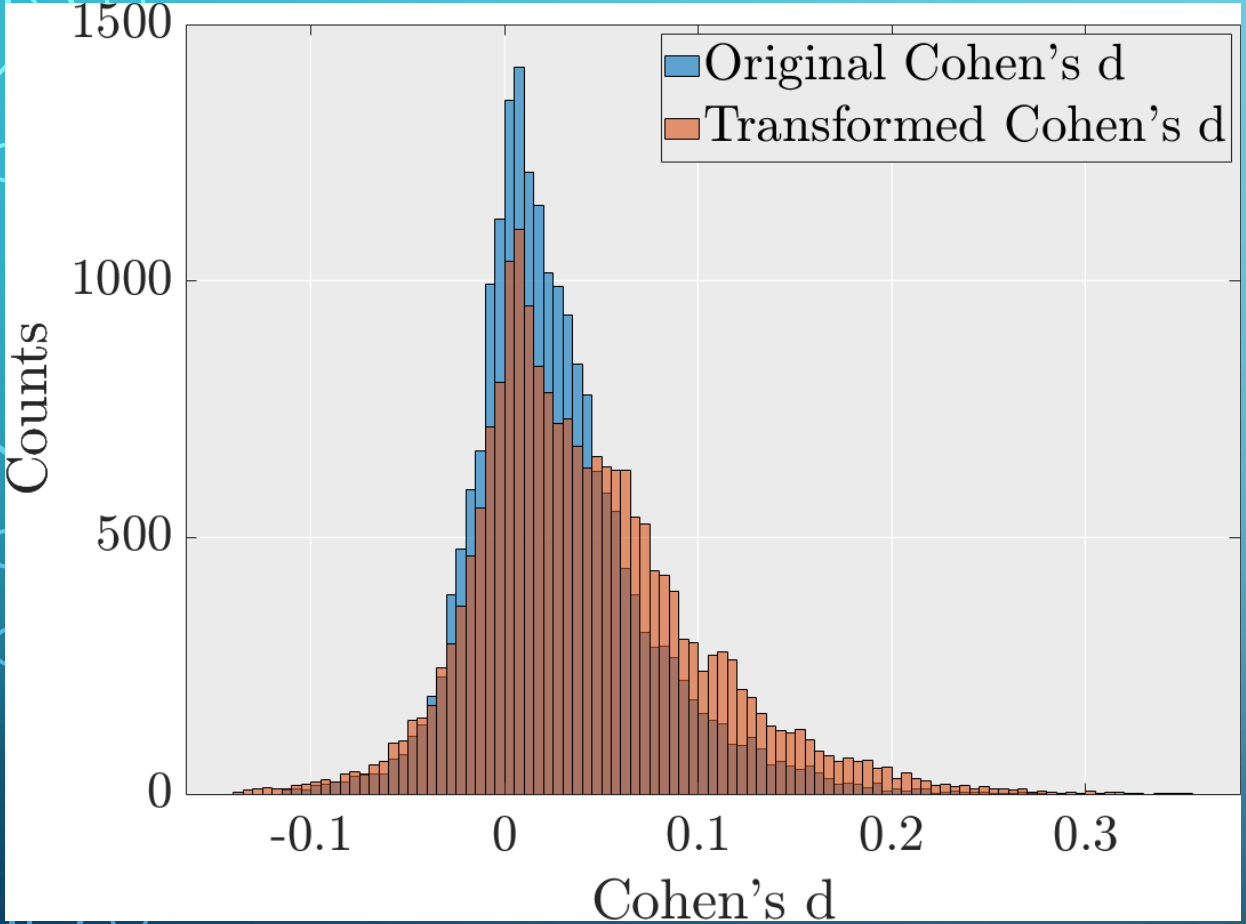


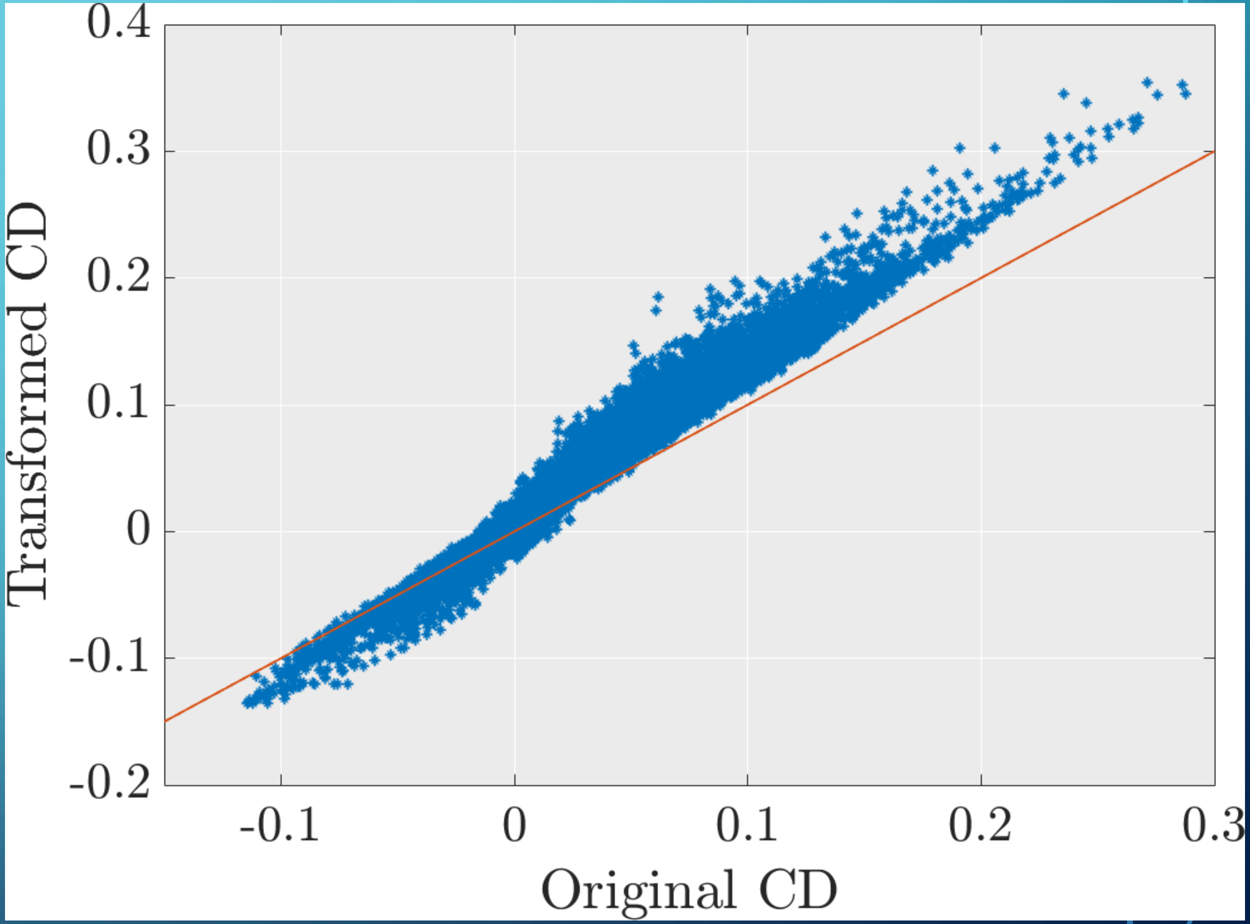


ABSOLUTE DIFFERENCE IN COHEN'S D



COMPARING COHENS D





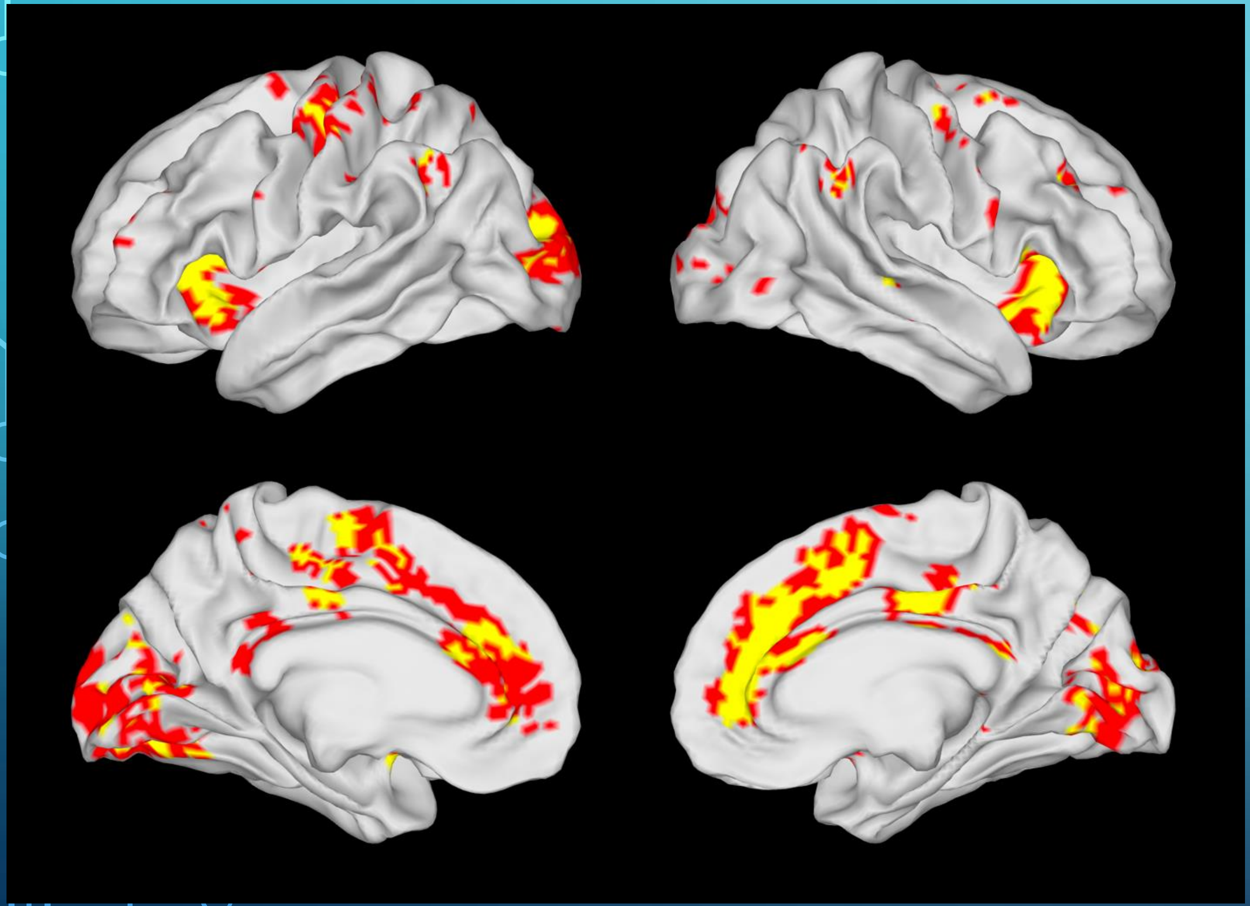
APPLICATION TO ABCD DATA

- We have n = 15000 subjects from the ABCD – MID task. We consider the constrast for antipating large vs small rewards.
- We have data at 18000 brain imaging vertices for each of these subjects.
- We consider inference with and without transformations.
- We use a transformation of $r = 1/5$ as this performed best in the simulations.
- We use the multiplier bootstrap to do directional inference. Bootstrapping is performed jointly over all voxels to ensure that the dependence structure is preserved.

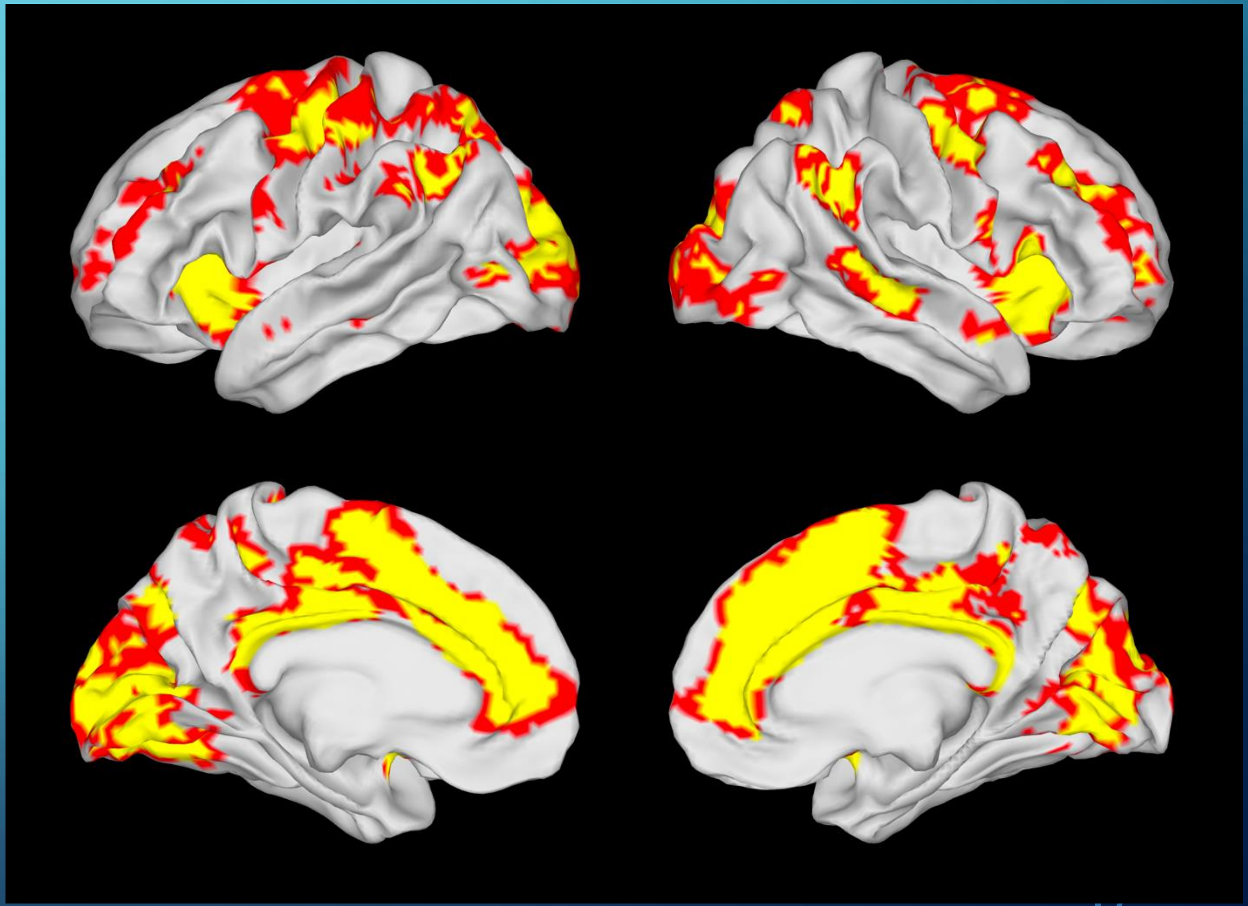
COMPARING DISCOVERIES

Vertices significant for both orginal and transformed

Vertices significant for transformed and not original



N = 1000

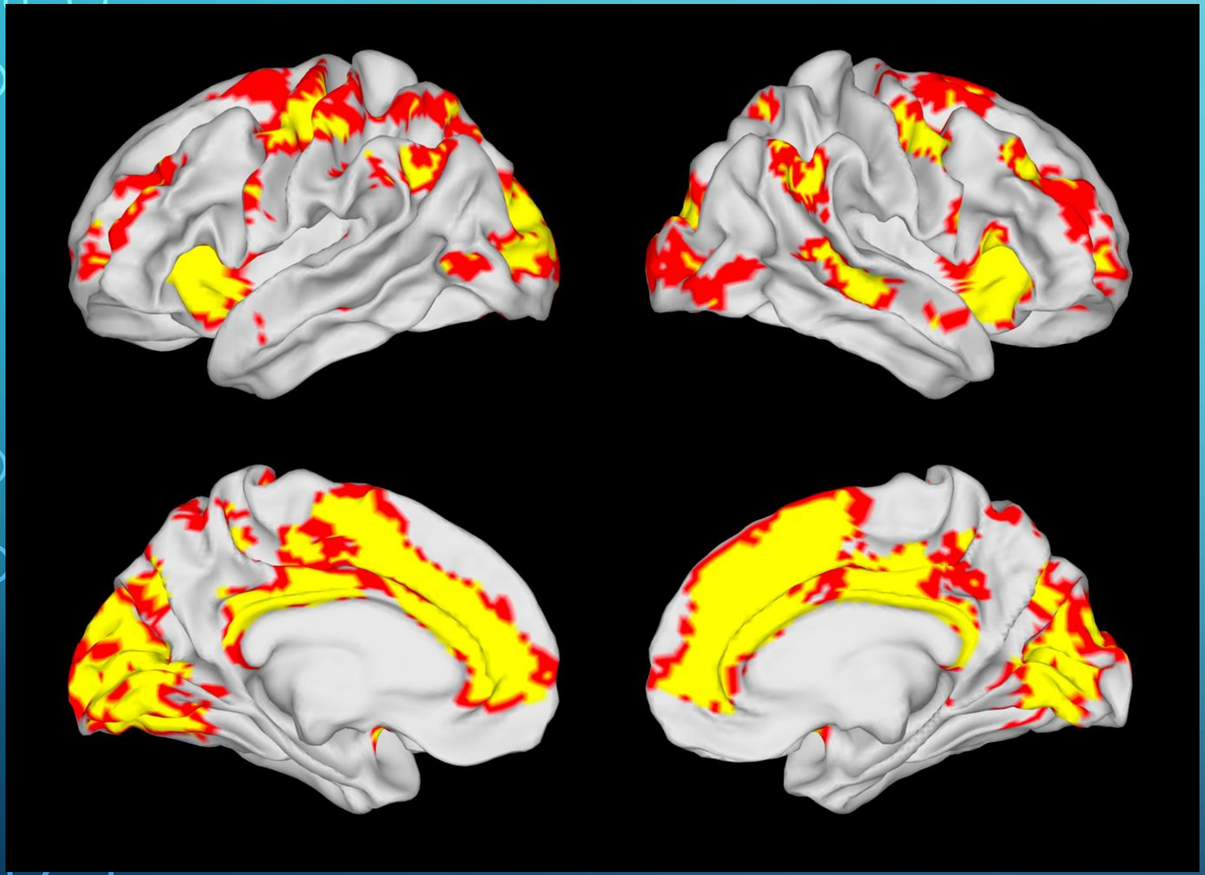


N = 4000

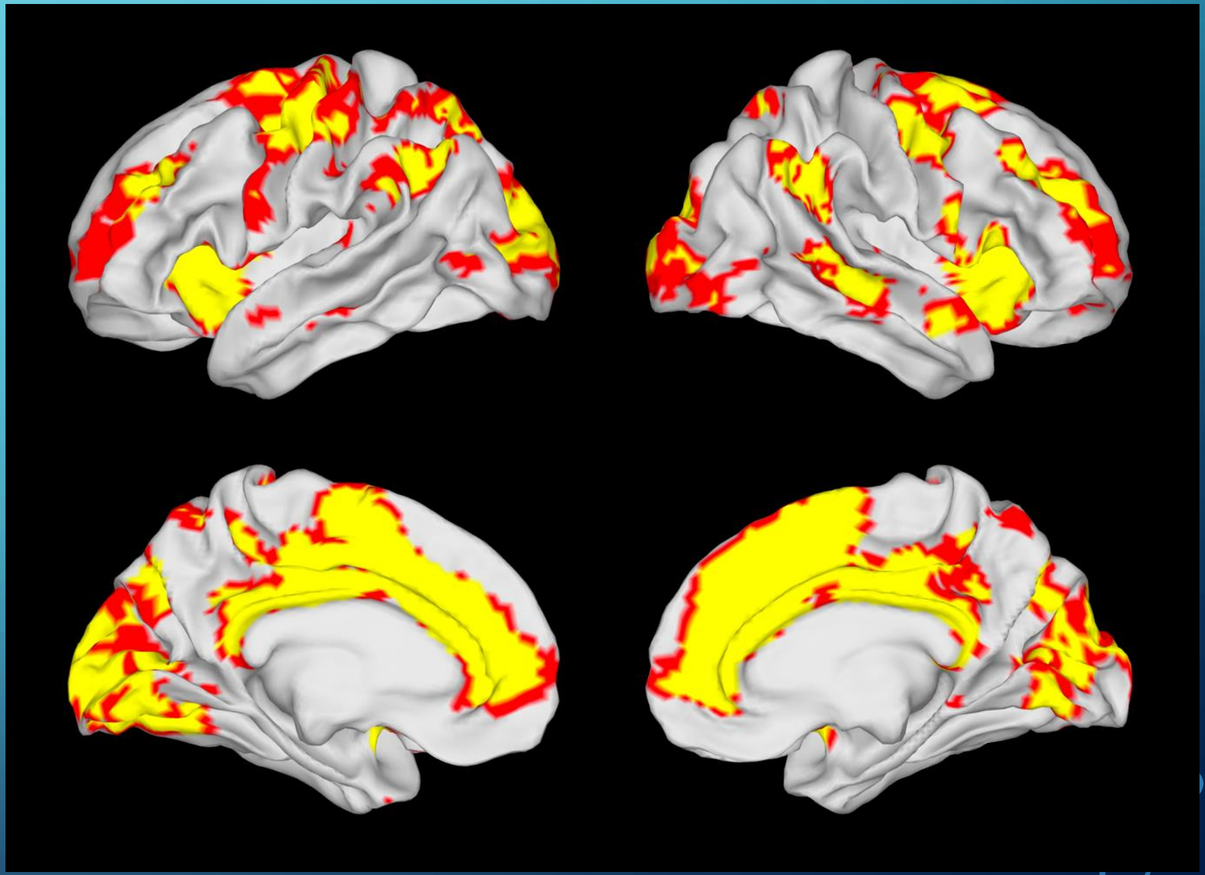
COMPARING DISCOVERIES

Vertices significant for both orginal and transformed

Vertices significant for transformed and not original



N = 3000

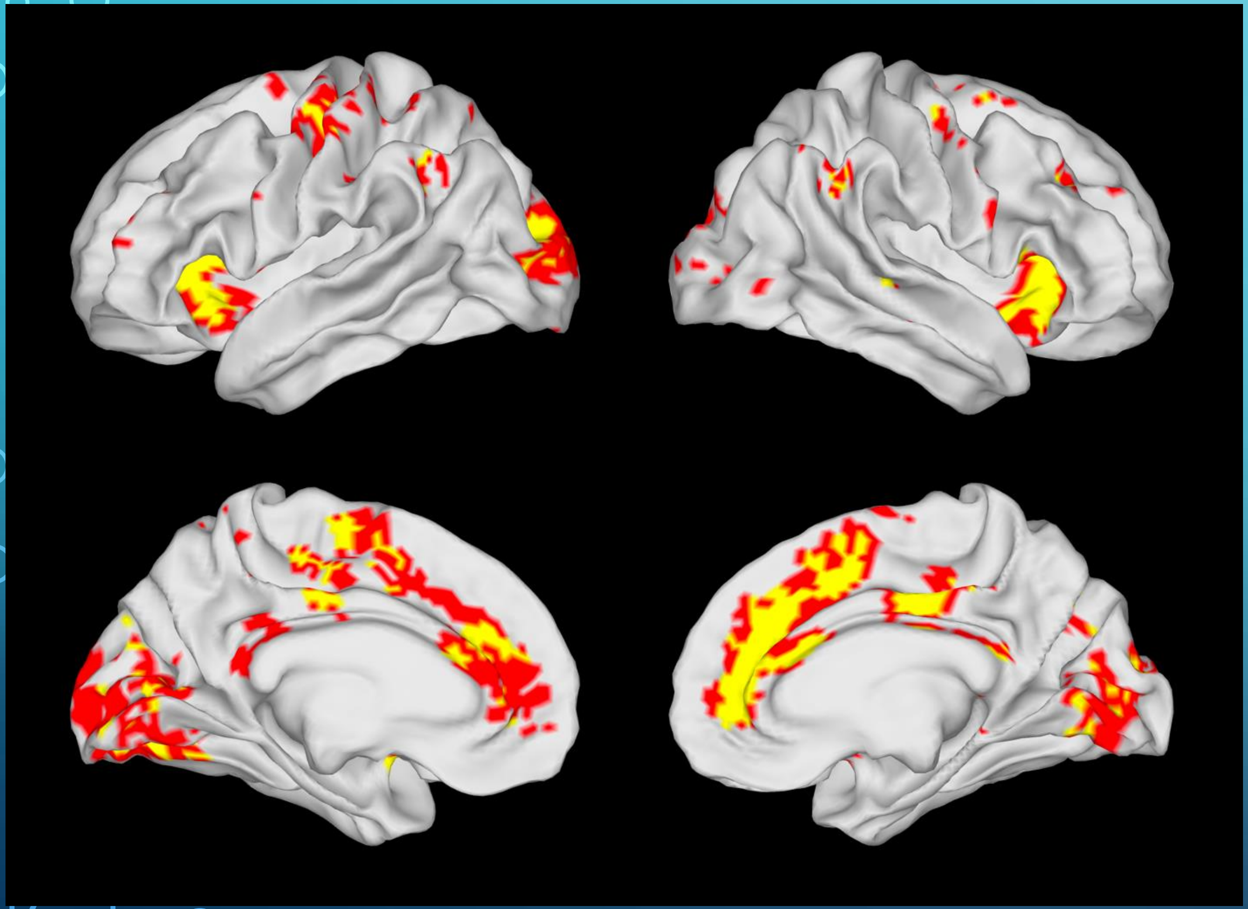


N = 6000

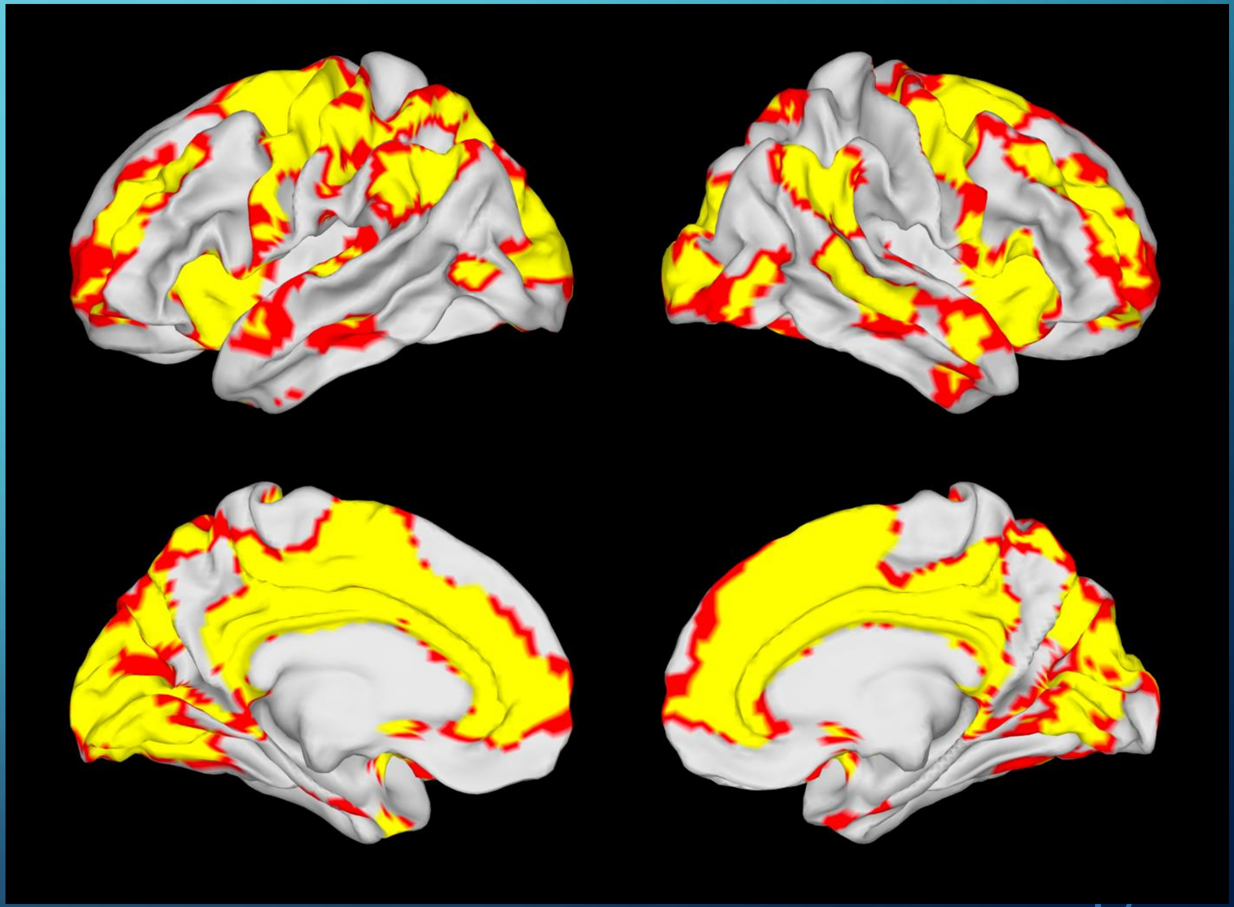
COMPARING DISCOVERIES

Vertices significant for both orginal and transformed

Vertices significant for transformed and not original

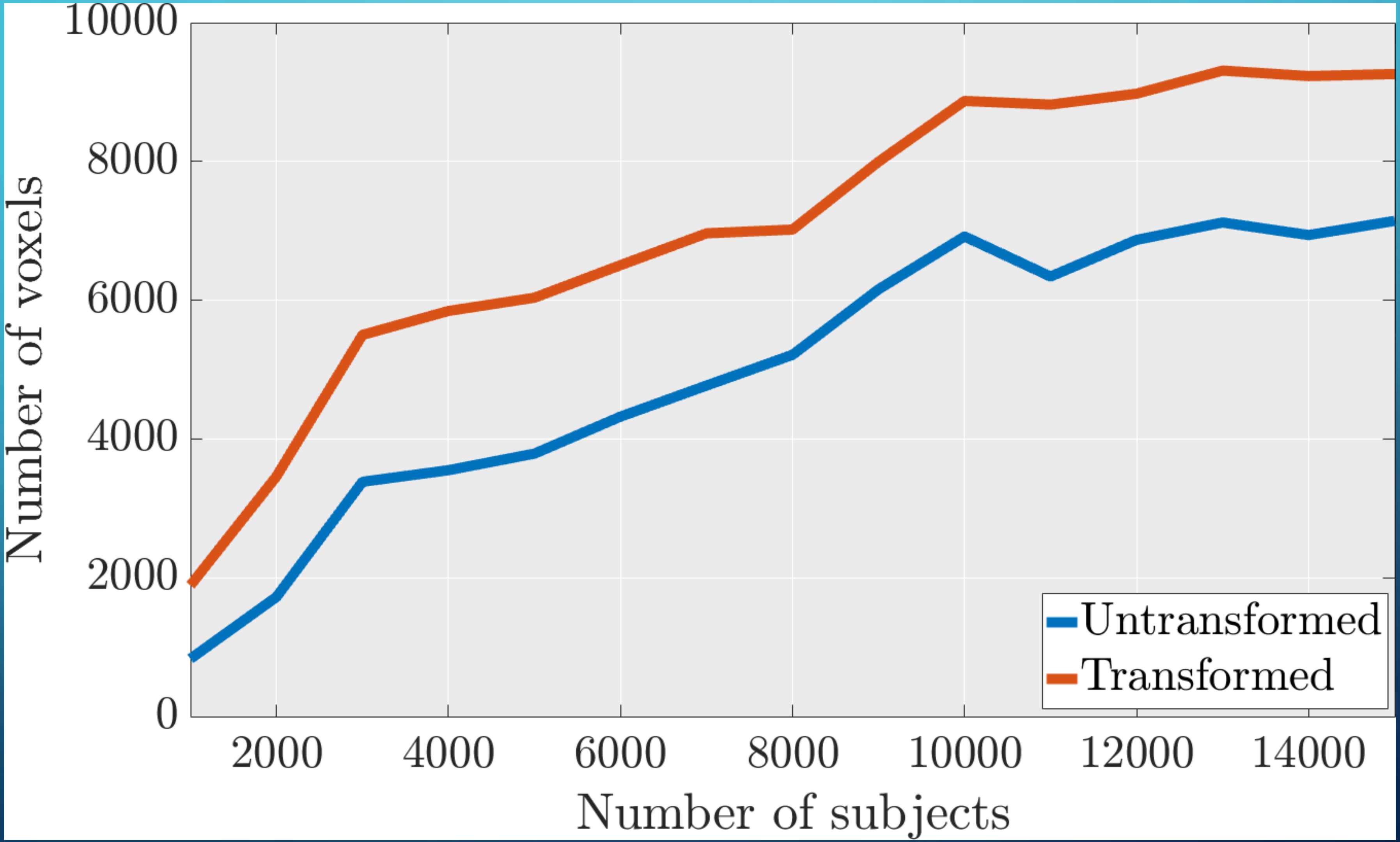


N = 1000



N = 15000

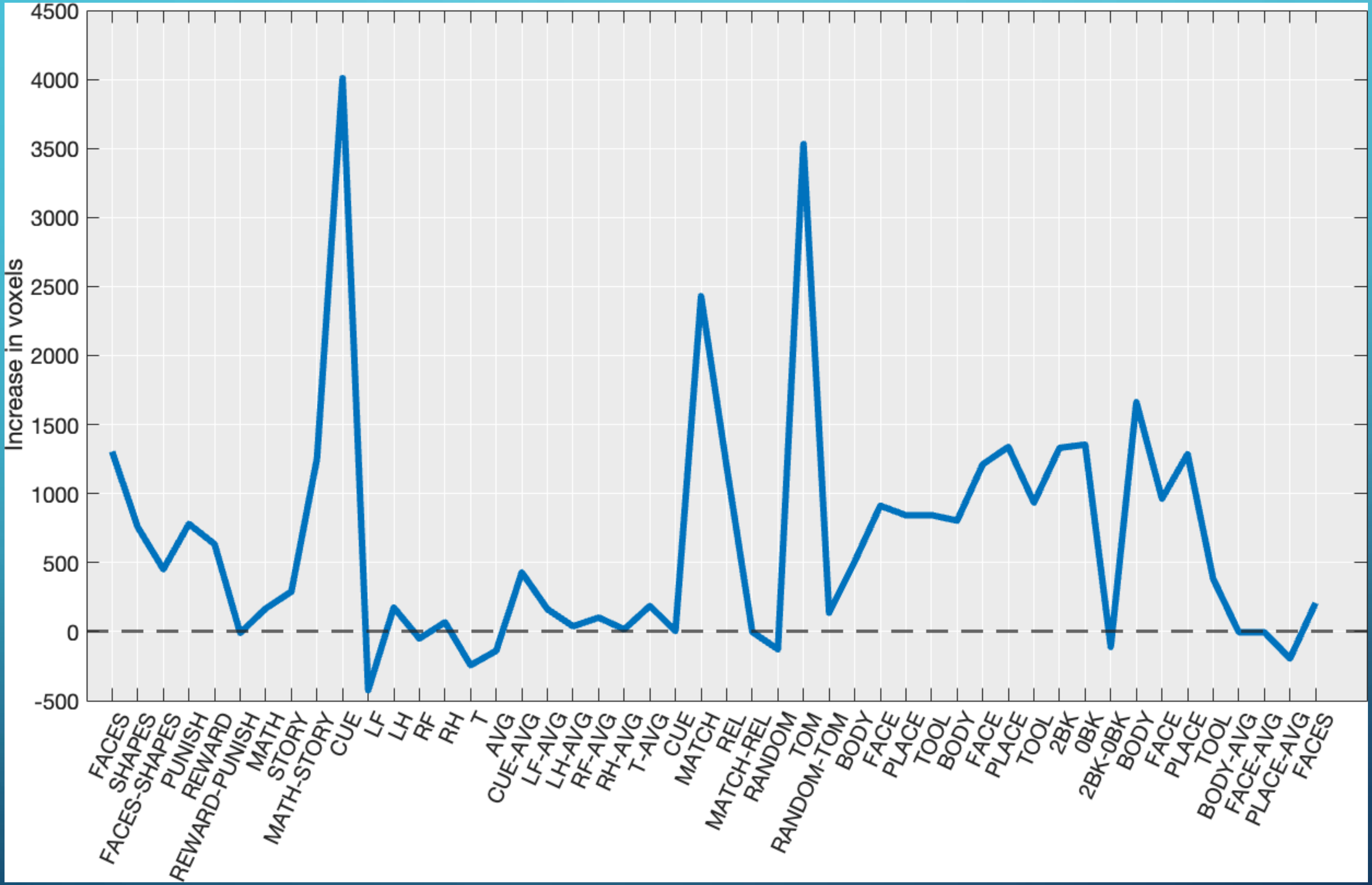
DISCOVERIES VS NUMBER OF SUBJECTS



CONCLUSIONS

- Transformations can be combined with non-parametric inference
- Transformations can be chosen to optimize power instead of Gaussianity
- The optimal transformation can be chosen based on the distribution of the data – in advance – in order to optimize power.
- fMRI data usually has a low sample size which can cause low power – this work helps to improve that
- We have focussed on $H_0^D(v) : \mu(v) \leq 0$ but testing $H_0^D(v) : \mu(v) \leq c$ is just as easy
- We focussed on the directional null however the point null can also be used and allows for exact inference based on sign-flipping.

INCREASE IN POWER ON THE HCP DATA



INCREASE IN POWER ON THE HCP DATA